# HealthInf 2017: Proceedings of the 10th International Conference on Health Informatics

**Book** · February 2017

CITATIONS
0

READS
6,710

**4 authors:**

Egon L. van den Broek
Utrecht University
**178** PUBLICATIONS **1,443** CITATIONS
SEE PROFILE

Ana L. N. Fred
University of Lisbon
**296** PUBLICATIONS **4,386** CITATIONS
SEE PROFILE

Hugo Gamboa
Universidade NOVA de Lisboa
**178** PUBLICATIONS **874** CITATIONS
SEE PROFILE

Mário A. P. Vaz
University of Porto
**286** PUBLICATIONS **1,269** CITATIONS
SEE PROFILE

# BIOSTEC 2017

*10TH INTERNATIONAL JOINT CONFERENCE ON BIOMEDICAL ENGINEERING SYSTEMS AND TECHNOLOGIES*

# PROCEEDINGS

## Volume 5: HEALTHINF

Porto, Portugal

21-23 February, 2017


EDITORS

Egon L. van den Broek, Ana Fred, Hugo Gamboa and Mário Vaz


http://www.biostec.org


SPONSORED BY

INSTICC

PAPERS AVAILABLE AT

SCITEPRESS DIGITAL LIBRARY

# BIOSTEC 2017

Proceedings of the
10th International Joint Conference on
Biomedical Engineering Systems and Technologies

Volume 5: HEALTHINF

Porto - Portugal

February 21 - 23, 2017

Sponsored by
**INSTICC - Institute for Systems and Technologies of Information, Control and Communication**

In Cooperation with
**ACM SIGKDD - ACM Special Interest Group on Knowledge Discovery and Data Mining**
**EUROMICRO**
**ISfTeH - International Society for Telemedicine & eHealth**
**AAAI - Association for the Advancement of Artificial Intelligence**
**ISCB - International Society for Computational Biology**
**BMES - Biomedical Engineering Society**

Edited by Egon L. van den Broek, Ana Fred, Hugo Gamboa and Mário Vaz

Printed in Portugal
ISBN: 978-989-758-213-4
Depósito Legal: 421067/17

http://www.healthinf.biostec.org
healthinf.secretariat@insticc.org

# Brief Contents

# INVITED SPEAKERS

**Bart M. ter Haar Romeny**
Eindhoven University of Technology (TU/e)
Netherlands

**Joyce Westerink**
Philips Research and Eindhoven University of Technology (TU/e)
Netherlands

**Kristina Höök**
Royal Institute of Technology
Sweden

**Bethany Bracken**
Charles River Analytics Inc.
United States

# SPECIAL SESSIONS CHAIRS

### SPECIAL SESSION ON SMART MEDICAL DEVICES - FROM LAB TO CLINICAL PRACTICE

Jan Sliwa, Bern University of Applied Sciences, Switzerland

Ivanoe De Falco, ICAR - CNR, Italy

Giovanna Sannino, Institute for High-Performance Computing and Networking, Italy

### SPECIAL SESSION ON ANALYSIS OF CLINICAL PROCESSES

Gert-Jan de Vries, Philips Research - Healthcare, Netherlands

# ORGANIZING COMMITTEES

### CONFERENCE CO-CHAIRS

Ana Fred, Instituto de Telecomunicações / IST, Portugal
Hugo Gamboa, LIBPHYS-UNL / FCT - New University of Lisbon, Portugal
Mário Vaz, INEGI LOME, FEUP, Portugal

### PROGRAM CHAIR

Egon L. van den Broek, Utrecht University, Netherlands

### ADVISORY BOARD

Joyce Westerink, Philips Research and Eindhoven University of Technology (TU/e), Netherlands

### SECRETARIAT

Vera Coelho, INSTICC, Portugal

### GRAPHICS PRODUCTION AND WEBDESIGNER

André Poeira, INSTICC, Portugal
Sara Santiago, INSTICC, Portugal

### WEBMASTER

João Francisco, INSTICC, Portugal
Carolina Ribeiro, INSTICC, Portugal

# PROGRAM COMMITTEE

**Anurag Agrawal**, CSIR Institute of Genomics and Integrative Biology, Center for Translational Research in Asthma & Lung, India

**Hesham Ali**, University of Nebraska at Omaha, United States

**Adil Alpkocak**, Dokuz Eylul University, Turkey

**Flora Amato**, Università degli Studi di Napoli Federico II, Italy

**Francois Andry**, Philips, United States

**Wassim Ayadi**, LERIA, University of Angers, France and LaTICE, University of Tunis, Tunisia, Tunisia

**Bert-Jan van Beijnum**, University of Twente, Netherlands

**Patrick Boissy**, Faculty of Medicine and Health Sciences, Université de Sherbrooke and Research Centre on Aging, Canada

**Sorana D. Bolboaca**, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

**Alessio Bottrighi**, Universita' del Piemonte Orientale, Italy

**Andrew D. Boyd**, University of Illinois at Chicago, United States

**Egon L. van den Broek**, Utrecht University, Netherlands

**Berry de Bruijn**, National Research Council, Canada

**Federico Cabitza**, Università degli Studi di Milano-Bicocca, Italy

**Eric Campo**, LAAS CNRS, France

**Guillermo Lopez Campos**, The University of Melbourne, Australia

**Marc Cavazza**, University of Kent, United Kingdom

**Philip K. Chan**, Florida Institute of Technology, United States

**Taridzo Chomutare**, University Hospital of North Norway, Norway

**Mihail Cocosila**, Athabasca University, Canada

**Miguel Coimbra**, Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Portugal

**Emmanuel Conchon**, University of Limoges, XLIM, France

**Carlos Costa**, Universidade de Aveiro, Portugal

**Andre Dekker**, MAASTRO Clinick, Netherlands

**Chrysanne DiMarco**, University of Waterloo, Canada

**Liliana Dobrica**, University Politehnica of Bucharest, Romania

**Stephan Dreiseitl**, Upper Austria University of Applied Sciences at Hagenberg, Austria

**José Fonseca**, UNINOVA, Portugal

**Daniel Ford**, Dell Research, United States

**Christoph M. Friedrich**, University of Applied Science and Arts Dortmund, Germany

**Ioannis Fudos**, University of Ioannina, Greece

**Henry Gabb**, University of Illinois at Urbana-Champaign, United States

**Hugo Gamboa**, LIBPHYS-UNL / FCT - New University of Lisbon, Portugal

**Mauro Giacomini**, University of Genoa, Italy

**Laura Giarré**, Università degli Studi di Palermo, Italy

**Alejandro Rodríguez González**, Centro de Tecnología Biomédica, Spain

**Yves Gourinat**, ISAE-SUPAERO, France

**David Greenhalgh**, University of Strathclyde, United Kingdom

**Jorge Henriques**, Universidade de Coimbra, Portugal

**Dragan Jankovic**, Faculty of Electronic Engineering Nis, Serbia

**Noreen Kamal**, University of Calgary, Canada

**Enkelejda Kasneci**, University of Tübingen, Germany

**Finn Kensing**, Professor of Computer Science, University of Copenhagen, Denmark, Denmark

**Irena Koprinska**, University of Sydney, Australia

**Sofia Kyratzi**, University of the Aegean, Greece

**Elyes Lamine**, University of Toulouse, CUFR J.F. Champollion, France

**Gondy Leroy**, University of Arizona, United States

**Giuseppe Liotta**, University of Perugia, Italy

**Guillaume Lopez**, Aoyama Gakuin University, Japan

**Martin Lopez-Nores**, University of Vigo, Spain

**Alda Marques**, University of Aveiro, Portugal

**Paloma Martínez**, Universidad Carlos III de Madrid, Spain

**Ken Masters**, Sultan Qaboos University, Oman

**Sally Mcclean**, University of Ulster, United Kingdom

**Gerrit Meixner**, Heilbronn University, Germany

**Mohyuddin Mohyuddin**, King Abdullah International Medical Research Center (KAIMRC), Saudi Arabia

**Christo El Morr**, York University, Canada

**Hammadi Nait-Charif**, Bournemouth University, United Kingdom

**Goran Nenadic**, University of Manchester, United Kingdom

**José Luis Oliveira**, Universidade de Aveiro, Portugal

**Rui Pedro Paiva**, University of Coimbra, Portugal

**Guy Pare**, HEC Montréal, Canada

**Vimla L. Patel**, Arizona State University, United States

**Sotiris Pavlopoulos**, ICCS, Greece

**Göran Petersson**, eHealth Institute, Dept of Medicine and Optometry, Linnaeus University, Sweden

**Enrico Maria Piras**, Fondazione Bruno Kessler - Trento, Italy

**Arkalgud Ramaprasad**, University of Illinois at Chicago, United States

**Grzegorz Redlarski**, Gdansk University of Technology, Poland

**Ita Richardson**, University of Limerick, Ireland

**Valter Roesler**, Federal University of Rio Grande do Sul, Brazil

**Elisabetta Ronchieri**, INFN, Italy

**Senjuti Basu Roy**, The University of Washington - Tacoma, United States

**Carolina Ruiz**, WPI, United States

**George Sakellaropoulos**, University of Patras, Greece

**Ovidio Salvetti**, National Research Council of Italy - CNR, Italy

**Akio Sashima**, AIST, Japan

**Jacob Scharcanski**, UFRGS - Universidade Federal do Rio Grande do Sul, Brazil

**Bettina Schnor**, Potsdam University, Germany

**Yuval Shahar**, Ben Gurion University, Israel

**Jan Sliwa**, Bern University of Applied Sciences, Switzerland

**George Spyrou**, The Cyprus Institute of Neurology and Genetics, Cyprus

**Jan Stage**, Aalborg University, Denmark

**Vicente Traver**, ITACA, Universidad Politécnica de Valencia, Spain

**Gary Ushaw**, Newcastle University, United Kingdom

**Aristides Vagelatos**, CTI, Greece

**Sumithra Velupillai**, KTH, Stockholm and King's College, London, Sweden

**Sitalakshmi Venkatraman**, Melbourne Polytechnic, Australia

**Francisco Veredas**, Universidad de Málaga, Spain

**Justin Wan**, University of Waterloo, Canada

**Szymon Wilk**, Institute of Computing Science, Poznan University of Technology, Poland

**Janusz Wojtusiak**, George Mason University, United States

**Clement T. Yu**, University of Illinois at Chicago, United States

**Xuezhong Zhou**, Beijing Jiaotong University, China

**André Zúquete**, IEETA, IT, Universidade de Aveiro, Portugal

# Auxiliary Reviewers

**Koundinya Desiraju**, Institute of Genomics and Integrative Biology, Mall Road, Delhi, India, India

**Angela Locoro**, Università degli Studi di Milano-Bicocca, Italy

**Imen MEGDICHE**, IRIT, France

**Sebastian Rauh**, Heilbronn University, Germany

**Kateryna Sergieieva**, Heilbronn University of Applied Science, Germany

# Special Sessions Program Committee

### Special Session on Smart Medical Devices - From Lab to Clinical Practice

**Emmanuel Benoist**, Bern University of Applied Sciences, Switzerland

**Pedro Brandão**, University of Porto, Portugal

**Xiao Dong**, Southeast University, China

**Ivanoe De Falco**, ICAR - CNR, Italy

**Ardelio Galletti**, Parthenope University of Naples, Italy

**Francesco Gargiulo**, ICAR CNR, Italy

**Jolanta Mizera-Pietraszko**, Institute of Informatics, Wroclaw University of Technology, Poland

**Raffaele Montella**, University of Napoli Parthenope, Italy

**Aanjhan Ranganathan**, ETH Zurich, Switzerland

**Giovanna Sannino**, ICAR (CNR), Italy

**Stefano Silvestri**, ICAR CNR, Italy

**Jan Sliwa**, Bern University of Applied Sciences, Switzerland

**Laura Verde**, University of Naples Parthenope, Italy

**Pascal Walliser**, Fluance AG, Switzerland

**Shuang Wang**, University of California San Diego, United States

### Special Session on Analysis of Clinical Processes

**Jennifer Caffarel**, Philips Research, Netherlands

**Sreejita Ghosh**, University of Groningen, Netherlands

**Rick van Veen**, De Montfort University, United Kingdom

**Gert-Jan de Vries**, Philips Research - Healthcare, Netherlands

**Thomas Wendler**, Philips Research, Germany

# SELECTED PAPERS BOOK

A number of selected papers presented at HEALTHINF 2017 will be published by Springer in a CCIS Series book. This selection will be done by the Conference Co-chairs and Program Chair, among the papers actually presented at the conference, based on a rigorous review by the HEALTHINF 2017 Program Committee members.

# FOREWORD

This book contains the proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017). This conference is sponsored by the Institute for Systems and Technologies of Information, Control and Communication (INSTICC), in cooperation with the ACM Special Interest Group on Knowledge Discovery and Data Mining (ACM SIGKDD), the EUROMICRO, the International Society for Telemedicine & eHealth (ISfTeH), the Association for the Advancement of Artificial Intelligence (AAAI), the International Society for Computational Biology (iSCB) and the Biomedical Engineering Society (BMES).

The purpose of BIOSTEC is to bring together researchers and practitioners interested in both theoretical advances and applications of information systems, artificial intelligence, signal processing, electronics and other engineering tools in knowledge areas related to biology and medicine.

BIOSTEC is composed of five complementary and co-located conferences, each specialized in at least one of the aforementioned main knowledge areas: - International Conference on Biomedical Electronics and Devices – BIODEVICES; - International Conference on Bioimaging – BIOIMAGING; - International Conference on Bioinformatics Models, Methods and Algorithms – BIOINFORMATICS; - International Conference on Bio-inspired Systems and Signal Processing – BIOSIGNALS; - International Conference on Health Informatics – HEALTHINF.

The purpose of the International Conference on Biomedical Electronics and Devices (BIODEVICES) is to bring together professionals from electronics, mechanical engineering and related disciplines, interested in studying and using models, equipment and materials inspired from biological systems and/or addressing biological requirements. Monitoring devices, instrumentation sensors and systems, biorobotics, micro-nanotechnologies and biomaterials are some of the technologies presented at this conference.

The International Conference on Bioimaging (BIOIMAGING) covers the complex chain of acquiring, processing and visualizing structural or functional images of living objects or systems, including extraction and processing of image-related information. Examples of image modalities used in bioimaging are many, including: X-ray, CT, MRI and fMRI, PET and HRRT PET, SPECT, MEG and so on. Medical imaging and microscope/fluorescence image processing are important parts of bioimaging referring to the techniques and processes used to create images of the human body, anatomical areas, tissues, and so on, down to the molecular level, for clinical purposes, seeking to reveal, diagnose, or examine diseases, or medical science, including the study of normal anatomy and physiology. Image processing methods, such as denoising, segmentation, deconvolution and registration methods, feature recognition and classification represent an indispensable part of bioimaging, as well as related data analysis and statistical tools.

The International Conference on Bioinformatics Models, Methods and Algorithms (BIOINFORMATICS) provides an interdisciplinary forum for computational and biomedical researchers as well as practitioners to discuss emerging topics associated with the application of computational systems and information technologies to various fields of biosciences, particularly molecular biology. The conference highlights the advantages of using algorithmic techniques and mathematical models to better understand biological processes and systems, with a focus on recent developments in genomics research and its impact on advancing biomedical research. Areas of interest for this community include sequence analysis, biostatistics, image analysis, scientific data management and data mining, machine learning, pattern recognition, computational evolutionary biology, translational genomics, bioinformatics tools and other related fields.

The goal of the International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS) is to bring together researchers and practitioners from multiple areas of knowledge, including biology, medicine, engineering and other physical sciences, interested in studying and using models and techniques inspired by or applied to biological systems. A diversity of signal types can be found in this area, including image, audio and other biological sources of information. The analysis and use of these signals is a multidisciplinary area including signal processing, pattern recognition and computational intelligence techniques, amongst others.

The International Conference on Health Informatics (HEALTHINF) is the premier forum for the exchange of interdisciplinary research and ideas, across a broad spectrum of design, development, adoption, and application of informatics-based innovations in healthcare services. The field deals with the resources, devices, and methods required to optimize the acquisition, storage, retrieval, and use of information in health care. Amongst many other topics, it includes the research, development, and application of decision support systems, clinical data mining and machine learning, physiological and behavioral modeling, software systems in medicine and public health, and mobile, wearable and assistive technologies in healthcare.

The conferences are complemented by three special sessions, respectively on: - Neuro-electrostimulation in Neurorehabilitation Tasks (NENT); - Smart Medical Devices - From Lab to Clinical Practice (SmartMedDev); - Analysis of Clinical Processes (ACP).

The BIOSTEC program also includes a Doctoral Consortium on Biomedical Engineering Systems and Technologies that brought together Ph.D. students within the biomedical field to discuss their research in an international forum.

In 2017, BIOSTEC features four invited talks delivered by internationally distinguished speakers: Bethany Bracken (Charles River Analytics Inc., United States), Kristina Höök (Royal Institute of Technology, Sweden), Bart M. ter Haar Romeny (Eindhoven University of Technology, The Netherlands) and Joyce H.D.M. Westerink (Philips Research and Eindhoven University of Technology, The Netherlands).

The BIOSTEC joint conference received 297 paper submissions from 56 countries in all continents, of which 21% were accepted as full papers. The submission's high quality imposed difficult choices during the review process. To evaluate each submission, a double blind paper review was performed by the Program Committee, whose members are highly qualified independent researchers in the five BIOSTEC Conferences' topic areas.

As in previous editions of BIOSTEC, based on the reviewers' evaluations and on the quality of the presentations, a short list of authors will be selected and invited to submit extended revised versions of their papers for a book that will be published by Springer with the best papers of BIOSTEC 2017.

We would like to express our thanks to all participants. First of all, to the authors, whose quality work is the essence of this joint conference. Next, we thank all the members of the program committee and the auxiliary reviewers for their diligence and expert reviewing. Also, we would like to deeply thank the invited speakers for their excellent contribution in sharing their knowledge and vision. Finally, special thanks to all the members of the INSTICC team whose collaboration was fundamental for the success of this Conference.

Finally, special thanks to all the members of the INSTICC team whose collaboration was fundamental for the success of this Conference.

We wish you all an inspiring Conference and an unforgettable stay in Porto, Portugal. We hope to meet you again next year for BIOSTEC 2018, details of which will soon be available at http://www.biostec.org.

**Egon L. van den Broek**
Utrecht University, Netherlands

**Ana Fred**
Instituto de Telecomunicações / IST, Portugal

**Hugo Gamboa**
LIBPHYS-UNL / FCT - New University of Lisbon, Portugal

**Mário Vaz**
INEGI LOME, FEUP, Portugal

# CONTENTS

## SPECIAL SESSION ON ANALYSIS OF CLINICAL PROCESSES

### FULL PAPERS



### SHORT PAPER

# Invited Speakers

# KEYNOTE SPEAKERS

# Brain-inspired Medical Image Analysis for Computer-aided Diagnosis

Bart M. ter Haar Romeny
*Eindhoven University of Technology (TU/e), The Netherlands*

Abstract: Discoveries on brain mechanisms have really taken off. Modern optical and new MRI technologies give insight in this spectacular organ, especially in the field of vision. Of mutual benefit are new developments in deep learning and neural network modeling, the mathematical understanding, and the availability of massively parallel computing power. The lecture will address a number of lessons to learn from the brain for medical computer-aided diagnosis, explain the mathematical intuition of a number of algorithms, and show some remarkable successes booked so far.

## BRIEF BIOGRAPHY

Bart M. ter Haar Romeny received the MSc degree in Applied Physics from Delft University of Technology in 1978, Ph.D. from Utrecht University in 1983 in biophysics. He became principal physicist of the Utrecht University Hospital Radiology Department. He was co-founder and associate professor at the Image Sciences Institute (ISI) of Utrecht University (1989-2001). From 2001, ter Haar Romeny holds the chair of Biomedical Image Analysis at the Department of Biomedical Engineering of Eindhoven University of Technology in the Netherlands, and since 2011 is appointed distinguished professor at Northeastern University, Shenyang, China. He closely collaborates with Philips Healthcare and Philips Research, other industries and (national and international) hospitals and research groups. Currently he is project leader of the Sino-Dutch RetinaCheck project, a large screening project for early detection of diabetic retinopathy in Liaoning, China.

He authored an interactive tutorial book on multi-scale computer vision techniques, edited a book on non-linear diffusion theory in computer vision and is involved in (resp. initiated) a number of international collaborations on these subjects. He is author/co-author of over 200 refereed journal and conference papers, 12 books and book chapters, and holds 2 patents. He supervised 29 PhD students, of which 4 graduated cum laude, and over 140 Master students. He is senior member of IEEE, associate member of the Chinese Brainnetome consortium, visiting professor at the Chinese Academy of Sciences in Beijing, member of the Governing Board of IAPR, Fellow of EAMBES, and chairman of the Dutch Society for Pattern Recognition and Image Processing.



Romeny, B.
Brain-inspired Medical Image Analysis for Computer-aided Diagnosis.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, page 5
ISBN: 978-989-758-213-4

# Enhancing Well-being through Psychophysiology

Joyce Westerink
*Philips Research and Eindhoven University of Technology (TU/e), The Netherlands*

Abstract: Wellbeing is of importance to all of us, yet it is under constant pressure in our 24/7 economy. Our mental state is reflected to some extent in bodily parameters like heart rate and skin conductance, and we can measure them in real life using wearable sensing technology. Therefore, the hope is that such psychophysiological measurements might help us find an optimal balance in our busy lives. The lecture will discuss a number of promising application areas. Among them is bringing stress awareness through monitoring physiology, e.g. showing a user his stress pattern as it develops over the day. Another use of the psychophysiological signals is for biofeedback, so that a user is optimally guided in specific relaxation exercises. By placing the user in the center of a closed optimization loop we might achieve such improvements in a completely effortless way. Finally, psychophysiological signals can also facilitate communication between persons, for instance allowing people to see whether they are in sync.

## BRIEF BIOGRAPHY

Joyce Westerink (1960) was originally trained as a physicist and afterwards expanded her horizon towards human-technology. She is affiliated with Philips Research (NL), where her research focuses on psychological topics in a technological context, such as visual perception & image quality, human factors & user interfaces, and more recently psychophysiology in the context of emotions, wellbeing and affective computing. She currently holds a chair on Wellbeing and Psychophysiology in Human-Technology Interaction at Eindhoven University of Technology. Written output of her work can be found in some 40 articles in books and international journals, and some15 US patents and patent applications.



Westerink, J.
Enhancing Well-being through Psychophysiology.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, page 7
ISBN: 978-989-758-213-4

# Designing for Somaesthetic Experiences
## *Focusing on Actuation Rather than Sensing?*

Kristina Höök
*Royal Institute of Technology, Sweden*

Abstract: In designing for bodily experiences, there has been a lack of theories that can provide the underpinnings we need to understand and deepen our design thinking. Despite all the work we have seen on designing for embodiment, the actual corporeal, pulsating, live, felt body has been notably absent from both theory and practical work. At the same time, digital products have become an integral part of the fabric of everyday life, the pleasures (and pains) they give, their contribution to our social identity, or their general aesthetics are now core features of their design. We see more and more attempts to design explicitly for bodily experiences with digital technology, but it is a notably challenging design task. With the advent of new technologies, such as biosensors worn on your body, interactive clothes, or wearable computers such as mobiles equipped with accelerometers, a whole space of possibilities for gesture-based, physical and body-based interaction is opened. How can we do a better job in interaction design involving our bodies? I will discuss how Shusterman's theories of somaesthetics might provide some inspiration, and the need to focus on actuation rather than sensing.

## BRIEF BIOGRAPHY

Kristina Höök is a professor in Interaction Design at the Royal Institute of Technology and also works part-time at SICS (Swedish Institute of Computer Science). She is the director of the Mobile Life centre. Höök has published numerous journal papers, books and book chapters, and conference papers in highly renowned venues. A frequent keynote speaker, she is known for her work on social navigation, seamfulness, mobile services, affective interaction and lately, designing for bodily engagement in interaction through somaesthetics. Her competence lies mainly in interaction design and user studies helping to form design. She has obtained numerous national and international grants, awards, and fellowships including the Cor Baayen Fellowship by ERCIM (European Research Consortium for Informatics and Mathematics), the INGVAR award and she is an ACM Distinguished Scientist. She has been listed as one of the 50 most influential IT-women in Sweden every year since 2008. She is an elected member of Royal Swedish Academy of Engineering Sciences (IVA).



Höök, K.
Designing for Somaesthetic Experiences - Focusing on Actuation Rather than Sensing?.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, page 9
ISBN: 978-989-758-213-4

# An Unobtrusive System to Measure, Assess, and Predict Cognitive Workload in Real-World Environments

Bethany Bracken
*Charles River Analytics Inc., U.S.A.*

Abstract: Across many careers, individuals face alternating periods of high and low attention and cognitive workload, which can result in impaired cognitive functioning and can be detrimental to job performance. For example, some professions (e.g., fire fighters, emergency medical personnel, doctors and nurses working in an emergency room, pilots) require long periods of low workload (boredom), followed by sudden, high-tempo operations during which they may be required to respond to an emergency and perform at peak cognitive levels. Conversely, other professions (e.g., air traffic controllers, market investors in financial industries, analysts) require long periods of high workload and multitasking during which the addition of just one more task results in cognitive overload resulting in mistakes.An unobtrusive system to measure, assess, and predict cognitive workload could warn individuals, their teammates, or their supervisors when steps should be taken to augment cognitive readiness. In this talk I will describe an approach to this problem that we have found to be successful across work domains includes: (1) a suite of unobtrusive, field-ready neurophysiological, physiological, and behavioral sensors that are chosen to best suit the target environment; (2) custom algorithms and statistical techniques to process and time-align raw data originating from the sensor suite; (3) probabilistic and statistical models designed to interpret the data into the human state of interest (e.g., cognitive workload, attention, fatigue); (4) and machine-learning techniques to predict upcoming performance based on the current pattern of events, and (5) display of each piece of information depending on the needs of the target user who may or may not want to drill down into the functioning of the system to determine how conclusions about human state and performance are determined. I will then focus in on our experimental results from our custom functional near-infrared spectroscopy sensor, designed to operate in real-world environments to be worn comfortably (e.g., positioned into a baseball cap or a surgeon's cap) to measure changes in brain blood oxygenation without adding burden to the individual being assessed.

## BRIEF BIOGRAPHY

Dr. Bethany Bracken, is a Senior Scientist at Charles River Analytics. Throughout her career, Dr. Bracken has used a variety of behavioral, physiological, cognitive, molecular, and neuroimaging methodologies in both humans and animals to answer questions about the neurobiology of behavior. At Charles River, she currently works on projects using neurophysiological and physiological sensing methods to assess human states such as stress, focused attention, and cognitive workload and to predict upcoming performance deficits to allow time to enact augmentation strategies to optimize that performance. Dr. Bracken has a B.S. in Psychology from Clarion University of Pennsylvania, and a Ph.D. in Neuroscience from Brandeis University. Before joining Charles River Analytics, Dr. Bracken completed a postdoctoral fellowship, quickly followed with a promotion to the faculty level, in the department of Psychiatry at McLean Hospital and Harvard Medical School.



Bracken, B.
An Unobtrusive System to Measure, Assess, and Predict Cognitive Workload in Real-World Environments.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, page 11
ISBN: 978-989-758-213-4

# PAPERS

# Full Papers

# Are Trustworthy Health Videos Reachable on YouTube?
## *A Study of YouTube Ranking of Diabetes Health Videos*

Randi Karlsen[1], José Enrique Borrás Morell[1] and Vicente Traver Salcedo[2]
[1]*Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway*
[2]*ITACA - Health and Wellbeing Technologies, Universidad Politècnica de València, Valencia, Spain*
*randi.karlsen@uit.no, jobormo@gmail.com, vtraver@itaca.upv.es*

Keywords: Consumer Health Information, Information Retrieval, Social Networks, YouTube, Health Video Retrieval, Ranking Evaluation, Diabetes, Personal Health.

Abstract: While health consumers are increasingly searching health information on the Internet, information overload is a serious obstacle for finding relevant and good-quality information among inaccurate, obsolete or incorrect health information. While a lot of information exists, information from credible sources, such as hospitals and health organisations, may be difficult to find. The aim of this study is to analyse ranking of diabetes health videos on YouTube over a time period, to learn whether videos from credible sources are ranked sufficiently high to be reachable to users. 19 diabetes-related queries were issued to YouTube each day over a 1.5-month period, and in total 2584 videos from credible sources was detected and their ranking position tracked. We found that only a small number of the tracked videos were in practice available to the user, as most videos were given a persistent low ranking. Also, since ranking is fairly stable, users cannot expect to find many new videos (from credible sources) when issuing a query multiple times. We conclude that new tools are needed that enable health video retrieval based on requirements concerning not only relevance and popularity, but also credibility of the sources and trustworthiness of the videos.

## 1 INTRODUCTION

Internet has, during the last years, become a major source of health information (AlGhamdi and Moussa, 2012; Griffiths et al., 2012; Madathil et al., 2015). Users are typically searching for information about specific diseases or symptoms, treatment side effects, second opinions, complementary or alternative medicines, search for others with similar health concerns and follow personal health experiences through blogs (de Boer et al., 2007; Diaz et al., 2002; Fox, 2011b; Powell et al., 2011). Also, online health information is used, not only by health consumers to gain knowledge about some health issue, but also by physicians, for clinical decision support and for education purposes (Hughes et al., 2009).

However, a general problem when searching the Internet, is the information overload and difficulty of finding relevant information satisfying the information need. Adding to this problem, too many websites have inaccurate, missing, obsolete, incorrect, biased or misleading information, and it may be difficult to distinguish between trustworthy and specious information (Briones et al., 2012; Madathil et al., 2015; Pant et al., 2012; Shabbir et al., 2013; Singh et al., 2012; Steinberg et al., 2010). When people are relying on online health information to take medical decisions or handle their health issues, it is obviously of highest importance that the health information provided to users is not only relevant, but also correct and trustworthy. Existing search engines select and rank information based on relevance to a search query and popularity. Evaluating quality aspects, such as reliability and validity of information, is currently left to the user. Thus, the overwhelming amount of health information together with the mixed quality, makes it difficult for users to identify good-quality health information on the Internet, especially when users are not familiar with new technologies or when their health knowledge is limited. Also, certification approaches, such as the ethical HON code, are not solving the issue (Diaz et al., 2002).

Health information on the Internet comes from different sources, including hospitals, health organisations, government, educational institutions, for-profit actors and private persons reporting on personal experiences with some disease. User studies have shown that the credibility of an information source is



Karlsen, R., Morell, J. and Salcedo, V.
Are Trustworthy Health Videos Reachable on YouTube? - A Study of YouTube Ranking of Diabetes Health Videos.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 17-25
ISBN: 978-989-758-213-4

one of the most powerful factors affecting information credibility (Freeman and Spyridakis, 2009). Users are for example more likely to trust health information published or authored by physicians or major health institutions than information provided by other sources (Dutta-Bergman, 2003; Moturu et al., 2008; Bermúdez-Tamayo et al., 2013). Such studies show that users show greater interest in health information published by professional sources, such as hospitals and health organisations, since these are considered more credible than the average health information on the Internet.

In our study we focus on health information provided through videos on YouTube and investigate to what extent health videos from professional sources, such as hospitals and health organisations, are available to the user. YouTube is today the most important video-sharing website on the Internet (Cheng et al., 2008). It has over a billion users (almost one-third of all people on the Internet) and every day people watch hundreds of millions of hours on YouTube and generate billions of views (YouTube, 2016). YouTube social media tools allow users to easily upload, view and share videos, and enable interaction by letting users rate videos and post comments.

YouTube is increasingly being used to share health information offered by a variety of sources (channels), including hospitals, organisations, government, companies and private users (Bennett, 2011). However, it may be difficult to find videos from credible channels, since YouTube video ranking is known to favour content from popular channels. This may cause for instance hospital videos, where social interaction through likes/dislikes and comments are not so common, to appear low in the ranked list. Also, YouTube ranking does not focus on trustworthiness, and both misleading and incorrect videos may well be popular and may therefore be given a high ranking (Briones et al., 2012; Shabbir et al., 2013).

A considerable amount of literature has been published on YouTube data analysis, such as studying relations between video ratings and their comments (Yee et al., 2009) or focusing on the social networking aspect of YouTube and social features (Cheng et al., 2008; Chelaru et al., 2012). Studies of YouTube performance have mainly focused on YouTube in general, rather than on specific domains, such as health. However, there have recently been some studies evaluating YouTube health video content with respect to their quality of information for patient education and professional training (Gabarron et al., 2013; Topps et al., 2013). Such studies, focusing on different areas of medicine, include the work of (Briones et al., 2012; Singh et al., 2012; Steinberg et al., 2010; Butler et al., 2013; Schreiber et al., 2013; Murugiah et al., 2011; Fat et al., 2011; Azer et al., 2013). In these studies, reviewers evaluate the quality or content of selected videos, and assess their usefulness as information source within their respective area.

This paper reports on a study where we tracked diabetes health videos on YouTube over a period of 1.5 month, to gain knowledge on how videos from professional channels are ranked on YouTube. The study was intended to answer the following questions: *"Where are videos from hospitals and health organisations ranked on YouTube?" "Are these videos ranked in positions that make them reachable to users?"* To the best of our knowledge, there has previously not been conducted a study where the availability of YouTube health videos has been tracked over time, as was done in our work.

The structure of the paper is the following. The next section presents the methodology used in our study. Section 3 presents the results of the work, while findings are discussed in Section 4. Section 5 concludes.

## 2 METHOD

This study is based on health videos obtained from YouTube through textual search queries on diabetes-related issues. We set up a test environment, where 19 diabetes-related queries were issued to YouTube each day over a period of 1.5 months, from March until April 2013. During this period, we daily collected the top 500 YouTube results for each query. Videos from white-listed (presumably credible) sources were identified and tracked during each day of the study, and their ranking position registered.

We implemented a system that for each day automatically issued the 19 queries and extracted information about the top 500 YouTube results. In addition to ranking position, we collected information such as video name and identifier, channel identifier, number of likes, dislikes and comments to the video. All 19 queries included the term "diabetes" and were focused towards different aspects concerning the disease. We used queries such as "diabetes a1", "diabetes glucose", "diabetes hyperglycemia" and "diabetes lada", and issued them as regular search queries on the YouTube home page using an anonymous profile (to avoid any bias) and with language option set to English. Video ranking was obtained by parsing the html of the result page, while video and channel information were collected through YouTube API version 2.0. All search queries can be seen in Table 1.

Through our study of YouTube health videos, we

Table 1: List of You Tube search queries.

| | | |
|---|---|---|
| diabetes type 1 | diabetes hyperglycemia | diabetes insulin |
| diabetes type 2 | diabetes hypoglycemia | diabetes injection |
| diabetes a1c | diabetes complications | diabetes glucose |
| diabetes food | diabetes retinopathy | diabetes mellitus |
| diabetes diet | diabetes ketoacidosis | diabetes education |
| diabetes obese | diabetes insulin pump | |
| diabetes lada | diabetes monitoring | |

identified a number of (assumed) credible health video sources, such as hospitals and health organisations. We organised these channels into a hospital white-list and a health organisations white-list, containing channel identifiers for hospitals and health organisations respectively. In the light of user-interests in peer-to-peer healthcare (Ziebland and Herxheimer, 2008; Fox, 2011a), we also generated a third white-list of channels, which includes users that are active and predominantly publishing diabetes videos. Our white-lists contained a total of 699 channels, where 651 were hospitals, 30 were organisations and 18 were active users. We used the Health Care Social Media List started by Ed Bennett (Bennett, 2011) as an initial white-list, and expanded with more channels that we identified during our studies (Karlsen et al., 2013; Morell et al., 2012).

# 3 RESULTS

Using the 19 search terms shown in Table 1, we tracked the rank position of a total of 2584 YouTube health videos from white-listed channels during the test period. The videos were uploaded from 73 hospital channels, 30 organisation channels and 18 user channels. Among these, 2372 videos were uploaded to YouTube before the study began, whereas 212 videos were uploaded while the study was performed.

For each day of the study, our system detected a number of new videos from white-listed channels (for which tracking started and continued to the end of the study). The number of new videos was large in the first days of the study, and after some days stabilised at around 10 new videos each day.

## 3.1 Ranking of Videos from White-listed Channels

A goal of this study is to identify the number of videos from hospitals, health organisations and active users that are in practice available to users. When a YouTube search returns over 600.000 ranked videos (which is the case for the "diabetes type 1" search), it is obvious that the lowest ranked videos are not very available. A question is: "How far down in the ranked list of videos is a user willing to browse in order to find a relevant video?" The answer may to some extent be a matter of how patient the user is, but testing several hundred videos are beyond what can be expected from an average user.

To characterise videos w.r.t availability, we have grouped the tracked videos using ranking position intervals that were chosen based on our perception of how available videos in the different groups are. We consider videos ranked in position 1-40 as *highly available*, position 41-100 as *available*, position 101-200 as *not very available* and position 201-500 as *almost unavailable*. In this work, we assume that videos ranked lower than position 500, are in practice unavailable, and we have therefore tracked only videos appearing in the top-500 ranking.

To learn where videos from hospitals, health organisations and active users were ranked, we examined, for each day, the rank positions for all videos from our white lists, and determined the number of videos that were ranked in position intervals (1-40), (41-100), (101-200), and (201-500). Based on this study, we found that only a small number of videos from white-listed channels were in practice available to the user. When examining the top-40 ranked videos, we found that on average, only 3.2% were from hospitals, 10.4% from health organisations and 3.6% from active users. This means that we on average will retrieve approximately 7 videos from white-listed channels among the top-40 ranked videos. In the next position interval (41-100), the average number of videos from white-listed channels will be approximately 6. The results for all rank intervals are seen in Figure 1. The numbers for the top-500 videos (not given in Figure 1) were 2.3% from hospitals, 6.4% from health organisations and 1.8% from active users.

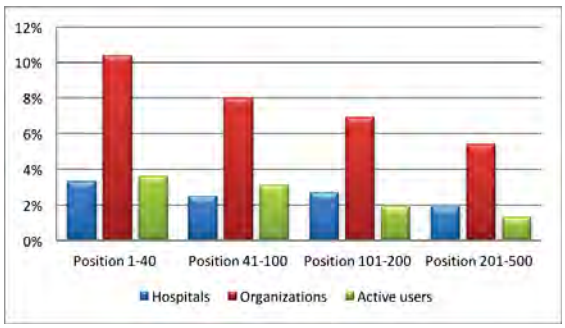



Figure 1: Average number (in percentage) of videos from white-listed channels within different rank intervals.

As our white-lists obviously do not contain every hospital or health organisation available, we took, for all 19 queries, the top-100 YouTube hits from one

Table 2: Classification of the tracked videos from white-listed channels, grouped according to their most frequent ranking position during the test period. The number of videos in each rank-position group are given (percentage between parentheses).

| | Group A videos (pos. 1-40) | Group B videos (pos. 41-100) | Group C videos (pos. 101-200) | Group D videos (pos. 201-500) | Group E videos (pos. $>$ 500) | Total no. of videos |
|---|---|---|---|---|---|---|
| Hospitals | 26 (1.0%) | 27 (1.1%) | 52 (2.0%) | 99 (3.8%) | 334 (12.9%) | 538 (20.8%) |
| Organizations | 80 (3.1%) | 89 (3.4%) | 127 (4.9%) | 279(10.8%) | 1017 (39.4%) | 1592(61.6%) |
| Active users | 30 (1.2%) | 33 (1.3%) | 33 (1.3%) | 55 (2.1%) | 303 (11.7%) | 454 (17.6%) |
| Total | 136(5.3%) | 149(5.8%) | 212(8.2%) | 433(16.7%) | 1654(64.0%) | 2584 |

day's search and manually checked the channel of each video. We found 15 new videos from hospitals not included in the white-list. This addition represents a modest 0.8% (of the 1900 videos), and does not significantly improve the amount of hospital videos given a high ranking.

## 3.2 Changes in Ranking Position

To investigate variations in video ranking, we first classified the tracked videos into five rank-position groups (Group A-E) according to their most frequent rank position during the test period. Occasional changes in ranking position were registered by counting the number of videos from each group that one or more days had a ranking position associated with a different group. We also calculated mean position and standard deviation for videos that were ranked within a top-500 position the whole period.

Among the 2584 videos from white-listed channels, we found 136 videos (5.3%) (26 hospital, 80 organisation and 30 user videos) that most frequently appeared in the top-40 results, while 1654 videos (64%) only occasionally appeared among the top-500 results. These were classified as Group A and Group E videos, respectively. Table 2 shows the number of videos within each of the five rank-position groups.

*Occasional change in ranking* is presented in Table 3, showing the proportion of videos classified as Group A-E that occasionally (i.e. one or more days) appeared in a different rank-position interval. For example, 38.2% of Group A videos appeared occasionally in position 41-100, while 14% occasionally appeared in position 101-200. We found that only a small proportion of videos from Group C, D and E were occasionally given a high-ranked position within 1 - 40 (11.8%, 2.5% and 4.4% respectively).

These results indicate that most of the videos appearing in low rank positions, are stuck in low positions, and will consequently remain (almost) out of reach to the user.

*Mean position and standard deviation (SD)* were calculated for the 175 videos that were ranked within top-500 the whole test period. Table 4 shows mean position and standard deviation for videos, grouped according to their most frequent rank position (i.e. rank-position group).

Table 3: Proportion of Group A-E videos (classified according to their most frequent ranking position) that occasionally changed ranking position to a different position interval.

| | Group A | Group B | Group C | Group D | Group E |
|---|---|---|---|---|---|
| Occ. pos 1-40 | - | 35.6% | 11.8% | 2.5% | 4.4% |
| Occ. pos 41-100 | 38.2% | - | 48.6% | 11.3% | 9.6% |
| Occ. pos 101-200 | 14.0% | 54.4% | - | 39.7% | 19.6% |
| Occ. pos 201-500 | 6.6% | 26.8% | 59.0% | - | 81.4% |
| Occ. pos $>$ 500 | 27.9% | 63.1% | 76.9% | 82.7% | - |

Table 4: Mean position and standard deviation for videos in different rank-position groups. Including the 175 videos that were ranked within top-500 the whole test period.

| | No. of videos | Mean position | Standard deviation |
|---|---|---|---|
| Group A | 71 | 15.5 | 7.0 |
| Group B | 43 | 71.0 | 20.9 |
| Group C | 37 | 145.9 | 38.2 |
| Group D | 24 | 272.3 | 48.0 |

We found that the highest ranked videos (Group A) had the lowest standard deviation, i.e. 7.0. These videos seemed to be established in a top-ranked position, and had in general less variation in rank-position than videos from other groups. In fact, stability in rank position seemed to be the case for all groups, even though the standard deviation for Group B, C and D is higher.

SD-values indicate that changes in rank position in general do not make Group D videos more accessible to users, while Group C and B videos may occasionally be given a more accessible ranking. As an example, take Group B videos having a mean position of 71 and an SD value of 20.9. This means that most

videos (about 68%, assuming a normal distribution) were ranked within position 50-92. Approximately 15% of the videos occasionally had a position within top-40, while approximately 15% were occasionally not included in the top-100. For Group C videos, less than 0.5% of the videos would occasionally have a position within the top-40, while approximately 10% would occasionally have a position within top-100.

The rank stability observed through these numbers, indicates that highly ranked videos remain available to users, while low ranked videos will almost always remain out of reach for the user.

## 3.3 Relevance of Videos

One could suspect that videos given a low-ranked position were not relevant to the query. To investigate this, we selected two queries ("diabetes hyperglycemia" and "diabetes retinopathy") and determined relevance of each tracked video by manually comparing keywords in the query to video title and description, and by watching the video to compare video content to query.

Over the test period, the system tracked 130 videos for the "diabetes hyperglycemia" query and 64 videos for the "diabetes retinopathy" query. Table 5 shows the number of videos that were i) relevant to the query, ii) relevant to diabetes in general and iii) not relevant to diabetes. For example, for Group E videos of the "diabetes hyperglycemia" query, we found that 50% were relevant to the query, an additional 47% were relevant to diabetes, while only 3% were not relevant to diabetes. For the "diabetes retinopathy" query, 55% of Group E videos were relevant to the query, an additional 18% were relevant to diabetes, while 27% were not relevant. For Group A and B videos (of both queries), every video was relevant to the query. In conclusion, we found that a large number of low-ranked videos were relevant to the query, implying that lack of relevance could not be the reason for their low ranking.

## 3.4 Video Properties

To detect possible correlations between video properties and ranking position, we compared video title and query terms, investigated social interaction by counting for each video the number of likes, dislikes, comments and views, and subsequently compared against the video's ranking position.

Having a *match between query terms and video title* is obviously an important criterion for considering the video relevant to the query. We found for Group A videos that 88% (120 of 136 videos) had a perfect match between video title and query (meaning that all terms in the query were found in the video title). The proportion of videos with a perfect query-title match was lower in the other groups, but there were still a large number of lower ranked videos that had a perfect query-title match. This shows that such a match is not sufficient for a high-ranked position.

Table 5: The number of videos relevant to i) the search query and ii) diabetes in general, for the two queries "diabetes hyperglycemia" and "diabetes retinopathy".

| Rank position | relevant to | Diabetes hyperglicemia | Diabetes retinopathy |
|---|---|---|---|
| Group A videos | query | 100% | 100% |
| | diabetes | - | - |
| Group B videos | query | 100% | 100% |
| | diabetes | - | - |
| Group C videos | query | 45% | 100% |
| | diabetes | 55% | - |
| Group D videos | query | 41% | 72% |
| | diabetes | 55% | 11% |
| | not relev | 4% | 17% |
| Group E videos | query | 50% | 55% |
| | diabetes | 47% | 18% |
| | not relev | 3% | 27% |

The average number of *likes/dislikes and comments* for Group A-E videos are displayed in Figure 2. The general trend was that the highest ranked videos had the highest number of social interactions. This coincides well with previous studies, which found that very few videos get the users' attention. This can be explained through the Yule process (or rich-get-richer principle), as the videos that appear in the first page are more likely to be viewed and interacted (Chelaru et al., 2012; Cha et al., 2009).

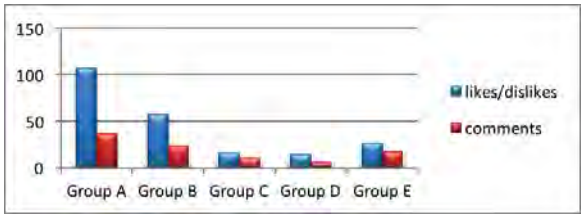



Figure 2: Average number of likes/dislikes and comments on videos.

However, when studying individual videos, we observed huge differences in the number of user interactions. We found for example that a number of videos without likes/dislikes and comments were highly ranked despite the lack of user activity. Table 6 shows the percentage of videos, within each rank-position group, that had zero likes, dislikes and comments. We see for example that 20% of Group A videos had no such social interaction.

Table 6: Number of videos without user interaction through likes/dislikes and comments.

| | Videos tracked all period (175 videos) | All videos (2584 videos) |
|---|---|---|
| Group A | 21% | 20% |
| Group B | 35% | 29% |
| Group C | 43% | 33% |
| Group D | 33% | 39% |
| Group E | - | 28% |

When examining the number of *views* for individual videos, we found a close correlation between views and ranking (see Figure 3). This seems obvious since users can easily find and access highly ranked videos, which then get a higher number of views compared to low ranked videos. However, there were also a few exceptions. For instance, one Group A video had only 18 views, zero likes, dislikes and comments.

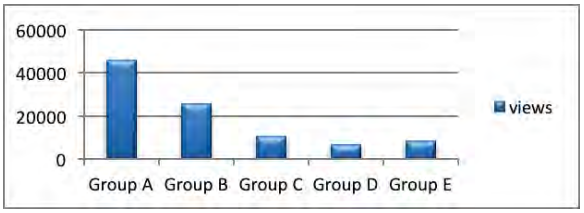



Figure 3: Average number of times a video has been viewed.

### 3.5 Videos From Non-white-listed Channels

To get an impression of the type of videos not tracked in our study, we also examined properties of videos published by non-white-listed channels. Because of the large number of videos, we restricted this investigation to the top-50 results of two queries: "diabetes type 1" and "diabetes injection". For all top-50 videos we manually examined relevance to query and channel type.

Relevance and channel type for the examined videos are shown in Table 7 and Table 8, respectively. Videos from both white-listed (WL) and non-white-listed channels are included in the tables. Among the non-white-listed videos, most came from private users (47 videos), while only 3 videos were from health organisations/centres. We further observed that white-listed channels had the highest proportion of relevant videos.

Table 7: Relevance of videos published by non-white-listed and white-listed channels.

| Relevant to | Diabetes type 1 | | Diabetes injection | |
|---|---|---|---|---|
| | non-WL-channels 35 videos | WL-channels 13 videos | non-WL-channels 38 videos | WL-channels 11 videos |
| query | 32 (91%) | 13 (100%) | 19 (50%) | 10 (91%) |
| diabetes | 3 (9%) | 0 | 9 (24%) | 1 (9%) |
| not relev. | 0 | 0 | 10 (26%) | 0 |

Table 8: Channels of videos published by non-white-listed and white-listed channels.

| | Diabetes type 1 | | Diabetes injection | |
|---|---|---|---|---|
| | non-WL-channels 35 videos | WL-channels 13 videos | non-WL-channels 38 videos | WL-channels 11 videos |
| Hospital | 0 | 3 (23%) | 0 | 0 |
| Health organization | 1 (3%) | 5 (38.5%) | 1 (2%) | 8 (73%) |
| Active users | 0 | 5 (38.5%) | 0 | 3 (27%) |
| Health center | 1 (3%) | - | 0 | - |
| Company | 8 (23%) | - | 5 (13%) | - |
| Private users | 24 (68%) | - | 23 (61%) | - |
| Others | 1 (3%) | - | 9 (24%) | - |

## 4 DISCUSSION

The goal of this study was to identify the number of videos from hospitals, health organisations and active user channels that are in practice available to users. On the positive side, the study shows that for each query, videos from white-listed channels are in fact available among the top-500 ranked videos. A problem, however, is that these videos represent a small proportion of the total number of retrieved videos, and that many of them are found in low ranked positions that make them in practice beyond reach for the user issuing the query. Thus, precision for videos from white-listed channels is not very good. Among the top-100 ranked videos for a diabetes related query, one can on average expect 15% to be from white-listed channels (2.8% from hospitals, 8.9% from health organisations and 3.3% from active users).

Of the 2584 tracked videos, 64% were most frequently ranked in a position lower than 500, only occasionally appearing within the top-500 results. This shows that many relevant videos from credible channels are in practice unreachable for users. Also, standard deviation values and observed ranking variations for individual videos show that the ranking of videos is fairly stable. This implies that only a small per-

centage of low ranked videos improved their ranking position sufficiently to be available to users and that users hardly obtain any new videos (from white-listed channels) by issuing a query multiple times. On the other hand, ranking stability also guarantees that top-ranked videos from white-listed channels are available to users over a period of time. This benefits new users that will have access to a few popular and potentially good quality health videos from credible channels.

One conclusion from our study is therefore that relevant diabetes-related health videos are available on YouTube, but too few are given a ranking that make them reachable for the user.

The YouTube ranking algorithm is based on video popularity among users. Previously the algorithm was based on view count of the video, while the current version (since 2012) is based on Watch Time, which is the amount of time on aggregate that viewers watch a video (Robertson, 2014). Even though Watch Time is a better measure of success for a video (since it rewards engaging videos that keep viewers watching), it still requires videos to be available to users in order to get sufficient attention and improve Watch Time. Also, there is no guarantee that an engaging, much watched video is trustworthy with respect to the health information it provides.

In our study, the investigation of correlation between ranking position and user attention in the form of social interactions, gave mixed results. There were on average a higher number of social interactions (i.e. likes/dislikes, comments and views) for the highest ranked videos, but we also saw many examples of videos that had a high-ranked position with no social interaction and very few views.

A critical factor in identifying relevant videos based on a textual query, is the accuracy of the metadata with respect to video content. When examining the correlation between video title and query terms, we found that a majority (88%) of the highest ranked videos (Group A videos), but also a large number of low ranked videos, had a perfect match between video title and query terms. However, by inspection, we also found many video descriptions that were very short and of such a general nature that they did not describe the video content. Video titles were also in many cases inaccurate with respect to video content.

An implication of these findings is that video publishers should make an effort in providing precise textual description of videos, where video title and description matches the video content as accurately as possible. This is a simple way of improving the likelihood for being selected as relevant and possibly ranked sufficiently high to be reachable. Allowing and even encouraging social interaction on videos may also help visibility of the video.

However, an accurate video title/description is only a step in the right direction for improving video rank position and precision. We believe there is a need for new video retrieval tools that not only focus on relevance and popularity as it is done today, but also retrieves health information based on requirements for credibility of the sources and trustworthiness of the videos. This provides topics for future research.

Some limitations to our work should be noted. Firstly, even though our white-lists of hospitals and health organisations include a large number of channels, they cannot include every hospital and health organisation available. The focus was not to track every relevant and trustworthy video in the result set from YouTube, but rather to track videos from specific channel types that are assumed to be of interest to health consumers. Also, it should be noted that the quality of each video was not assessed in this study. We base the study on the assumption that videos from hospitals, health organisations and also active users are of interest and therefore worthwhile investigating. We are fully aware that videos from other channels (not tracked in our study) may provide useful and trustworthy information. Furthermore, for each query we only examined the top-500 ranked videos from YouTube. When some queries return over 600.000 videos, this is a small number. However, we believe that a position over 500 is not significant in terms of availability to users.

# 5 CONCLUSION

To gain knowledge about how health videos are ranked on YouTube, we have tracked diabetes health videos on YouTube every day over a period of 1.5 month. We focused on videos published by credible channels, such as hospitals, health organisations and users actively publishing diabetes-related videos. Our findings show that most videos from these channels are given a persistent low ranking that makes them in practice unavailable to users. Additionally, since ranking position of videos is fairly stable, users receive the same videos over and over again if issuing a query multiple times. Thus, users may find it difficult to obtain new information from YouTube. A conclusion from this work is that research is needed to provide users with new tools that enable health video retrieval based on requirements concerning not only relevance and popularity, but also credibility of the sources and trustworthiness of the videos. Mechanisms for alternative ranking or less stable ranking

could also be useful for making a larger number of relevant videos available to the user.

## ACKNOWLEDGEMENTS

The authors appreciate support from UiT The Arctic University of Troms through funding from Troms Research Foundation, and the important contributions of the ITACA-TSB Universitat Politècnica de València group.

## REFERENCES

AlGhamdi, K. M. and Moussa, N. A. (2012). Internet use by the public to search for health-related information. *International journal of medical informatics*, 81(6):363–373.

Azer, S. A., AlGrain, H. A., AlKhelaif, R. A., and AlEshaiwi, S. M. (2013). Evaluation of the educational value of youtube videos about physical examination of the cardiovascular and respiratory systems. *Journal of medical Internet research*, 15(11):e241.

Bennett, E. (2011). Social media and hospitals: From trendy to essential. *Futurescan 2001: Healthcare Trends and Implications 2011-2016*, pages 43–48.

Bermúdez-Tamayo, C., Alba-Ruiz, R., Jiménez-Pernett, J., García Gutiérrez, J.-F., Traver-Salcedo, V., and Yubraham-Sánchez, D. (2013). Use of social media by spanish hospitals: perceptions, difficulties, and success factors. *Telemedicine and e-Health*, 19(2):137–145.

Briones, R., Nan, X., Madden, K., and Waks, L. (2012). When vaccines go viral: an analysis of hpv vaccine coverage on youtube. *Health communication*, 27(5):478–485.

Butler, D. P., Perry, F., Shah, Z., and Leon-Villapalos, J. (2013). The quality of video information on burn first aid available on youtube. *Burns*, 39(5):856–859.

Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.-Y., and Moon, S. (2009). Analyzing the video popularity characteristics of large-scale user generated content systems. *IEEE/ACM Transactions on Networking (TON)*, 17(5):1357–1370.

Chelaru, S. V., Orellana-Rodriguez, C., and Altingovde, I. S. (2012). Can social features help learning to rank youtube videos? In *International Conference on Web Information Systems Engineering*, pages 552–566. Springer.

Cheng, X., Dale, C., and Liu, J. (2008). Statistics and social network of youtube videos. In *Quality of Service, 2008. IWQoS 2008. 16th International Workshop on*, pages 229–238. IEEE.

de Boer, M. J., Versteegen, G. J., and van Wijhe, M. (2007). Patients use of the internet for pain-related medical information. *Patient education and counseling*, 68(1):86–97.

Diaz, J. A., Griffith, R. A., Ng, J. J., Reinert, S. E., Friedmann, P. D., and Moulton, A. W. (2002). Patients' use of the internet for medical information. *Journal of general internal medicine*, 17(3):180–185.

Dutta-Bergman, M. (2003). Trusted online sources of health information: differences in demographics, health beliefs, and health-information orientation. *Journal of medical Internet research*, 5(3):e21.

Fat, M. J. L., Doja, A., Barrowman, N., and Sell, E. (2011). Youtube videos as a teaching tool and patient resource for infantile spasms. *Journal of Child Neurology*, 26(7):804–809.

Fox, S. (2011a). *Peer-to-peer healthcare*. Pew Internet & American Life Project.

Fox, S. (2011b). *The social life of health information, 2011*. Pew Internet & American Life Project Washington, DC.

Freeman, K. S. and Spyridakis, J. H. (2009). Effect of contact information on the credibility of online health information. *IEEE Transactions on Professional Communication*, 52(2):152–166.

Gabarron, E., Fernandez-Luque, L., Armayones, M., and Lau, A. Y. (2013). Identifying measures used for assessing quality of youtube videos with patient health information: a review of current literature. *Interactive Journal of Medical Research*, 2(1):e6.

Griffiths, F., Cave, J., Boardman, F., Ren, J., Pawlikowska, T., Ball, R., Clarke, A., and Cohen, A. (2012). Social networks–the future for health care delivery. *Social science & medicine*, 75(12):2233–2241.

Hughes, B., Joshi, I., Lemonde, H., and Wareham, J. (2009). Junior physician's use of web 2.0 for information seeking and medical education: a qualitative study. *International journal of medical informatics*, 78(10):645–655.

Karlsen, R., Morell, J. E. B., Fernández-Luque, L., and Salcedo, V. T. (2013). A domain-based approach for retrieving trustworthy health videos from youtube. In *MedInfo*, page 1008.

Madathil, K. C., Rivera-Rodriguez, A. J., Greenstein, J. S., and Gramopadhye, A. K. (2015). Healthcare information on youtube: a systematic review. *Health informatics journal*, 21(3):173–194.

Morell, J. E. B., Karlsen, R., Salcedo, V. T., and Luque, L. F. (2012). Analysis of hospital youtube videos as health reliable sources. *IVWorkshop on on Technology for Healthcare and Healthy Lifestyle*.

Moturu, S. T., Liu, H., and Johnson, W. G. (2008). Trust evaluation in health information on the world wide web. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1525–1528. IEEE.

Murugiah, K., Vallakati, A., Rajput, K., Sood, A., and Challa, N. R. (2011). Youtube as a source of information on cardiopulmonary resuscitation. *Resuscitation*, 82(3):332–334.

Pant, S., Deshmukh, A., Murugiah, K., Kumar, G., Sachdeva, R., and Mehta, J. L. (2012). Assessing the credibility of the youtube approach to health informa-

tion on acute myocardial infarction. *Clinical cardiology*, 35(5):281–285.

Powell, J., Inglis, N., Ronnie, J., and Large, S. (2011). The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. *Journal of Medical Internet Research*, 13(1):e20.

Robertson, M. R. (2014). Watch time: A guide to youtubes top search ranking factor. *Internet: http://www. reelseo. com/youtube-watch-time/, Einsicht*, 8:2015.

Schreiber, J. J., Warren, R. F., Hotchkiss, R. N., and Daluiski, A. (2013). An online video investigation into the mechanism of elbow dislocation. *The Journal of hand surgery*, 38(3):488–494.

Shabbir, S. A., Fernandez-Luque, L., Jian, W.-S., Li, Y.-C., Crain, S., Hsu, M. H., Wang, Y. C., Khandregzen, D., Chuluunbaatar, E., Nguyen, P. A., et al. (2013). Misleading health-related information promoted through video-based social media. *Journal of Medical Internet Research*, 15(2).

Singh, A. G., Singh, S., and Singh, P. P. (2012). Youtube for information on rheumatoid arthritisa wakeup call? *The Journal of rheumatology*, 39(5):899–903.

Steinberg, P. L., Wason, S., Stern, J. M., Deters, L., Kowal, B., and Seigne, J. (2010). Youtube as source of prostate cancer information. *Urology*, 75(3):619–622.

Topps, D., Helmer, J., and Ellaway, R. (2013). Youtube as a platform for publishing clinical skills training videos. *Academic Medicine*, 88(2):192–197.

Yee, W. G., Yates, A., Liu, S., and Frieder, O. (2009). Are web user comments useful for search. *Proc. LSDS-IR*, pages 63–70.

YouTube (2016). *Statistics*. https://www.youtube.com/yt/press/statistics.html, Accessed: 2016-09-26.

Ziebland, S. and Herxheimer, A. (2008). How patients experiences contribute to decision making: illustrations from dipex (personal experiences of health and illness). *Journal of nursing management*, 16(4):433–439.

# MonAT: A Visual Web-based Tool to Profile Health Data Quality

Monica Noselli[1], Dan Mason[2], Mohammed A. Mohammed[2,3] and Roy A. Ruddle[1]
[1]*Department of Computing, University of Leeds, EC Stoner Building, LS2 9JT, Leeds, U.K.*
[2]*Bradford Institute for Health Research, Bradford Royal Infirmary, Bradford, U.K.*
[3]*Department of Research, University of Bradford, Bradford, U.K.*
{*scmno, r.a.ruddle*}*@leeds.ac.uk, dan.mason@bthft.nhs.uk, m.a.mohammed5@bradford.ac.uk*

Keywords: Data Quality, Visualization, Health Data, Longitudinal Data.

Abstract: Electronic Health Records (EHRs) are an important asset for clinical research and decision making, but the utility of EHR data depends on its quality. In health, quality is typically investigated by using statistical methods to profile data. To complement established methods, we developed a web-based visualisation tool called MonAT Web Application (MonAT) for profiling the completeness and correctness of EHR. The tool was evaluated by four researchers using anthropometric data from the Born in Bradford Project (BiB Project), and this highlighted three advantages. The first was to understand how missingness varied across variables, and especially to do this for subsets of records. The second was to investigate whether certain variables for groups of records were sufficiently complete to be used in subsequent analysis. The third was to portray longitudinally the records for a given person, to improve outlier identification.

## 1 INTRODUCTION

Clinical patient data are stored digitally in EHRs and contains a wide range of information. These data are becoming a powerful resource for secondary uses such as investigating and developing decision support systems (Coorevits et al., 2013). Clinical research results and decisions depend on the quality of EHR data.

There are three different aspects of data quality assessment: *Data profiling* defines issues in data quality, *Data wrangling* prepares the data for further investigation by transforming it into a required structure, and *Data cleansing* analyses and corrects the data (Gschwandtner et al., 2014). This paper focuses on data profiling, for which most health researchers typically adopt a statistically driven workflow and make little use of interactive data visualization.

The aim of our research is to provide new methods for profiling health data. This paper describes the design and formative evaluation of a new interactive visualization tool called MonAT, which provides users with rich functionality for data profiling, leveraging human cognitive capabilities.

## 2 RELATED RESEARCH

New frameworks and tools have been created to define and assess data quality (Stausberg et al., 2015). A new framework to assess health data quality (Weiskopf and Weng, 2013) provides a definition of data quality in five dimensions: completeness, correctness, concordance, plausibility and currency. In a related work (Weiskopf et al., 2013), the authors demonstrated that the completeness depends on the type of tasks. Moreover, completeness and correctness are considered the important dimensions to assess first since the others depend on the quality of these (Dungey et al., 2014).

A data quality ontology has been defined (Johnson et al., 2015) based on the above data quality definition (Weiskopf and Weng, 2013). The ontology describes concepts and measures of data quality (Table 1 and 2).

Another aspect of the completeness is the *Missing Data Mechanisms* (Rubin, 1976). Data are missing for several reasons. Understanding the reasons for missingness is important for evaluating the rest of the data and for generating accurate results (Nakagawa and Freckleton, 2008), (Farhangfar et al., 2008).

Visualization techniques are powerful tools for knowledge discovery. In the work (West et al., 2014), the authors state that “EHR data are complicated by missing values, inaccurate data entry, and mixed data



Noselli, M., Mason, D., Mohammed, M. and Ruddle, R.
MonAT: A Visual Web-based Tool to Profile Health Data Quality.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 26-34
ISBN: 978-989-758-213-4

Table 1: Concept and Description of (Johnson et al., 2015)'s ontology for Completeness Dimension.

| | Concept | Description |
|---|---|---|
| a | DomainCoverage | Data represents Domain Concepts |
| b | DomainComplete | Presence or absence of expected information |
| c | RepresentationComplete | Data is not missed without consider the Domain |
| d | Sufficiency | Longitudinal data is sufficient to investigate a given Task |
| e | Relevance | Data complies the Domain and is sufficient to investigate a given Task |
| f | TaskCoverage | Variables used to investigate a given Task appears in the data |
| g | Flexibility | Data is sufficient to investigate different Tasks |

Table 2: Concept and Description of (Johnson et al., 2015)'s ontology for Correctness Dimension.

| | Concept | Description |
|---|---|---|
| h | RepresentationIntegrity | Data is not corrupted, no data entry errors |
| i | Reliability | The data can be used because it is correct |
| l | RepresentationCorrectness | The data is complete and accurate |

types that must be considered in developing visualization techniques" and "users wants to see both categorical and numerical data when interactively exploring the data, and they like to look at the detail in the record".

There is a huge amount of data quality investigations specifically in health care and medical research (Stausberg et al., 2015) but tackling data quality assessment with a visual analytic perspective is quite novel (Kohlhammer et al., 2011).

Two examples of visual analytic tools are TimeCleanser (Gschwandtner et al., 2014) and Profiler (Kandel et al., 2012). TimeCleanser is a visual analytics system that helps to clean time-oriented data. Profiler adopts visualization and statistical analysis for data quality assessment. The cleansing job is driven by summaries and automated detection of anomalies and then visualizes and interacts with the results.

MonAT aims to help experts detect and flag completeness and correctness data quality issues by providing interactive and connected visualizations.

We designed MonAT to cater to the (Weiskopf and Weng, 2013)'s data quality definitions and considering the (Johnson et al., 2015)'s ontology (Table 1 and Table 2). MonAT offers a visualization that can help to classify the data in the (Rubin, 1976)'s categories and considers the longitudinal characteristic of health data.

# 3 BORN IN BRADFORD

Born in Bradford (BiB) is a longitudinal multi-ethnic birth cohort study aiming to examine the impact of environmental, psychological and genetic factors on maternal and child health and wellbeing (Raynor, 2008). Bradford is a city in the North of England with high levels of socio-economic deprivation and ethnic diversity. Women were recruited at the Bradford Royal Infirmary at 26-28 weeks gestation. The full BiB cohort recruited 12,453 women comprising 13,776 pregnancies between 2007 and 2010 and the cohort is broadly characteristic of the city's maternal population. The mothers and children are followed up longitudinally through direct surveys and linkage to routine healthcare records. Ethical approval for the data collection was granted by the Bradford Research Ethics Committee (Ref 07/H1302/112).

We obtained access to a subset of the BiB children's growth variables (Wright et al., 2013) in two flat files. The first file contains 500 records with children's raw observations (Table 3). There are 360 children. The greatest number of records per child is 21. The greatest number of height observations per child is 3. The greatest number of weight observations per child is 20. On average, there are 0.51 height observations and 1.38 weight observations per child. The second file contains the same information cleaned by statisticians following an established cleansing method (see Section 3.1).

## 3.1 Data Cleansing Process

We met several times with BiB statisticians. The aim of the meetings was to understand their cleansing method, detect deficiencies, and establish requirements for our tool.

The BiB method uses STATA and goes through the following steps to clean height and weight variables:

- Implausible values. Weight observations greater than 25kgs are removed.

Table 3: Variables in the anthropometric data set that was used to evaluate MonAT.

| Variables | | | |
|---|---|---|---|
| Name | Type | Description | #Categories |
| ChildID | Encoding | Child's ID to identify longitudinal data | 360 |
| age | Numerical (Integer) | Child's age calculate in days | |
| weight | Numerical (Float) | Child's weight calculated in kilograms | |
| height | Numerical (Float) | Child's height calculated in centimeter | |
| gender | Categorical | Child's sex | 2 |
| eth0ethall | Categorical | Child's ethnicity category | 22 |
| eth0eth9gp | | | 9 |
| eth0ethgrp | | | 6 |
| ethgrp4 | | | 4 |
| eth0eth3gp | | | 3 |
| source | Categorical | Information about the data provenance | 3 |

- Outliers. Data are divided by baby's sex and mother's ethnicity group. Four Standard Deviation (4SD) from mean is calculated separately for the groups over intervals of two months. Values greater than +4SD and smaller than -4SD are considered as outliers and deleted from the data set.
- Negative Increment. Weight and height observations are checked by comparing two consecutive numerical observations and deleting the second observation if it does not follow certain criteria. Weight observations are also compared with the subsequent measurement. Weight values between the first two weeks of age with a drop of more than 10% and weight and height values with a drop of more than 5% in the other intervals are removed.
- Plot. Any further outlier is analyzed by plotting weight and height against age by sex and the three ethnicity categories.
- Combine height and weight observations. Another plot shows weight against height to check for further outliers.
- Ponderal Index (a ratio of a person's height to weight) is calculated and plotted against age by sex and ethnicity to check for further outliers.

### 3.2 Limitations

Our analysis highlighted three key limitations in the BiB data cleansing process. First, the process does not consider missing data at all, which prevents feedback being given to data providers and places the onus for investigating missingness solely on the users of BiB data.

Second, the process only cleans a subset of the variables that are in a given data set. For example, BiB only cleaned the weight and height variables in the anthropometric dataset. There may also be data quality issues with the age and ethnicity variables. An assumption that underlies categorical variables such as ethnicity is that they do not change over the time. However, a child's ethnicity may not be present for each data record, or ethnicity may be recorded differently over time, indicating that some data are not consistent.

Third, numerical longitudinal variables (weight and height, in the case of the anthropometric data set) are only cleaned in a pair-wise fashion, rather as a whole. The sequence of points in a plot defines the overall weight (or height) profile of a child. Inspection of individual points in the context of the children's overall profile can help researchers to identify outliers. Such profiles are even more useful for investigating negative increments, because a visual check can identify which point is suspicious rather than always assuming that the second observation in a pair is wrong, as is the case with the current BiB process.

## 4 MonAT

This section list the requirements and describes the design and implementation of novel visualization tool, MonAT, which allows researchers to profile data by combining basic computations and human cognitive capabilities.

### 4.1 Requirements

Based on the limitations discussed in 3.2, we identify two key requirements for our tool.

*Profiling Completeness*. A normal statistical approach to deal with missingness is to remove records containing missing data. However, this can lead to biased results. Missing data profiling (Table 1) can help

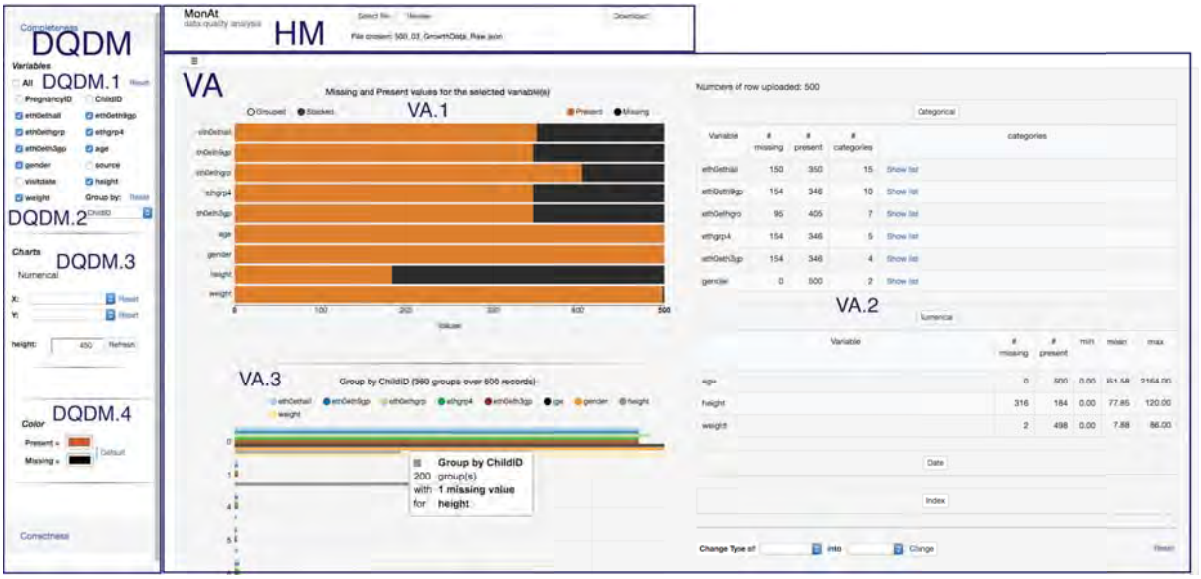



Figure 1: Screen shot showing the three main areas of MonAT. The *Horizontal Menu* (*HM*) is on the top. The *Data Quality Dimension Menu* (*DQDM*) on the left shows the Completeness sub-menu expanded. It contains the four filters: the *List of Variables Filter* (DQDM.1), the *Group Filter* (DQDM.2), the *Chart Filter* (DQDM.3), and the *Color Filter* (DQDM.4). The *Visualization Area* (VA) in the middle-right-bottom shows *Present-Missing Overview* bar chart (VA.1), *Tables overview* (VA.2), and *Grouped Present-Missing Data* (VA.3).

to mitigate bias when researchers are defining data inclusion criteria.

*Profiling Correctness*. Abnormal observations can lead to misleading findings. Outliers can occur in categorical variables due to erroneous free input text or data processing errors. In numeric variables outliers can be due to mistyping, measurement errors, or data processing errors. When observations are related to each other, longitudinal data profiling can improve outlier identification.

Other, general requirements were as follows. First, a tabular view showing summary statistics (e.g., minimum, maximun and mean of numerical observations, and the number of categories in categorical variables) is important to provide an overview of data. Second, MonAT needs to be accessible to researchers without specialist knowledge of statistical tools and programming languages (e.g., STATA and R). Third, the interface should be user friendly and intuitive, and support workflows that are easy to learn and follow. Fourth, researchers should find that MonAT provides added value for data profiling.

## 4.2 Design and Implementation

MonAT is a web-base tool implemented using Django Rest Framework, AngularJS, and Krispo's nvd3 (D3) reusable charts.

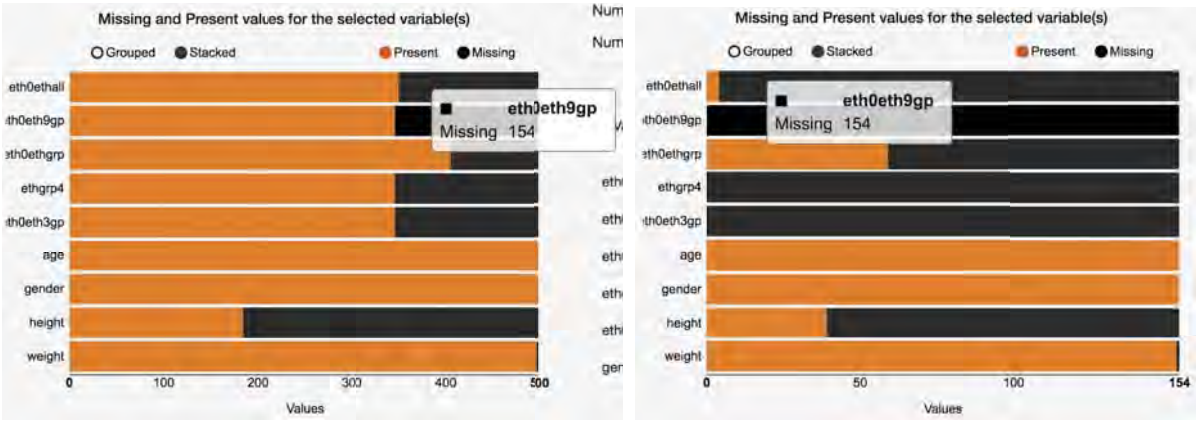
The architecture of MonAT follows the concepts of Single Page Application (SAP) and Model View Control (MVC). MonAT web interface (Figure 1) is divided into three main areas: the *Horizontal Menu* (*HM*), the *Data Quality Dimension Menu* (*DQDM*), and the *Visualization Area* (*VA*).

The *HM* contains a function to select a local flat file. Statistical information and some additional variables for managing page layout are automatically calculated.

This menu stores the data in local memory in JSON format object. The type of variable (numerical, categorical, index or date), number of present and missed observations per each variable are precalculated and stored in the JSON object. For numerical variables, minimum, mean, and maximum are calculated. For categorical variables, a list of categories is created. The scope of all this information is the root: the variables can be accessed from every component. Some additional variables are added with a root scope to allow flexible layouts to be created.

The *DQDM* is a dynamic menu that maps the dimensions of data quality (Weiskopf and Weng, 2013) to sub-menus. The application automatically creates this menu when user upload a file. Currently there are two sub-menus - one for 'Completeness' and one for 'Correctness'.

The *VA* is the core of MonAT, and where it shows visualization and tabular output in response to users' interactions. We designed different *VA* components for profiling completeness and correctness.




(a) (b)

Figure 2: The *Present-Missing Overview* bar chart, showing the number of present and missing observations (orange and black bars, respectively) for nine variables. The plots show missingness for: (a) all 500 records in the data set, and (b) the 154 records where the 'eth0eth9grp variable is missing (selected by clicking on the black bar for that variable in (a)).

#### 4.2.1 Completeness

Interactions with the Completeness *DQDM* creates *VA* components to explore different levels of missingness (Table 1).

The checkbox *List of Variables Filter* (Figure 1 - DQDM.1) represents the Domain of the data (Table 1.a). It allows users to understand what tasks can be answered with the set of variables (Table 1.f).

Interactions with the checkbox list create a *Present-Missing Overview* (Figure 1 - VA.1), and *Tables Overview* (Figure 1 - VA.2).

The *Present-Missing Overview*, a grouped/stacked interactive bar chart, shows missing and present values, calculated over all the entries (Table 1.b). It allows analysis of data quantity (Table 1.c and Figure 2).

The *Tables Overview* shows four tables, one for each type of variable (numerical, categorical, index, and date) with the number of present and missing observations, categories and statistical information.

The *Group Filter* (Figure 1 - DQDM.2) is a drop-down menu for selecting a variable to group the data. It creates the *Grouped Present-Missing Data* (Figure 1 - VA.3), a grouped bar chart, which shows information about number of groups that miss variables selected in the *List of Variables Filter* (Figure 1 - DQDM.1). The *y* axis represents the number of missing observations and the bar length represents a variable. Each barsize represents the number of groups missing a variable. It is useful for inclusion criteria giving information of sufficiency and relevant observations (Table 1.b and 1.e).

The *Chart Filter* (Figure 1 - DQDM.3) is a combination of two drop-down menu allowing selection of two numerical variables that create a plot (Figure 3.a). The two variables are respectively the *x* and *y* axes of the plot. The filter automatically creates a contextual menu (Figure 3.b), a series of checkboxes that allows users to visualize *Data in Context* and *Distributions*.

The *Data in Context* (Figure 3.a) can be analyzed to evaluate sufficiency, task coverage and variable flexibility (Table 1.d, 1.f, and 1.g) by interacting with the contextual menu (Figure 3.b) and *Logic Function Filter* (3.c).

The *Logic Function Filter* allows investigation of variables' missingness with 'AND' and 'OR' operations.

The *Distribution* (Figure 3.d and 3.e) shows the number of missing values of selected variable, in respect of the two axes of the scatterplot to investigate correlation of missingness between variables. It can be used to evaluate the *Missing Data Mechanisms*.

#### 4.2.2 Correctness

Users may analyze correctness by interacting with the Correctness *DQDM*. The set of filters is similar to the Completeness *DQDM*, but the *VA* components change.

Interactions with the *Distribution Filter*, a checkbox list of variables, shows a boxplot for each selected variable, allowing users to detect outliers for numerical values (points above or below the upper and lower whiskers) (Table 2.h and 2.i).

The *Chart Filter* creates a *Data in Context* view to let users detect outliers and analyze negative increment on longitudinal data. It helps to analyze correctness and suitability of the data for a given task (Table 2.i and Table 2.l). It can also be used to spot variation

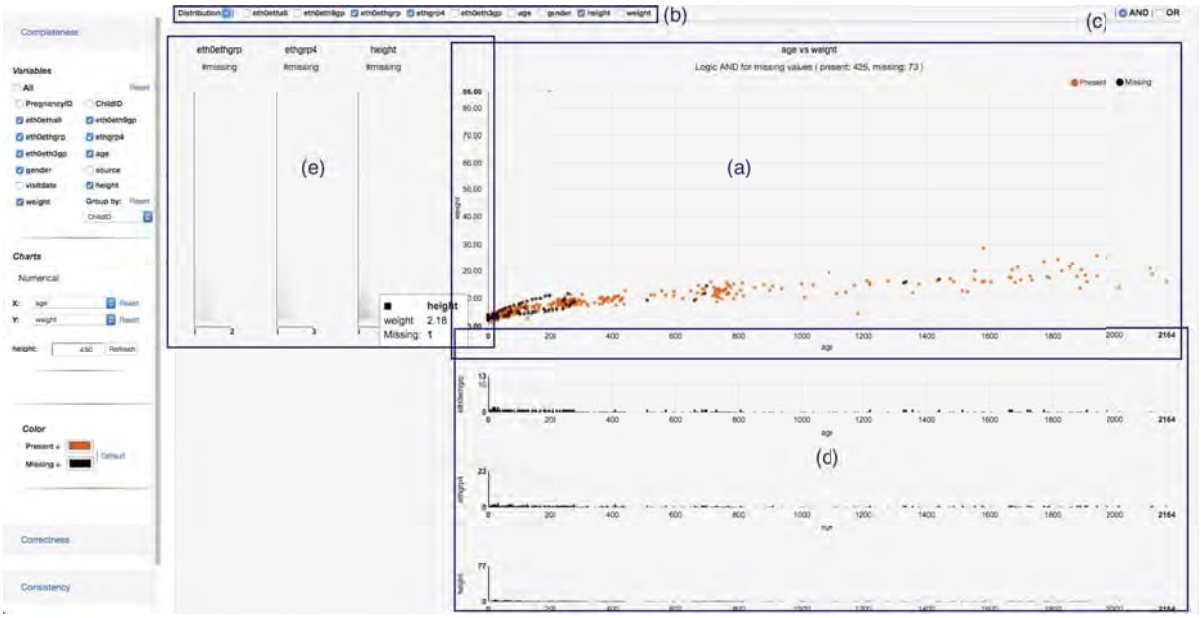



Figure 3: The *VA* shows *Data in Context*. A contextual menu (b) allows show missing and present observations of selected variables (eth0thgrp, ethgrp4, and height) in the age against weight scatterplot (a) with an *AND* operation (*Logic Function Filter* c). The contextual menu also contains a check box button to show/hide *Distributions* (d, e) of the selected variables.

in categories (i.e., variation of ethnicity category for a given child) (Table 2.h).

# 5 MonAT EVALUATION

This section describes a formative evaluation of MonAT, using the BiB Project anthropometric data set. We carried out a formative evaluation to list a minimum set of functions that end-users would benefit while using the tool.

The aims of the evaluation were to assess the value of MonAT's existing functionality for investigating data completeness and correctness, and to identify important additional functionality for the tool.

## 5.1 Method

This section presents information on recruited participants, materials and procedure for the evaluation.

### 5.1.1 Participants

We recruited four participants aged between 28 and 51. One of them had worked at the Bradford Institute for Health Research (BIHR) for two years and was knowledgeable about the BiB Data Collection (BiB DC). Two of them were post-doctoral researchers, with experience of visual analytics and data analysis, respectively. The fourth participant was a PhD student with experience in data analysis.

### 5.1.2 Materials

A set of exercises were created with the aim to highlight issues in the tool and identify new functions for the release version of MonAT. The exercises tested the main functions of Completeness (4.2.1) and Correctness (4.2.2) dimensions. Examples of the Visualization Components are shown in Figures 1 and 3.

### 5.1.3 Procedure

At the start of the evaluation, participants were given an information sheet about the aims of the evaluation, a description of the data set, and a brief description of the MonAT functionality. Then participants were asked to think aloud while they completed some data completeness and correctness exercises. If they were uncertain about anything, participants were encouraged to ask the experimenter questions during the exercises. At the end of the exercises the experimenter conducted a semi-structured interview to gain further feedback about MonAT and identify additional functionality that participants thought would be beneficial.

The first set of exercises required the use of MonAT to investigate the completeness functionality.

Participants started by selecting a subset of variables listed in the completeness *DQDM* and evaluated the *Present-Missing Overview* and the *Tables*

*Overview*. Then, they used the *Group Filter* to investigate missingness in the groups.

The next steps were to plot a chart, select some variables in the *Contextual Menu* and evaluate the differences between visualizing values in *logic AND* and *logic OR*. Finally, they visualized the *Distribution* charts.

The second set of exercises focused on the correctness functionality. Participants interacted with the Correctness *DQDM* to dynamically create the visualizations and detect outliers.

## 5.2 Results

Overall, participants were impressed by the novel capabilities of MonAT. This section reports participants main comments and suggestions.

Participants found the *Present-Missing Overview* (Figure 1 VA.1) useful for understanding how missingness varied across variables, and especially having the possibility to select a subset of the records. For example, the chart shows that there is a similar amount of missing data for the ethnicity variables (first five black bars). By selecting one of these bars, the participant can explore that similarity, and an example is shown in Figure 2. Participants preferred stacked bars for that rather than grouped bars.

The participants liked the *Tables Overview* (Figure 1 VA.2). However they recommended that it was shown in a separate panel to leave the focus on the visual chart.

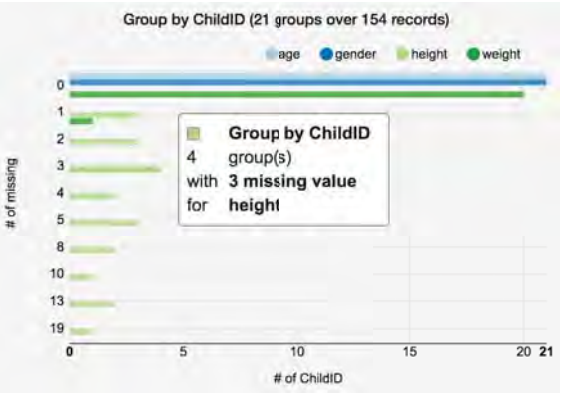



Figure 4: The x axis shows the number of groups having missing data for selected variables (age, gender, height, weight). In this example there are 21 children (groups) over 154 records. The y axis shows the number of missing values per variable. In this example, there are four children missing exactly three observations for the height. A mouse over on bar displays a tooltip that shows related information.

The *Grouped Present-Missing Data* (Figure 1 VA.3 and Figure 4) shows the number of missing observations for selected variables (in this case eth0eth9gp and height), grouped by the value of another variable (in this case 'childID'). Participants initially found it difficult to understand the meaning of this visualization but, once they did, they considered it to be useful for investigating whether groups had sufficient values for a given task (Table 1.d and 1.e).

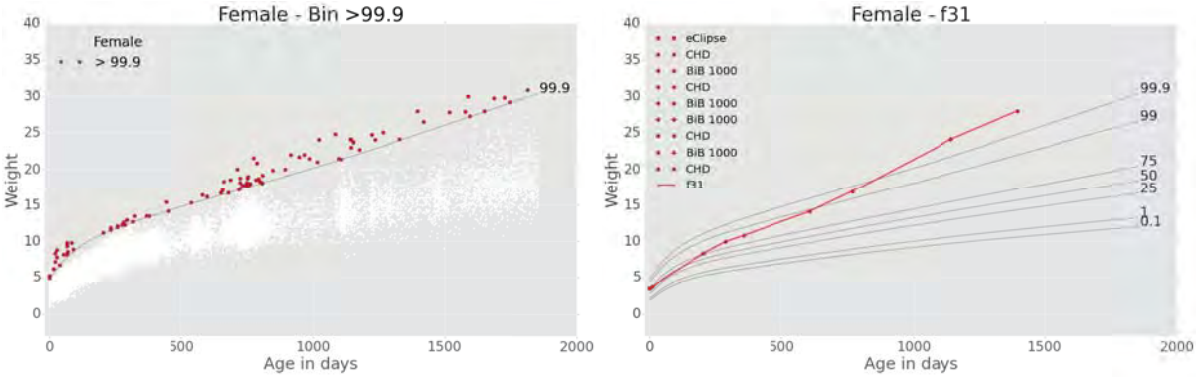
Participants suggested being able to interact with bars in the *Grouped Present-Missing Data* (as it is for the *Present-Missing Overview*) to show frequency distributions of missing values. They also said they would like to be able to select more than one group. For example to include in the further analysis groups having no missing values or not more than one value missed for a given variable.

The scatterplot (Figures 3) shows missing and present *Data in Context*. The contextual menu (at the top of the Figure 3) allows users to include variables in the scatterplot. Participants found it useful to visualize how categorical missing observations relate to the scatterplot variables. For example (Figure 3a) the three selected variables (eth0ethgr, ethgrp4 and height) shown in AND, present missing data mostly in early children days (from 0 to 300). Participants considered it useful to be able to 'AND' or 'OR' the selected variables, and suggested that 'XOR' (only one of the selected variables is missing) would also be useful.

Participants made a number of other comments, which were as follows. The *Distributions* that are shown in the bar charts of Figure 3 are difficult to see, which could be addressed by binning the data to reduce the number of bars that are drawn.

Referring to the correctness exercises, the scatterplot and the scatterplot with lines (Figure 5) show *Data in Context* to reveal outliers. Investigating the data as single points may lead to some of them being defined as outliers because they lie above the 99.9th percentile (Figure 5a). However, a longitudinal visualization indicates that some of those points are correct because they are from a single child and follow a reasonable curve (Figure 5b). Participants found these scatterplots useful to visualize the longitudinal data to explore correctness. However, to improve the legibility of a plot they suggested the use of bins to visualize a subset of the data. The number of groups (children) should be low to avoid confusion. Moreover, the longitudinal groups should be smartly selected to avoid overlaps in the visualization.

Overall, participants found MonAT useful for profiling the quality of the data. They suggested adding 'print' and 'download' function for each visualization, so the visualizations can be easily included in presentations and reports. Moreover, the tool can be




(a) (b)

Figure 5: *Data in Context* for outlier identification, showing: (a) all of the height data for female children (the red points lie above the 99.9th percentile), and (b) the height profile of a single child (the red line and points).

used to create a subset of data set that satisfies inclusion criteria for further investigations.

## 6 CONCLUSIONS

This paper describes the design and formative evaluation of a novel tool called MonAT for visual data profiling. The tool was developed in conjunction with researchers from the BiB Project, and designed to address limitations in the statistical data profiling methods that are commonly used with EHR.

MonAT involves automatic and human analysis to compute the complex task of profiling different dimensions of data quality of medical data. The tool is novel in providing multiple and connected interactive visualizations to investigate completeness and correctness in longitudinal data such as EHRs. It answers the need of profiling data for different tasks and different dimensions of data quality.

The tool was positively received, and provided three advantages as a complement to statistical methods. These were to understand how missingness varied across variables, investigate whether certain variables were sufficiently complete to be used in subsequent analysis, and improve outlier identification.

A revised version of MonAT is being developed and will be deployed for use in the BiB Project. Further functionality is planned, to combine interactive visualisation with a data mining capability.

## ACKNOWLEDGEMENTS

Born in Bradford (BiB) is only possible because of the enthusiasm and commitment of the Children and Parents in BiB. We are grateful to all the participants, health professionals and researchers who have made BiB happen.

We gratefully acknowledge the contribution of TPP and the TPP ResearchOne team in completing study participant matching to GP primary care records and in providing ongoing informatics support.

This research was funded by the Bradford Institute for Health Research.

## REFERENCES

Coorevits, P., Sundgren, M., Klein, G. O., Bahr, A., Claerhout, B., Daniel, C., Dugas, M., Dupont, D., Schmidt, A., Singleton, P., De Moor, G., and Kalra, D. (2013). Electronic Health Records: New Opportunities for Clinical Research. *Journal of Internal Medicine*, 274:547–560.

Dungey, S., Beloff, N., Puri, S., Boggon, R., Williams, T., and Rosemary, A. (2014). A Pragmatic Approach for Measuring Data Quality in Primary Care Databases. In *Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on*, pages 797–800.

Farhangfar, A., Kurgan, L., and Dy, J. (2008). Impact of Imputation of Missing Values on Classification Error for Discrete Data. *Pattern Recognition*, 41(12):3692–3705.

Gschwandtner, T., Aigner, W., Miksch, S., Gärtner, J., Kriglstein, S., Pohl, M., and Suchy, N. (2014). Time-cleanser: A visual analytics approach for data cleansing of time-oriented data. In *Proceedings of the 14th international conference on knowledge technologies and data-driven business*, page 18. ACM.

Johnson, S. G., Speedie, S., Simon, G., Kumar, V., and Westra, B. L. (2015). A Data Quality Ontology for the Secondary Use of EHR Data. *AMIA 2015 Annual Symposium Proceedings*, pages 1937–1946.

Kandel, S., Parikh, R., Paepcke, A., Hellerstein, J. M., and Heer, J. (2012). Profiler : Integrated Statistical Analysis and Visualization for Data Quality Assessment. *Proceedings of Advanced Visual Interfaces, AVI*, pages 547–554.

Kohlhammer, J., Keim, D., Pohl, M., Santucci, G., and Andrienko, G. (2011). Solving Problems With Visual Analytics. In *Procedia Computer Science*, volume 7, pages 117–120.

Nakagawa, S. and Freckleton, R. P. (2008). Missing Inaction: the Dangers of Ignoring Missing Data. *Trends in Ecology and Evolution*, 23(11):592–596.

Raynor, P. (2008). Born in Bradford, a Cohort Study of Babies Born in Bradford, and their Parents: Protocol for the Recruitment Phase. *BMC public health*, 8:327.

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, 63:581–592.

Stausberg, J., Nasseh, D., and Nonnemacher, M. (2015). Measuring Data Quality: A Review of the Literature between 2005 and 2013. *Studies in health technology and informatics*, 210:712–6.

Weiskopf, N. G., Hripcsak, G., Swaminathan, S., and Weng, C. (2013). Defining and Measuring Completeness of Electronic Health Records for Secondary Use. *Journal of Biomedical Informatics*, 46:830–836.

Weiskopf, N. G. and Weng, C. (2013). Methods and Dimensions of Electronic Health Record Data Quality Assessment: Enabling Reuse for Clinical Research. *Journal of the American Medical Informatics Association : JAMIA*, 20:144–51.

West, V. L., Borland, D., and Hammond, W. E. (2014). Innovative Information Visualization of Electronic Health Record Data: a Systematic Review. *Journal of the American Medical Informatics Association : JAMIA*, pages 1–7.

Wright, J., Small, N., Raynor, P., Tuffnell, D., Bhopal, R., Cameron, N., Fairley, L., A Lawlor, D., Parslow, R., Petherick, E. S., Pickett, K. E., Waiblinger, D., and West, J. (2013). Cohort profile: The Born in Bradford Multi-Ethnic Family Cohort Study. *International Journal of Epidemiology*, 42:978–991.

# Comparison of Different Implementations of a Process Limiting Pharmaceutical Expenditures Required by German Law

Timo Emcke[1], Thomas Ostermann[2], Marc Heidbreder[3] and Reinhard Schuster[4]
[1]*Chair of Department of Prescription Analysis, Association of Statutory Health Insurance Physicians, Bismarckallee 1-6, 23812 Bad Segeberg, Germany*
[2]*Chair of Research Methodology and Statistics in Psychology, Witten/Herdecke University, Alfred-Herrhausen-Straße 50, 58448 Witten, Germany*
[3]*Medical Advisory board of Statutory Health Insurance in Northern Germany, Katharinenstraße 11, 23554 Lübeck, Germany*
[4]*Chair of Department of Health Economics, Epidemiology and Medical Informatics, Medical Advisory Board of Statutory Health Insurance in Northern Germany, Katharinenstraße 11, 23554 Lübeck, Germany*
*timo.emcke@kvsh.de, thomas.ostermann@uni-wh.de, {reinhard.schuster, marc.heidbreder}@mdk-nord.de*

Keywords: Controlling Measures, Outpatient Care, Morbidity Adjusted Benchmark, Prescription Data, MRG, Big Data.

Abstract: German legislation demands controlling measures for outpatient drug costs. As of 2017 the health insurance actors rose to a challenge to reform the benchmark system on the federal state level. We look at the previous system applied until 2015, the improvements in 2016 and the method the regional parties agree on for 2017. After discussing hard- and software systems and the underlying data briefly we describe the flaws of the old approach and develop a general model for controlling measures in the outpatient field. Finally we present the first real world applications of the new model: a patient type classification system leading to target costs and a derived distance structure of physicians regarding their prescription behaviour.

## 1 INTRODUCTION

In Europe and especially in Germany rising pharmaceutical expenditures put the health service at risk. Every modern health care system has to ensure the quality and equity of care while keeping the cost down. Therefore controlling measures were established by the German legislation as early as in 1993. Since 1995 this is subject to regional negotiations between Statutory Health Insurances (SHI) and SHI associated physicians. This type of regulation aims to limit expenditures per patient without restricting the necessary treatment.

Of the exiting two types of instruments, the first one puts German patients/cases in certain more or less morbidity related cost groups, the other promotes or restricts drug classes with different economic characteristics but same curative effects. We will look at those using health insurance data of the German Federal State Schleswig-Holstein in 2015.

In the years from 1995 till 2015 physician groups got three different treatment case budgets for each insurance status defined by statutory health insurance (member [M], dependent coverage [D] and retired [R]). Some regions merged status [M] and [D]. Several expensive drug substances and pharmaceuticals regulated by treatment regimen are excluded resulting in internal inconsistencies and uncertainties regarding all participating players.

Budgets are calculated using expenditure shares for the mentioned case groups per physician group in a reference period (last year) and the negotiated target volume of expenditure for the resent year.

In December 2013 the social welfare court of Dresden passed the sentence that guide values/budgets have to be based on age groups. Additionally the Federal Social Court judged that authorities have an obligation to regulate atypical drug prescriptions. As an immediate consequence regarding the budget calculation for 2016 four age groups superseded insurance status: 0-15, 16-49, 50-64 and 65 and above. Those groups, utilized in all statutory health insurances, have a very poor age resolution for this field of application in general.

From 2017 on, the federal legislator made regional negotiated far-reaching reforms of controlling measures possible (Versorgungsstärkungsgesetz = Supply Support Act). A new system developed in this context is expenditure controlling by Morbidity Related Groups (MRG). MRG is an adaption of



Emcke, T., Ostermann, T., Heidbreder, M. and Schuster, R.
Comparison of Different Implementations of a Process Limiting Pharmaceutical Expenditures Required by German Law.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 35-40
ISBN: 978-989-758-213-4

the Diagnosis-Related-Group-System (DRG) used for classification and compensation of hospital cases and put into effect in 2003 by German legislation. It is based on similar systems elsewhere: Since the first use for hospital payment in the United States in the early 1980s, DRG-type systems have become the main method of hospital payment in the majority of OECD countries. The German version (G-DRG) is based on the Australian DRG-system (AR-DRG).

Hereinafter we will compare the systems based on insurance status, age groups and MRG, including some new results for MRG.

## 2 MATERIAL AND METHODS

For comparing the previous and the new controlling measures we analyze detailed prescription data of the Statutory Health Insurance in Schleswig-Holstein of quarter two in 2015. There's no benefit using annual data due to the stable prescription framework. The data on the prescription level are combined with master data containing drug classes (ATC [anatomic-therapeutic-chemical] with some additions for products not classified), physician groups and drugs to be excluded. Treatment cases of the Association of Statutory Health Insurance Physicians are also added. Obtaining results required the processing of large amounts of data.

The hardware used is a dedicated Debian GNU/Linux Server administered by the Medical Advisory board of Statutory Health Insurance in Northern Germany also used to generate consultation materials from the same data.

It runs a LAMP configuration (Debian GNU/Linux, Apache 2.4, MYSQL Community Edition 5.7 [extensive use of partitioning] and PHP 7 [with PEAR framework esp. for spreadsheet output]). The inexpensive open source/free software setup makes the cooperation of different administrative bodies possible. The coding was done using the Perl programming language.

The previous model used till 2016 applies prescription data, treatment cases and status defined by statutory health insurance/age groups. The implementation is straight forward. Treatment cases in a certain age/status group get their share of the negotiated volume of expenditure based on the development of last years' expenditures and treatment cases.

The new MRG-model requires prescription data, the ATC classification and physician group information depending on the model configuration. It can be defined as follows:

$B =$ set of physicians/practices

$F =$ set of physician groups

There is a transformation mapping physicians/practices to groups: $f = f(b)$ while splitting up practices containing different physician groups.

$P(b)$ = patients of $b \in B$, $b = b(p)$ is the mapping of patients and physicians whereas the transformation $D = D(p)$ maps patients $p \in P(b)$ to the prescribed drugs. Multiple prescriptions of one drug are counted repeatedly. $o(d)$ is a quantity factor for $d \in D$ representing the ration of package size of the prescription drug in relation to the biggest one available. A pharmaceutical classification system (e.g. ATC4) as transformation: $a = a(p), a \in A$ used identification of similar medicinal products. The drugs $d \in D$ are linked to costs by the cost function: $k = k(d), k \in \mathbf{R}, k > 0$.The age of the patient is defined by: $t = t(p)$ in five-yearly stages. A MRG is a pair $d = (c,s)$ [c:basic MRG, s:degree of severity] with $c \in A, s \in \mathbf{Z}, 0 \le s \le 9$.

$$\text{Cost per ATC} = \bar{k}(p,a^*) = \sum_{d \in D(p), a(p)=a^*} k(d)$$

ATC with the highest costs = basic MRG is characterized by $\bar{k}(p,c) \ge \bar{k}(p,a)$ for all $a \in A$. In case of the occurrence of several $c_i$ the lexicographically dominating element is chosen. $c = c(p)$ is the transformation to determine patients basic MRG. Number of ATC4 groups per patient (multimedication) is defined as:

$$v(p) = \#\{a \in A : \bar{k}(p,a) > 0\}.$$

The number of prescriptions for patient $p \in P$ assigned to basic MRG $c(p)$ is represented by:

$$\bar{o}(p) = w\Big(\sum_{d \in D(p): a(d)=c(p)} o(d)\Big) \text{ with}$$

$$w(x) = \begin{cases} x, \text{ if } x \in \mathbf{Z} \\ \lfloor x \rfloor, \text{ if } x \notin \mathbf{Z}. \end{cases}$$

We define threshold values for subgroups:

$$(v_0, \cdots, v_9) = (0.5, 0.75, 1.25, 1.5, 2.0, 2.5, 5, 10)$$

$i(v) = i$ is true if $v_i \le v < v_i + 1$. $m(X)$ shall be the mean of $x \in X$. The costs of basic MRG $c \in A$ in the physician group are defined as:

$$k_1^*(c^*, f^*) = m_{p \in P: f(b(p))=f^*, c(p)=c^*}\Big(\bar{k}(p,c^*)\Big)$$

and adding the age dimension the term changes to:

$$k_1^*(c^*, f^*, t^*) = m_{p \in P: f(b(p))=f^*, c(p)=c^*, t(p)=t^*}\Big(\bar{k}(p,c^*)\Big)$$

whereby the age related severity is given by:

$$i_1(c,f,t) = i\Big(k_1^*(c,f,t)/k^*(c,f)\Big).$$

Costs differentiated by multimorbidity are expressed by the formula:

$k_2^*(c^*,f^*,j^*) = m_{p\in P: f(b(p))=f^*, c(p)=c^*, v(p)=j^*}\left(\bar{k}(p,c^*)\right)$
with the corresponding degree of severity:

$$i_2(c,f,j) = i\Big(k_2^*(c,f,j)/k^*(c,f)\Big).$$

The same can be done by looking at prescription intensity:

$k_3^*(c^*,f^*,j^*) = m_{p\in P: f(b(p))=f^*, c(p)=c^*, \bar{o}(p)=j^*}\left(\bar{k}(p,c^*)\right)$

$$i_3(c,f,j) = i\Big(k_3^*(c,f,j)/k^*(c,f)\Big).$$

Total degree of severity is given by:

$i(p) = max\Big(i_1(c(p),f(b(p),t(p)),i_2(c(p)),f(b(p)),v(p)),$

$i_3(c(p)),f(b(p),\bar{o}(p))\Big).$

The MRG including severity levels is recalculated with respect to physician groups:

$k_g^*(c^*,f^*,j^*) = m_{p\in P: f(b(p))=f^*, c(p)=c^*, i(p)=j^*}\left(\bar{k}(p,c^*)\right).$

Thereby we get the target cost for benchmarking the physician:

$$k^{\#} = \sum_{p=P(b)} k_g^*\Big(c(p),f(b(p)),i(p)\Big).$$

In our setting we look for the group with the highest drug costs within a quarter for each consulted physician for a certain patient. This group should strongly be related to the morbidity of the patient and we will call it therefore Morbidity Related Group (MRG). One considers the costs as a proxy for the severity of drug treatment and could also take other weight functions instead of cost. The following is an example regarding a diabetes patient who belongs to the basic group A10A (Insulins and analogues) with total patient cost of 1,536.75 €:

Table 1: Example of (basic) MRG determination.

| cost | nr | ATC | substance | drug | amount |
|---|---|---|---|---|---|
| 320.74 | 1 | B01AF01 | Rivaroxaban | XARELTO 15 mg | 98 |
| 272.61 | 1 | N06AX21 | Duloxetine | CYMBALTA 60 mg | 98 |
| 248.02 | 2 | A10AD04 | Insulin Lispro | LIPROLOG Mix 25 | 10X3 |
| 208.25 | 7 | V04CA02 | Glucose | CONTOUR Test-streifen | 50 |
| 159.35 | 1 | N02AA55 | Oxycodone | TARGIN 10 mg/5 mg | 100 |
| 124.01 | 1 | A10AD04 | Insulin Lispro | LIPROLOG Mix 50 | 10X3 |
| 112.35 | 1 | N02AA55 | Oxycodone | TARGIN 5 mg/2.5 mg | 100 |
| 23.97 | 1 | C10AA01 | Simvastatin | SIMVA ARISTO 40 mg | 100 |
| 19.22 | 1 | C03CA04 | Torasemide | TORASEMID AL 20 mg | 100 |
| 16.27 | 1 | N02BB02 | Metamizole Sodium | NOVAMINSULFON 1A | 100 |
| 15.41 | 1 | H03AA01 | Levothyroxine Sodium | L-THYROX HEXAL 125 | 100 |
| 13.98 | 1 | C07AB12 | Nebivolol | NEBIVOLOL Glenmark 5 mg | 100 |

As an initial adjustment factor the age of patients can be applied. In each 5 year group of patients the ratio of costs per patient in the subgroup compared to the whole MRG was considered. If the ratio lies in certain intervals (0-0.5, 0.5-0.75, 0.75-1.25,..., 10) the age severity level 0,1,...,9 were assigned. The same can be conducted with respect to other factors correlated with morbidity. By using subgroup structures a risk adjustment can be accomplished. All of this has not to be precise on the level of the patient but on the physicians level. Regarding the considered MRG A10A (Insulins and analogues) seven degrees of severity in the range of 101.27 € up to 1,385.61 € resulted:

Table 2: Example of severity levels of MRG A10A.

| degree | cost in Euro | number of patients |
|---|---|---|
| 2 | 101.27 | 21 |
| 3 | 273.68 | 60,634 |
| 4 | 517.87 | 16,840 |
| 5 | 707.95 | 20,904 |
| 6 | 995.74 | 2,085 |
| 7 | 1,385.61 | 1,954 |

We divide the (basic) MRG into several severity levels that will be analysed by Lorenz curves and the corresponding Gini coefficients: Additionally the

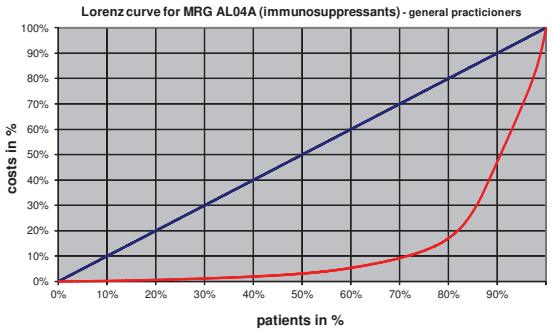



Figure 1: Lorenz curve of MRG AL04A.

Shannons entropy $(-\sum p_i log(p_i))$ can be applied to the patient type structure in each physicians group with respect to the MRG basic groups:

Table 3: Shannon entropy per physician group (1).

| enthropy | number of patients | physician group |
|---|---|---|
| 3.9985 | 994,220 | general practitioners |
| 3.8156 | 95,054 | paediatricians |
| 3.5635 | 3,548 | non-specialised internists |
| 3.4234 | 4,991 | nephrologists |
| 3.1437 | 1,955 | haematologist/oncologists |
| 3.1149 | 3,285 | cardiologists |
| 2.9632 | 2,961 | gastroenterologists |

Table 4: Shannon entropy per physician group (2).

| enthropy | number of patients | physician group |
|---|---|---|
| 2.7608 | 84,660 | dermatologists |
| 2.7552 | 16,155 | surgeons |
| 2.7327 | 40,045 | neck nose ear physician |
| 2.6253 | 91,868 | gynaecologists |
| 2.4939 | 37,199 | urologists |
| 2.2263 | 49,653 | neurologists |
| 2.1716 | 1,325 | endocrinologists |
| 2.1210 | 3,851 | rheumatologists |
| 2.0282 | 40,149 | orthopaedic |
| 1.6552 | 2,123 | anaesthetists |
| 1.4179 | 51,819 | ophthalmologists |
| 1.3517 | 26,392 | pulmonologists |
| 1.1722 | 7,003 | psychiatrists |

# 3 RESULTS

The application of the treatment case oriented approaches over the last decades showed that these systems are incapable of considering age and progress related increase of prescription costs. Recent analysis of the age distribution of treatment cases in each Statutory Health Insurance status group shows that the applied age groups might be too coarse and unsuited as the insurance status for the morbidity related depiction of prescription costs per patient:

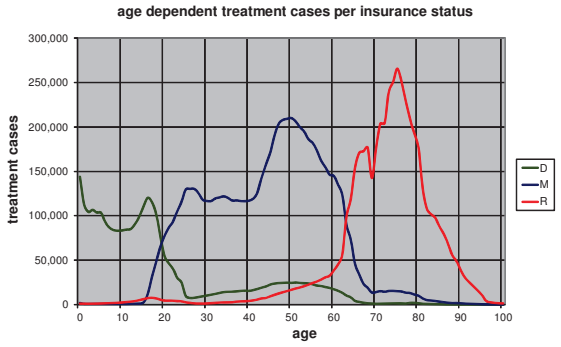



Figure 2: Age dependent number of treatment cases per insurance status/age groups.

The high correlation of the results of the two methods applied until 2016 confirms that shifting to age groups on the physicians level had practically no benefit:

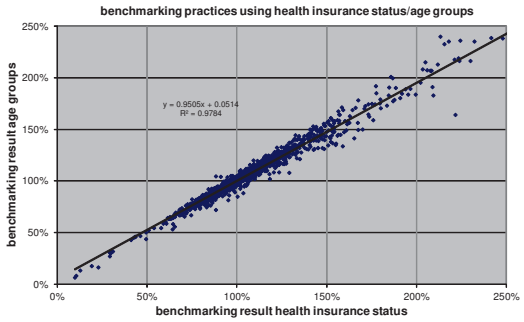



Figure 3: Correlation of benchmarks using health insurance status vs. age groups.

Hence, a new system based on MRG is introduced in 2017. There is little correlation between the results obtained by the previous and the new results on the practitioners level. That‘s due to the fact that many factors were disregarded in the past and inconsistencies were compensated by “manual intervention“:

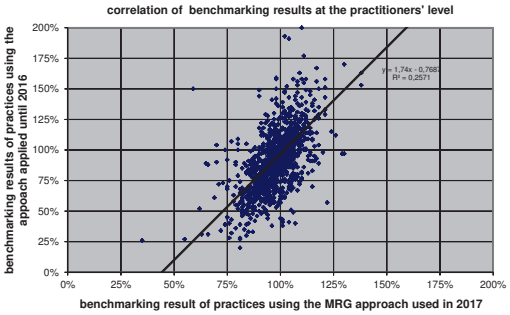



Figure 4: Correlation of benchmarks applying the previous (until 2016) and the new approach (2017).

Sorting the practices in ascending order for all affected groups due to their MRG benchmarking result and comparing those to the outcomes of the older system demonstrates the progress in model adaption made:

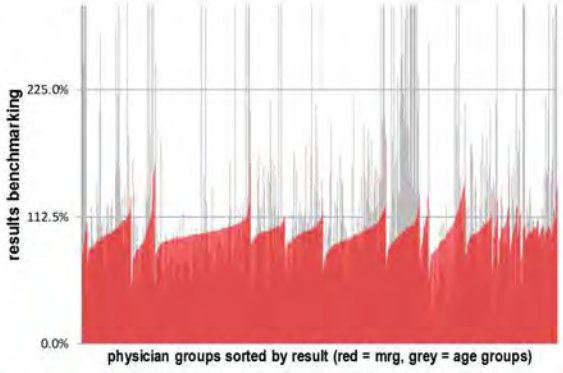



Figure 5: Results of MRG vs. system based on treatment cases in 2016 (each line one practice).

In those new MRG models all patients of a certain practice are classified and a specific structure for each practice is the result. As an example, we consider a physician with 14.0 % of his patients in the MRG A10A (Insulins and analogues) and 11.8 % patients in the MRG V04C (other diagnostic agents = test strips measuring glucose). In this group of general practitioners (GP) patients in those groups only account for 3.8 % in these two groups. The physician can thereby be identified as a diabetologist:

Table 5: Patient structure of a diabetologist.

| nr | MRG | nr. pat. | cost per patient | prop. doc. | prop. group | drug droup |
|---|---|---|---|---|---|---|
| 1 | A10A | 193 | 463.12 | 14.0% | 2.4% | Insulins and analogues |
| 2 | V04C | 162 | 307.76 | 11.8% | 1.4% | Other diagnostic agents |
| 3 | H03A | 86 | 22.28 | 6.3% | 4.5% | Thyroid preparations |
| 4 | A10B | 82 | 185.78 | 6.0% | 2.7% | Oral blood glucose lowering drugs |
| 5 | A02B | 73 | 60.91 | 5.3% | 7.2% | Drugs for peptic ulcer and gastro-oesophageal reflux disease (gord) |
| 6 | B01A | 53 | 366.21 | 3.9% | 4.0% | Antithrombotic agents |
| 1 | J01D | 22 | 31.70 | 1.6% | 2.4% | Other beta-lactam antibacterials |
| 19 | C10A | 17 | 58.30 | 1.2% | 3.0% | Cholesterol and triglyceride reducer |
| 20 | N03A | 16 | 275.63 | 1.2% | 1.4% | Antiepileptics |

After the formation of groups for all practices/physicians one can compare the MRG distributions and values in each group:

Table 6: MRG patient shares of orthopaedics.

| frac. pat. | cost p. pat. | MRG | drug group |
|---|---|---|---|
| 42.9% | 19.95 | M01A | Antiinflammatory and antirheumatic products, non-steroids |
| 13.6% | 26.26 | H02A | Corticosteroids for systemic use, plain |
| 12.4% | 23.81 | N02B | Other analgesics and antipyretics |
| 8.2% | 134.66 | M05B | Drugs affecting bone structure and mineralization |
| 6.1% | 123.78 | N02A | Opioids |
| 3.6% | 92.18 | B01A | Antithrombotic agents |
| 3.3% | 40.71 | M03B | Muscle relaxants, centrally acting agents |
| 2.5% | 33.75 | A02B | Drugs for peptic ulcer and gastro-oesophageal reflux disease (gord) |
| 1.1% | 2,515.02 | L04A | Immunosuppressive agents |
| 1.0% | 304.63 | L01B | Antimetabolites |

Regarding orthopedics we observe a patient type structure, in which 42.9 % of all patients belong to the MRG M01A (antiinflammatory and antirheumatic products, non-steroids). The 10 leading positions cover 94.6 % of the costs. Costs again depend mainly on the medical discipline. In oncology average costs per patient are 15,288.17€ in the MRG L04A (immunosuppressive agents including all the other drugs for the patient) versus 2,515.02€ for orthopedics. In urology the top ten positions with respect to the number of patients cover 83.6 % of the costs. In the case of GP these costs are only 44.2 %:

Table 7: MRG patient shares of urologists.

| frac. pat. | cost p. pat. | MRG | drug group |
|---|---|---|---|
| 34.8% | 44.67 | G04C | Drugs used in benign prostatic hypertrophy |
| 16.7% | 136.12 | G04B | Other urologicals, incl. antispasmodics |
| 9.9% | 19.84 | J01M | Quinolone antibacterials |
| 6.8% | 618.61 | L02A | Hormones and related agents |
| 4.9% | 30.42 | J01X | Other antibacterials |
| 3.1% | 33.47 | J01D | Other beta-lactam antibacterials |
| 2.0% | 154.56 | G03B | Androgens |
| 1.9% | 22.13 | J01E | Sulfonamides and trimethoprim |
| 1.7% | 27.71 | D01A | Antifungals for topical use |
| 1.7% | 4,122.10 | L02B | Antimetabolites |

Table 8: MRG patient shares of general practitioners.

| frac. pat. | cost p. pat. | MRG | drug group |
|---|---|---|---|
| 7.3% | 62.55 | A02B | Drugs for peptic ulcer and gastro-oesophageal reflux disease (gord) |
| 5.4% | 42.04 | C07A | Beta blocking agents |
| 5.1% | 34.67 | M01A | Antiinflammatory and antirheumatic products, non-steroids |
| 4.7% | 24.25 | H03A | Thyroid preparations |
| 4.2% | 185.14 | R03A | Adrenergics, inhalants |
| 3.8% | 316.40 | B01A | Antithrombotic agents |
| 3.8% | 124.59 | C09D | Angiotensin II antagonists, combinations |
| 3.4% | 85.10 | C09C | Angiotensin II antagonists, plain |
| 3.4% | 30.68 | C09A | Ace inhibitors, plain |
| 3.2% | 88.42 | N06A | Antidepressants |

The MRG patient shares can be utilized to generate distance measures for the clustering of all practices/physicians. Let $p_m^k$ be the fraction of patients with MRG $m$ $(m \in M)$ for the physician $k$ $(k \in P)$. With respect to the medical discipline $s$ $(s \in S)$ and let $q_m^s$ be the respective fraction. Let $r$ and $s$ be such fractions for physicians or medical disciplines we can use a Manhattan distance:

$$\sum_{i \in M} |r_i - s_i|$$

Alternatively we can apply spherical distances on the n-dimensional sphere where n is the number of MRG classes with respect to the points:

$$\frac{r_m}{\sqrt{\sum_{j \in M} r_j^2}} \text{ and } \frac{s_m}{\sqrt{\sum_{j \in M} s_j^2}} \text{ or } \sqrt{r_m} \text{ and } \sqrt{s_m}$$

The spherical distances are differentiable with respect to the components of r and s and thereby is more suitable for optimization procedures.

We can define the discipline $t \in S$ of a physician $k \in P$ by the value $s \in S$ for which:

$$\sum_{m \in M} |p_k^m - q_m^s|$$

has a minimal value. The distance of a physician to a group measures to which extent he is typical or not. Extreme values may be a hint for the need for special considerations. One can use cluster methods in order to receive a classification of physicians without the use of their medical discipline which is primarily determined by admission law.

# 4 CONCLUSIONS

The 2016 switch from health insurance status to age groups did not eliminate the flaws of the old benchmarking/budget approach. New promising ideas on the regional level like MRG have a huge potential still to be researched and utilized. The necessary data is provided, hard-/software and knowledge are available. Steady change and especially new form of health care require adapting benchmarking systems on a sound data and legal foundation. Therefore MRG seems to be a highly suitable approach meeting the criteria.

# REFERENCES


Bratzke, B. Spies, K.-P. and Krebs, S. (2012). Morbiditätskomponente bei Arznei- und Heilmittelbudgets einführen. *Deutscher Ärztetag*, Drucksache V I-37.

Breyer, F. and Ulrich, F. (2000). Gesundheitsausgaben, Alter und medizinischer Fortschritt: Eine Regressionsanalyse. *Jahrbücher für Nationalökonomie und Statistik*, 220:1–17.

Busse, R. Panteli, D. and Krebs, S. (2015). Arzneimittelversorgung in der GKV und 15 anderen europäischen Gesundheitssystemen: Ein systematischer Vergleich. *Working papers in health policy and management. Universitätsverlag der TU Berlin*, 11.

GKV Versorgungsstärkungsgesetz (GKV-VSG) (2015). *BGBl.*, I:1211.

Gottwald, N. (2015). Damoklesschwert Richtgrößenprüfung. Im Dschungel der 25-Prozent-Regel. *HNO-Nachrichten. Springer*, 3.

Korzilius, N. (2015). Arzneimittelrichtgrößen: Ärzte sollen angstfrei verordnen. *Dtsch Arztebl*, 112.

Schuster, R. (2015). Morbidity Related Groups (MRG) and drug economic index - a new concept after the age of Richtgrößen benchmarks in Germany. *GAA 2015*.

Schuster, R. and von Arnstedt, E. (2012). Mediziner-technischer Fortschritt und demografischer Wandel bei den GKV-Ausgaben im Vertragärztlichen Bereich. *GMDS 2011*.

Schuster, R. and Voss, H. (2012). Mediziner-technischer Fortschritt und demografischer Wandel bei den GKV-Ausgaben im DRG-Bereich im Krankenhaus. *GMDS 2012*.

Schuster, R. Emcke, T. and Schuster, F. (2016). Vergleich der Richtgrößenauswertungen für Arzneimittel im Vertragsärztlichen Bereich mit Statusgruppen und Altersgruppen. *HEC 2016*.

Sinowatz, F. (2015). Ressourcenalokation in der hausärztlichen Versorgung: eine empirische Untersuchung. *Dissertation LMU München: Medizinische Fakultät*.

Spitzenverband der Krankenkassen Bund und Kassenärztliche Bundesvereinigung (2015). Rahmenvorgaben nach 84 Abs. 7 SGB V -Arzneimittel- für das Jahr 2015.

Wersborg, T. (2006). Morbiditätsbezogene Richtgrößen zur Steuerung einer bedarfsgerechten und wirtschaftlichen Arzneimittelversorgung innerhalb der gesetzlichen Krankenversicherung in Deutschland. *Dissertation LMU München: Medizinische Fakultät*.

# Deviations in Birth Rates with Respect to the Day of the Week and the Month for a 100 Year Period Regarding Social and Medical Aspects in Explaining Models

Fabian Schuster[1], Thomas Ostermann[2], Reinhard Schuster[3] and Timo Emcke[4]
[1]*Faculty of Law, European University Viadrina Frankfurt (Oder), 15230 Frankfurt (Oder), Germany*
[2]*Chair of Research Methodology and Statistics in Psychology, Witten/Herdecke University, 58313 Herdecke, Germany*
[3]*Chair of Department of Health Economics, Epidemiology and Medical Informatics, Medical Advisory board of Statutory Health Insurance in Northern Germany (MDK), 23554 Lübeck, Germany*
[4]*Chair of Department of Prescription Analysis, Association of Statutory Health Insurance Physicians, 23795 Bad Segeberg, Germany*
*fabian.schuster@biomathematik.de, thomas.ostermann@uni-wh.de, reinhard.schuster@mdk-nord.de, timo.emcke@kvsh.de*

Keywords: Sunday Births, Long Time Considerations, Social Aspects, Instable Problems, Shannon Entropy, Gini Coeffient.

Abstract: During the last hundred years the birth rates on Sundays changed dramatically with a neutral point around 1955. Modern birth regulation is considered as the main reason for that. Medical backgrounds for this situation were discussed in the 1970s. Prior to that no analysis has relevant case numbers. The time from conception to birth measured in days is divisable by 7. The time of conception is relevant in relation to social aspects. Conception rates can be determined under the assumption that we can split up the population in a low and a high risk share. This consideration principally leads to an instable problem on a discrete cyclic space. But using some limiting considerations we get a numerically stable solution with feasible characteristics. For observing long time changes we need a relevant smoothing operator. In numerical calculations we look for a quadratic minimum solution or alternatively a linear program. For the discussion of inequality the concept of Shannon entropy as well as and Lorenz curve and Gini coefficient are relevant.

## 1 INTRODUCTION

We will consider, how the birth rate per weekday has changed in the last hundred years using data of the statutory health and care insurances. Reduced birthrates at weekends are usually discussed in the context of elective interventions. Larger birth rates at Sundays at the beginning of the 20th century should be discussed in the social context. One has to take into account that it is not possible to measure real birth rates from 1900-1950 but only the component related to insurance benefits decades later. Even survival rates may depend on weekday of birth. On the other hand the benefits of health insurance may depend on the underlying risk structure. Even the health status (”medical age”) may also depend on the weekday of birth.
Next we consider different daily birth rates and health costs with respect to the month of birth during different decades of the last century. Social and medical influences cause short and long term changes. In order to avoid large variations we use a 5 year smoothing of data. Interesting points are the day of birth and day of fertilization 100-85 years ago with varying social background. Large amounts of data are required to determine significant statistical effects. For this time

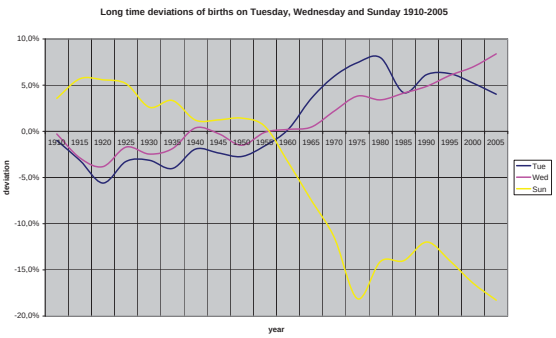


Figure 1: Long term development of birth rates on Tuesday, Wednesday and Sunday 1910-2005.



Schuster, F., Ostermann, T., Schuster, R. and Emcke, T.
Deviations in Birth Rates with Respect to the Day of the Week and the Month for a 100 Year Period Regarding Social and Medical Aspects in Explaining Models.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 41-47
ISBN: 978-989-758-213-4

period no register data are digitally available in the extend needed. One has to take into account the extensive migration movements during the last 100 years. A possible solution might be given by aggregated digital administrative data of health and care insurances. But precise resolution (day) is rarely available after aggregation has been done for other reasons. The first discussion of the influence of the weekday of birth on a large data base was given in (Macfarlane, 1978) and (Mathers, 1983) using birth data of the seventies, our data focuses on some decades before. Furthermore the number of births with respect to the weekday differs much from the current pattern. Related backgrounds are discussed in the stated references (cf.(Kibele et al., 2013), (Klein et al., 2001), (Klein and Unger, 2002), (Lampert and Kroll, 2014), (Ma et al., 2012), (Mackenbach, 2006), (Schnell and Trappmann, 2006), (Schuster and Emcke, 2016), (Ostermann and Schuster, 2015)).

## 2 MATERIAL AND METHODS

We use health and care insurance data from a German federal state. With respect to sufficient statistical significance in the care insurance field we can go as far back as people born in 1905 by using data from 1998 till 2006, in the health insurance data from 2006 one can track back until 1920. Although we only need aggregated data, such data with a weekday resolution are rarely available.

We use the script language perl in order to aggregate data and for the association of day of the week and date. If we refer to birth rates with respect to months we have to take into account their different lengths. Gender was only available for the care insurance data. The detailed insurance can be identified by a 9-digit identification code (IK-number). We used a reference table containing the insurance type in order to get a known social indication.

If we use drug data, there is information about additional private payment of patients. Patients with low social status have an additional payment exemption. There is also a mixed status in which patients get an additional payment exemption after having payed a certain amount themselves. We are interested in the social circumstances during birth, but we measure the social status many years later. A Markov model for transition of states would be useful. But there is no real information about transition rates. If we assume that the states are stable, we underestimate social effects.

Another type of analysis could combine low and high risk at birth with a survival in the following categories: first three days after birth and mothers with an age under or over 50 years. A derived, more detailed refinement could lead to mortality tables in dependence of the day and month of birth. Due to the low availability of historical information this remains a modeling challenge.

The time from the last menstrual period (LMP) to childbirth is usually taken as 40 weeks or 280 days. Pregnancy from conception to childbirth is 38 weeks or 266 days long. But there are no large scale measurements for mean values and standard deviations and in particular about deviations from normal distribution. We can divide the population into two subsets with respect to high and low pregnancy risk: $X = X_1 + X_2$ as random variables. Let $s(X)$ be the standard deviation of $X$. We use $s(X_1) < s(X_2)$. It is known from literature that we have $9 < s(X) < 13$. We use $s(X_1) = 1,2,3$. $X_1$ leads to increasing peaks, $X_2$ gives a nearly uniform variation to all days. If fertilization data would be given, the distribution of the random variable length of pregnancy would be a smoothing parameter on cyclic space (with discretization to days of week). But if we have given the birth data and want to derive the weekday distribution of the fertilization we get an inversion operator which tends to be instable. Constraints lead to numerical stabilization. We start with a quadratic-deviations model. Let f(i) be the observed deviation from 1/7 for likelihood of birth at day i (i = 0, 1,...,6) and w(i) the fertilization deviation pattern at day i (i = 0, 1,..., 6). Than $d_s(j)$ shall be the translation of j days by normal distribution with standard deviation s using integer intervals. We look for the quadratic minimum:

$$\sum_{i=0}^{6} \left( f(i) - \sum_{j=-30}^{30} d(i-j)w(j) \right)^2 \longrightarrow \text{Min!}$$

with the constraints $-1 < -a < w(i) < b < 1$. Practically we use $a = b = 1/(7*5)$ in order to limit the deviation for each day with respect to the mean of the week to 20 %. Alternatively we could use linear programming:

$$\left| f(i) - \sum_{j=-30}^{30} d(i-j)w(j) \right| < s, s \longrightarrow \text{Min!}$$

For calculations we use Microsoft Excel and Mathematica from Wolfram Research.

In order consider the different deviations during the considered time period we use the concept of Shannon entropie $\sum_{i=0}^{6} -p_i \ln(p_i)$ for the birth rates $p_i$ at day i. The same considerations we can adopt to months instead of the weekdays. Alternative measures of the inequality are given by the Lorenz Curve and the related Gini coefficient. In order to quan-

tify the deviation from the equal distribution we define $x_i = p_i - 1/7$ and from $\sum_{i=0}^{6} p_i = 1$ it follows $\sum_{i=0}^{6} x_i = 0$. The function $(1/7+x)\,ln(1/7+x)$ has the Taylor series: $-ln(7)/7 + x(1-ln(7)) + 7x^2/2 - 49x^3/6 + O(x^4)$ . As result we get a constant if we sum up the index i from 0 to 6 (with respect to the weekdays, with respect to the months we have to sum up from 0 to 11). Therefore the entropy reflects a quadratic (non-linear) property with respect to the $p_i$. The Gini coefficient is in contrast to that linear in the $p_i$ with weighting coefficients depending of the order up to a constant $\sum_{i=0}^{6}(1-p)p_i$ with monotonically increasing $p_i$. First we consider an empiric connection between Shannon entropy and Gini coefficient looking at 5 year periods. Second we compare the Taylor series to the quadratic term of the function $(1/7+x)\,ln(1/7+x)$:

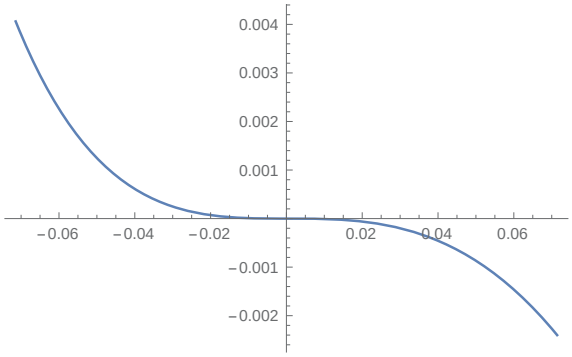



Figure 2: Difference of term in the Shannon sum and its quadratic Taylor series representation.

## 3 RESULTS

If we use data of the care insurance from 1998-2006, we can consider deviations of the birth rates back to 1905 in Figure 3.
On Saturdays and Sundays we have increased birth rates, lower ones on Tuesdays and Wednesdays. The


Figure 3: Deviations of birth rates in dependence of the weekday (care insurance data).

other weekdays are somewhere between with instabilities with respect to time periods. One has to take into consideration that only about 20% of the people ever get benefits of care insurance. In contrast to this the great majority of older people gets at least one drug each year. If we use drug prescription data of 2006 we get the distribution of birth rates in Figure 4. There

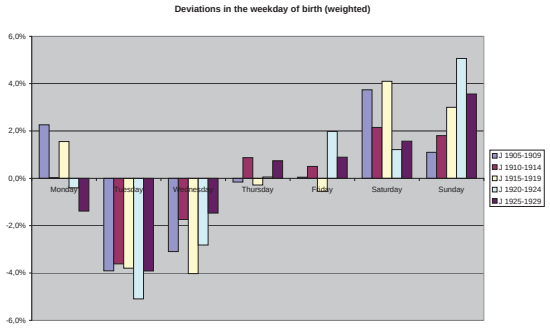



Figure 4: Deviations of birth rates in dependence of the weekday (health insurance data).

are only small differences if we use drug prescription data of 2007 or 2008. At Saturdays the birth rates are less increased compared to care insurance, the rates at Sundays are even larger. The reduced birth rates on Tuesdays and Wednesdays correspond with the results from the care insurance analysis.
If we compare the drug costs of the patients born between 1920 and 1924 with those born between 1925 and 1929 we find an average annual increase of 1.51%. For such considerations it is important to use an age group with monotonously increasing drug costs. Having regard to that we create subgroups with respect to the weekday of birth, cf. Figure 5.

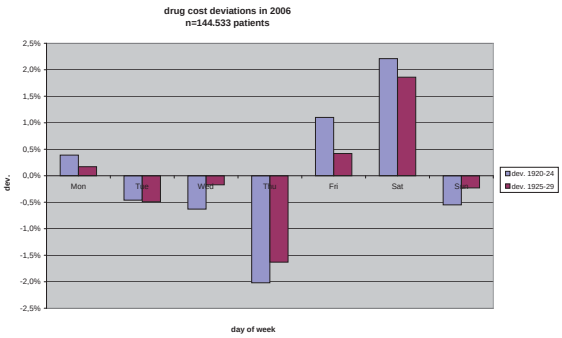



Figure 5: Drug costs in dependence of the weekday of birth for age groups born 1920-24 and 1925-29.

The weekdays with increased and reduced costs do not match those of increased and reduced birth rates. The 1.51% increased drug costs of patients born on Saturdays can be interpreted as having a one year higher biological age than calendar age. On the other hand the people born on Thursdays are one year

younger biologically.
Using the data of care insurance, we find a relevant gender dependent difference in the birth rates on Sundays, cf. Figure 6.

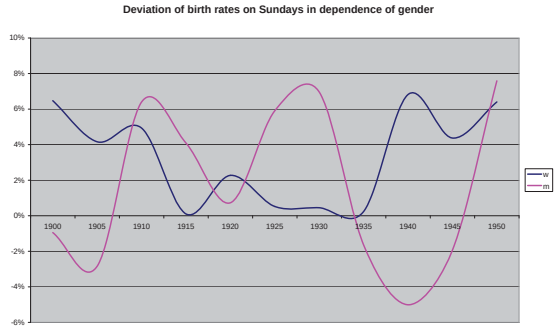



Figure 6: Deviations of the birth rates on Sundays with respect to gender.

Next we consider subgroups with respect to the social status. We use additional payment as a proxy.

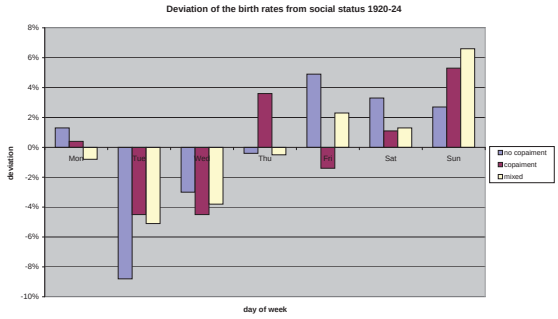



Figure 7: Birth rates in dependence of the social status and the weekday of birth for the period 1920-24.

The group of patients of births for the 1925-29 period and a socially week status show a lower increase in rates on Sundays and higher reduction on Tuesdays, cf. Figure 7 . In the next five year period the situation is quite different, cf. Figure 8.

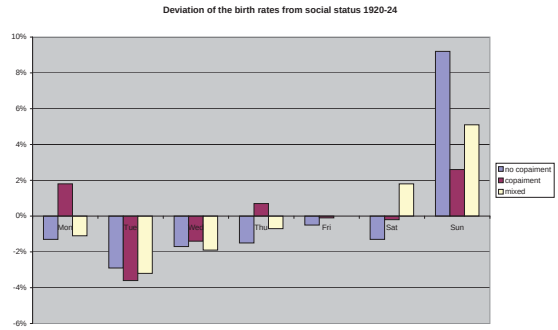



Figure 8: Birth rates in dependence of the social status and the weekday of birth for the period 1925-29.

Social week patients show an even higher increase in birth rates on Sundays but no significant differences in reduced birth rates on Tuesdays. We can compare the rate changes on Thursdays and Sundays directly, cf. Figure 9 and 9.

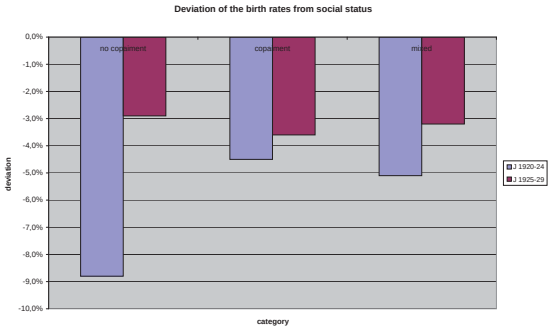



Figure 9: Birth rates on Tuesdays in dependence of the social status in 1920-24 vs. 1925-29.

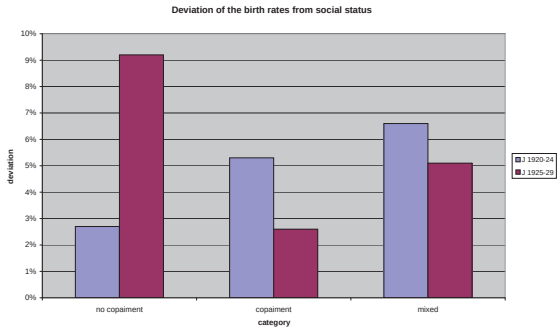



Figure 10: Birth rates on Sundays in dependence of the social status in 1920-24 vs. 1925-29.

Additionally we can use social information using the type of insurance, cf. Figure 11.

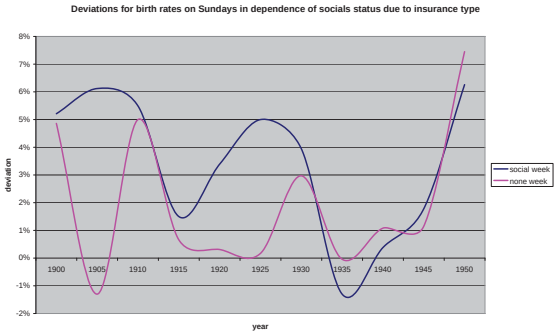



Figure 11: Birth rates on Sundays in dependence of the social status using health insurance type.

We will consider the low risk population and calculate the different fertility rates by the considered quadratic-deviations model with standard deviations 1, 2 and 3 and a limitation of the rate deviations by 20%, cf. Figure 12.

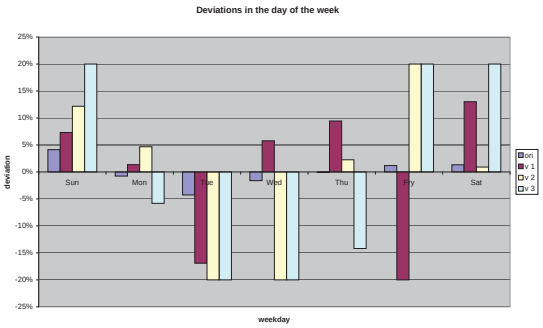



Figure 12: Calculated deviation of the fertility rates with standard deviations 1, 2 and 3.

We see that in general the effects at Saturday and Sundays are increased, the effects at Tuesdays and Wednesdays are reduced. We have used the 20% value in order to limit instabilities. If we would use values from 10% to 25%, we would get the same result for the distribution to the weekdays. Unfortunately we get no further information about a true limit value.
Shannon entropy and Gini coefficients have the same behavior with respect to local maxima and minima. Additionally we can use social information using the type of insurance, cf. Figure 13.

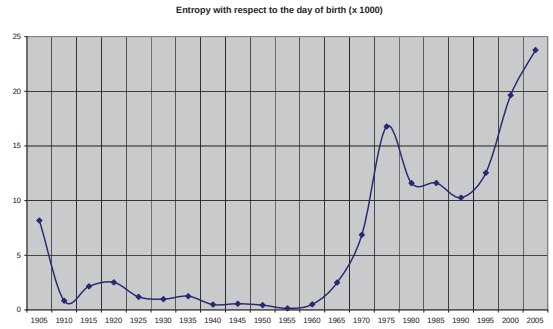



Figure 13: Shannon entropy with respect to the weekdays of birth in dependence of 5 year periods.

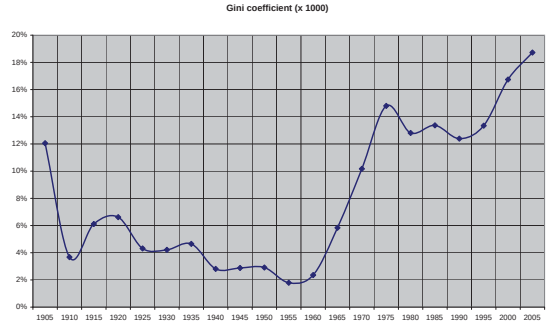



Figure 14: Gini coefficient with respect to the weekdays of birth in dependence of 5 year periods.

Both results show the global minimum for the year 1955. We remember that this year separates the age of increased and that of reduced Sunday birth rates. There is a different resolution between the entropy and the Gini result. The Shannon entropy result uses a nonlinear effect but does not order the used rates, the Gini result is linear but uses ordered rates. Thereby it is interesting that both results coincide so much.
Till now we have considered the weekday period. It is also interesting to consider months.

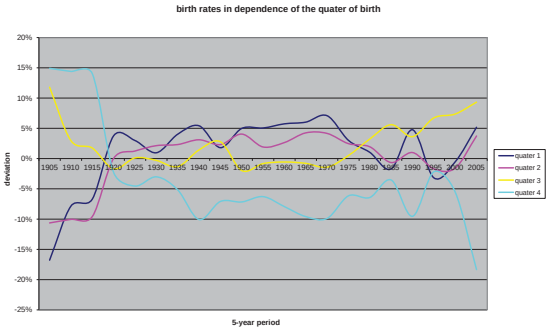



Figure 15: Birth rates in dependence of the quarter of birth.

In 1920-1980 for the first six months there are increased birth rates. Reduced birth dates we have since 1920. Before 1920 the situation is quite different. The mean costs in dependence of the month of birth are quite heterogeneous with respect to different historical periods, cf. Figure 16.

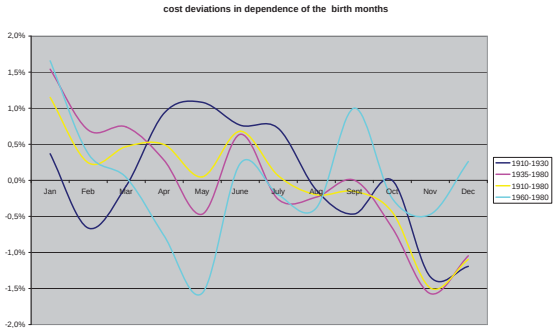



Figure 16: Deviations in drug costs in dependence of the month of birth during different historical periods.

If we consider the period from 1920 till 1980, we have increased costs during the first half of the year and reduced costs during the second half of the year. One explanation could be, that the month of birth has different influences due to the historical period of birth. On the other hand the effect can depend on the age of the persons. We compare the mean effect for birth rates and drug costs from 1920 till 1980 with respect to the quarters of the year, we see that generally increased birth rates and health status measured by drug costs behave reverse, cf. Figure 17.

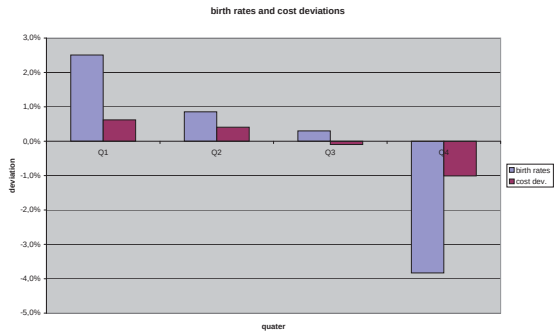



Figure 17: Deviations in birth rates and drug costs with in dependence of the quarters of the year.

Last we compare drug costs in dependence of the quarter of the year for the two groups born from 1910 till 1930 versus the group born between 1960 an 1980, cf. Figure 18.

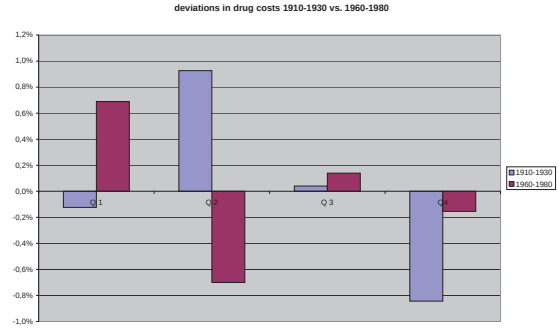



Figure 18: Deviations in drug costs in dependence of the month of birth during different historical periods.

The highest difference we have at quarter two. It can be a consequence of different historical health conditions near to birth. An other explanation would be an age dependent effect.

# 4 CONCLUSIONS

In order to consider the time between birth and measurements using data of health and care insurance the following statements and guesses can be made regarding the results.
In scenario 1 more births measured in insurance data can be caused by more real births in the considered time 1915-1930. That can be due to different conception and/or fertilization possibilities depending on the day of the week. A bias may be caused by migration.
In scenario 2 the day of birth may causes different survival expectancies in the critical first three days after birth and the related health conditions during these days. That is why we analyze drug costs in dependence of the day of birth. As we already stated, drug costs increase in the mean by 1.5 % per year between the considered two age groups. As a modeling consideration one can use drug costs as a proxy for biological age, comparing it with calendar age. Due to the considered age dependent drug cost increase we can suspect a strong connection to the residual life expectancy. Thursday births around 90 years ago have a one year higher residual life expectancy. Saturday births have a one year lower residual life expectancy, Sunday births have 4 months higher residual life expectancy. In contrast to the situation stated in Macfarlane (1978) lower perinatal mortality rates at weekends can be caused by the fact that quality of care was higher due to family background. In those times specialist obstetric services have been less common compared to later decades. It is quite important, that the psycho-social near birth circumstances 90 years ago may induce significant differences today.

# REFERENCES

Kibele, E. U., Jasilionis, D., and Shkolnikov, V. M. (2013). Widening socioeconomic differences in mortality among men aged 65 years and older in Germany. *J Epidemiol Community Health*, 67(1):453–457.

Klein, T., Schneider, S., and Löwel, H. (2001). Bildung und Mortalität. die Bedeutung gesundheitsrelevanter Aspekte des Lebensstils. *Zeitschrift für Soziologie*, 30(5):384–400.

Klein, T. and Unger, R. (2002). Aktive Lebenserwartung in Deutschland und in den USA. *Zeitschrift für Gerontologie und Geriatrie*, 35:528–539.

Lampert, T. and Kroll, L. (2014). Soziale Unterschiede in der Mortalitt und Lebenserwartung. *Hrsg. Robert Koch-Institut, Berlin*, 5(2).

Ma, J., Xu, J., Anderson, R., and Jemal, A. (2012). Widening educational disparities in premature death rates in twenty six states in the United States, 1993-2007. *PloS one*, 7(7).

Macfarlane, A. (1978). Variations in number of births and perinatal mortality by day of week in England and Wales. *British Medical Journal*, 2:1670–1673.

Mackenbach, J. (2006). Health inequalities: Europe in profile. An independent expert report commissioned by the UK Presidency of the EU. *Department of Health, London*.

Mathers, C. (1983). Births and perinatal deaths in Australia: variations by day of week. *Journal of Epidemiology and Community Health*, 37:57–62.

Ostermann, T. and Schuster, R. (2015). An Information-theoretical Approach to Classify Hospitals with Respect to Their Diagnostic Diversity using Shannon's Entropy. *HEALTHINF*.

Schnell, R. and Trappmann, M. (2006). Konsequenzen der Panelmortalitt im SOEP fr Schätzungen der

Lebenserwartung. Arbeitspapier. *Zentrum für Quantitative Methoden und Surveyforschung, Mannheim, Konstanz.*, 2.

Schuster, R. and Emcke, T. (2016). Unterjährige Schwankungen bei den Arzneimittelverschreibungen und bei der Anzahl der Patienten pro Arzt: eine Baseline für den vertragsärztlichen Bereich. *HEC 2016. München*.

# Pattern Recognition Application in ECG Arrhythmia Classification

Soodeh Nikan, Femida Gwadry-Sridhar and Michael Bauer
*Department of Computer Science, University of Western Ontario, London, ON, Canada*
*{snikan, fgwadrys, bauer}@uwo.ca*

Keywords: Arrhythmia Classification, Pattern Recognition, Beat Segmentation, 1-D LBP, ELM Classification.

Abstract: In this paper, we propose a pattern recognition algorithm for arrhythmia recognition. Irregularity in the electrical activity of the heart (arrhythmia) is one of the leading reasons for sudden cardiac death in the world. Developing automatic computer aided techniques to diagnose this condition with high accuracy can play an important role in aiding cardiologists with decisions. In this work, we apply an adaptive segmentation approach, based on the median value of R-R intervals, on the de-noised ECG signals from the publically available MIT-BIH arrhythmia database and split signal into beat segments. The combination of wavelet transform and uniform one dimensional local binary pattern (1-D LBP) is applied to extract sudden variances and distinctive hidden patterns from ECG beats. Uniform 1-D LBP is not sensitive to noise and is computationally effective. ELM classification is adopted to classify beat segments into five types, based on the ANSI/AAMI EC57:1998 standard recommendation. Our preliminary experimental results show the effectiveness of the proposed algorithm in beat classification with 98.99% accuracy compared to the state of the art approaches.

## 1 INTRODUCTION

One of the primary cause of sudden death globally is cardiovascular disease. The improper life style by having an unhealthy diet, tension and stress, tobacco consumption and insufficient exercise leads to cardiovascular disease. Atrial and ventricular arrhythmias are concurrent side effects arises from cardiovascular disease. Arrhythmia is abnormal changes in the heart rate due to improper heart beating which causes failure in the blood pumping. The abnormal electrical activity of the heart can be life threatening. Arrhythmias are more common in people who suffer from high blood pressure, diabetes and coronary artery disease.

Electrocardiograms (ECGs) are the recordings of electrical activities of the heart. Each heart beat in an ECG record is divided into P, QRS and T waves which indicate the atrial depolarization, ventricular depolarization and ventricular repolarisation, respectively. Electrocardiograms are used by cardiologists to detect abnormal rhythms of the heart. Cardiologists must deal with challenges in the diagnosis of arrhythmia due to the effect of noise in ECG signals and the nonstationary nature of the heart beat signal. Automatic interpretation of ECG data using time-frequency signal processing techniques and pattern recognition approaches could be helpful to both cardiologists and patients for improved diagnostics (Thomas et al., 2015; Elhaj et al., 2016).

Although in the past few years, several computer-aided methods for early prediction of the risk of cardiovascular disease have been investigated, it is still an extremely challenging problem. There are many pattern recognition techniques in the literature to recognize and classify ECG beats. Particle swarm optimization (PSO) and radial basis functional neural network (RBFNN) were employed in the proposed beat classification algorithm in (Korurek and Dogan, 2010). In (Khoshnoud and Ebrahimnezhad, 2013), an accuracy of 92.9% was obtained where linear predictive coefficients (LPC) were adopted as beat features and normal and abnormal beat types were classified using probabilistic neural networks. In (Inan et al., 2006), beats are classified with an accuracy of 95.16% using the combination of time–frequency features, using wavelet transform, time domain information and the use of an artificial neural network (ANN) as a classifier. In (Martis et al., 2013a) a combination of a linear DWT feature extraction and principal component analysis (PCA), as dimensionality reduction technique, and neural



Nikan, S., Gwadry-Sridhar, F. and Bauer, M.
Pattern Recognition Application in ECG Arrhythmia Classification.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 48-56
ISBN: 978-989-758-213-4

network classifier leads to 98.78% classification accuracy between normal and abnormal beats. In (Kadambe and Srinivasan, 2006), normal and abnormal time domain features, P, QRS and T waves from American Heart Association database were classified with the accuracy of 96%, 90% and 93.3%, respectively, by discretizing the wavelet basis function using an adaptive sampling scheme. Adaptive parameters of wavelet non-linear functions and the relative weight of each basis function were estimated using a neural network. An accuracy of 99.65% for arrhythmia recognition was reported in (Yu and Chen, 2007) using wavelet transform and a probabilistic neural network. However, a small subset of MIT-BIH ECG database (only 23 records) was employed for evaluation. In order to classify ECG beats, the authors in (Ebrahimzadeh and Khazaee, 2009) adopted statistical features from Lyapunov exponents and wavelet coefficients power spectral density (PSD) values of eigenvectors and achieved a 94.64% accuracy for eight records from MIT-BIH arrhythmia database. Due to the effect of noise, and the nonstationary nature of ECG signal, nonlinear techniques appear to be more effective in extracting distinctive and hidden characteristics of ECG signals. In (Martis et al., 2013b), higher order spectra (HOS) bi-spectrum cumulants and PCA dimensionality reduction approach were adopted to represent ECG signals and feed-forward neural network and least square support vector machine (SVM) were used for classifying different types of beats with an accuracy of 93.48%. In (Khalaf et al., 2015), a cyclostationary analysis was proposed as a feature extraction approach to reduce the effect of noise and also to reveal hidden periodicities of ECG beats where spectral correlation coefficients were utilised as statistical signal characteristics and passed through SVM for classification; this results in an accuracy of 98.6% for 30 records of the MIT-BIH Arrhythmia database. Authors in (Oster et al., 2015) proposed a switching Kalman filter technique for arrhythmia classification, and automatic selection of beat type. This method also includes a beat type for unknown morphologies "X-factor" which incorporates a form of uncertainty in classification for the case of indecision on the beat type. The classification F1 score of the algorithm on MIT-BIH arrhythmia database was 98.3%. Employing the fusion of linear and nonlinear features has benefits the advantages of handling noise and a more effective description of the signal. In (Elhaj et al., 2016), a combination of linear (PCA of DWT coefficients) and nonlinear (high order statistics, cumulants and independent component analysis) features were proposed for heart beat representation. An accuracy of 98.91% was achieved using the fusion of SVM and radial basis function classifiers to classify five types of arrhythmia. The combination of fourth and fifth scale dual-tree complex wavelet transform (DTCWT) coefficients, AC power, kurtosis, skewness and timing information were adopted in (Thomas et al., 2015) as QRS characteristics. Multi-layer back propagation neural network was proposed to classify five types of ECG beats of MIT-BIH Arrhythmia database with the accuracy of 94.64%.

As discussed, encouraging results on the arrhythmia classification have been obtained in previous research. However, more applicable and fully automatic techniques with high accuracy and low complexity need to be developed. In particular, developing automatic computer aided segmentation of ECG signal into heart beats is very important as the first stage in beat classification. In previous research (Thomas et al., 2015; Khalaf et al., 2015), R peaks were located using an annotated file which makes the techniques semi-automatic. In contrast, in the approach proposed in this paper, R peaks are detected automatically based on a parabolic fitting algorithm. Moreover, a novel adaptive segmentation technique used in our work reduces the probability of beat misclassification and the risk of misdiagnosis due to the interference of adjacent beats which may occur when a constant beat size was used as in previous works (Martis et al., 2013a; Elhaj et al., 2016). As well, the chosen feature extraction technique has significant role in the accuracy of diagnosis. By discovering hidden patterns and extracting distinctive features from the ECG signal, which are less sensitive to noise, the accuracy of arrhythmia classification can be improved without requiring very complicated classifiers. Uniform 1-D local binary pattern (LBP), used in our work, has the advantage of less sensitivity to noise and effectiveness in extracting hidden and salient information from non-stationary ECG signals and due to low computational complexity, it can be employed in real-time applications (Kaya et al., 2014).

The proposed arrhythmia recognition approach in this paper is based on beat classification by adopting the fusion of wavelet transform and uniform 1-D LBP feature extraction of ECG signal and extreme learning machine (ELM) classification. The ECG signal is pre-processed to remove the unwanted effect of noise. Then, the de-noised signal

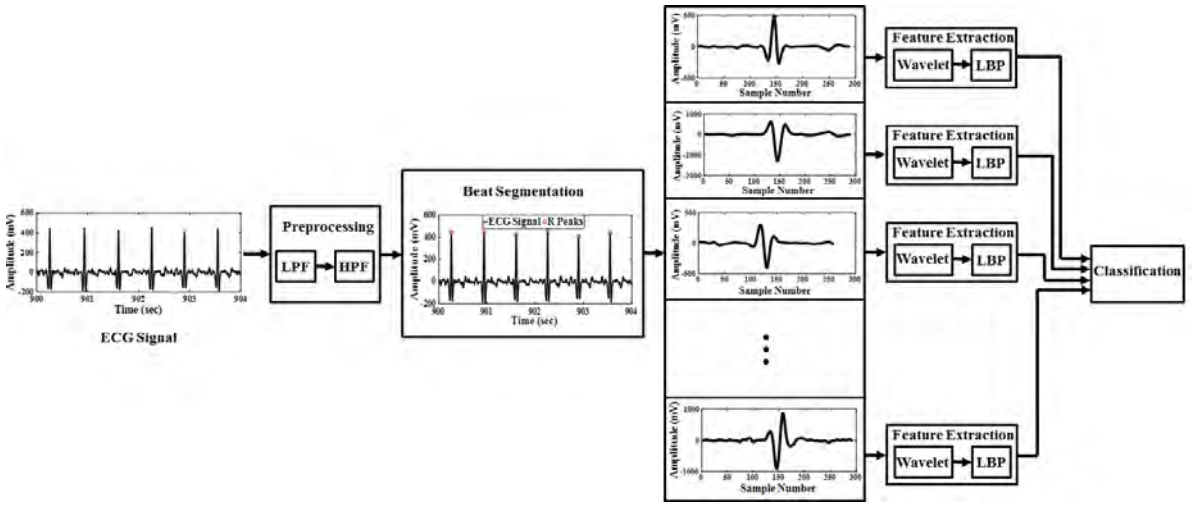



Figure 1: Block diagram of the proposed arrhythmia recognition and classification algorithm.

is divided into heart beats using the proposed adaptive segmentation technique in this paper, which is based on the detected R peaks and the median value of R-R intervals. Each segment of the ECG signal is transformed into time-frequency space by applying digital wavelet transform (DWT). Wavelet coefficients of signal go through a one-dimensional version of LBP which is a histogram-based signal descriptor and extracts hidden and distinctive characteristics of a signal. By transforming the feature space into histograms, the dimensionality of the feature space is reduced from the number of signal samples to the number of histogram bins. In this paper, we just keep uniform patterns which contain useful information about the one-dimensional signal, such as sudden changes, edges, end of lines and corners. The beat segments are divided into training and testing sets. The extracted features of training set are fed to an ELM classifier for the training procedure. The remaining feature vectors are used to test the beat classification algorithm. Figure 1 shows the block diagram of the proposed algorithm. The rest of paper is organized as follows: Section 2 describes the adopted ECG database and Section 3 provides mathematical details of the pre-processing techniques. Section 4 explains the proposed beat segmentation approach. Section 5 and 6 discuss feature extraction and classification techniques and Section 7 provides an evaluation through experimental results. Finally, the paper is concluded in Section 8.

## 2 MATERIALS

In this paper, we consider the ECG signals which are available online from PhysioNet that offers free web access to a large-scale dataset of recorded physiologic signals. The MIT-BIH arrhythmia database (Moody and Mark, 2001; Goldberger et al., 2000) is used to evaluate the arrhythmia recognition and classification technique which has been proposed in this paper. There are 48 ECG records, with the length of a little more than 30 minutes, in the MIT-BIH collection and the sampling frequency of each ECG signal is 360 Hz. Twenty-three of recordings were routine clinical ECGs selected from 4000 ambulatory records at Boston's Beth Israel Hospital and the remaining 25 ECG signals were collected from the same set to include other less common significant arrhythmia types that may not be represented well in a small randomly selected group. Each beat in the ECG signal shows one cycle of electrical activity of the heart. The irregular heart rhythms are considered as ectopic beats. The entire MIT-BIH database is grouped into five beat types based on the ANSI/AAMI EC57:1998 standard recommendation (Martis et al., 2013a). The five classes include normal beats (N), fusion beats (F), supra-ventricular ectopic beats (S), ventricular ectopic beats (V) and unknown or unreadable beats (U) as shown in Fig 2. In this paper, we adopted the entire 48 ECG records in the database including (90,580) N, (2973) S, (7707) V, (1784) F and (7050) U beats (110094 beats, totally).

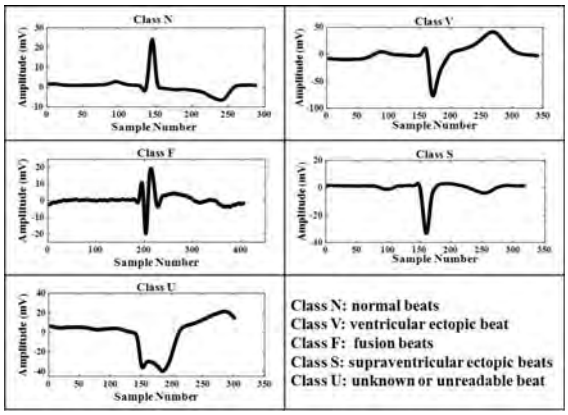



Figure 2: Five categories of ECG beat classes based on the ANSI/AAMI EC57-1998 standard.

# 3 PREPROCESSING

The effect of noise on the ECG signal reduces the accuracy of recognition of arrhythmia in the ECG records and therefore, the precision of diagnosis of cardiovascular disease will be decreased. Various categories of noise are associated with the ECG signal, such as powerline interference, device noise, muscle noise, motion noise, contact noise and quantization noise (Elhaj et al., 2016). In order to increase the accuracy of disease detection, pre-processing is required to be applied on the ECG signal to reduce the effect of noise and improve the signal to noise ratio. In this paper, we applied a digital elliptic band-pass filter with passband of 5-15 Hz (maximizes the QRS energy), which is constructed by cascading a low-pass and high-pass filters, to remove muscle noise and baseline wander (Pan and Tompkins, 1985) as follows.

## 3.1 Low-pass Filter

The adopted low-pass filter has the following transfer function and amplitude response, respectively (Pan and Tompkins, 1985).

$$H(z) = \frac{(1-z^{-6})^2}{(1-z^{-1})^2}, \quad (1)$$

$$|H(\omega T)| = \frac{sin^2(3\omega T)}{sin^2(\frac{\omega T}{2})} \quad \text{where T is sampling period} \quad (2)$$

## 3.2 High-pass Filter

The transfer function of the high-pass filter is based on the subtraction of the output of a first-order low-pass filter from an all-pass filter as follows (Pan and Tompkins, 1985).

$$H(z) = z^{-16} - \frac{(1-z^{-32})}{(1-z^{-1})}. \quad (3)$$

The proposed high-pass filter has the following amplitude response.

$$|H(\omega T)| = \frac{[256 + sin^2(16\omega T)]^{0.5}}{cos(\frac{\omega T}{2})}. \quad (4)$$

# 4 BEAT SEGMENTATION

In order to recognize arrhythmia, we had to compute a beat classification by dividing each ECG signal into beat segments and classify different types of beats. The segmentation process consists of R peak detection and isolation of beats based on the duration of R-R intervals.

## 4.1 R Peak Detection

R peaks are the largest deviation of ECG signals from the baseline. The proposed algorithm for R peak detection in this work is based on the parabolic fitting algorithm (Jokic et al., 2011). By adopting two polynomial functions (PFs) of degree 3, we modelled the R peak. A signal $x$ of length $M$ is defined as follows.

$$x\,(m)\colon x\,(1), x\,(2), \dots, x\,(M) \quad (5)$$

where $x\,(i)$ is the $i^{th}$ sample of the signal. The approximation of signal $x$ using the polynomial function $\hat{x}$ of order $d$ is denoted by the following equation.

$$\hat{x}\,(m) = c_0 + c_1\,m + c_2\,m^2 + \cdots + c_d\,m^d,$$
$$where \qquad m = 1, 2, \dots, M. \quad (6)$$

By minimizing the least square error (the square of $l_2$ norm of the residual), we can calculate the $c_k$ coefficients as follows.

$$er^2 = \|\hat{x}\,(m) - x(m)\|_2^2 = \sum_{m=1}^{M} (\hat{x}\,(m) - x(m))^2, \quad (7)$$

$$\frac{\partial er^2}{\partial c_k} = 0. \quad (8)$$

In order to find R peak, a differentiator is first used to highlight the high inclines of the ECG signal. Then, the PFs are fitted from the Q peak to the R peak (through the ascending R leg) and from R peak to the S peak (through the descending R leg) (Jokic et al., 2011).

### 4.2 Segmentation

After detection of the R peaks we need to split ECG signal into beat segments. The segmentation technique which is proposed in this paper starts from each R peak and separates beats by choosing some samples from the left and right side of the R peak without inclusion of the former or latter beats. In previous work in the literature (Thomas et al., 2015; Elhaj et al., 2016) a constant number of samples are selected from both signal sides. Therefore, the length of all beat segments is equal. However, due to the non-stationary and aperiodic nature of ECG signal, beat lengths for all of the ECG records are not equal-sized. Therefore, determining a constant size for all beats may lead to inclusion of adjacent beats in each segment. In this paper, in order to reduce the effect of beat interference, we employ a novel adaptive segmentation approach. For each ECG record we calculate the consecutive R-R intervals and find the median value of R-R durations for each ECG signal as the adaptive beat duration. Therefore, from each R peak, we select the number of samples equal to the half of the beat duration from the left and right sides of the R peak.

## 5 FEATURE EXTRACTION TECHNIQUES

In this section, we describe how we find the distinctive characteristics of ECG beats to feed to classification stage for beat recognition. The cascade combination of wavelet transform and uniform 1-D LBP is applied on beat segments to extract sudden variances and sparse hidden patterns from signal.

### 5.1 Wavelet

Discrete wavelet transform is a viable and powerful feature extraction technique to analyse ECG signals locally in multi-resolution manner in time and frequency domain simultaneously and separate the signal frequency sub-bands. A signal can be displayed with different scaling and wavelet basis functions (Emadi et al., 2012). DWT extracts the approximation (low frequency components) and detailed coefficients (high frequency components) as shown in Fig 3 ($A_i$ and $D_i$ are approximation and detail coefficients and $i = 1, 2\ and\ 3$). A continuous wavelet transform is generated by a series of translations and dilations of mother wavelet $\varphi(.)$ as follows (Ródenas et al., 2015).

$$\varphi_{\alpha,\beta}(t) = |\alpha|^{-\frac{1}{2}}\varphi\left(\frac{t-\beta}{\alpha}\right) \quad (9)$$

where, $\alpha$ and $\beta$ are scaling and shift parameters, respectively. DWT is the sampled version of continuous wavelet as follows.

$$\varphi_{\alpha,\beta}[n] = 2^{-\frac{\alpha}{2}}\varphi[2^{-\alpha}\, n - \beta]. \quad (10)$$

The wavelet transform of a signal, $x\,[n]$ of length N, is the correlation between the wavelet function $\varphi_{\alpha,\beta}[n]$ and signal as shown by the following set of wavelet coefficients (Ródenas et al., 2015).

$$CW[\alpha,\beta] = \sum_{n=1}^{N} x\,[n]\varphi_{\alpha,\beta}[n]. \quad (11)$$

In this paper, we use 8 level wavelet decomposition and adopt the approximation and detail coefficients as the extracted features. Therefore, the size of wavelet features for each beat, depending on the beat size, is different.

### 5.2 1-D LBP

Two-dimensional local binary pattern (2-D LBP) is one of the most successful feature extractors, which extracts texture features of the 2-D images by comparing each signal sample (image pixel) with its neighbour samples in a small neighbourhood. There is no training requirement which makes the feature extraction fast and easy to integrate into the new data sets. Furthermore, due to the application of

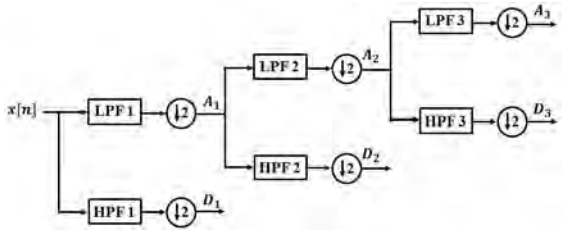



Figure 3: Three-level wavelet decomposition.

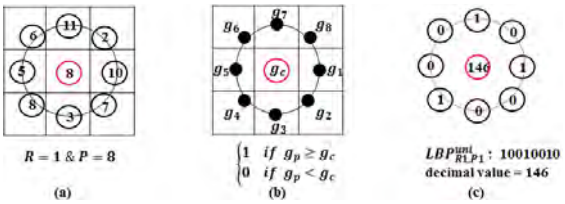



Figure 4: 2-D LBP for a sample point of a 2-D signal: a) choosing P neighbours on the neighbourhood of radius R around a centre sample point, b) comparing signal values for centre and neighbour points and c) creating the binary pattern and associated decimal value for the centre sample.

histograms as the feature sets, the image-size dimension of the feature space can be reduced to the number of histogram bins (Ahonen et al., 2004). $R$ is the radius of the neighbourhood and $P$ is the number of neighbour samples which are compared with the centre pixel as shown in Fig 4. If the value of the neighbour sample is greater than or equal to the centre sample, a 1 is assigned to that neighbour and if it is less than the centre pixel a 0 is assigned to that sample. Therefore, we have a $P$-bit binary pattern for each pixel at $(r, c)$ location and the decimal value (DV) associated with the binary pattern is calculated as follows.

$$DV(r,c) = \sum_{k=1}^{P} G(g_k - g_c).2^{P-1}, \quad (12)$$

$$G(u) = \begin{cases} 1 & if\ u \geq 0 \\ 0 & if\ u < 0 \end{cases}. \quad (13)$$

Decimal values are used to make the histogram for the 2-D signal. Therefore, the size of feature vector which is extracted from the 2-D image is equal to the number of histogram bins ($2^P$). In order to reduce the size of features and remove redundant information, we ignore non-uniform patterns due to the fact that considerable amount of discriminating information (important local textures such as spots, line ends, edges and corners) is preserved by taking only uniform patterns into consideration (Ahonen et al., 2004). The binary pattern is uniform if there are at most two bitwise transitions from 0 to 1 or 1 to 0. Each histogram has $P(P-1)+2$ bins for uniform and 1 bin for all non-uniform patterns, in total there are $P(P-1)+3$ bins. Therefore, the computational complexity is also reduced (Nikan and Ahmadi, 2015).

The one-dimensional version of LBP can be adopted to extract distinctive characteristics from ECG signals. The same procedure is applied on each sample point of the signal by comparing P/2 neighbours from right and left side of centre sample to create the P-bit pattern as shown in Fig 5 (Kaya et al., 2014). In this paper, the uniform 1-D LBP with neighbourhood size of 8 points is applied on wavelet coefficients from the previous section. Therefore, a histogram of 59 bins (based on the above formulations $P(P-1)+3 = 8(8-1)+3 = 59$ bins) is created as the feature vector for each beat segment. This technique not only discovers local sudden variances and hidden patterns from ECG signal but also has the advantage of having less sensitivity to noise, extracting sparser characteristics, and is computationally effective. Furthermore, all feature vectors regardless of the beat size, have equal length of feature sets.

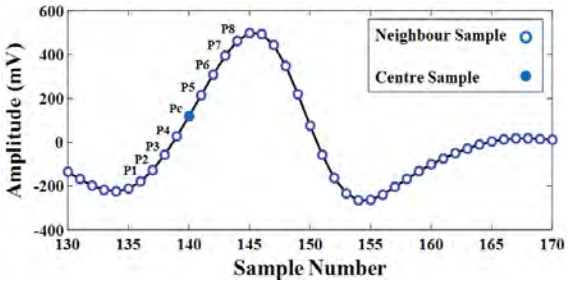



Figure 5: Neighbouring around one sample point of ECG signal for 1-D LBP feature extraction.

# 6 CLASSIFICATION

In this section, we describe our approach for training a classifier to learn the set of arrhythmia classes from a set of the extracted features from ECG beat segments. We then use the remaining features for testing the classifier to predict the class labels of beat segments; we apply 10-fold cross validation (to keep consistency with reference works). The feature set of all ECG beat segments is divided into two randomly selected subsets for training and validation, for 10 times, and the classification approach is applied every time to predict the arrhythmia class labels for the test set. Each time, 90% of the dataset is devoted to the training subset and the rest forms the testing subset. The final accuracy is the average of 10 folds. We employ an extreme learning machine as the proposed classification approach. Feed-forward neural networks are used extensively as classification strategies in medical pattern recognition applications due to their capability in approximating the nonlinear mappings in the data. In order to tune the weights and biases of the network, traditional learning mechanisms such as gradient decent method are employed. However, due to very slow iterative tuning by a gradient decent technique and its convergence into local minima, feed-forward neural networks suffer from slow learning and poor scalability. Extreme learning machine (ELM), as a learning algorithm for single hidden layer feed-forward neural network (FF-NN), is a faster technique. An ELM classifier is generalized single hidden layer neural network with random hidden nodes and determined hidden layer weights without iterative weight tuning (Huang et al., 2006). For N distinct training samples, the single hidden layer FF-

NN with $N_h$ random hidden neurons, L input and K output nodes are modelled as follows.

$$\sum_{j=1}^{N_h} \bar{\lambda}_j F_j(\bar{x}_i) = \sum_{j=1}^{N_h} \lambda_j F\left(\bar{w}_j . \bar{x}_i + \mu_j\right) = \bar{o}_i, \quad where\ i = 1, 2, \dots, N \tag{14}$$

where, $\bar{x}_i = [x_{i1}, x_{i2}, \dots, x_{iL}]^T$ and $\bar{y}_i = [y_{i1}, y_{i2}, \dots, y_{iK}]^T$ are input and output nodes, $F(.)$ is the activation function of network, $\mu_j$ is threshold of $j^{th}$ hidden node and $\bar{w}_j = \left[w_{j1}, w_{j2}, \dots, w_{jL}\right]^T$ and $\bar{\lambda}_j = \left[\lambda_{j1}, \lambda_{j2}, \dots, \lambda_{jK}\right]^T$ denote the weight vectors between the $j^{th}$ hidden node and the input and output nodes, respectively. $N$ samples can be approximated to have zero error means such that,

$$\sum_{j=1}^{N_h} \lambda_j F\left(\bar{w}_j . \bar{x}_i + \mu_j\right) = \bar{y}_i \tag{15}$$

where, (15) can be denoted as follows.

$$H\Lambda = Y \tag{16}$$

where, $\Lambda = \left[\bar{\lambda}_1{}^T, \bar{\lambda}_2{}^T, \dots, \bar{\lambda}_{N_h}{}^T\right]^T$ and $Y = \left[\bar{y}_1{}^T, \bar{y}_2{}^T, \dots, \bar{y}_N{}^T\right]^T$ and $H$ is the hidden layer matrix, the $l^{th}$ column of which is the output of $l^{th}$ hidden node. It is proven in (Huang et al., 2006) that if $F(.)$ is infinitely differentiable, then we can assign random values to the weights and biases and the required hidden layer nodes is $N_h \leq N$. Therefore, in the ELM technique, rather than tuning the weights and biases iteratively in gradient descent method, they are randomly assigned in the beginning of learning. Then, $H$ is calculated and output weights are obtained through the following minimum norm least squares solution of (16),

$$\widehat{\Lambda} = H^{\dagger} Y \tag{17}$$

where, $H^{\dagger}$ is the Moore-Penrose generalized inverse of $H$ (Huang et al., 2006).

# 7 EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed algorithm for arrhythmia recognition and classification, cross validation is applied on the entire MIT-BIH arrhythmia database (110094 beat segments). Sensitivity and precision of classification of each beat type are calculated using true positive ($TP$), false positive ($FP$) and false negative ($FN$) as follows and shown in Table 1.

$$Sensitivity\% = \frac{TP}{TP + FN} \times 100, \tag{18}$$

$$Precision\% = \frac{TP}{TP + FP} \times 100. \tag{19}$$

Table 2 shows the total accuracy of the proposed arrhythmia classification approach compared to the previous works in the literature, using the following equation.

$$Accuracy\% = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{20}$$

where, $TN$ is the true negative value of the classification. As shown in Table 2, our proposed method outperforms other reference techniques in the accuracy of beat classification. In the presented work, we adopted the same dataset as what was employed in the studies that are used for comparison, except for the work in (Khalaf et al., 2015), were only 30 ECG recordings were adopted which is much smaller than the 48 recordings in our study.

Our proposed algorithm is fully automatic, compared to the semi-automatic techniques in (Thomas et al., 2015; Khalaf et al., 2015). Based on

Table 1: Sensitivity and Precision of the proposed algorithm for classifying each beat type.

| Beat Class | Sensitivity % | Precision % |
|---|---|---|
| **N** | 97.86 | 98.50 |
| **V** | 96.20 | 98.63 |
| **F** | 92.73 | 96.35 |
| **S** | 90.50 | 94.06 |
| **U** | 78.66 | 82.36 |

Table 2: Comparison of the accuracy of different algorithms for arrhythmia classification.

| Method | Total Accuracy % |
|---|---|
| **DTCWT+Morphological-ANN (Thomas et al., 2015)** | 94.64 |
| **PDHI-SVM/RBF (Elhaj et al., 2016)** | 98.91 |
| **SC-SVM (Khalaf et al., 2015)** | 98.60 |
| **DWT+PCA-NN (Martis et al., 2013a)** | 98.78 |
| **Proposed algorithm** | 98.99 |

the results in Table 2, the classification accuracies in (Martis et al., 2013a) and (Elhaj et al., 2016) are very close to the result of our proposed approach. However, the proposed adaptive segmentation method in this paper reduces the interference of adjacent beats, which is caused by using fixed beat size as in (Martis et al., 2013a) and (Elhaj et al., 2016). Our proposed technique outperforms those approaches by 232 and 89 less misclassifications, respectively.

# 8 CONCLUSIONS

The proposed arrhythmia classification approach introduces a novel adaptive beat segmentation method based on the median value of the R-R intervals which reduces the misclassification due to the inclusion of adjacent beats in each segment. Moreover, applying uniform 1-D LBP on the wavelet coefficients not only reduces the dimensionality of feature space to 59 bins, which makes the proposed algorithm computationally effective, but also extracts local sudden variances and sparser hidden patterns from the ECG signal and has the advantage of having less sensitivity to noise. ELM classification leads to 98.99% accuracy of beat classification of ECG records in the MIT-BIH arrhythmia database, based on the ANSI/AAMI EC57:1998 standard recommendation, which outperforms the performance of the state of the art arrhythmia recognition algorithms in the literature. These types of algorithms create opportunities for automatic methods that can be applied to ECG readings to help cardiologists assess the risk of arrhythmias that may result in sudden cardiac death. This, given the shortage of cardiologists, can enhance our ability to screen people at risk.

# REFERENCES


Ahonen, T., Hadid, A., & Pietikainen, M. (2004, May). Face recognition with local binary patterns. in *Proceedings of 8th European Conference on Computer Vision (ECCV)*, Prague, 469-481.

Ebrahimzadeh, A., & Khazaee, A. (2009). An efficient technique for classification of electro-cardiogram signals. *Advances in Electrical and Computer Engineering*, 9(3), 89-93.

Elhaj, F. A., Salim, N., Harris, A. R., Swee. T. T., & Ahmed, T. (2016). Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals. *Computer Methods and Programs in Biomedicine*, 127, 52-63.

Emadi, M., Khalid, M., Yusof, R., & Navabifar, F. (2012). Illumination normalization using 2D wavelet. *Procedia Engineering*, 41, 854-859.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., …, Stanley, H. E. (2000). PhysioBank, PhysioToolkit and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23), E215-E220.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3), 489-501.

Inan, O. T., Giovangrandi, L., & Kovacs, G. T. (2006). Robust neural network based classification of premature ventricular contraction using wavelet transform and timing interval feature. *IEEE Transactions on Biomedical Engineering*, 53(12), 2507-2515.

Jokic, S., Delic, V., Peric, Z., Krco, S., & Sakac, D. (2011). Efficient ECG modeling using polynomial functions. *Electronics and Electrical Engineering*, 4(110), 121-124.

Kadambe, S., & Srinivasan, P. (2006). Adaptive wavelets for signal classification and compression. *International Journal of Electronics and Communications*, 60(1), 45-55.

Kaya, Y., Uyar, M., Tekin, R., & Yıldırım, S. (2014). 1D-local binary pattern based feature extraction for classification of epileptic EEG signals. *Applied Mathematics and Computation*, 243, 209-219.

Khalaf, A. F., Owis, M. I., & Yassine, I. A. (2015). A novel technique for cardiac arrhythmia classification using spectral correlation and support vector machines. *Expert Systems with Applications*, 42(21), 8361-8368.

Khoshnoud, S., & Ebrahimnezhad, H. (2013). Classification of arrhythmias using linear predictive coefficients and probabilistic neural network. *Applied Medical Informatics*, 33(3), 55-62.

Korurek, M., & Dogan, B. (2010). ECG beat classification using particle swarm optimization and radial basis function neural network. *Expert Systems with Applications*, 37(12), 7563-7569.

Martis, R. J., Acharya, U. R., & Min, L. C. (2013a). ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. *Biomedical Signal Processing and Control*, 8(5), 437-448.

Martis, R. J., Acharya, U. R., Mandana, K. M., Raya, A. K., & Chakraborty, C. (2013b). Cardiac decision making using higher order spectra. *Biomedical Signal Processing and Control*, 8(2), 193-203.

Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3), 45-50.

Nikan, S., & Ahmadi, M. (2015). Local gradient-based illumination invariant face recognition using LPQ and multi-resolution LBP fusion. *IET Image Processing*, 9(1), 12-21.

Oster, J., Behar, J., Sayadi, O., Nemati, S., Johnson, A. E., & Clifford, G. D. (2015). Semisupervised ECG ventricular beat classification with novelty detection

based on switching kalman filters. *IEEE Transactions on Biomedical Engineering*, 62(9), 2125-2134.

Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3), 230-236.

Ródenas, J., García M., Alcaraz, R., & Rieta, J. J. (2015). Wavelet entropy automatically detects episodes of atrial fibrillation from single-lead electrocardiograms. *Entropy*, 17(9), 6179-6199.

Thomas, M., Das, M. K., & Ari, S. (2015). Automatic ECG arrhythmia classification using dual tree complex wavelet based features. *International Journal of Electronics and Communications*, 69(4), 715-721.

Yu, S. N., & Chen, Y. H. (2007). Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network. *Pattern Recognition Letters*, 28(10), 1142-1150.

# Using HL7 and DICOM to Improve Operational Workflow Efficiency in Radiology

Thusitha Mabotuwana and Christopher S. Hall
*Radiology Solutions, Philips Healthcare, Seattle, WA, U.S.A.*
*{thusitha.mabotuwana, christopher.hall}@philips.com*

Keywords: DICOM, HL7, Medical Informatics Applications, PACS, Radiology Analytics, Radiology Informatics, Radiology Workflow.

Abstract: Radiology departments are increasingly asked to do more with less annual budget and to remain competitive while managing bottom lines. Identifying opportunities to improve workflow efficiency is an important aspect of managing a department and reducing associated costs. Workflow enhancement tools can be built by making use of HL7 and DICOM messages that are directly related to various workflow steps. In this paper, we discuss the importance of using both HL7 and DICOM to determine more accurate metrics related to granular workflow operations, such as distinguishing between billing and operational exam volumes. Using a production dataset, we also demonstrate how visualization can be used to provide better visibility into routine radiology operations.

## 1 INTRODUCTION

For many years, a hospital's radiology department has functioned as a key profit center. In 2007, radiology accounted for 37% of outpatient profit, defined as revenue less direct costs, making imaging the most valuable hospital outpatient service line (The Advisory Board Company 2008). However, with significant increases to healthcare related spending in recent years, projected to be close to 20 percent of the US GDP by 2024 (Centers for Medicare & Medicaid Services), there has been a strong emphasis towards moving away from the traditional fee-for-service model to alternative reimbursement models.

In the traditional fee-for-service payment model, providers are reimbursed by insurers for each service provided. Unnecessary imaging alone is reported to waste at least $7 billion annually in the US (peer60). Since each service gets reimbursed, there is no major incentive for hospitals to minimize costs associated with these tests while the insurer has an open-ended economic risk. On the other hand, with capitated payment models, the economic risk shifts to the hospital since the hospital only gets reimbursed a fixed amount to treat a specific condition (Centers for Medicare & Medicaid Services). With specific healthcare reforms currently underway in the US, there has been a strong focus toward integrated care delivery while reducing costs – for instance, under the new Accountable Care Organization payment model, starting from 1st April 2016, hip and knee replacement payments will be based not only on the procedures performed, but on the quality of care delivered as well (Centers for Medicare & Medicaid Services). Similarly, starting from around 2011, various radiology procedures have been getting paid under 'bundled codes' when two or more related studies are performed together.

The American College of Radiology routinely monitors changes to radiology-related payments and recently reported that the bundled code payments are falling short of the payment levels of the predecessor codes and values; for instance, computed tomography (CT) abdomen-pelvis without contract exams were paid at $418.43 prior to using bundled codes; in 2013, under the bundled payment model this was reduced to $306.05 and in 2014, this was further reduced to $241.79. With such changes to reimbursements, and in an attempt to reduce costs associated with unnecessary imaging, radiology has gradually been shifting from one of the primary profit-centers for a hospital to a cost-center. Radiology departments are increasingly being asked to do more with less annual budget and to remain competitive and manage bottom lines. Radiology departments need to optimize quality of care, patient experience, outcomes, efficiency and throughput while reducing costs.



Mabotuwana, T. and Hall, C.
Using HL7 and DICOM to Improve Operational Workflow Efficiency in Radiology.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 57-65
ISBN: 978-989-758-213-4

An important aspect of managing a radiology department is to have meaningful insights into the routine operations. This could include fairly straightforward metrics such as the total number of billable exams and exams by modality over a particular time period. However, to identify workflow improvement opportunities it is important to gain visibility into the more granular metrics, such as the difference between billing and operational volume, total patient encounter duration, imaging systems utilization and number of scans by hour of day and/or day of week.

In this paper we discuss a generic approach using two established healthcare information exchange standards, Health Level Seven (HL7) and Digital Imaging and Communications in Medicine (DICOM), to determine metrics important to the operations in a radiology department. The main contribution is the linking of HL7 and DICOM to determine granular workflow steps and the discussion around specific radiology workflow nuances.

# 2 METHODS

## 2.1 Background

Healthcare vendors have embraced the rapid uptake of technology in healthcare and as a result, most hospitals have clinical systems from different vendors to accommodate the needs of various departments – for instance, a computerized physician order entry system (CPOE) may be used for order entry, a hospital information system (HIS) for patient registration, a radiology information system (RIS) for radiology specific functions, an EMR for medical records, a scheduler for scheduling appointments, a billing system for accounting purposes, dictation systems for creating reports and a picture archiving and communication system (PACS) for imaging related tasks. To provide integrated patient care, these different clinical systems need to communicate with each other. HL7 messaging standard is arguably the most widely implemented standard for interoperability in healthcare across the world and allows for the exchange of clinical data between disparate systems (HL7 2016). Similarly, DICOM (The DICOM Standard) is the de facto standard for exchanging medical images. Although system-to-system direct communication may be possible, hospitals often use an HL7 interface engine (HL7 2016) to facilitate information exchange. Figure 1 shows a typical hospital configuration, with a focus on radiology – often, mammography requires dedicated workstations compared to other modalities, such as X-ray (XR) and CT, and as such, is shown separately. In-house systems would typically provide some form of aggregated patient view that combines information from RIS, HIS and laboratory information system.

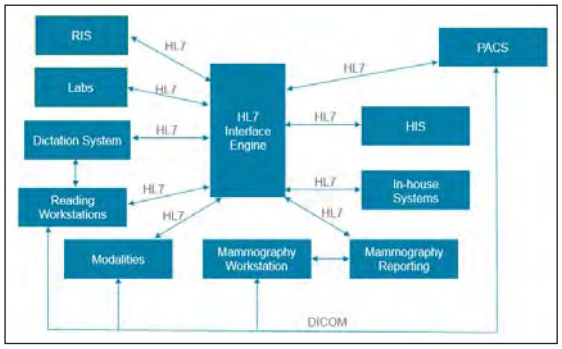



Figure 1: Overview of communication between various clinical systems.

With recent incentives towards increased system interoperability, facilitated by healthcare reforms (e.g., Meaningful Use Stage 2 (HealthIT.gov 2015)), hospitals have been moving towards enterprise electronic health record systems (EHRs) to improve patient care by facilitating sharing of patient data that is typically distributed across multiple clinical systems, and also improve workflow efficiency (e.g., EHRs have a single sign-on where disparate systems will require multiple sign-ons for the same user). However, most clinical systems are 'closed-systems' where the data is not directly accessible to external parties, and often, even to hospital IT administrators. As such, developing tools based directly on HL7 and DICOM can have widespread applicability irrespective of the individual hospital setting.

## 2.2 Reporting for Operational Excellence

There are various systems already in place to provide routine operational reports to radiology department managers, often at a cost center level to which budgets are allocated – definition of a cost center can vary depending on the hospital, but for radiology, it is usually one or more modalities. For instance, high volume modalities such as CT would be a standalone cost center whereas ultrasound (US), nuclear medicine and vascular imaging could be combined into a single cost center. Routine reports may not always be sufficient for operational purposes; for instance, it may be useful to know the machine utilization of a shared US machine and using a cost

center based approach will not capture all exams performed on this resource. Additionally, there are often exams which are split into two or more billing codes although they occupy one scheduled slot.

Karami discusses a comprehensive list of metrics important for radiology across seven main categories (Karami 2014) while other investigators (Morgan, Branstetter et al. 2008, Cook and Nagy 2014) have discussed the importance of analytics and other business intelligence software for radiology. The underlying data source for systems that provide such capabilities can be broadly categorized as:

1. Systems used directly in workflow – these systems are used during routine operations and would include systems such as the EHR, RIS, HIS and PACS. Data is entered directly into these systems.
2. Third-party software that subscribe to HL7 messages – these systems are often setup as a 'listener node' where a copy of all, or a selected subset, of HL7 messages will be sent to, often via the HL7 interface engine. Having an interface engine is not so common for DICOM since PACS is often the only destination for images.
3. Third-party software that integrate with systems used in workflow – these systems often have closely-coupled integration with systems used directly in workflow. For instance, a new CT dose monitoring software application may be installed in a hospital as a new DICOM node and all CT machines can be configured to forward a copy of DICOM structured report (which is a way to analyze dose-related data for CT) to this node.

Due to the specialized nature of clinical software, most of the systems often consume only HL7 or DICOM. However, as discussed later in the paper, there are significant benefits to linking data from these two sources for more accurate metric calculation.

## 2.3 Overview of HL7

An HL7 message is composed of a series of segments with each segment identifying the type of information the message contains (e.g., patient demographics, lab/observation result, diagnosis, insurance and next of kin). In turn, each segment includes one or more composites (also referred to as "fields") that contain the actual information (such as names and result values). Composites can contain sub-composites (or sub-fields) – for instance, patient name is a composite within the 'PID' segment and can contain over six sub-composites (such as family name, given name, middle name and suffix). Composites are typically separated by a "|" character, while sub- composites are usually separated using "^".

Each HL7 message starts with a message header, corresponding to segment MSH, and defines the message's source, purpose, destination, and other syntax specifics like composite delimiters. MSH field 9, denoted by MSH-9, is particularly important since this specifies the type of message that is being transmitted (such as ADT, ORM, ORU, ACK and so on (HL7 2016)). The segments present in a given message vary depending on the type of message that is being transmitted. For instance, Figure 2 shows the composition of an ADT message (used to convey information related to patient admission, discharge

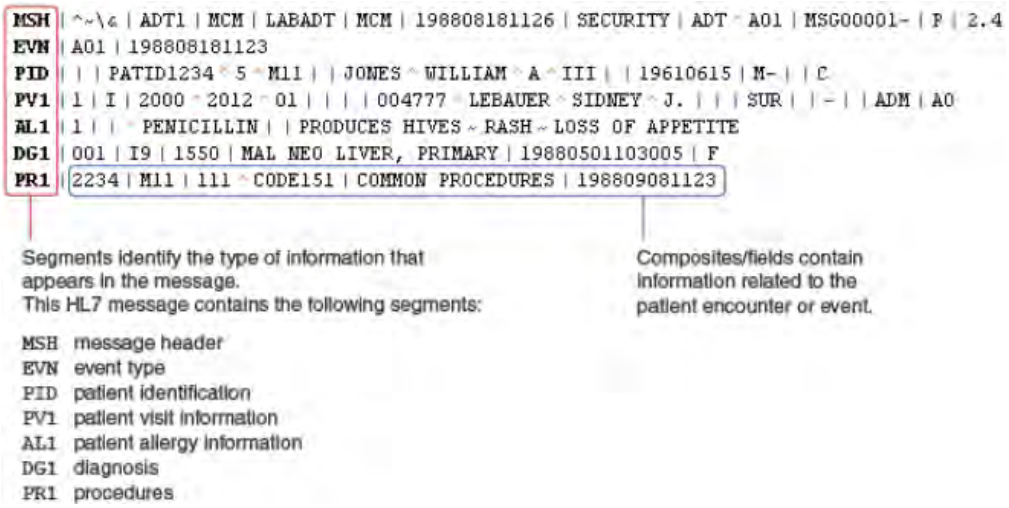



Figure 2: Components of an HL7 ADT message for a fictitious patient (Altova 2016).

and transfers) containing seven segments (MSH, EVN, PID and so on).

Similar to the number of segments within a message type, the number of fields present within a segment can vary as well. For instance, the PID segment can contain over 30 different fields, although it is common for the segments to terminate after the last non-empty field (corresponding to value “C” in Figure 2).

## 2.4 Overview of DICOM

DICOM is a specification for creation, transmission, and storage of medical images and report data (The DICOM Standard). In addition to the binary pixel data, all DICOM files contain metadata related to patient (e.g., name, gender and date of birth), acquisition setup (e.g., type of equipment used and settings such as source IP address and machine name), and study (such as study description). Metadata is contained in the DICOM header which is essentially a list of key-value pairs – the keys are standardized values in hexadecimal. As an example, tag (0008,1030) corresponds to the study description.

## 2.5 Typical Radiology Workflow and Information extraction from HL7 and DICOM

At a high level, once a referring physician has ordered an imaging exam, the exam gets scheduled (after necessary pre-procedure steps are completed, such as pre-authorization from insurance). Each imaging exam will be associated with one imaging order. When the patient arrives at the radiology department, the front desk staff would typically ‘arrive’ the patient in the EHR (could be the RIS or some other system depending on the hospital configuration). At this point, the technologist knows that the patient has arrived for the scan (this could be by looking at a ‘technologist view’ in the EHR/‘modality worklist’, or some other means, such as the front desk staff printing out a ‘patient requisition form’ and handing over to a technologist). When the technologist is ready for the patient, he/she will go to the patient waiting area and call for the patient. After explaining the process, the technologist will start preparing the patient for the scan, for instance, by giving oral contrast if needed. Once ready, the patient will move into the scanning room and around the same time, the technologist will ‘start exam’ in the EHR. The DICOM images get acquired at this point and sent to a modality workstation. The RIS/EHR/PACS systems typically work independent of the modality workstation. At the end of the scan, the technologist will review and push the images from workstation to the PACS and then ‘end exam’ in the EHR. At this point, the images are ready to be reviewed by a radiologist. All these workflow steps trigger HL7 messages. The end-to-end radiology workflow from order-to-report is more extensive as discussed by McEnery (McEnery 2013), but the image acquisition process is where combining data from HL7 and DICOM is most relevant. As such, we focus only on this part of the workflow.

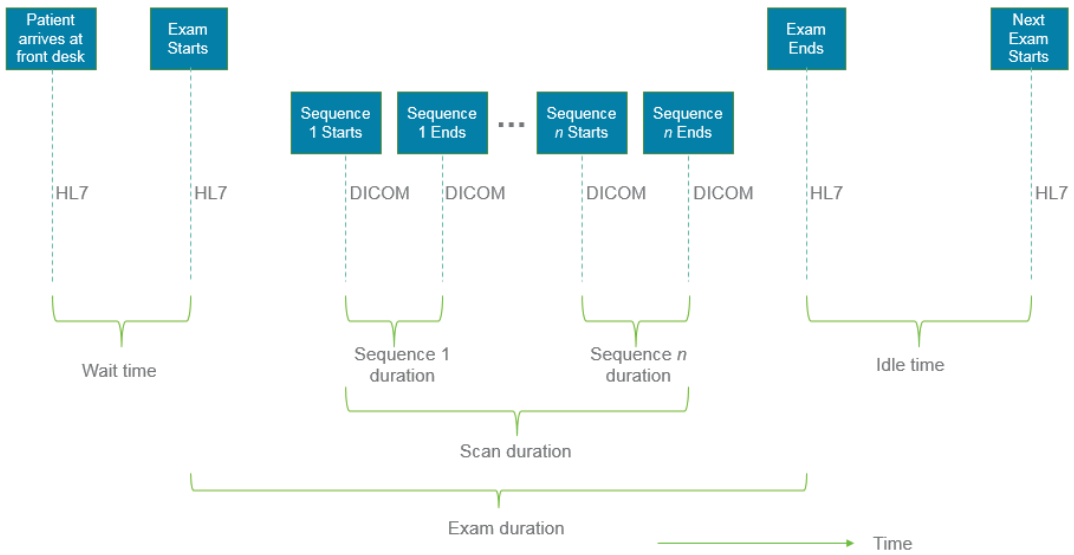



Figure 3: Status messages that get triggered during radiology workflow.

Table 1: Events required to determine workflow metrics for a CT Abdomen-Pelvis exam.

| Metric | Data Source | Event(s) / Segment | Comments |
|---|---|---|---|
| Patient wait time | HL7 | ORM^001: ORC-5 (order status=ARRIVED); ORM^001: ORC-5 (order status=BEGIN) | Value is difference between the two events; e.g., 25 minutes |
| Total scan time | DICOM | Acquisition times from 1st and last image in PACS using (0008,0032) | Value is difference between the two timestamps of images; e.g., 18 minutes |
| Sequence time | DICOM | Acquisition times from 1st and last image in PACS using (0008,0032) for each Series grouped by series UID (0020,000E) | Value is difference between the two timestamps of images for each series; e.g., 4 minutes for Series 1; 7 minutes for Series 2 |
| Begin-to-End time | HL7 | ORM^001: ORC-5 (order status=BEGIN); ORM^001: ORC-5 (order status=END) | Value is difference between the two events; e.g., 23 minutes |
| Arrive-to-End time | HL7 | ORM^001: ORC-5 (order status=ARRIVED); ORM^001: ORC-5 (order status=END) | Value is difference between the two events; e.g., 48 minutes |
| Billing exam volume | HL7 | Accession number count based on ORM^001: OBR-3 | For a CT Abdomen-Pelvis exam, 2 orders will be placed; exam volume is 2 |
| Operational volume | DICOM | Accession number count using (0008,0050) | For a CT Abdomen-Pelvis exam, only 1 physical scan is performed. |
| Machine utilization | DICOM | Performed machine is identified using AE Title tag (0073,1003) | Calculated using some interval (e.g., 1hr) minus sum of total scan times per AE Title |
| Technologist productivity | HL7 + DICOM | ORM, OBR-34 – operator name; accession from DICOM and HL7 | Calculated using operational volume per some interval (e.g., 1hr) per technologist |

The various status messages that get triggered during the different steps of the radiology workflow are shown in Figure 3. Table 1 shows a few important metrics most radiology department track along with the HL7/DICOM field(s) that can be used to calculate the value.

A radiology exam is identified by a unique accession number. This can be determined using the value in HL7 ORM^001 OBR-3 segment or DICOM (0008,0050) tag. Accession number is then used to join between HL7 and DICOM data to determine the accurate value using one or both data sources.

## 2.6 Dataset

Through a product co-creation agreement with an integrated care delivery network, we had access to a database that stored all HL7 and DICOM traffic that was sent from the radiology department to the PACS since June-2015. The database was within the hospital premises in a secure data center with restricted user access. All metrics computed were at an aggregate level with no PHI exposed, and no data left the hospital environment. As of 31-May-2016, the database contained over 13 million HL7 messages over 120 million DICOM records.

# 3 RESULTS

## 3.1 Workflow Considerations

Here we discuss seven important aspects that need to be considered when specific metrics are calculated for operational purposes, with a focus on the power of combining data from HL7 and DICOM.

### 3.1.1 Billed vs Performed Exams

Study volume is essentially the number of unique accession numbers. This is the fundamental chargeable unit for a radiology department, and as such, many clinical systems will produce this volume report on a scheduled basis (typically weekly or monthly). These reports are often driven by financial reporting requirements, and as such, will contain only the billing exam volume. As illustrated in Table 1, this means that a CT abdomen-pelvis study where the images are acquired in a single scan will get reported as two billable exams since there will be two orders associated with the scan. However, it is important to know the operational study volume as well since this can have a significant impact on metrics such as number of exams performed on a machine and the number of scans a given technologist has performed

– it takes significantly longer, in fact nearly twice as long, to perform two CT abdomen exams on two patients (due to various changeover and documentation times) than to perform two scans on one patient. As a result, from an operations point of view, it may not be accurate to say that one technologist who has performed two billable exams on the same patient has been as efficient as another technologist who has performed two exams on two different patients (assuming everything else is comparable).

Distinguishing between billable and performed exams may or may not have a significant impact depending on the study mix performed at a given institute. For instance, in our dataset, for a certain day, there were 891 total billable exams based on HL7 messages whereas there were only 829 exams based on DICOM. In general, the difference was between 5-10%.

It should be noted that the ability to use the accession count from DICOM to determine operational volume depends on the particular hospital's workflow. Some hospitals, including the one in our study, typically scan all images under a single accession number, push them to the PACS, and then either split, or link the images to the accession numbers associated with the different orders. Alternatively, the splitting can happen at the modality workstation itself, in which case two accession numbers (in the CT abdomen-pelvis example) will be seen in DICOM. In this case, the reporting engine will need to perform some logic, such as 'same patient, same acquisition times for different accession numbers' to determine which studies should be merged for operational reporting purposes.

### 3.1.2 Exams with Multiple Modalities

Studies where multiple modalities are involved are identified using the same accession number. A few examples of such studies are PET-CT, PET-MR and interventional radiology exams (which may often involve XR and/or ultrasound and/or CT). In each instance, the complete exam will often be billed under a single accession number, although from an operations point of view, two (or more) resources were utilized to perform the exam. Images acquired from different modalities can be determined using DICOM Source AE Title tag. These exams need to be correctly accounted for when determining relevant metrics (such as operational volume, technologist productivity and machine utilization).

### 3.1.3 Shared Resources

It is common practices for different departments within radiology to share resources. For instance, a PET/CT machine may be used mainly for PET scans, but due to low PET volumes, the CT department may often make use of this resource to perform certain CT exams during busy periods. If PET and CT are different cost centers, PET and CT volumes will be shown separately for each departments, but for machine utilization, both volumes need to be accounted for.

### 3.1.4 Manual vs Automated Timestamps

Care must be taken when calculating various turnaround times using timestamps. For instance, per Figure 3, scan duration is calculated using times from the DICOM header. These times will often be reliable since these are machine generated timestamps. On the other hand, depending on the clinical system, exam start and end HL7 messages may be trigged manually. This flexibility is provided often for valid practical reasons, for instance, after acquiring all images for a CT exam, a technologist may have time to 'end exam' in the system only after scanning a new emergency patient (i.e., back-time the value for the previous exam). Similarly, 'start exam' time may be entered manually and may depend on the individual technologist – some technologists may consider the start of exam to be when they call the patient from the waiting room, some may consider the start to be when the patient walks into the scanning room, while others may consider start of the exam when the patient is on the scanner itself. As such, it is important to standardize the terminology associated with granular workflow steps. If the workflow can be standardized so that all technologists start the exam when they go to get the patient from the waiting room, then the time difference between 'patient arrived' and 'exam start' HL7 messages will accurately reflect patient wait time while the difference between 'exam start' HL7 message and 'first DICOM image' timestamp will show the overhead associated with getting the patient on the scanner (which could be significant for obese and/or immobile patients) and adjusting the scanner settings prior to image acquisition.

### 3.1.5 Same Information in HL7 and DICOM

Some data can be available in both HL7 and DICOM. Either source can be used if the value in both sources is the same (such as the accession number), but there could be instances where same data is entered slightly

differently depending on the clinical system in use. For instance, when a technologist completes an exam in the EHR/RIS, the resulting HL7 'end exam' message will contain the complete operator name. On the other hand, the technologist also needs to enter the name into the modality workstation; however, if all reporting is EHR/RIS driven, technologists will often enter only their initials into DICOM since this information is not used anywhere. Therefore, it is important to identify the right data source and merge data from either HL7 or DICOM after identifying the study based on accession number.

### 3.1.6 Site-specific Business Logic

It is important to give priority to any site-specific business logic since these are used in routine operations. For instance, at the DICOM level, the modality for X-ray may be CR or DR (indicating analog vs digital X-ray respectively) whereas operational managers may consider all of them to be XR. Similarly, cancelled exams and historical data imports should not count towards exam volume, although HL7/DICOM traffic related to such exams may be visible on the network. It is important to accurately capture and implement such site specific business logic when making inferences from raw data.

### 3.1.7 Workflow Related Nuances

Given the diversity and complexity of various radiology exams, there could be various workflow specific nuances. For instance, certain MR exams may require post-processing of images, which can take up to an hour (post-processing usually happens on a separate machine while the technologist is start scanning the next patient). Radiologists can typically start reading exams as soon as a technologist has ended an exam. If a technologist ends the exam after post-processing is complete, and uses the current time as the end exam time, then it would appear as if the exam took a long time to complete. On the other hand, if the technologists back-times the study end time to when the exam truly ended (ignoring all the post-processing time), it would appear as if the exam has been waiting in the reading queue for a long time which affects the report-turnaround time. As such, it is important to agree upon how to interpret the turnaround times in context.

## 3.2 Identifying Workflow Improvement Opportunities

Using HL7 end exam messages, we determined the monthly study volumes (Figure 4) as well as the volume by day of week and hour of day (Figure 5) for MR, CT and XR.

For CT and XR, the heatmap representation indicates the times when most scans are completed – as expected, this is during normal business hours – Monday to Friday between 8am to 6pm. On the other hand, the MR heatmap suggests that there is some unusually high activity happening later in the day, between 10 and 11 pm for the months of March and April.

Upon investigation, the MR technologists confirmed that they routinely end exams only towards the end of the day, typically during the shift change.

| | MR | | | | CT | | | | XR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hour | January | February | March | April | January | February | March | April | January | February | March | April |
| 0 | 26 | 30 | 33 | 24 | 48 | 78 | 82 | 88 | 115 | 94 | 95 | 98 |
| 1 | 32 | 28 | 39 | 40 | 66 | 73 | 60 | 57 | 79 | 84 | 83 | 91 |
| 2 | 49 | 31 | 26 | 36 | 69 | 55 | 58 | 64 | 45 | 50 | 66 | 56 |
| 3 | 14 | 17 | 25 | 30 | 40 | 49 | 52 | 56 | 129 | 145 | 210 | 181 |
| 4 | 8 | 13 | 21 | 20 | 53 | 33 | 65 | 47 | 521 | 389 | 389 | 418 |
| 5 | 31 | 31 | 23 | 35 | 68 | 42 | 39 | 63 | 439 | 284 | 287 | 266 |
| 6 | 45 | 27 | 34 | 38 | 52 | 45 | 56 | 37 | 175 | 150 | 207 | 191 |
| 7 | 45 | 57 | 45 | 44 | 117 | 104 | 113 | 98 | 233 | 228 | 233 | 195 |
| 8 | 65 | 50 | 62 | 55 | 153 | 126 | 145 | 156 | 467 | 410 | 501 | 484 |
| 9 | 67 | 75 | 72 | 68 | 186 | 180 | 190 | 184 | 655 | 619 | 700 | 613 |
| 10 | 71 | 58 | 60 | 65 | 232 | 202 | 224 | 181 | 747 | 769 | 871 | 768 |
| 11 | 77 | 80 | 77 | 75 | 255 | 226 | 264 | 240 | 799 | 786 | 896 | 774 |
| 12 | 91 | 74 | 92 | 77 | 263 | 252 | 310 | 288 | 841 | 779 | 913 | 818 |
| 13 | 101 | 80 | 75 | 88 | 278 | 267 | 300 | 256 | 749 | 726 | 857 | 743 |
| 14 | 95 | 97 | 95 | 89 | 243 | 258 | 310 | 252 | 778 | 730 | 824 | 742 |
| 15 | 78 | 85 | 79 | 88 | 260 | 250 | 268 | 229 | 661 | 665 | 705 | 698 |
| 16 | 101 | 85 | 76 | 75 | 233 | 228 | 278 | 252 | 552 | 507 | 579 | 556 |
| 17 | 83 | 80 | 103 | 82 | 228 | 221 | 220 | 236 | 384 | 432 | 416 | 399 |
| 18 | 86 | 91 | 85 | 80 | 189 | 183 | 172 | 210 | 348 | 308 | 362 | 361 |
| 19 | 91 | 89 | 57 | 78 | 158 | 189 | 166 | 142 | 243 | 291 | 276 | 281 |
| 20 | 66 | 74 | 64 | 67 | 149 | 151 | 150 | 135 | 223 | 207 | 205 | 217 |
| 21 | 65 | 66 | 80 | 71 | 102 | 106 | 109 | 145 | 169 | 189 | 191 | 212 |
| 22 | 68 | 85 | 95 | 92 | 105 | 99 | 129 | 137 | 144 | 142 | 159 | 166 |
| 23 | 71 | 48 | 40 | 65 | 70 | 84 | 88 | 82 | 92 | 123 | 126 | 113 |

Figure 4: Monthly exam volume by hour of day.

| | MR | | | | | | | CT | | | | | | | XR | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hour | Sun | Mon | Tue | Wed | Thu | Fri | Sat | Sun | Mon | Tue | Wed | Thu | Fri | Sat | Sun | Mon | Tue | Wed | Thu | Fri | Sat |
| 0 | 6 | 9 | 18 | 25 | 22 | 18 | 15 | 49 | 36 | 39 | 40 | 46 | 40 | 46 | 56 | 44 | 53 | 62 | 61 | 50 | 76 |
| 1 | 20 | 16 | 20 | 18 | 30 | 22 | 13 | 34 | 38 | 33 | 34 | 25 | 48 | 44 | 44 | 47 | 57 | 50 | 45 | 44 | 50 |
| 2 | 8 | 8 | 11 | 33 | 29 | 24 | 29 | 50 | 37 | 31 | 38 | 26 | 24 | 40 | 22 | 29 | 27 | 34 | 36 | 37 | 32 |
| 3 | 9 | 6 | 12 | 11 | 23 | 12 | 13 | 26 | 32 | 31 | 43 | 16 | 28 | 21 | 111 | 87 | 103 | 74 | 78 | 95 | 117 |
| 4 | 15 | 9 | 5 | 7 | 4 | 14 | 8 | 28 | 29 | 29 | 19 | 29 | 26 | 38 | 219 | 252 | 239 | 198 | 249 | 286 | 274 |
| 5 | 6 | 12 | 13 | 22 | 27 | 20 | 20 | 36 | 34 | 44 | 34 | 17 | 28 | 19 | 204 | 194 | 176 | 192 | 159 | 175 | 176 |
| 6 | 12 | 12 | 30 | 23 | 24 | 24 | 19 | 28 | 33 | 27 | 37 | 17 | 24 | 24 | 82 | 75 | 134 | 118 | 123 | 100 | 91 |
| 7 | 5 | 33 | 29 | 44 | 40 | 35 | 5 | 20 | 80 | 101 | 72 | 64 | 77 | 18 | 29 | 130 | 191 | 159 | 173 | 168 | 39 |
| 8 | 13 | 51 | 40 | 40 | 37 | 35 | 16 | 29 | 99 | 99 | 109 | 106 | 100 | 38 | 119 | 279 | 404 | 326 | 326 | 265 | 143 |
| 9 | 22 | 39 | 50 | 41 | 53 | 49 | 28 | 38 | 136 | 107 | 136 | 112 | 152 | 59 | 132 | 438 | 501 | 487 | 422 | 428 | 179 |
| 10 | 27 | 44 | 34 | 40 | 37 | 41 | 31 | 57 | 152 | 141 | 147 | 130 | 142 | 70 | 130 | 535 | 590 | 631 | 593 | 509 | 167 |
| 11 | 16 | 53 | 51 | 54 | 53 | 51 | 31 | 73 | 171 | 151 | 174 | 161 | 174 | 81 | 154 | 544 | 636 | 597 | 605 | 562 | 157 |
| 12 | 51 | 45 | 56 | 52 | 49 | 38 | 43 | 68 | 182 | 190 | 198 | 211 | 186 | 78 | 143 | 573 | 622 | 676 | 625 | 547 | 165 |
| 13 | 46 | 50 | 49 | 41 | 46 | 54 | 58 | 62 | 187 | 214 | 202 | 173 | 191 | 72 | 133 | 533 | 619 | 558 | 605 | 484 | 143 |
| 14 | 44 | 55 | 51 | 58 | 60 | 55 | 53 | 65 | 189 | 201 | 172 | 189 | 185 | 62 | 146 | 502 | 658 | 623 | 580 | 431 | 134 |
| 15 | 43 | 46 | 43 | 46 | 47 | 41 | 64 | 74 | 172 | 167 | 185 | 181 | 156 | 72 | 135 | 485 | 538 | 556 | 467 | 392 | 156 |
| 16 | 54 | 47 | 40 | 45 | 47 | 47 | 57 | 56 | 158 | 192 | 185 | 180 | 163 | 57 | 120 | 352 | 415 | 447 | 395 | 321 | 144 |
| 17 | 47 | 51 | 49 | 55 | 47 | 46 | 53 | 62 | 170 | 166 | 139 | 140 | 162 | 66 | 112 | 303 | 297 | 302 | 268 | 238 | 111 |
| 18 | 44 | 46 | 43 | 45 | 47 | 51 | 66 | 76 | 126 | 124 | 118 | 119 | 117 | 74 | 110 | 254 | 229 | 215 | 251 | 201 | 119 |
| 19 | 37 | 43 | 53 | 38 | 52 | 39 | 53 | 54 | 97 | 110 | 125 | 94 | 125 | 50 | 80 | 186 | 197 | 188 | 173 | 159 | 108 |
| 20 | 12 | 39 | 50 | 55 | 44 | 45 | 26 | 46 | 103 | 88 | 91 | 87 | 86 | 84 | 90 | 130 | 128 | 127 | 133 | 126 | 118 |
| 21 | 16 | 41 | 46 | 60 | 50 | 42 | 27 | 54 | 77 | 57 | 56 | 68 | 96 | 54 | 105 | 116 | 115 | 130 | 116 | 96 | 83 |
| 22 | 13 | 56 | 62 | 60 | 67 | 61 | 21 | 46 | 63 | 71 | 80 | 67 | 80 | 63 | 82 | 91 | 72 | 92 | 88 | 98 | 88 |
| 23 | 15 | 34 | 42 | 43 | 37 | 41 | 12 | 51 | 47 | 58 | 53 | 32 | 49 | 34 | 63 | 61 | 70 | 65 | 61 | 78 | 56 |

Figure 5: Exam volume by day of week and hour of day.

# 4 DISCUSSION

In this paper we have discussed the importance of using both HL7 and DICOM to determine various metrics related to the operations of a radiology department. While using a single data source may be a good approximation, it is important to take an integrated approach in order to get better visibility into more granular operations as well as determine more accurate values for the metrics of interest.

A radiology department needs to create clear definitions of metrics; even the seemingly obvious terms such as “start of exam” need to be explicitly tied to workflow steps and the electronic measurements using HL7 and DICOM. This “data governance” is an important aspect of the data analytics and process improvement approach. Data governance should define clearly the metrics, agree on the measurement methodology, understand the exceptions cases where the methodology might be imperfect, and serve as a governing body to increase the acceptance of the process improvement initiatives.

In the context of the MR workflow, we have discussed a specific example where technologists were routinely ending exams towards the end of the shift. This may be acceptable for practical reasons, but at the same time, this affects the data quality, which in turn affect the various metrics that are based on this data. As such, it is important for radiology administrators and Department Chairs to proactively set forth suitable guidelines and educate the technologists on the importance of adhering to such guideline. Providing visible feedback to the technologists on a regular basis on the performance of the department may help improve compliance to such requests.

Despite having access to a large dataset, the current study has one main limitation – the dataset is from a single institution, and as such, the DICOM tags we have used may not always be generalizable. Although vendors are expected to follow the standard, they often use private tags (to exchange vendor-specific information that is not covered by the DICOM standard) instead of public tags, and sometimes populate different public tags instead of the commonly used tags; as such, the mapping may need to be modified depending on the site.

Having access to tools to provide visibility into granular workflow operations is crucial for the success of radiology departments. However, as discussed, developers of such tools need to keep in mind the various nuances associated with hospital workflows in order for such tools to be meaningful and widely adopted by all stakeholders.

## REFERENCES


Altova. (2016). "HL7 Technology Overview." Retrieved Sep 7, 2015, from http://www.altova.com/HL7_technology_primer.html.

Centers for Medicare & Medicaid Services. "Comprehensive Care for Joint Replacement Model." Retrieved Sep 17, 2016, from https://innovation.cms.gov/initiatives/cjr.

Centers for Medicare & Medicaid Services. "Health Care Payment Learning and Action Network." Retrieved Sep 15, 2016, from https://www.cms.gov/newsroom/mediareleasedatabase/fact-sheets/2015-fact-sheets-items/2015-02-27.html.

Centers for Medicare & Medicaid Services. "NHE Fact Sheet." Retrieved Sep 17, 2016, from https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nhe-fact-sheet.html.

Cook, T. S. and P. Nagy (2014). "Business intelligence for the radiologist: making your data work for you." J Am Coll Radiol 11(12 Pt B): 1238-1240.

HealthIT.gov. (2015). "Meaningful Use Definition & Objectives." Retrieved Sep 7, 2015, from http://www.healthit.gov/providers-professionals/meaningful-use-definition-objectives

HL7. (2016). "HL7 Interface Engine." Retrieved Sep 8, 2015, from http://www.hl7.com/interface-engine.html.

HL7. (2016). "HL7 Messaging Standard Version 2.7." Retrieved Sep 7, 2015, from http://www.hl7.org/implement/standards/product_brief.cfm?product_id=146.

Karami, M. (2014). "A design protocol to develop radiology dashboards." Acta Inform Med 22(5): 341-346.

McEnery, K. (2013). Radiology Information Systems and Electronic Medical Records. IT Reference Guide for the Practicing Radiologist, American College of Radiology.

Morgan, M. B., B. F. t. Branstetter, D. M. Lionetti, J. S. Richardson and P. J. Chang (2008). "The radiology digital dashboard: effects on report turnaround time." J Digit Imaging 21(1): 50-58.

peer60. "Unnecessary Imaging: Up to $12 Billion Wasted Each Year." Retrieved Sep 17, 2016, from http://research.peer60.com/unnecessary-imaging/.

The Advisory Board Company. (2008). "Imaging Marketing Strategy: Lessons from Progressive Centers on Securing Share." Retrieved Sep 2, 2016, from https://www.advisory.com/-/media/Advisory-com/Research/IPP/Research-Study/2008/Imaging-Marketing-Strategy/17272_IPP_Imaging_Marketing_Strategy.pdf.

The DICOM Standard. "DICOM Homepage." Retrieved Sep 2, 2016, from http://dicom.nema.org/standard.html.

# Technologies for Ageing in Place to Support Home Monitoring of Patients with Chronic Diseases

Alexandra Queirós[1], Luís Pereira[2], Ana Dias[3] and Nelson Pacheco Rocha[4]
[1]*Health Sciences School, University of Aveiro, Campo Universitário de Santiago, 3810-193 Aveiro, Portugal*
[2]*Medtronic Portugal, Torres de Lisboa, Rua Tomás da Fonseca, Torre E - 11º andar, 1600-209 Lisboa, Portugal*
[3]*Department of Economicas, Management, Industrial Engineering and Tourism, University of Aveiro, Campo Universitário de Santiago, 3810-193 Aveiro, Portugal*
[4]*Department of Medical Sciences, University of Aveiro, Campo Universitário de Santiago, 3810-193 Aveiro, Portugal*
*alexandra@ua.pt, luis.pereira@medtronic.com, anadias@ua.pt, npr@ua.pt*

Keywords: Technologies for Ageing in Place, Home Monitoring, Chronic Diseases, Older Adults.

Abstract: Objectives - This study aims to identify: i) the most relevant chronic diseases in terms of the use of technologies for ageing in place to support home monitoring; and ii) types, outcomes and impacts of technologies for ageing in place being used to support home monitoring. Methods - A systematic review of reviews and meta-analysis was performed based on a search of the literature. Results - A total of 35 reviews and meta-analysis across 4 chronic diseases, diabetes, congestive heart failure, chronic obstructive pulmonary disease, and hypertension, were retrieved. These studies compare home monitoring supported by different technologies with usual care. Conclusion - Home monitoring has positive effects in various health related outcomes, but further research is required to allow its incorporation in the clinical practice.

## 1 BACKGROUND

The active ageing concept refers not only to the ability to be physically active or have an occupation, but also to be able to participate in social, economic, cultural, civil or spiritual matters (Kickbusch and Nutbeam, 1998; World Health Organization, 2002). Therefore, the older adults, even when having some kind of pathology or disability, should continue to contribute actively in social terms, together with their family, friends and community (Kickbusch and Nutbeam, 1998). In this context, information technologies have a key role in the promotion of human functioning and in the mitigation of limitations, particularly the ones resulting from the natural ageing process (Queirós, 2013; 2015).

Technological solutions emerge as potentially cost-effective to meet the needs of citizens and to promote the services reorganization (Genet et al., 2011), which are the aims of concepts such as Medicine 2.0 (Eysenbach, 2008), connected health (Kvedar, Coye, and Everett, 2014), or holistic health (Mori et al., 2013; Koch, 2013). In particular, technologies for ageing in place (Connelly, Mokhtari and Falk, 2014) can overcome multiple impairments, including declines in cognitive and functional abilities (Teixeira et al., 2013; Cruz et al., 2013; 2014) and, consequently, can allow older adults to live safely, independently, autonomously, and comfortably, without being required to leave their own residences, but with the necessary support services to their changing needs (Pastalan, 1990).

The present study is part of a medium term project that aims to systematize current evidence of technologies for ageing in place. Particularly, a systematic review of reviews and meta-anaysis was perform to identify technologies being used to support home monitoring of patients with chronic diseases, not specifically designed for older adults, but that can be used by this population, and to analyse how these tecnologies impact health related outcomes.

There are several reviews of reviews related to home care of patients with chronic diseases (Househ, 2014; McBain, Shipley and Newman, 2015; Kitsiou, Paré and Jaana, 2015; Slev, 2016). However, these reviews focus on specific technologies (e.g. short message services (Househ, 2014), or specific pathologies (e.g. congestive heart failure (Kitsiou, Paré and Jaana, 2015)). Therefore, the broad analysis of the study reported in the present article is useful to inform: i) the practitioners



Queirós, A., Pereira, L., Dias, A. and Rocha, N.
Technologies for Ageing in Place to Support Home Monitoring of Patients with Chronic Diseases.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 66-76
ISBN: 978-989-758-213-4

about the available home monitoring solutions; and ii) the researchers about home monitoring issues that are being object of research.

# 2 METHODS

Considered the aforementioned objective, the systematic review of reviews and meta-analysis reported in the present article was informed by the following research questions:

- What are the most relevant chronic diseases in terms of the use of technologies for ageing in place to support home monitoring?
- What are the types, outcomes and impacts of technologies for ageing in place being used to support home monitoring?

In order to determine the most appropriate search strategy, an initial scoping study was conducted. The outcomes of this process were discussed with various researchers and captured in a review protocol with explicit descriptions of the methods to be used and the steps to be taken.

The resources considered to be searched were two general databases (i.e. Web of Science and Scopus) and two specific databases (i.e. PubMed, a medical sciences database, and IEEE Explorer, a technological database).

The list of keywords for the systematic review was created through three steps:

- First, health related and technological terms were selected for a draft search strategy based on the terminology that the authors were familiar due to their background readings. A preliminary search with the identified keywords was tested by two authors.
- Afterwards, the two authors carried out a hand search of the table of contents of three relevant journals: Journal of Telemedicine and Telecare, Telemedicine and Ehealth and Journal of Medical Internet Research.
- Finally, new keywords were introduced in order to gather articles of the mentioned journals that were not retrieved in the previous queries.

The queries that resulted from these successive refinements intended to include: i) all the reviews where any of the keywords ‘telecare’, ‘telehealth’, ‘telemedicine’, ‘homecare’, ‘telemonitoring’, ‘home monitoring’, ‘remote monitoring´, ehealth´, ‘telerehabilitation’, ‘mobile health’, ‘mhealth’ or ‘assisted living’ were presented in the title or abstract; and ii) all the reviews where any the keywords ‘technology-based’, ‘information technology’, ‘information and communication’, ‘internet-based’, ‘web-based’, ‘on-line’, ‘smartphones’, ‘mobile apps’, ‘mobile phone’, ‘monitoring devices’ or ‘consumer health information’ were presented in the title or abstract together with any of the keywords ‘healthcare’, ‘health care’, ‘patient’, ‘chronic disease’, ‘older’ or ‘elderly’.

The search was limited to articles in English, but conducted in any country, and performed on 30 of April of 2016, to include reviews published during the preceding 10 years.

## 2.1 Inclusion and Exclusion Criteria

The study reported in the present article included reviews and meta-analysis related to technological solutions that can be used to support home monitoring of older adults living with a chronic disease. Chronic disease is defined as an illness that is prolonged in duration, has a non-self-limited nature, is rarely cured completely and is associated with persistent and recurring health problems (Thrall, 2005; Australian Institute of Health and Welfare, 2006).

Since the scientific literature presents a large number of articles that report studies related to home monitoring, it was planned to include systematic reviews or meta-analysis only.

The authors excluded all the articles not published in English or that report systematic reviews of reviews. Furthermore, the authors also excluded all the reviews and meta-analysis reporting solutions that: i) are not focused on the monitoring of health conditions; ii) target more than one chronic condition (e.g. diabetes together with congestive heart failure); iii) target long-term health condition not related to older patients (e.g. paediatric conditions); iv) do not target the patients (i.e. studies that were clinicians focused or were intended primary to deal with the problems of caregivers rather than the patients); and v) were designed to be used in an institutional environment and not in the domicile of the patients.

## 2.2 Review Selection

After the removal of duplicates and articles not published in English, the selection of the remainder articles was performed by two authors in three steps:

- First, the authors assessed all titles for relevance and those clearly not meeting the inclusion criteria were removed.

- Afterwards the abstracts of the retrieved articles were assessed against the inclusion and exclusion criteria.
- Finally, authors assessed the full text of the articles according to the outlined inclusion and exclusion criteria.

In all these three steps any disagreement between the two authors was discussed and resolved by consensus.

### 2.3 Data Extraction

The following characteristics of the retrieved articles were extracted: i) authors, title and year of publication; ii) aims of the review or meta-analysis; iii) target chronic disease; iv) technologies being used; v) search strategy; vi) inclusion and exclusion criteria; vii) quality assessment; viii) data extraction procedure; ix) total number of primary studies; x) total number of random clinical trials (RCT); xi) total number of participants; xii) primary outcomes; xiii) secondary outcomes; xiv) author's interpretations; and xv) author's conclusions.

The relevant data were extracted and recorded independently by two authors. Once more, any disagreement between the two authors was discussed and resolved by consensus.

## 3 RESULTS

The present study comprises a narrative synthesis of the retrieved systematic reviews and meta-analyses and followed the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher, 2009). Figure 1 presents the respective flowchart.

A total of 2681 articles were retrieved from the initial searches on PubMed (822 articles), Web of Science (1263 articles), Scopus (550 articles) and IEEE Explorer (46 articles). The initial screening yielded 1429 articles by removing the duplicates (1210 articles) or the articles without abstracts or without the names of the authors (42 articles). After exclusions based on title alone 563 articles were retrieved. Additionally, 315 articles were eliminated based upon review of their abstracts.

The full texts of the 248 remaining articles were assessed and 213 articles were eliminated, due to the following reasons: i) the studies target multiple chronic diseases - 102 articles; ii) the main goals of the studies are health promotion related to general population - 48 articles; iii) the studies are not focused on home monitoring of patients with chronic diseases - 31 articles; iv) the studies are not systematic literature reviews or meta-analysis - 12 articles - or are reviews of reviews - 3 articles; v) the target users are not the patients but the caregivers - 8 articles; vi) the reported solutions are to be used in an institutional environment or are related to acute conditions - 4 articles; vi) the studies were not reported in English - 5 articles.

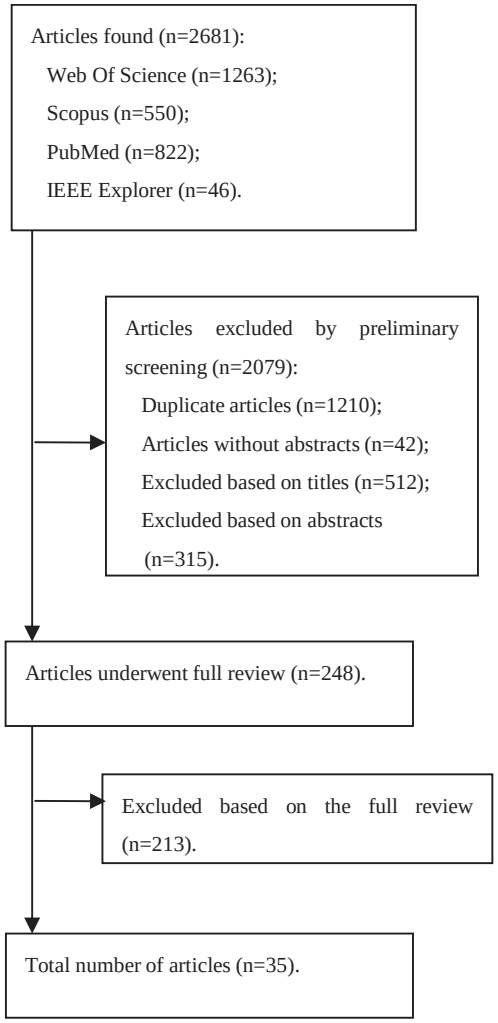



Figure 1: PRISMA Flowchart.

### 3.1 Characteristics of the Studies

The 35 resulting articles from the filtered queries synthesize evidence of home monitoring to support patients with chronic diseases. After an analysis of the full text of the retrieved articles, they were categorized into 4 clinical domains: i) diabetes - 20 articles; ii) congestive heart failure - 9 articles; iii) chronic obstructive pulmonary disease - 5 articles; and iv) hypertension - 1 article. The following

subsections present the results of these 4 categories.

## 3.2 Diabetes

Of the 35 retrieved articles, 20 dealt with home monitoring of patients with diabetes. A significant number of articles focuses both type 1 and type 2 diabetes (Jaana and Paré, 2007; Verhoeven et al., 2007; Baron, McBain and Newman, 2012; El-Gayar, 2013; van Vugt et al., 2013; Or and Tao, 2014; Huang et al., 2015; Tildesley, Po and Ross, 2015; Riazi et al., 2015; Garabedian, Ross-Degnan and Wharam, 2015). Others articles focus type 2 diabetes (Jackson, 2006; Ramadas et al., 2011; Frazetta, Willet and Fairchild, 2012; Cassimatis and Kavanagh, 2012; Tao and Or, 2013; Huang et al., 2015; Hunt, 2015; Ortiz, Felix and Sosa, 2015; Arambepola, 2016). Only one of the retrieved studies focuses exclusively on type 1 diabetes (Peterson, 2014).

By principle, the articles of the diabetes category include primary studies with high quality scientific evidence. All the 20 retrieved articles considered RCT primary studies and 11 of them considered RCT as one of the inclusion criteria (Jaana and Paré, 2007; Verhoeven et al., 2007; Cassimatis and Kavanagh, 2012; Baron, McBain and Newman, 2012; Pal et al., 2013; Tao and Or, 2013; van Vugt et al., 2013; Or and Tao, 2014; Huang et al., 2015; Tildesley, Po and Ross, 2015; Arambepola, 2016). On the other hand, aggregating all the primary studies included in the 20 studies of the diabetes category it is evident that the number of the involved patients is relatively significant (e.g. 1 article reports the involvement of 3578 patients (Pal et al., 2013) and other reports the involvement of 3798 patients (Huang et al., 2015)).

In technological terms, several articles (Ramadas et al., 2011; Frazetta, Willet and Fairchild, 2012; El-Gayar, 2013; Pal et al.; Tao and Or, 2013; van Vugt et al., 2013; Huang et al., 2015; Tildesley, Po and Ross, 2015; Riazi et al., 2015; Hunt, 2015) refer web-based applications (Table 1). In general, these applications allow synchronous (e.g. instant messaging or chat) and asynchronous (e.g. electronic mail or bulletin board) communications together with web pages to register clinical parameters (e.g. weight or blood pressure) and medication.

Besides web-based applications, there are other technological solutions reported in different articles:

- Computer-assisted applications integrating the management of clinical data with electronic practice guidelines, reminder systems, and feedback to the patients (Jackson, 2006; El-Gayar, 2013).
- Smartphones (i.e. standalone smartphones and smartphones integrating specific devices such as glucometers for automatic glucose level upload) (Frazetta, Willet and Fairchild, 2012; Cassimatis and Kavanagh, 2012; Baron, McBain and Newman, 2012; El-Gayar, 2013; Pal et al., 2013; Peterson, 2014; Tildesley, Po and Ross, 2015; Garabedian, Ross-Degnan and Wharam, 2015; Hunt, 2015; Ortiz, Felix and Sosa, 2015; Arambepola, 2016).
- Automatic patient data transmission by means of monitoring devices (i.e. devices to monitor vital signals or devices to monitor behaviour outcomes such as pedometers or accelerometers connected by wireless communications to monitor physical activity (Jaana and Paré, 2007)).
- Video-conference (Verhoeven et al., 2007; El-Gayar et al., 2013).
- Telephone calls (Riazi et al., 2015).

The main outcome of most of the articles included in the diabetes category is the control of glycaemia by using glycosylated haemoglobin (HbA1c) as a proxy. However, in all the studies, this aim is complemented with other health related outcomes (e.g. health related quality of life (Verhoeven et al., 2007; Ramadas et al., 2011; Pal et al., 2013; van Vugt et al., 2013), weight (Ramadas et al., 2011; Pal et al., 2013; Huang et al., 2015; Garabedian, Ross-Degnan and Wharam, 2015), depression (Pal et al., 2013), blood pressure (Verhoeven et al., 2007; Or and Tao, 2014; Riazi et al., 2015; Garabedian, Ross-Degnan and Wharam, 2015), cholesterol level (Ramadas et al., 2011; Or and Tao, 2014), triglycemius level (Or and Tao, 2014), fluctuation index (Ramadas et al., 2011)), behaviour outcomes (e.g. physical activity) (Jackson, 2006; Verhoeven et al., 2007; Ramadas et al., 2011; Cassimatis and Kavanagh, 2012; van Vugt et al., 2013; Riazi et al., 2015; Garabedian, Ross-Degnan and Wharam, 2015; Hunt, 2015; Arambepola, 2016), patient self-motivation (Tildesley, Po and Ross, 2015), patient-clinician communication (Tildesley, Po and Ross, 2015), medication adherence (Cassimatis and Kavanagh, 2012; Hunt, 2015) ), and structural outcomes related to care coordination (Jaana and Paré, 2007; Verhoeven et al., 2007).

Table 1: Articles that focus diabetes.

| Study | Technology | (*) |
|---|---|---|
| Jackson et al., 2006 | Web-based applications, computer assisted applications and standard telephone calls | 26/14 |
| Jaana et al., 2007 | Automatic patient data transmission by means of monitoring devices | 17/17 |
| Verhoeven et al., 2007 | Video-conference | 39/39 |
| Ramadas et al., 2011 | Web-based applications | 13/8 |
| Frazettaet al., 2012 | Smartphones | 7/7 |
| Cassimatis et al., 2012 | Smartphones | 13/13 |
| Baron, et al., 2012 | Smartphones | 24/24 |
| El-Gayar et al., 2013 | Web-based applications, smartphones, computer assisted applications and video-conference | 104/60 |
| Pal et al., 2013 | Web-based applications and smartphones | 16/16 |
| Tao et al., 2013 | Web-based applications | 43/43 |
| van Vugt et al., 2013 | Web-based applications | 13/13 |
| Or et al., 2014 | Web-based applications | 67/67 |
| Peterson | Smartphones | 14/1 |
| Huang et al., 2015 | Web-based applications and standard telephone calls | 18/18 |
| Tildesley et al., 2015 | Web-based applications and smartphones | 22/22 |
| Riazi et al., 2015 | Web-based applications and standard telephone calls | 67/52 |
| Garabedian et al., 2015 | Smartphones | 1479 |
| Hunt, 2015 | Web-based applications and smartphones | 14/9 |
| Ortiz et al., 2015 | Smartphones | 8/4 |
| Arambepola et al., 2016 | Smartphones | 13/13 |

(*) Number of RCT included in the review / Number of primary studies include in the review.

Most of the articles of the diabetes category report moderate to large significant reduction of HbA1c when compared with usual care (Jackson, 2006; Jaana and Paré, 2007; Frazetta, Willet and Fairchild, 2012; Cassimatis and Kavanagh, 2012; Tao and Or, 2013; Or and Tao, 2014; Peterson, 2014; Huang et al., 2015; Tildesley, Po and Ross, 2015; Riazi et al., 2015; Garabedian, Ross-Degnan and Wharam, 2015; Hunt, 2015; Ortiz, Felix and Sosa, 2015; Arambepola, 2016). However, several studies are not conclusive about the reduction of HbA1c (Verhoeven et al., 2007; Ramadas et al., 2011; Baron, McBain and Newman, 2012; Pal et al., 2013). In particular, computer-based diabetes self-management interventions (Pal et al., 2013) and consultations supported by video-conference (Verhoeven et al., 2007) appear to have a small beneficial effect on glycaemia control.

An article (El-Gayar et al., 2013) reporting research gaps of the technological approaches identifies the need to improve the usability of the applications as well the need for more comprehensive solutions, including real-time feedback to the patients and the integration of electronic health records systems supporting the service providers.

## 3.3 Congestive Heart Failure

The number of RCT and non-RCT primary studies included in the 9 articles dealing with congestive heart failure varies from 9 to 42 (Table 2). The majority of the articles (i.e. 6 articles (Chaudhry et al., 2007; Dang, Dimmick and Kelkar, 2009; Polisena et al., 2010a; Ciere, Cartwright and Newman, 2012; Conway, Inglis and Clark, 2014; Nakamura, Koga and Iseki, 2014)) considered RCT as one of the inclusion criteria.

Considering the supporting technologies (Table 2), automatic patient data transmission by means of monitoring devices is being used together with video-conference and standard telephone calls to allow the assessment of symptoms and vital signs, as well as the transmission of automatic alarms.

In terms of clinical outcomes, the main concerns are the impacts of home monitoring in heart failure-related hospitalizations and all-cause mortality (Conway, Inglis and Clark, 2014) when compared with usual care. However, several secondary outcomes are also considered such as self-care behaviour (e.g. adherence to prescribed medication, daily weighing or adherence to exercise recommendations (Ciere, Cartwright and Newman, 2012)).

Table 2: Articles that focus congestive heart failure.

| Study | Technology | (*) |
|---|---|---|
| Martínez et al., 2006 | Automatic patient data transmission by means of monitoring devices | 42/13 |
| Chaudhry et al., 2007 | Automatic patient data transmission by means of monitoring devices and standard telephone calls | 9/9 |
| Clark et al., 2007 | Automatic patient data transmission by means of monitoring devices and standard telephone calls | 14/14 |
| Dang et al., 2009 | Automatic patient data transmission by means of monitoring devices | 9/9 |
| Polisena et al., 2010a | Automatic patient data transmission by means of monitoring devices | 9/9 |
| Ciere et al., 2012 | Automatic patient data transmission by means of monitoring devices and standard telephone calls and video-conference | 12/7 |
| Grustam et al., 2014 | Automatic patient data transmission by means of monitoring devices and standard telephone calls and video-conference | 32/21 |
| Conway et al. | Automatic patient data transmission by means of monitoring devices and standard telephone calls | 25/25 |
| Nakamura et al., 2014 | Automatic patient data transmission by means of monitoring devices, including external, wearable, or implantable electronic devices | 13/13 |

(*) Number of RCT included in the review / Number of primary studies include in the review.

Accordingly the reviewed articles home monitoring has a positive effect on clinical outcomes in community dwelling patients with congestive heart failure. Home monitoring reduces mortality when compared with usual care and it also helps to lower both the number of hospitalizations and the use of other health care services (Dang, Dimmick and Kelkar, 2009; Polisena et al., 2010a; Conway, Inglis and Clark, 2014; Nakamura, Koga and Iseki, 2014).

However, there is a need for high-quality trials (Chaudhry et al., 2007). Additionally, Grustam et al. (2014) state that evidence from the scientific literature related to home monitoring to support congestive heart failure patients is still insufficient. Also, more full economic analyses are needed to reach a sound conclusion. This means that further research is required in terms of comparisons of home monitoring with usual care of patients with congestive heart failure.

## 3.4 Chronic Obstructive Pulmonary Disease

All the retrieved articles dealing with chronic obstructive pulmonary disease analyse RCT primary studies (Table 3). In particular, 3 of them considered RCT as one of the inclusion criteria (Polisena et al., 2010b; Pedone and Lelli, 2015; Lundell et al., 2015).

Home monitoring is supported by commercially available devices to measure and transmit different types of information (e.g. weight, temperature, blood pressure, oxygen saturation, spirometry parameters, symptoms, medication usage or steps in 6-minutes walking distance). In some cases the automatic data acquisition is complemented by clinical staff using questionnaires in telephone interviews (Polisena et al., 2010b; Pedone and Lelli, 2015). Video-conference can also be used to provide feedback to the patients (Lundell et al., 2015).

Table 3: Articles that focus chronic obstructive pulmonary disease.

| Study | Technology | (*) |
|---|---|---|
| Polisena et al., 2010b | Automatic patient data transmission by means of monitoring devices and standard telephone calls | 10/10 |
| Bolton et al. 2011 | Automatic patient data transmission by means of monitoring devices | 6/2 |
| Pedone et al., 2015 | Automatic patient data transmission by means of monitoring devices and standard telephone calls | 12/12 |
| Lundell et al., 2015 | Automatic patient data transmission by means of monitoring devices and video-conference | 9/9 |

(*) Number of RCT included in the review / Number of primary studies include in the review.

In what concerns the primary and secondary outcomes, 3 studies (Polisena et al., 2010b; Bolton et al. 2011; Pedone and Lelli, 2015) compare home monitoring with usual care of patients with chronic obstructive pulmonary disease, considering mortality, admissions to hospital or other health care utilization as primary outcomes. Secondary outcomes include, among others, health related quality of life, patient satisfaction, physical capacity and dyspnea.

Home monitoring was found to reduce rates of hospitalization and emergency department visits, while the findings related to hospital bed days of care varied between studies (Polisena et al., 2010b; Pedone and Lelli, 2015). However, 1 study (Polisena et al., 2010b) reports a greater mortality in a telephone-support group compared with usual care. Additionally, there is evidence that home monitoring has a positive effect on physical capacity and dyspnea (Lundell et al., 2015) and it is similar or better than usual care in terms of quality of life and patient satisfaction outcomes (Polisena et al., 2010b).

The evidence systematized by the articles of the category related to chronic obstructive pulmonary disease does not allow drawing definite conclusions, as the studies are small. The benefit of home monitoring of patients with chronic obstructive pulmonary disease is not yet proven and further research is required before wide-scale implementation be supported.

### 3.5 Hypertension

Finally, concerning patients with hypertension 1 article systematizes the results of 12 RCT using devices with automated data transmission, and video-conference.

The article reports improvements in the proportion of participants with controlled blood pressure compared to those who received usual care, but the authors conclude that more interventions are required and cost-effectiveness of the intervention should also be assessed (Chandak and Joshi, 2015).

## 4 DISCUSSION

According to the findings of the systematic review reported in the present article, diabetes, congestive heart failure, chronic obstructive pulmonary disease and hypertension are the most relevant chronic diseases in terms of the use of technologies for ageing in place to support home monitoring (i.e. the first research question of the present study).

Type 1 and type 2 diabetes stand out from other chronic conditions with a total of 20 studies, which constitute 57.1% of the articles that were retrieved. In order of relevance, the second chronic condition is congestive heart failure (i.e. 28.6% of the articles that were retrieved), which was followed by chronic obstructive pulmonary disease (i.e. 11.4% of the articles that were retrieved). Furthermore, one article reporting a systematic review related to home monitoring of patients with hypertension was also included in the present systematic review.

Self-management of diabetes requires patient adherence to best practice recommendations (e.g. glucose monitoring, dietary management or physical activity) (Or and Tao, 2014), congestive heart failure has a high rate of hospital readmission (Bonow, 2005; Joe and Demiris, 2013) and key aspects of the natural history of the chronic obstructive pulmonary disease are episodes of acute exacerbations, which are considered related to a faster disease progression, presence of comorbidities, and worse functional prognosis (Calvo, 2014). Therefore, the results of the present systematic review are in line with the current strong motivation for using technological solutions as a way to monitor patients with chronic diseases at home and to promote an increasing compliance of self-care.

In terms of types, outcomes, and impacts of technologies supporting home monitoring of patients with chronic diseases (i.e. the second research question of the study reported in the present article), the results show that:

- The technological solutions being used include web-based applications, computer assisted applications, smartphones, automatic patient data transmission by means of monitoring devices, video-conference and standard telephone calls (Tables 1-3).
- In general, the systematic reviews compare home monitoring with usual care and the primary outcomes depend of the type of the patients being considered (e.g. glycaemia control for patients with diabetes, patient's readmissions and mortality for patients with congestive heart failure and patients with chronic obstructive pulmonary disease, or control of the blood pressure of patients with hypertension).
- Secondary outcomes are quite diverse and include health related quality of life, weight, depression, blood pressure, behaviour outcomes, self-management, care knowledge, medication adherence, patient-clinician

communication, or structural outcomes related to care coordination.

- The analysis of the retrieved articles suggest that home monitoring has positive effects with a moderate to large improvements of different outcomes when compared with usual care of patients with diabetes, congestive heart failure, chronic obstructive pulmonary disease and hypertension (although in this case the evidence is not as robust – only 1 article – as it is in the other 3 chronic diseases). However, some studies are not conclusive about this positive impact and only report small beneficial effects.

Despite a high level of technological innovation and implementation, one of the findings is that telephone calls are still an important channel for the communication between patients and care providers.

Furthermore, it seems that important aspects are neglected during the technological developments, since there are reports of usability drawbacks as well as reports of the need for more comprehensive solutions, including provision of real-time feedback and the integration of the electronic health records systems being used by the care providers (El-Gayar et al., 2013).

Therefore, the results show that not only disruptive technological solutions have a key role when dealing with home monitoring, since practical and robust solutions are required, which means that the integration and the interoperability of existing technologies assume a great importance.

In general, the retrieved studies suggest positive effects of home monitoring, but evidence provided for the real benefit of home monitoring in some aspects was not totally convincing. Further research, including large scale RCT trials with consistent primary and secondary outcomes, and robust analysis about long-term sustainability, is required to allow the full incorporation of home monitoring in the clinical practice.

# 5 CONCLUSION

Considering the large amount of articles that report studies related to home monitoring of patients with chronic diseases, the authors decided to perform a review of reviews and meta-analysis.

Although the authors tried to be as elaborate as possible in methodological terms to guarantee that the review selection and the data extraction were rigorous, it should be acknowledged that this study has limitations, namely the weaknesses inherent to secondary analyses (i.e. review of reviews and meta-analysis), the limitations related to the dependency on the keywords and the databases selected, or the assumption that the retrieved articles have a homogeneous quality, which was not verified.

Despite these possible biases, the authors believe that the systematically collected evidence contributes to the understanding of the use of technologies for ageing in place to support home monitoring of patients with chronic diseases.

In parallel with this study, the authors have used similar methods to analyse the role of technologies for ageing in place to empower patients with chronic diseases. However, due to several limitations, the authors decide not to report the results in this article. Therefore, these results will be the object of a future publication. Also, as a future work, further studies will be implemented, namely to analyse how technologies for ageing in place are being used to support daily activities and promote the participation in social, economic, cultural, civil or spiritual matters of older adults.

# ACKNOWLEDGEMENTS

This work was partially supported by COMPETE - Programa Operacional Competitividade e Internacionalização (COMPETE 2020), Sistema de Incentivos à Investigação e Desenvolvimento Tecnológico (SI I&DT), under the project Social Cooperation for Integrated Assisted Living (SOCIAL).

# REFERENCES

Arambepola, C., Ricci-Cabello, I., Manikavasagam, P., Roberts, N., French, D.P. and Farmer, A., 2016. The impact of automated brief messages promoting lifestyle changes delivered via mobile devices to people with type 2 diabetes: a systematic literature review and meta-analysis of controlled trials. Journal of medical Internet research, 18(4).

Australian Institute of Health and Welfare, 2006. Chronic Diseases and Associated Risk Factors in Australia, 2006. Australian Institute of Health and Welfare.

Baron, J., McBain, H. and Newman, S., 2012. The impact of mobile monitoring technologies on glycosylated hemoglobin in diabetes: a systematic review. Journal of diabetes science and technology, 6(5), pp.1185-1196.

Bolton, C.E., Waters, C.S., Peirce, S. and Elwyn, G., 2011. Insufficient evidence of benefit: a systematic

review of home telemonitoring for COPD. Journal of evaluation in clinical practice, 17(6), pp.1216-1222.

Bonow, R.O., Bennett, S., Casey, D.E., Ganiats, T.G., Hlatky, M.A., Konstam, M.A., Lambrew, C.T., Normand, S.L.T., Pina, I.L., Radford, M.J. and Smith, A.L., 2005. ACC/AHA clinical performance measures for adults with chronic heart failure: a report of the American College of Cardiology/American Heart Association task force on performance measures (writing committee to develop heart failure clinical performance measures) endorsed by the Heart Failure Society of America. Journal of the American College of Cardiology, 46(6), pp.1144-1178.

Calvo, G.S., Gómez-Suárez, C., Soriano, J.B., Zamora, E., Gónzalez-Gamarra, A., González-Béjar, M., Jordán, A., Tadeo, E., Sebastián, A., Fernández, G. and Ancochea, J., 2014. A home telehealth program for patients with severe COPD: the PROMETE study. Respiratory medicine, 108(3), pp.453-462.

Cassimatis, M. and Kavanagh, D.J., 2012. Effects of type 2 diabetes behavioural telehealth interventions on glycaemic control and adherence: a systematic review. Journal of telemedicine and telecare, 18(8), pp.447-450.

Chandak, A. and Joshi, A., 2015. Self-management of hypertension using technology enabled interventions in primary care settings. Technology and Health Care, 23(2), pp.119-128.

Chaudhry, S.I., Phillips, C.O., Stewart, S.S., Riegel, B., Mattera, J.A., Jerant, A.F. and Krumholz, H.M., 2007. Telemonitoring for patients with chronic heart failure: a systematic review. Journal of cardiac failure, 13(1), pp.56-62.

Ciere, Y., Cartwright, M. and Newman, S.P., 2012. A systematic review of the mediating role of knowledge, self-efficacy and self-care behaviour in telehealth patients with heart failure. Journal of telemedicine and telecare, 18(7), pp.384-391.

Clark, R.A., Inglis, S.C., McAlister, F.A., Cleland, J.G. and Stewart, S., 2007. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. Bmj, 334(7600), p.942.

Connelly, K., Mokhtari, M. and Falk, T.H., 2014. Approaches to Understanding the impact of technologies for aging in place: a mini-review. Gerontology, 60(3), pp.282-288.

Conway, A., Inglis, S.C. and Clark, R.A., 2014. Effective technologies for noninvasive remote monitoring in heart failure. Telemedicine and e-Health, 20(6), pp.531-538.

Cruz, V.T., Pais, J., Bento, V., Mateus, C., Colunas, M., Alves, I., Coutinho, P. and Rocha, N.P., 2013. A rehabilitation tool designed for intensive web-based cognitive training: Description and usability study. Journal of Medical Internet Research, 15 (12), p. e59. Doi: 10.2196/resprot.2899.

Cruz, V.T., Pais, J., Alves, I., Ruano, L., Mateus, C., Barreto, R., Bento, V., Colunas, M., Rocha, N. and Coutinho, P., 2014. Web-based cognitive training: patient adherence and intensity of treatment in an outpatient memory clinic. Journal of medical Internet research, 16(5), p. e122. Doi: 10.2196/jmir.3377.

Dang, S., Dimmick, S. and Kelkar, G., 2009. Evaluating the evidence base for the use of home telehealth remote monitoring in elderly with heart failure. Telemedicine and e-Health, 15(8), pp.783-796.

El-Gayar, O., Timsina, P., Nawar, N. and Eid, W., 2013. A systematic review of IT for diabetes self-management: are we there yet?. International journal of medical informatics, 82(8), pp.637-652.

Eysenbach, G., 2008. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. Journal of medical Internet research, 10(3), p.e22.

Frazetta, D., Willet, K. and Fairchild, R., 2012. A systematic review of smartphone application use for type 2 diabetic patients. Online Journal of Nursing Informatics (OJNI), 16(3).

Garabedian, L.F., Ross-Degnan, D. and Wharam, J.F., 2015. Mobile phone and smartphone technologies for diabetes care and self-management. Current diabetes reports, 15(12), pp.1-9.

Genet, N., Boerma, W.G., Kringos, D.S., Bouman, A., Francke, A.L., Fagerström, C., Melchiorre, M.G., Greco, C. and Devillé, W., 2011. Home care in Europe: a systematic literature review. BMC health services research, 11(1), p.1.

Grustam, A.S., Severens, J.L., van Nijnatten, J., Koymans, R. and Vrijhoef, H.J., 2014. Cost-effectiveness of telehealth interventions for chronic heart failure patients: a literature review. International journal of technology assessment in health care, 30(01), pp.59-68.

Househ, M., 2014. The role of short messaging service in supporting the delivery of healthcare: An umbrella systematic review. Health informatics journal, p.1460458214540908.

Huang, Z., Tao, H., Meng, Q. and Jing, L., 2015. MANAGEMENT OF ENDOCRINE DISEASE: Effects of telecare intervention on glycemic control in type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. European Journal of Endocrinology, 172(3), pp.R93-R101.

Hunt, C.W., 2015. Technology and diabetes self-management: An integrative review. World journal of diabetes, 6(2), pp.225-233.

Jaana, M. and Paré, G., 2007. Home telemonitoring of patients with diabetes: a systematic assessment of observed effects. Journal of evaluation in clinical practice, 13(2), pp.242-253.

Jackson, C.L., Bolen, S., Brancati, F.L., Batts?Turner, M.L. and Gary, T.L., 2006. A systematic review of interactive computer?assisted technology in diabetes care. Journal of general internal medicine, 21(2), pp.105-

Joe, J. and Demiris, G., 2013. Older adults and mobile phones for health: A review. Journal of biomedical informatics, 46(5), pp.947-954.

Kickbusch, I. and Nutbeam, D., 1998. Health promotion glossary. Geneva: World Health Organization.

Kitsiou, S., Paré, G. and Jaana, M., 2015. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. Journal of medical Internet research, 17(3), p.e63.

Koch, S., 2013. Achieving holistic health for the individual through person-centered collaborative care supported by informatics. Healthcare informatics research, 19(1), pp.3-8.

Kvedar, J., Coye, M.J. and Everett, W., 2014. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Affairs, 33(2), pp.194-199.

Lundell, S., Holmner, Å., Rehn, B., Nyberg, A. and Wadell, K., 2015. Telehealthcare in COPD: a systematic review and meta-analysis on physical outcomes and dyspnea. Respiratory medicine, 109(1), pp.11-26.

Martínez, A., Everss, E., Rojo-Álvarez, J.L., Figal, D.P. and García-Alberola, A., 2006. A systematic review of the literature on home monitoring for patients with heart failure. Journal of telemedicine and telecare, 12(5), pp.234-241.

McBain, H., Shipley, M. and Newman, S., 2015. The impact of self-monitoring in chronic illness on healthcare utilisation: a systematic review of reviews. BMC health services research, 15(1), p.1.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine, 151(4), pp.264-269.

Mori, A.R., Mazzeo, M., Mercurio, G. and Verbicaro, R., 2013. Holistic health: predicting our data future (from inter-operability among systems to co-operability among people). International journal of medical informatics, 82(4), pp.e14-e28.

Nakamura, N., Koga, T. and Iseki, H., 2014. A meta-analysis of remote patient monitoring for chronic heart failure patients. Journal of telemedicine and telecare, 20(1), pp.11-17.

Or, C.K. and Tao, D., 2014. Does the use of consumer health information technology improve outcomes in the patient self-management of diabetes? A meta-analysis and narrative review of randomized controlled trials. International journal of medical informatics, 83(5), pp.320-329.

Ortiz, P.M.V., Felix, P.M. and Sosa, E.S.G., 2015. Text messaging interventions to glycemic control in type 2 diabetes adults: systematic review/Mensajes de texto para el control glucémico en adultos con diabetes tipo 2: revisión sistemática. Enfermería Global, 14(1), p.445.

Pal, K., Eastwood, S.V., Michie, S., Farmer, A.J., Barnard, M.L., Peacock, R., Wood, B., Inniss, J.D. and Murray, E., 2013. Computer based diabetes self management interventions for adults with type 2 diabetes mellitus. The Cochrane Library.

Pastalan, L.A., 1990. Aging in place: The role of housing and social supports (Vol. 6, No. 1-2). Psychology Press.

Pedone, C. and Lelli, D., 2015. Systematic review of telemonitoring in COPD: an update. Pneumonol Alergol Pol, 83, pp.476-8484.

Peterson, A. (2014). Improving Type 1 Diabetes Management With Mobile Tools A Systematic Review. Journal of diabetes science and technology, 8(4), 859-864.

Polisena, J., Tran, K., Cimon, K., Hutton, B., McGill, S., Palmer, K. and Scott, R.E., 2010. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. Journal of Telemedicine and Telecare, 16(2), pp.68-76.

Polisena, J., Tran, K., Cimon, K., Hutton, B., McGill, S., Palmer, K. and Scott, R.E., 2010. Home telehealth for chronic obstructive pulmonary disease: a systematic review and meta-analysis. Journal of telemedicine and telecare, 16(3), pp.120-127.

Queirós, A., Carvalho, S., Pavão, J. and da Rocha, N.P., 2013. AAL Information based Services and Care Integration. In HEALTHINF 2013 - Proceedings of the International Conference on Health Informatics, pp. 403-406.

Queirós, A., Silva, A., Alvarelhão, J., Rocha, N.P. and Teixeira, A., 2015. Usability, accessibility and ambient-assisted living: a systematic literature review. Universal Access in the Information Society, 14(1), pp.57-66. Doi: 10.1007/s10209-013-0328-x.

Ramadas, A., Quek, K.F., Chan, C.K.Y. and Oldenburg, B., 2011. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. International Journal of Medical Informatics, 80(6), pp.389-405.

Riazi, H., Larijani, B., Langarizadeh, M. and Shahmoradi, L., 2015. Managing diabetes mellitus using information technology: a systematic review. Journal of Diabetes & Metabolic Disorders, 14(1), p.1.

Slev, V.N., Mistiaen, P., Pasman, H.R.W., Verdonck-de Leeuw, I.M., van Uden-Kraan, C.F. and Francke, A.L., 2016. Effects of eHealth for patients and informal caregivers confronted with cancer: A meta-review. International journal of medical informatics, 87, pp.54-67.

Tao, D. and Or, C.K., 2013. Effects of self-management health information technology on glycaemic control for patients with diabetes: a meta-analysis of randomized controlled trials. Journal of telemedicine and telecare, 19(3), pp.133-143.

Teixeira, A., Pereira, C., e Silva, M.O., Alvarelhão, J., Silva, A.G., Cerqueira, M., Martins, A.I., Pacheco, O., Almeida, N., Oliveira, C. and Costa, R., 2013. New telerehabilitation services for the elderly. I. Miranda, M. Cruz-Cunha, Handbook of Research on ICTs and Management Systems for Improving Efficiency in Healthcare and Social Care, IGI Global, pp.109-132. Doi: 10.4018/978-1-4666-3990-4.ch006.

Thrall, J.H., 2005. Prevalence and Costs of Chronic Disease in a Health Care System Structured for

Treatment of Acute Illness 1. Radiology, 235(1), pp.9-12.

Tildesley, H.D., Po, M.D. and Ross, S.A., 2015. Internet Blood Glucose Monitoring Systems Provide Lasting Glycemic Benefit in Type 1 and 2 Diabetes: A Systematic Review. Medical Clinics of North America, 99(1), pp.17-33.

van Vugt, M., de Wit, M., Cleijne, W.H. and Snoek, F.J., 2013. Use of behavioral change techniques in web-based self-management programs for type 2 diabetes patients: systematic review. Journal of medical Internet research, 15(12), p.e279.

Verhoeven, F., van Gemert-Pijnen, L., Dijkstra, K., Nijland, N., Seydel, E. and Steehouder, M., 2007. The contribution of teleconsultation and videoconferencing to diabetes care: a systematic literature review. J Med Internet Res, 9(5), p.e37.

World Health Organization, 2002. Active ageing: a policy framework: a contribution of the World Health Organization to the Second United Nations World Assembly on Ageing. Madrid (ES): World Health Organization.

# Generating a Distilled N-Gram Set
## *Effective Lexical Multiword Building in the SPECIALIST Lexicon*

Chris J. Lu[1,2], Destinee Tormey[1,2], Lynn McCreedy[1,2] and Allen C. Browne[1]
[1]*National Library of Medicine, Bethesda, MD, U.S.A.*
[2]*Medical Science & Computing, LLC, Rockville, MD, U.S.A.*
*{chlu, lmccreedy, albrowne}@mail.nih.gov, dtormey@gmail.com*

Keywords: MEDLINE N-Gram Set, Multiwords, Medical Language Processing, Natural Language Processing, the SPECIALIST Lexicon.

Abstract: Multiwords are vital to better Natural Language Processing (NLP) systems for more effective and efficient parsers, refining information retrieval searches, enhancing precision and recall in Medical Language Processing (MLP) applications, etc. The Lexical Systems Group has enhanced the coverage of multiwords in the Lexicon to provide a more comprehensive resource for such applications. This paper describes a new systematic approach to lexical multiword acquisition from MEDLINE through filters and matchers based on empirical models. The design goal, function description, various tests and applications of filters, matchers, and data are discussed. Results include: 1) Generating a smaller (38%) distilled MEDLINE n-gram set with better precision and similar recall to the MEDLINE n-gram set; 2) Establishing a system for generating high precision multiword candidates for effective Lexicon building. We believe the MLP/NLP community can benefit from access to these big data (MEDLINE n-gram) sets. We also anticipate an accelerated growth of multiwords in the Lexicon with this system. Ultimately, improvement in recall or precision can be anticipated in NLP projects using the MEDLINE distilled n-gram set, SPECIALIST Lexicon and its applications.

## 1 INTRODUCTION

This section introduces: first, the SPECIALIST Lexicon; second, the importance of multiwords in NLP; third, the background and purpose of developing a new n-gram-based system for building lexical multiwords.

### 1.1 The SPECIALIST Lexicon

The SPECIALIST Lexicon, distributed in the Unified Medical Language System (UMLS) Knowledge Sources by the National Library of Medicine (NLM), is a large syntactic lexicon of biomedical and general English, designed and developed to provide the lexical information needed for the SPECIALIST Natural Language Processing System (McCray et al., 1993). Lexical records are used for part-of-speech (POS) tagging, indexing, information retrieval, concept mapping, etc. in many NLP projects, such as Lexical Tools (McCray et al., 1994), MetaMap (Aronson, 2001; Aronson and Lang, 2010), cTAKES (Savova, 2010), Sophia (Divita et al., 2014), gSpell (Divita et al., 2000), STMT (Lu and Browne, 2012), SemRep, UMLS Metathesaurus, ClinicalTrials.gov, etc. It has been one of the richest and most robust NLP resources for the NLP/MLP community since its first release in 1994. It is important to keep the Lexicon up to date with broad coverage to ensure the success of NLP applications that use it.

Each lexical entry in the Lexicon records the syntactic, morphological, and orthographic information needed by the SPECIALIST NLP System. Terms must meet 3 requirements to qualify as lexical entries: 1) part-of-speech, 2) inflections, and 3) a special unit of lexical meaning by themselves. Linguists in the Lexical Systems Group (LSG) look at the usage of candidate terms from various sources to add terms into the Lexicon if the above three requirements are met. Terms (base forms and inflectional variants) may be single words or multiwords - namely words that contain space(s). If it is a multiword, such as "*ice cream*" or "*hot dog*", it is called a lexical multiword (LexMultiword or LMW). Single words in the Lexicon have increased 2.6 times from 180,468 in 2002 to 468,655 in 2016. These Lexicon single words cover only about 10.62% of unigrams (single words) from titles and abstracts in



Lu, C., Tormey, D., McCreedy, L. and Browne, A.
Generating a Distilled N-Gram Set - Effective Lexical Multiword Building in the SPECIALIST Lexicon.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 77-87
ISBN: 978-989-758-213-4

MEDLINE.2016. However, single-word Lexicon terms comprise 98.42% of MEDLINE unigrams if the word count (WC) is taken into consideration. In other words, the current Lexicon has a very high recall rate of single words in MEDLINE, because most frequently used single words in MEDLINE are covered. As for LMWs, we observe a continuous growth in the Lexicon from 88,324 (32.86%) in 2002 to 446,928 (48.81%) in 2016. Both the high coverage of existing single words and the trend of increasing growth of LMWs in the Lexicon lead to our position that multiword acquisition is key for future Lexicon building.

## 1.2 Multiwords in NLP

Multiwords are vital to the success of high quality NLP applications (Sag et al., 2002; Fraser, 2009). First, multiwords are ubiquitous. Technical terminologies in many specialized knowledge domains, particularly in areas like medicine, computer science and engineering, are often created as Multiword Expressions (MWEs) (Frantzi et al., 2000; Green et al., 2013; Ramisch, 2014). Second, MWEs are hard to deal with in NLP tasks, such as identification, parsing, translation, and disambiguation, not only because MWEs have a large amount of distinct phenomena, but also due to the absence of major syntactic theories and semantic formalisms. Our Lexicon with multiwords remedies these issues. For example, most NLP applications on word segmentations are word-oriented (tokenization), relying on POS taggers, stemmers, and chunkers to segment each MWE as a phrasal unit from the sentence. This process can be improved if multiwords can be identified as a phrasal unit directly (such as through a Lexicon lookup) and not processed further by taggers, e.g. phrasal preposition ("*because of*", "*due to*"), and adverbs ("*on time*"). Thus, POS ambiguity can be reduced through identifying the POS of these MWEs. Third, non-decomposable MWEs, such as fixed phrases ("*kingdom come*", "*by and large*") and idioms ("*kick the bucket*", "*shoot the breeze*"), are very challenging tasks for NLP syntactically as well as semantically. While syntactic aspects of idiom usage necessitates a beyond-Lexical-level solution to those non-decomposable MWEs, fixed phrases are handled well as LMWs in our Lexicon. NLP techniques, such as Query Expansion, do not work well on fixed-phrase MWEs for concept mapping, unless they are seen as LMWs. For example, "*hot dog*" should not be expanded as "*high temperature canine*" to find its concept. Instead, a direct Lexicon look up of "*hot dog*" (E0233017) without further query expansion resolves issues caused by fixed-phrase MWEs. Furthermore, the Metathesaurus concept associated with a sentence often coincides with the longest multiword in the sentence. This idea is implemented in MetaMap by identifying the longest LMWs in sentences for mapped concept ranking. Accordingly, a comprehensive lexicon with a rich resource of MWEs is an essential component to a more precise, effective, and efficient NLP system.

## 1.3 MWEs and LMWs

Research on Multiword Expressions (MWEs) has been growing since the late 1990s. State of the art methods including statistical association measures (Silva and Lopes, 1999; Fazly et al., 2009; Pecina, 2010), machine learning (Boukobza and Rappoport, 2009; Tsvetkov and Wintner, 2011; Green et al., 2011), syntactic patterns (Seretan and Wehrli, 2009; Kim and Baldwin, 2010; Green et al., 2013), web queries (Takahashi and Morimoto, 2013), semantic analysis (Pearce, 2001; Baldwin et al., 2003), and a combination of the above methods (Calzolari et al., 2002; Bejček et al., 2013; Sangati and Cranenburgh, 2015) are used in MWE research for acquisition, identification, interpretation, disambiguation and other applications. Despite a great deal of research on MWEs, there is no approach that fits perfectly for building LMWs in the SPECIALIST Lexicon. LMWs are a subset of MWEs due to our requirements that a legitimate Lexical entry must have a POS, inflections, and be a unit of meaning. In short, the broader notion of MWEs are distinguished from LWMs in four ways. First, a collocation (an arbitrary statistically significant association between co-occurring items) is not necessarily a LMW because it is not necessarily qualified as a Lexical entry. For example, "*undergoing cardiac surgery*" occurs frequently (3,770 hits in 3,062 documents) in the 2016 MEDLINE n-gram set, but it is not a LMW because it is not functioning as a special unit of meaning by itself. Moreover, this collocation is sometimes, but not always, a single POS. On the other hand, its subterm, "*cardiac surgery*", which occurs frequently (37,171 hits in 22,404 documents) in MEDLINE, is a LMW. In other words, frequency alone is not sufficient to determine if a term is a LMW. For the same reason, some phrases are not LMWs. For example, "*in the house*" is not a LMW while "*in house*" is. Second, verb particle constructions are handled by complementation types (Browne et al., 2000) in Lexical records to coordinate lexical meaning with syntactic characteristics of the verb.

For example, "*beat someone up*" can be constructed from the Lexical record of "*beat*", as shown in Figure 1. Similarly, light verbs that are covered within Lexical records, such as "*make love*" and "*give birth*", are included in the Lexical records of "*make*" (E0038623) and "*give*" (E0029785), respectively. The information on these types of MWEs is stored inside the Lexical records and they are not considered LMWs (not a base form or inflectional variants of a Lexical entry). However, they can be retrieved/identified by a parser based on the Lexicon. Third, non-decomposable idioms are beyond the scope of the Lexicon, such as "*kick the bucket*" and "*shoot the breeze*". Aligning the syntactic analysis of idiomatic phrases with their semantic interpretations is beyond the scope of what a lexicon can accomplish. Thus, they are not under consideration here. Fourth, due to the complicated nature of multiwords, much previous MWE research only focuses on bi-grams or tri-grams, which do not meet the requirement of including up to five-grams to reach an estimated recall value of 99.47% of multiwords (Lu et al., 2015).

```
{base=beat
entry=E0012175
        cat=verb
        variants=irreg|beat|beats|beat|beaten|beating|
        intran
        intran;part(about)
        tran=np
        tran=np;part(back)
        tran=np;part(up)
        tran=np;part(down)
        tran=np;part(in)
        link=advbl
        cplxtran=np,advbl
        nominalization=beating|noun|E0219216
}
```

Figure 1: Lexical record of the verb "beat".

## 2 MOTIVATION

Previously, an element word approach (Lu et al., 2014) was used to build the Lexicon by linguists through a web-based computer-aided tool, LexBuild (Lu et al., 2012). Unigrams with high frequency (WC) from the MEDLINE n-gram set that are not in the Lexicon were used as element words for finding new LMW candidates through the Essie search engine (Ide et al., 2007). There are several issues with this approach: 1) it is time consuming; 2) multiwords associated with high frequency element words do not necessarily have high frequency; 3) new multiwords associated with processed element words are missed. If we use the mean value, 65%, as an estimated multiword ratio based on the empirical measurement (Ramisch, 2014), it will take more than 21 years for current LSG staff to add all multiwords to the Lexicon by this approach (on average, 20,000 terms are added by LSG staff annually). And this estimate does not account for the fact that many new multiwords are continuously being created by biomedical researchers and English users in general. Thus, we decided to develop a new system to effectively build LMWs to the Lexicon by using an n-gram approach. MEDLINE was chosen as the corpus because it is the biggest and most commonly used resource in the biomedical domain. The MEDLINE n-gram set (MNS) was generated by the following steps: 1) English titles and abstracts from MEDLINE documents are collected and then tokenized to sentences and words (tokens); 2) by requirements, the MNS includes up to 5-grams with information of associated document count (DC) and word count (WC); 3) n-gram and DC|WC are used as key and values in Java HashMap class for n-gram retrieval; 4) due to the large scale, the computer program for retrieving n-grams exceeds the maximum keys in the Java HashMap class ($2^{30}$-1) when n > 3. Thus, a new model is developed to resolve this issue. This model involves processes of splitting, grouping, filtering, combining and sorting (Lu et al., 2015). The MNS is generated by above processes and has been distributed annually to the public since 2014. The MNS provides comprehensive raw n-gram data from titles and abstracts of MEDLINE. Due to its large-scale size (> 19M n-grams), it is difficult to be handled by computer programs with complicated algorithms. So, a distilled MEDLINE n-gram set (DMNS), with reduced size, higher precision and similar recall in terms of LMWs, is required for the multiword acquisition task and useful for NLP applications.

## 3 APPROACH - FILTERS

Filters (exclusive filters) are designed to be applied on the MNS to generate the DMNS by trapping invalid LMWs. The design goal of these filters is set to keep the similar (high) recall rate by not trapping valid LMWs. Ideally, all valid multiwords should pass through these filters. The precision of the filtered n-gram set can be improved significantly by applying a series of filters with high recall rate. Exclusive filters are developed based on empirical models with heuristic rules in this task. They are categorized into three types as described below. Patterns and trapped

examples are illustrated for each filter in the format of [pattern] and "*example*" in this paper, respectively.

## 3.1 General Exclusive Filters

This type of filter is intuitive and based on surface features of terms. Terms composed merely of certain characters/words, such as punctuation, digits, numbers, spaces and stopwords do not meet the requirement of having a special unit of lexical meaning to themselves. They are used for the general purpose of filtering out invalid LMWs:

- Pipe Filter: A term that contains pipe(s) is trapped because a pipe is used as a field separator in most NLP systems. Trapped examples include: "*(|r|*", "*Ag|AgCl*", etc.
- Punctuation or Space Filter: A term that contains nothing but punctuation or space(s) is trapped. Trapped examples include: "=", "+/-", "<", "*(%)*" and "-->".
- Digit Filter: A term that contains nothing but digit(s), punctuation, and space(s) is trapped. Trapped examples include: "*2000*", "*95%*", "*3-5*", "*$1,500*", "*(+/10.05)*", "*192.168.1.1*" and "*[192, 168]*".
- Number Filter: A term that contains nothing but number(s) is trapped. This filter can be considered as a domain filter because all numbers are already recorded in the Lexicon. Trapped examples include: "*two*", "*first and second*", "*one third*", "*twenty-eight*", "*Four hundred and forty-seven*" and "*half*".
- Digit and Stopword Filter: A term that contains nothing but digit(s) or stopword(s) is trapped. Trapped examples include: "*50% of*", "*of the*", "*1, 2, and*", "*2003 to 2007*", "*for >=50%*" and "*OR-462*".

## 3.2 Pattern Exclusive Filters

This type of filter looks for certain matching patterns in a term for trapping. Computer programs are implemented based on observed empirical patterns. Some filters require sophisticated algorithms.

- Parenthetic Acronym Pattern (PAP) Filter: A parenthetic acronym is a conventional way of representing an acronym expansion with the associated acronym. The pattern is an acronym expansion followed by an acronym within a closed parenthesis, e.g., [acronym-expansion (ACRONYM)]. The expansions of acronyms are usually valid multiwords. A term that contains this pattern is trapped because it contains a potential multiword plus the associated acronym and thus cannot be a valid LMW. Trapped examples include: "*magnetic resonance imaging (MRI)*", "*imaging (MRI)*", "*magnetic resonance (MR) imaging*" and "*(CREB)-binding protein (CBP)*".
- Indefinite Article Filter: A lowercased term that starts with an indefinite article and a space, [a ], without other n-grams that match as its spelling variants (spVar) pattern in the corpus (n-gram set) is trapped. Patterns of [a-XXX] and [aXXX] are used as the spVar pattern of indefinite articles of [a XXX], where XXX represents any term. Trapped examples include: "*a significant*", "*a case*", "*a case of*", "*a dose-dependent*" and "*a delivery rate per*".
- UPPERCASE Colon Filter: A term that contains the pattern of [UPPERCASE:] is trapped. In MEDLINE, this is a conventional usage for this pattern, such as [CONCLUSION:], [RESULTS:], [OBJECTIVE:], [METHODS:], [MATERIALS AND METHODS:], and [BACKGROUND:]. Trapped examples include "*MATERIALS AND METHODS: The*", "*95% CI:*" and "*PHPT:*"
- Disallowed Punctuation Filter: A term that contains disallowed punctuation is trapped. Disallowed punctuation includes: {}_!@#*\;"?~=|<>$`^. Trapped examples include: "*(n =*", "*(P < 0.05)*", "*N^N*", "*group (n=6) received*" and "*CYP3A7*1C*".
- Measurement Pattern Filter: A term that contains a measurement pattern is trapped. A measurement pattern is [number + unit], including age ("*4-year-old*", "*4 year-old*", "*four year-old*", "*4 year-olds*" and "*4 years or older with*"), time ("*four months*", "*1 January 1991*", "from May 2002" and "*6 hours plus*"), range ("*2-3 days*" and "*1-2 tablets*"), temperature ("*at -5 degrees*"), dosage ("*10 cigarettes per day*" and "*0.1-2.3 mg/day*") and others ("*60 inches*", "*0.5 mg*", "*3 mg/EE*", "*10 mg/kg*" and "*50 mg/kg/day*").
- Incomplete Pattern Filter: A term that contains an incomplete pattern is trapped. A valid multiword should have completed parentheses or brackets. Incomplete patterns are terms that do not have an even number of left and right parentheses or square brackets or they are not closed. Trapped examples include: "*II (Hunter syndrome*", "*0.05) higher*", "*bond]C-C[triple*", "*(chi(2)*" and "*interval [95%*".

### 3.3 Lead-End-Term Exclusive Filters

LMWs do not start with certain terms, such as auxiliaries ("*be*", "*do*", etc.), complementizers ("*that*"), conjunctions ("*and*", "*or*", "*but*", etc.), determiners ("*a*", "*the*", "*some*", etc.), modals ("*may*", "*must*", "*can*", etc.), pronouns ("*it*", "*he*", "*they*", etc.), and prepositions ("*to*", "*on*", "*by*", etc.). They are called invalid lead-terms. Similarly, multiwords do not end with words in the above-listed categories. N-grams ending in them are invalid LMWs. They are used in exclusive filters to exclude invalid multiwords. Terms from the Lexicon with any of the above seven categories are used as invalid lead-end-term (ILET) candidates. ILETs only comprise 0.05% (488) of total forms in Lexicon.2016 (915,583). Notably, ILET candidates are considered static because no new terms in the above 7 categories have been added since 2010. Please refer to LSG web documents on Lead-End-Term filter models for details (National Library of Medicine, Lexicon: Lead-End-Terms Model, 2015).

- Absolute Invalid Lead-Term Filter: A term that leads with an absolute invalid lead-term (AILT) is trapped. There are 382 AILTs derived from the Lexicon, such as [the], [from], [is] and [of]. Trapped examples include: "*The results*", "*from the*", "*is a*" and "*of a*".
- Absolute Invalid End-Term Filter: A term that ends with an absolute invalid end-term (AIET) is trapped. There are 407 AIETs derived from the Lexicon, such as [with], [the] and [that]. Trapped examples include: "*patients with*", "*at the*" and "*suggest that*".
- Lead-End-Term Filter: A term that leads with an ILET and also ends with an ILET is trapped. Trapped examples include: "*in a*", "*to be*", "*with a*" and "*as a*".
- Lead-Term No SpVar Filter: A term that leads with a valid lead-term (VLT) without any other term matching its spVar pattern in the same corpus is trapped. There are 52 VLTs derived from the Lexicon, such as [to], [as], [for] and [plus]. Trapped examples include: "*to determine*", "*as a result*", "*for example*" and "*plus LHRH-A*".
- End-Term No SpVar Filter: A term that ends with a valid end-term (VET) without any other term matching its spVar pattern in the same corpus is trapped. There are 27 VETs derived from the Lexicon, such as [of], [to], [in] and [more]. Trapped examples include: "*effects of*", "*was used to*", "*(HPV) in"* and "*loss of two or more*".

## 4 TESTS AND RESULTS

The evaluation of each individual filter, the combination of all filters, and the distilled MEDLINE n-gram set are discussed in this section. The 2016 release of the Lexicon and MEDLINE n-gram set are used in this paper, unless specified otherwise.

### 4.1 Recall Test of Filters

A recall test model has been established for testing each developed filter individually. Recall is defined as: TP / (TP + FN), where T is true, F is false, P is positive, N is negative. Terms (915,583) in the Lexicon are used to test exclusive filters. All Lexicon terms are valid (relevant) and should pass through filters for preserving high recall rate. In this test, the pass-through terms are counted as TP (retrieved, relevant) while the trapped terms are FN (not retrieved, relevant) for the filtered set.

Columns 4 and 5 in Table 1 list the recall rate and number of trapped terms (FN) for this recall test. The results show that all filters meet the design goal to have very high recall rates. The lowest recall rate (99.9913%) is at filter 15, Lead-Term No SpVar Filter.

### 4.2 The Distilled N-gram Set

The distilled MEDLINE n-gram set is generated by applying these high recall filters to the MEDLINE n-gram set in the same sequential order of the first column (ID) in Table 1. Let's say X filters are applied to all MEDLINE n-grams. The number of valid LMWs (TP) and number of invalid LMWs (FP) of the filtered MEDLINE n-gram set after ith filter are $TP_i$ and $FP_i$, respectively, where i = 0, 1, 2, … X. The number of valid LMWs are about the same ($TP_0 \cong TP_1 \cong TP_2 \cong \ldots \cong TP_X$) if high recall filters are used. The number of invalid LMWs is reduced ($FP_0 > FP_1 > FP_2 > \ldots > FP_X$) from the original MEDLINE n-gram set to the final distilled MEDLINE n-gram set after applying filters. Accordingly, the distilled MEDLINE n-gram set (X) has higher precision $P_X$ and similar recall $R_X$ to the MEDLINE n-gram set (0), as shown in equations 1 and 2, respectively, where the number of $FN_i$ (not retrieved, relevant) is a constant.

$$P_X = TP_X / (TP_X + FP_X) \cong TP_0 / (TP_0 + FP_X) > TP_0 / (TP_0 + FP_0) \quad (1)$$

$$R_X = TP_X / (TP_X + FN_X) = TP_X / (TP_X + FN_0) \cong TP_0 / (TP_0 + FN_0) \quad (2)$$

Sixteen high recall rate filters are applied to the MEDLINE n-gram set in the same sequential order as the first column in Table 1 to filter out invalid LMWs. Columns 6, 7 and 8 in Table 1 list the number of trapped terms, the passing rate (PR) and cumulative passing rate (cum. PR) for all filters applied on the MEDLINE n-gram set. The passing rate of the ith filter is the pass through terms/total terms when applying the ith filter on the MNS individually. The pass through terms equals the total terms minus the trapped terms. The cum. PR of ith filter is the cumulative passing rate after applying i filters in the sequential order of the first column in Table 1 to the MNS. In other words, the trapped number is the sum of trapped terms by filters that apply before the ith filter. As a result (i = 16), the distilled MEDLINE n-gram set, after filtering out the majority (11,922,490) of invalid LMWs by these 16 filters, contains about 38.31% (7,402,848) n-grams of the MEDLINE n-gram set (19,325,338). Figure 2 shows a schematic diagram for generating the distilled MNS by applying these filters on the MNS. These filters are designed to independently trap invalid lexMultiwords, so the order of filter application does not affect the final results. These filters are generic and can be used by different NLP projects if they meet the project requirements. The Lead-End-Term filters (ID: 12-16) have higher efficiency (trapped terms/total terms) by trapping more n-grams in this process while the recall rate is above 99.99%. Theoretically, the distilled MEDLINE n-gram set, preserves valid terms in the MNS and thus has higher precision and similar recall compared to the MNS. The size of DMNS is reduced to 38% of MNS, making it possible for complicated computer programs to work in a reasonable time frame in practice, such as the SpVar Pattern matcher (please see section 5.1).

Table 1: Results of applying exclusive filters on Lexicon recall test and the MEDLINE n-gram set.

| ID | Filter Type | Filter Name | Lexicon Recall Test | | Applied on the MEDLINE N-gram Set | | |
|---|---|---|---|---|---|---|---|
| | | | Recall | Trapped | Trapped | PR | Cum. PR |
| 1 | General Filters | Pipe | 100.0000% | 0 | 7 | 100.0000% | 100.0000% |
| 2 | | Punctuation or Space | 100.0000% | 0 | 425 | 99.9978% | 99.9978% |
| 3 | | Digit | 99.9999% | 1 | 132,650 | 99.3136% | 99.3114% |
| 4 | | Number | 99.9953% | 43 | 4,326 | 99.9775% | 99.2890% |
| 5 | | Digit and Stopword | 99.9991% | 8 | 157,786 | 99.1777% | 98.4725% |
| 6 | Pattern Filters | Parenthetic Acronym - (ACR) | 100.0000% | 0 | 197,022 | 98.9647% | 97.4530% |
| 7 | | Indefinite Article | 99.9986% | 13 | 344,403 | 98.1713% | 95.6709% |
| 8 | | UPPERCASE Colon | 99.9999% | 1 | 113,936 | 99.3838% | 95.0813% |
| 9 | | Disallowed Punctuation | 99.9986% | 13 | 135,508 | 99.2625% | 94.3801% |
| 10 | | Measurement | 99.9920% | 73 | 336,112 | 98.1572% | 92.6409% |
| 11 | | Incomplete | 100.0000% | 0 | 166,356 | 99.0708% | 91.7801% |
| 12 | Lead-End-Term Filters | Absolute Invalid Lead-Term | 99.9943% | 52 | 4,712,162 | 73.4329% | 67.3967% |
| 13 | | Absolute Invalid End-Term | 99.9997% | 3 | 2,710,470 | 79.1897% | 53.3713% |
| 14 | | Lead-End-Term | 99.9992% | 7 | 2,687 | 99.9739% | 53.3573% |
| 15 | | Lead-Term No SpVar | 99.9913% | 80 | 1,450,394 | 85.9342% | 45.8522% |
| 16 | | End-Term No SpVar | 99.9968% | 29 | 1,458,246 | 83.5433% | 38.3064% |

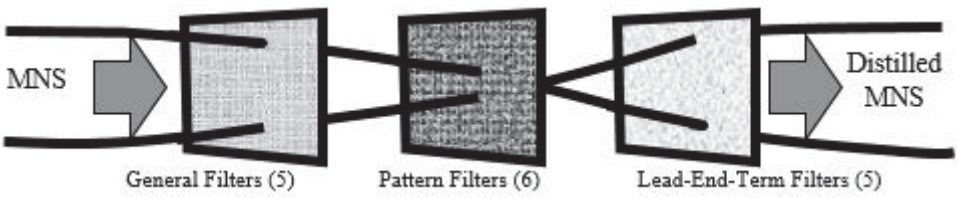



Figure 2: Schematic diagram of the MNS, filters and the distilled MNS.

## 4.3 Evaluation of DMNS

We further verify the DMNS by comparing the performance of the MNS and the DMNS. A smaller test set is set up by retrieving LMW candidates from the Parenthetic Acronym Pattern (PAP) matcher. Matchers (inclusive filters) are designed to retrieve LMWs from MEDLINE n-grams by trapping valid multiwords that match valid LMW patterns. In other words, terms trapped by matchers should be valid LMWs. The design goal of matchers is set to generate high precision LMW candidates. On the other hand, the recall of matchers might decrease because not all valid LMWs are trapped.

Acronym expansions are good patterns for a matcher because they have a high possibility of generating valid LMWs. The PAP matcher model is implemented as follows. First, apply Parenthetic Acronym Pattern Filter on the MEDLINE n-gram set to retrieve terms matching the pattern of [acronym expansion (ACRONYM)]. For example, "*computed tomography (CT)*", "*magnetic resonance imaging (MRI)*", "*Unified Health System (SUS)*", etc. are retrieved from the n-gram set. Second, retrieve expansions if they match the associated acronym. Heuristic rules are implemented, such as checking the initial characters of first and last words of the expansion to match the first and last characters of the associated acronym. For example, the expansion of "*Unified Health System (SUS)*" is identified as an invalid LMW because the first initial of the expansion (U) does not match the first character of acronym (S). Third, remove terms if the expansion is a subterm of other expansions in the list. For example, both n-grams of "*cell sarcoma (CCA)*" and "*clear cell sarcoma (CCA)*" pass the first two steps. The invalid LMW of "*cell sarcoma*" is removed in this step because it is a subterm of the valid LMW "*clear cell sarcoma*".

We applied the PAP matcher to the MNS to retrieve LMW candidates. The lowercased core-terms of these candidates are collected as the test set. Core-term normalization is to normalize an n-gram to its core form by stripping the leading and ending punctuation. For example, "in details,", "- in details" and "- in details," have the same core-term form of "in details". Core-terms might have punctuation internally, such as "in (5) details". It is a useful normalization to cluster terms with the same core together from the n-gram set in multiword acquisition. As a result, 17,707 LMW candidates are retrieved by this process. They are tagged by LSG linguists and are added to the Lexicon if they are valid LMWs. 15,850 candidates in this set are tagged as valid LMWs to reach 89.51% precision for this PAP matcher, where precision is defined as: TP/(TP+FP), as shown in case 1 in Table 2. The recall cannot be found because all LMWs from MEDLINE cannot be identified in real practice. The result of this PAP matcher is used as the baseline for performance test to compare the results of other filters and matchers. Accordingly, recall in case 1 is set to 1.00 for the purpose of comparison. F1 score is defined as: (2 x precision x recall) / (precision + recall), is calculated and shown in the last column in Table 2.

We repeat the same process by applying the PAP matcher to the DMNS to retrieve LMWs. The results (case 2) show an improvement on F1 score with better precision and almost the same recall. This confirms the theoretic conclusion and the result of the recall test on these filters, that the distilled MEDLINE n-gram contains almost the same amount of valid multiwords as the MEDLINE n-gram set while its size is reduced to 38%. Furthermore, the cumulative recall rates of these 16 filters on the recall test (0.9996, multiple product of recall column in table 1) and the recall rate of case 2 in Table 2 (0.9994) are almost identical. This confirms that the approach of applying these filters results in a similarly high recall rate for both the Lexicon and the test set from PAP matcher. Similar results of the Lexicon recall test and DMNS in Table 1 and the performance test of the PAP matcher on the MNS and the DMNS in Table 2 for 3 releases (2014 to 2016) of the Lexicon and MEDLINE are found to confirm the consistency of this approach.

Table 2: Performance comparison of MNS, DMNS and SVP matchers on a test set with 17,707 terms.

| Case | Test Case - Model | TP | FP | FN | TN | Precision | Recall | F1 |
|---|---|---|---|---|---|---|---|---|
| 1 | PAP matcher on MNS (baseline) | 15,850 | 1,857 | 0 | 0 | 0.8951 | (1.0000) | 0.9447 |
| 2 | PAP matcher on DMNS (16 filters) | 15,840 | 1,299 | 10 | 558 | 0.9242 | 0.9994 | 0.9603 |
| 3 | SVP matcher on case 2 | 8,094 | 499 | 7,756 | 1,358 | 0.9419 | 0.5107 | 0.6623 |

# 5 APPLICATIONS ON DMNS

Despite the high precision of the PAP matcher, it only retrieves a small amount of LMW candidates. Other matchers have been developed to retrieve more LMW candidates for Lexicon building.

## 5.1 Spelling Variant Pattern Matcher

The Spelling Variant Pattern (SVP) matcher model with a complicated algorithm was developed to retrieve large amount of LMW candidates. As we observed, an n-gram is a good LMW candidate if it has spelling variants existing in the same corpus (n-gram set). A sophisticated computer algorithm was developed to identify all n-grams that have potential spVars. First, a special normalization program was developed to normalize spVars into their canonical forms by converting non-ASCII Unicode to ASCII (e.g. "*Labbé*" to "*Labbe*"), synonym substitution (e.g. "*St. Anthony's fire*" to "*Saint Anthony's fire*"), rank substitution (e.g. "*Vth nerve*" to "*5th nerve*"), number substitution (e.g. "*12-lead*" to "*twelve-lead*"), Roman numeral substitution (e.g. "*BoHV-I*" to "*BoHV-1*"), strip punctuation (e.g. "*lamin-A*" to "*lamin A*"), stripping genitive (e.g. "*Laufe's forceps*" to "*Laufe forceps*"), converting to lowercase, and removing any space(s). All terms that have the same normalized spVar canonical form are identified as spVars to each other. The Lexicon.2015 has 379,269 spVars (including inflectional spelling variants) in 867,728 (unique) inflectional variants, and was used to test this model. As shown in the recall column in Table 3, 80.50% of all spVars in the Lexicon are identified by spVar normalization (step 1). All identified spVars are grouped in spVar classes for further NLP processing. Second, a MES (Metaphone, Edit distance, and Sorted distance) model is developed to improve recall. The MES model is composed of an algorithm of Metaphone phonetic code (Philips, 1990), edit distance (the minimum number of operations required to transform one term into the other), and minimum sorted distance. Sorted distance is the distance between two terms in an alphabetic sorted list of a set of terms. It is used to measure the similarity of two terms compared to other terms in the set. All terms having the same phonetic code and an edit distance (ED) less than a specified value are collected and sorted. The pair with the minimum sorted distance (the closest pair) is identified as spVars to each other. For example, "*yuppie flu*" and "*yuppy flu*" have different spVar canonical forms of "*yuppieflu*" and "*yuppyflu*", respectively, and thus are not identified as spVars in the step 1, normalization. They are identified as spVars in step 2 (MES model), because they have the same Metaphone code of [YPFL], edit distance of 2, and the minimum sorted distance. This step identifies more spVars that cannot be identified by normalization in step 1. The recall is increased to 97.92% (Table 3). Third, an ES (Edit distance and Sorted distance) model is developed for further improvement of recall. Terms with an edited distance less than a specified value are collected and sorted. The pair with the minimum sorted distance is identified as being spVars. For example, "*zincemia*" and "*zincaemia*" are identified as spVars by the ES model with an edit distance of 1, while they were not identified as spVars in the previous steps, because they have different spVar canonical forms of "*zincemia*" and "*zincaemia*" and also have different Metaphone codes of [SNSM] and [SNKM], respectively. By relaxing the value of edit distance in both models repeatedly, our program reaches 99.72% recall on spVar identification in six steps in this test, as shown in Table 3. Precision (Prec.), recall, F1, accuracy, and running time (RT) of each step in this SVP matcher model are shown in Table 3, where accuracy is defined as: (TP + TN) / (TP + FP + FN + TN).

For testing purposes, we applied this SVP matcher model to the test set from the PAP matcher (case 2 in Table 2). The results indicate improvement in precision while recall dropped, as shown in case 3 in Table 2. This confirms the design characteristics of matchers.

Table 3: Performance analysis of the SVP matcher model.

| Step | Algorithm | ED | TP | FP | FN | TN | Prec. | Recall | F1 | Accuracy | RT |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | SpVarNorm | N/A | 305,309 | 3,495 | 73,960 | 484,964 | 0.9887 | 0.8050 | 0.8874 | 0.9107 | 1 min |
| 2 | MES | 2 | 371,385 | 156,648 | 7,884 | 331,811 | 0.7033 | 0.9792 | 0.8187 | 0.8104 | 7 hr |
| 3 | ES | 1 | 376,646 | 270,881 | 2,623 | 217,578 | 0.5817 | 0.9931 | 0.7336 | 0.6848 | 23 hr |
| 4 | MES | 3 | 377,004 | 285,046 | 2,265 | 203,413 | 0.5694 | 0.9940 | 0.7241 | 0.6689 | 8 min |
| 5 | ES | 2 | 378,134 | 337,461 | 1,135 | 150,998 | 0.5284 | 0.9970 | 0.6907 | 0.6098 | 26 hr |
| 6 | MES | 4 | 378,211 | 340,105 | 1,058 | 148,354 | 0.5265 | 0.9972 | 0.6892 | 0.6068 | 2 min |

The next step is to apply this SVP matcher model to the MNS to generate LMW candidates from MEDLINE. The running time of this model on the Lexicon took over 56 hours (sum of the RT column in Table 3) even with a powerful computer with 192 GB memory. The running time will be exponentially increased when applying the SVP model on the MNS, which is over 22 times the size of the Lexicon. This is impractical and not feasible in real practice. Thus, the smaller size (38%) DMNS is chosen as input to replace the MNS for reducing the processing time without sacrificing recall. Further purification processes of core-term normalization and frequency threshold restriction (WC > 150) are also applied to reduce the size of the n-gram set for better performance. As a result, 752,920 spVars in 269,871 spVar classes are identified by running this computer program for 20 days and are used for LMWs building in the SPECIALIST Lexicon.

### 5.2 More Filters and Matchers

Other filters and matchers have also been developed to apply to the DMNS to further improve LMW building. For example, domain filters exclude terms that are in a certain domain, such as single word, frequency, and existing in the current Lexicon.

By requirement, a valid LMW must have a meaning. Thus, a term with valid concept(s) has a better possibility of being a valid LMW. We utilized UMLS Metathesaurus concepts to create one such matcher, the Metathesaurus CUI Pattern (MCP) matcher. The Synonym Mapping Tool (SMT) in STMT (Lu and Browne, 2012) is used to retrieve Metathesaurus concepts (CUIs) in this model to generate LMW candidates. The SMT is set up to find concepts within 2 subterm substitutions by their synonyms. The default synonym list in SMT is used. In addition, an End-Word Pattern (EWP) matcher was also developed. In the biomedical domain, multiwords often end with certain words (End-Words), such as [syndrome] (e.g. "*migraine syndrome*", "*contiguous gene syndrome*"), [disease] (e.g. "*Fabry disease*", "*Devic disease*"), and so on. An End-Word candidate list composed of the top 20 frequency End-Words for LMWs has been derived from the Lexicon. These End-Words are used in the EWP matcher to retrieve LMW candidates.

The combining of filters and matchers improves precision. This work focuses on generating high precision LMW candidates for effective LMW building. On the other hand, the recall of the matchers is not emphasized because there are too many multiwords yet to be found.

## 6 CONCLUSIONS

A set of high recall rate filters has been developed. These filters are used to derive the distilled MEDLINE n-gram set, resulting in reducing its size to 38%, with better precision and similar recall to that of the MEDLINE n-gram set. These filters and the distilled n-gram set have been tested against the Lexicon and a test set of terms retrieved from MNS by PAP matchers. The distilled MEDLINE n-gram set is needed for further NLP processes with complicated algorithms, such as the SVP matcher model, to reduce the running time for retrieving more LMW candidates for Lexicon building.

Other matchers have also been developed and evaluated. Combinations of filters and matchers have been used to generate high precision LMW candidates for effectively building the Lexicon. The LSG plans to continuously enhance and develop filters and matchers for further improvement. The filters and matchers we have developed are generic and can be used independently or in combination for different research purposes. The approach of generating the distilled MEDLINE n-gram set is also generic and can be applied to other n-gram sets for reducing size and better precision without sacrificing recall. Most importantly, this approach provides a modular and extendable framework for more and better filters and matchers for LMW acquisition and NLP research.

Multiwords are pervasive, challenging and vital in NLP. The LSG aims to provide a lexicon with high coverage of multiwords matching that of single words. We believe the impact of enriched multiword acquisition will enhance the precision, recall, and naturalness of NLP applications. The SPECIALIST Lexicon, the MEDLINE n-gram set and the distilled MEDLINE n-gram set (National Library of Medicine, Lexicon: The MEDLINE n-gram set, 2016) are distributed by the National Library of Medicine (NLM) annually via an Open Source License agreement.

## ACKNOWLEDGEMENTS

This research was supported by the Intramural Research Program of the NIH, National Library of Medicine. The authors would like to thank Mr. Guy Divita for his valuable discussions and suggestions.

## REFERENCES


Aronson, A.R., 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In proceedings of AMIA 2001 Annual Symposium, Wash., DC, USA, Nov. 3-7, pages 17-21.

Aronson, A.R. and Lang, F.M., 2010. An Overview of MetaMap: Historical Perspective and Recent Advances. JAMIA, vol. 17, pages 229-236.

Baldwin, T., Bannard, C., Tanaka, T., Widdows, D., 2003. An Empirical Model of Multiword Expression Decomposability. In proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment, Sapporo, Japan, July 12, pages 89-96.

Bejček, E., Straňák, P., Pecina, P., 2013. Syntactic Identification of Occurrences of Multiword Expressions in Text using a Lexicon with Dependency Structures. In proceedings of the 9th Workshop on Multiword Expressions, Atlanta, Georgia, USA, June 13-14, pages 106–115.

Boukobza, R., Rappoport, A., 2009. Multi-Word Expression Identification Using Sentence Surface Features. In proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Singapore, August 6-7, pages 468–477.

Browne, A.C., McCray, A.T., Srinivasan, S., 2000. The SPECIALIST LEXICON. Lister Hill National Center for Biomedical Communications, National Library of Medicine, Bethesda, Maryland, USA, June, pages 30-49.

Calzolari, N., Fillmore, C.J., Grishman, R., Ide, N., Lenci, A., MacLeod, C., Zampolli, A., 2002. Towards Best Practice for Multiword Expressions in Computational Lexicon. In proceedings of the Third International Conference on Language Resources and Evaluation (LREC), Las Palmas, Canary Islands, Spain, May 29-31, pages 1934-1940.

Divita, G., Browne, A.C., Tse, T., Cheh, M.L., Loane, R.F., Abramson, M., 2000. A Spelling Suggestion Technique for Terminology Servers. In proceedings of AMIA 2000 Annual Symposium, Los Angeles, CA, USA, Nov. 4-8, page 994.

Divita, G., Zeng, Q.T., Gundlapalli, A.V., Duvall, S., Nebeker, J., and Samore, M.H., 2014. Sophia: An Expedient UMLS Concept Extraction Annotator. In proceedings of AMIA 2014 Annual Symposium, Wash., DC, USA, Nov. 15-19, pages 467-476.

Fazly, A., Cook, P., Stevenson, S., 2009. Unsupervised Type and Token Identification of Idiomatic Expressions. Computational Linguistics, vol. 35, no. 1, pages 61-103.

Frantzi, K., Ananiadou, S., Mima, H., 2000. Automatic Recognition of Multi-Word Terms: the C-value/NC-value Method. International Journal on Digital Libraries, vol. 3, no. 2, pages 115-130.

Fraser, S., 2009. Technical vocabulary and collocational behaviour in a specialised corpus. In proceedings of the British Association for Applied Linguistics (BAAL), Newcastle University, Sep. 3-5, pages 43-48.

Green, S., de Marneffe, M.C., Bauer, J., and Manning, C.D., 2011. Multiword Expression Identification with Tree Substitution Grammars: A Parsing tour deforce with French. In proceedings of EMNLP, Edinburgh, Scotland, UK, July 27-31, pages 725–735.

Green, S., de Marneffe, M.C., Manning, C.D., 2013. Parsing models for identifying multiword expressions. Computational Linguistics. vol. 39, no. 1, pages 195–227.

Ide, N.C., Loane, R.F., Fushman, D.D., 2007. Essie: A Concept-based Search Engine for Structured Biomedical Text. JAMIA, vol. 14, no. 3, May/June, pages 253-263.

Kim, S.N. and Baldwin, T., 2010. How to pick out token instances of English verb-particle constructions. Language Resources and Evaluation, April, vol. 44, no. 1, pages 97-113.

Lu, C.J. and Browne, A.C., 2012. Development of Sub-Term Mapping Tools (STMT). In proceedings of AMIA 2012 Annual Symposium, Chicago, IL, USA, Nov. 3-7, page 1845.

Lu, C.J., McCreedy, L., Tormey, D., and Browne, A.C., 2012. A Systematic Approach for Automatically Generating Derivational Variants in Lexical Tools Based on the SPECIALIST Lexicon. IEEE IT Professional Magazine, May/June, pages 36-42.

Lu, C.J., Tormey, D., McCreedy, L., Browne, A.C., 2014. Using Element Words to Generate (Multi)Words for the SPECIALIST Lexicon. In proceedings of AMIA 2014 Annual Symposium, Wash., DC, USA, Nov. 15-19, page 1499.

Lu, C.J., Tormey, D., McCreedy, L., Browne, A.C., 2015. Generating the MEDLINE N-Gram Set, In proceedings of AMIA 2015 Annual Symposium, San Francisco, CA, USA, Nov. 14-18, page 1569.

McCray, A.T., Aronson, A.R., Browne, A.C., Rindflesch, T.C., Razi, A., Srinivasan, S., 1993. UMLS Knowledge for Biomedical Language Processing. Bull. Medical Library Assoc., vol. 81, no. 2, pages 184-194.

McCray, A.T., Srinivasan, S., Browne, A.C., 1994. Lexical Methods for Managing Variation in Biomedical Terminologies. In proceedings of the 18th Annual Symposium on Computer Applications in Medical Care, pages 235-239.

National Library of Medicine, Lexicon, 2016. Lead-End-Terms Model. Available from: <https://lsg3.nlm.nih.gov/LexSysGroup/Projects/lexicon/current/docs/designDoc/UDF/multiwords/leadEndTerms/index.html>.

National Library of Medicine. Lexicon, 2016. The MEDLINE n-gram set. Available from: <https://lsg3.nlm.nih.gov/LexSysGroup/Projects/nGram/index.html>.

Pearce, D., 2001. Using Conceptual Similarity for Collocation Extraction. In proceedings of the 4th UK Special Interest Group for Computational Linguistics (CLUK4), University of Sheffield, Sheffield, UK, January 10-11, pages 34–42.

Pecina, P., 2010. Lexical association measures collocation extraction. Language Resources and Evaluation, vol. 44, pages 137-158.

Philips, L., 1990. Hanging on the Metaphone. Computer Language, December, vol. 17, no. 12, pages 39-43.

Ramisch, C., 2014. Multiword Expressions Acquisition: A Generic and Open Framework (Theory and Applications of Natural Language Processing). Springer, 2015th Edition, pages 4, 9, 37.

Sag, I., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D., 2002. Multiword expressions: A pain in the neck for NLP. In proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002), Mexico City, Mexico, pages 1-15.

Sangati, F., Cranenburgh, A.V., 2015. Multiword Expression Identification with Recurring Tree Fragments and Association Measures. In proceedings of Annual conference of North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Denver, Colorado, May 31-June 5, pages 10-8.

Savova, G.K., Masanz, J.J., Ogren, P.V., Zheng, J., Sohn, S., Kipper-Schuler, K.C., Chute, C.G., 2010. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. JAMIA, vol. 17, no. 5, pages 507-513.

Seretan, V. and Wehrli, E., 2009. Multilingual collocation extraction with a syntactic parser. Language Resources and Evaluation, March, vol. 43, no. 1, pages 71-85.

Silva, J.F. and Lopes, G.P., 1999. A Local Maxima method and a Fair Dispersion Normalization for extracting multi-word units from corpora. In proceedings of the Sixth Meeting on Mathematics of Language (MOL6), Orlando, FL, USA, pages 369-381.

Takahashi, S. and Morimoto, T., 2013. Selection of Multi-Word Expressions from Web N-gram Corpus for Speech Recognition. In proceedings of International Symposium on Natural Language Processing (SNLP), Phuket, Thailand, Oct. 28-30, pages 6-11.

Tsvetkov, Y. and Wintner, S., 2011. Identification of Multi-word Expressions by Combining Multiple Linguistic Information Sources. In proceedings of the Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland, UK, July 27-31, pages 836–845.

# Spectral Data Fusion for Robust ECG-derived Respiration with Experiments in Different Physical Activity Levels

Iman Alikhani[1], Kai Noponen[1], Arto Hautala[1], Rahel Ammann[2] and Tapio Seppänen[1]
[1]*Physiological Signal Analysis Group, Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland*
[2]*Swiss Federal Institute of Sport, Magglingen, Switzerland*
{*iman.alikhani, kai.noponen, arto.hautala, tapio.seppanen*}*@oulu.fi, rahel.ammann@baspo.admin.ch*

Keywords: Respiratory Sinus Arrhythmia, Heart Rate Variability, Spectral Fusion, R-peak Amplitude, QRS Morphological Shape, Time-frequency Analysis, Robustness, Single-channel ECG.

Abstract: In this paper, we study instant respiratory frequency extraction using single-channel electrocardiography (ECG) during mobile conditions such as high intensity exercise or household activities. Although there are a variety of ECG-derived respiration (EDR) methods available in the literature, their performance during such activities is not very well-studied. We propose a technique to boost the robustness and reliability of widely used and computationally efficient EDR methods, aiming to qualify them for ambulatory and daily monitoring. We fuse two independent sources of respiratory information available in ECG signal, including respiratory sinus arrhythmia (RSA) and morphological change of ECG time series, to enhance the accuracy and reliability of instant breathing rate estimation during ambulatory measurements. Our experimental results show that the fusion method outperforms individual methods in four different protocols, including household and sport activities.

## 1 INTRODUCTION

Respiratory frequency is a vital physiological signal used for a variety of diagnostic and clinical purposes. Often, it is not measured just by itself but together with other vital signals using a multitude of sensors to judge correlations between a patient's physiology and different diseases. Especially in ambulatory monitoring, where the measurements are made during regular daily activities, the sensors, however, might interfere with and change the breathing rhythms of subjects and cause discomfort. Since instantaneous breathing rate can be estimated indirectly using ECG signal, development of ECG-derived respiration (EDR) software tools could decrease the cost and facilitate making long-term measurements in a more pleasant and true-to-life manner.

The concept of EDR was proposed decades ago in (Moody et al., 1985) and clinically validated in (Moody et al., 1986). Basically, the estimation of EDR is enabled by two physiological phenomena:

- The heart rate (HR) is modulated by the respiration such that R-R intervals (RRI) shorten during inhale and elongate during exhale, which is known as the respiratory sinus arrhythmia (RSA).
- The mechanical effects of chest movement during breathing modulates the observed ECG morphology, which is especially visible in the QRS-complex part and can be measured, e.g. as either R-peak amplitude (RPA) or the morphological scale variation of QRS complexes (MSV).

These derived quantities – RRI, RPA and MSV – are employed widely in published EDR methods that operate on single-channel ECG (Cysarz et al., 2008; Thayer et al., 2002; Schäfer and Kratky, 2008; Correa et al., 2008; Orphanidou et al., 2013; Noponen et al., 2012). A proportion of publications contributing in EDR area, attempt to extract the respiratory time series waveform (Correa et al., 2008; Cysarz et al., 2008; Widjaja et al., 2012), while, some other studies, including our own paper, focus on acquiring breathing rate of subjects regardless of the respiratory waveform morphology (Schäfer and Kratky, 2008; Thayer et al., 2002; Orphanidou et al., 2013).

In 2008, the three following comparison papers were published in the EDR area. The first paper (Schäfer and Kratky, 2008) studied various EDR methods based on RSA and proposed a time-domain counting method to estimate instantaneous respiratory frequency. In the second paper (Correa et al., 2008), the authors compared breathing morphology derived from RRI, RPA and area under R wave (AUR) sig-



Alikhani, I., Noponen, K., Hautala, A., Ammann, R. and Seppänen, T.
Spectral Data Fusion for Robust ECG-derived Respiration with Experiments in Different Physical Activity Levels.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 88-95
ISBN: 978-989-758-213-4

nals with the reference breathing phase recorded from both plethysmography and nasal air flow signals. The third paper (Cysarz et al., 2008), compares the performance of the estimated breathing rate derived from RPA and RRI signal with the actual values and experimented on a database consisting of sleep and wake data. The first paper claims that spectral-based EDR estimation cannot produce accurate results and time-domain advanced counting method is a more stable alternative for the EDR problem which could offer higher correlation. The second paper, concludes that the correlation coefficient between AUR based EDR and plethysmography signal is 0.56 after temporal delay matching, making AUR superior to both the RPA and the RSA by more than 0.05. One limitation of the study is the small data set that they have experimented with. The authors of the third paper found fairly decent agreement (concordance correlation coefficient of 0.41 to 0.79 within different age groups) between the EDR estimation and reference respiration. They also noted that RSA components strength weaken for people over 50 years old which is in agreement with the findings of the first paper.

Although fairly good correspondence has been achieved with rather inactive subjects, morphological variation of ECG that is generated by respiration-related chest movement can be contaminated by a person's movement, body position changes, and upper/lower limb motions; especially during sport activities. Consequently, there is a clear need for a robust EDR method that can tolerate such interference and remain reliable.

The aforementioned EDR methods estimate the respiration from the ECG derived quantities rather directly. More advanced techniques have been proposed in search for a more dependable EDR method to be able to assess breathing frequency or rhythm more robustly. Decomposition-based EDR methods, including principal component analysis (PCA) and kernel PCA, were investigated in (Widjaja et al., 2012). In addition, adapted PCA and independent component analysis (ICA) were utilized in (Tiinanen et al., 2009; Tiinanen et al., 2015) to extract the RSA component from RRI signal and subsequently estimate the breathing rate of subjects. Recently, a robust technique was published in (Orphanidou et al., 2013), which they proposed a selective framework that evaluates both RPA and RRI signals selecting the one with more strongly modulated respiratory component. They represented the filtered RPA and RRI signals with auto regressive (AR) models. An AR model considers the signal to be characterized by a number of spectral poles. They picked the pole within the respiratory band with higher magnitude as the estimated breathing frequency. This model assumes that pole magnitude is a metric for the signal quality and apparently higher quality signal contains distinctive respiratory fluctuation. They claimed that this selective framework outperforms EDR estimations compared to individual RPA and RRI signal by 0.5 and 0.02 unit of mean absolute error (MAE), respectively, in young subjects and 0.14 and 0.13 unit within elderly subjects.

In order to address EDR throughout sport activities, it should be noted that apart from potential artifacts during such measurements, physiological cardiolocomotion coupling (CLC) component is introduced in the heart activity. It is already reported that there is a coupling between locomotor, respiration and heart (Nomura et al., 2003). CLC is caused by the cadence of running, walking or activities which involves limb movement and exterior impacts from floor coming toward body. The rhythmic cadence of subjects alters HRV signal in a periodic manner, by influencing muscle pump on the circulatory system and cardiac reflexes (Novak et al., 2007). This makes EDR estimation an even more challenging problem during mobile activities, particularly at higher exercise intensities.

Although EDR is an old-standing topic in physiological signal analysis, the performance of methods during typical daily activities, specifically household and sport activities, is not very well-examined in the literature. Our goal in this paper is to evaluate widely-used EDR techniques and propose a preliminary robust framework for instant breathing rate estimation throughout uncontrolled ambulatory condition where a mobile subject performs daily activities. We analyze the signals in a robust way to extract the most correlated component to the instant respiratory frequency. To this extent, we are using a fusion technique to make the system redundant, in case artifact, CLC or any other components alter the extracted instantaneous respiratory estimation.

## 2 MATERIALS AND METHODS

In this section, the construction of signals expected to contain respiratory frequency information is explained. The modeling of signal's spectral content using AR time-frequency model and their adjustments, as well as the fusion of information acquired by individual sources are described.

## 2.1 Calculation of RRI, RPA and MSV

Single-channel ECG is recorded concurrently with a spirometer signal. In the subsequent ECG processing, there are some steps that are followed before extraction of RRI, RPA and MSV signals.

Firstly, baseline wander reduction has been performed using a second-order Savitzky-Golay filter with one second window. The filter fits a polynomial function with the given order on each frame of ECG data minimizing the least-squares error, and replaces the considered sample with the fitted value. We assume that the fitted function follows the baseline of the ECG signal closely. Therefore, we subtract the baseline from the ECG.

Secondly, the R-peaks of the ECG are detected. We passed the signal through a 20th-order high-pass Butterworth IIR filter using a zero-phase digital filtering technique to suppress P and T waves and keep R waves steady. The output of filter is explored (using MATLAB findpeaks function) for peaks with specific criteria, including passing a minimum amplitude, minimum peak prominence and peak separation. Applied constraints filter the detected local maximas to R-peaks.

Thirdly, the results of R-peak detection are used in the calculation of derived quantities. The RRI signal is obtained from the successive R-peak time stamp differences as $RR_i$, where the $i$th value of this signal is equal to the difference of $i$th and $i+1$th time stamp of R-peaks. However, the detection of some ectopic beats or false beats is likely when the subjects are performing exercise and daily activities freely. To counteract this, ectopic/false beat detection and editing is conducted using the following procedure:

1. Detect and reject the obvious outliers ($RR_i \leq 200$ ms or $RR_i \geq 2000$ ms). Such intervals are rare in a healthy subject's heart rhythm due to physiological considerations.
2. Estimate short term RRI level by fitting a smoothing spline function on the signal, and consider the absolute difference between the actual RRI values and the smoothed one as an indicator on how much the RRI deviates from the local level.
3. Detect ectopic/false beats as those with $RR_i$s with large deviation from the short term RRI level. More precisely, mark the ones which deviate both more than 50 ms and also more than $95th$ percentile of all deviations from the spline.
4. Interpolate over ectopic/false and outlier beats with spline interpolation, to preserve the number of beats of the initial detection.

Using the aforementioned procedure, it is assured that less than 5% of the $RR_i$s are edited. Accordingly, the RPA signal can be constructed from the ECG amplitudes at R-peak instants. The construction of MSV is explained comprehensively in (Noponen et al., 2012). Initially, a 50 ms window is defined symmetrically around detected R-peak locations. It is followed by the collection of all QRS complexes within the window in a matrix and a definition of a statistical mean shape as a template for the scale variation measure. The candidate QRS complexes are projected into the template and the scale difference between them is considered as MSV signal.
Finally, we use an 8 Hz spline interpolation to have uniformly sampled signal in further spectral analysis. A sample of ECG segment as well as constructed and interpolated signals are depicted in Figure 1.

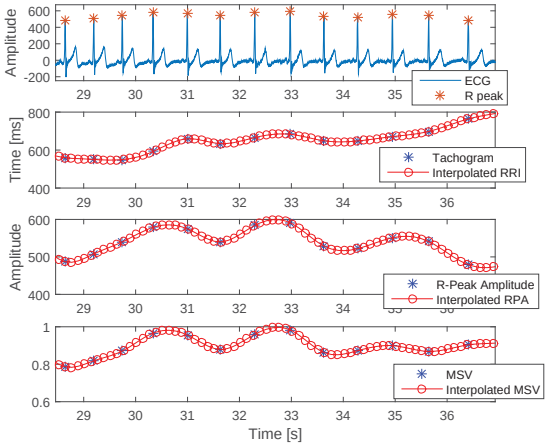



Figure 1: A segment of ECG with constructed signals. Top most sub-figure shows the ECG segment in blue and R-peak annotations are expressed by red stars. The second sub-figure shows original RRI samples. RPA signal is depicted in the third sub-figure and MSV time series in the fourth sub-figure. In the last three sub-figures, the interpolated samples are marked in red, and original values in blue.

## 2.2 Spectral Analysis

We are aiming to extract instant breathing frequency from the constructed signals. Usually, breathing rate of a healthy subject in normal condition varies roughly between 10 to 36 breaths per minute (bpm), while in our application where the subjects are exercising or performing activities, the breathing rate might be over 60 bpm. Therefore, we use a relatively wide band-pass filter with low and high cut-off frequencies of 0.15 Hz and 1.2 Hz corresponding to 9 bpm and 72 bpm. Figure 2 shows the frequency response of the band-pass filter.

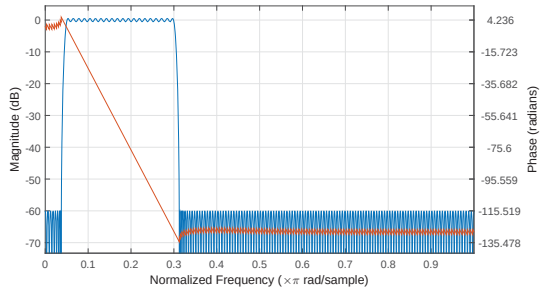



Figure 2: The magnitude and phase response of the FIR filter used to keep the spectral content of signals within the possible respiratory frequency. 60 dB and 1 dB is adjusted as the stop- and pass-band attenuation. Sampling frequency is 8 Hz and the x-axis is the normalized frequency.

### 2.2.1 AR Model

Physiological signals are principally non-stationary, which requires specific tools for spectral analysis. In this study, according to (Thayer et al., 2002; Orphanidou et al., 2013) recommendation, we have used 12th-order AR model on 20-second segments of data which has 19 seconds overlap with adjacent windows. This model considers an all-pole transfer function to describe the frequency distribution of signal. The higher the order of AR model, the more poles are used for the signal description. A sample AR model outcome is illustrated in Figure 3.

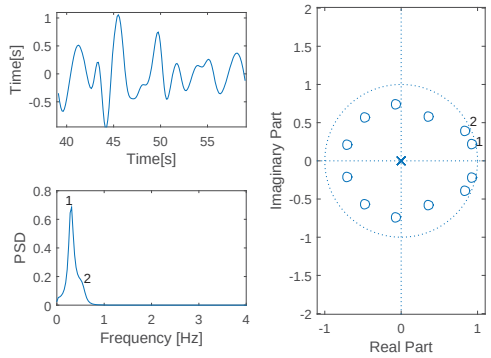



Figure 3: A 20-second segment of interpolated and filtered RRI signal is fed to the AR model and output zero-pole plot is depicted in the right unity circle. Lower left sub-figure illustrates the power spectral density (PSD) of the upper left RRI time series. The local trends of the PSD labeled by 1 and 2 are constructed as a result of poles labeled by 1 and 2 in the $z$-plane. These two poles are higher in magnitude and closer to unity circle which means they have stronger effect on the construction of PSD curve.

To derive the respiratory frequency from the spectrum, in the case of a single source EDR using RRI, RPA or MSV, we find the frequency bin having highest power spectral density. For instance in the lower left sub-figure of Figure 3, the frequency value of signal marked as 1 is considered as the respiratory frequency for this segment of data.

## 2.3 Fusion of PSDs

According to the EDR literature, the power spectra of RRI, RPA and MSV signals within a short time window are expected to contain energy at or near the instantaneous respiratory frequency. In addition, the respiratory component is expected to usually be strong in the sense that spectral power levels around the respiratory frequency are higher than the background levels in each spectrum. However, the spectra are expected to also contain other peaks rising from noise/artifacts, CLC, nonlinearities, and also side lobes induced by amplitude and/or frequency modulation by physiological feedback loops such as the control of heart rate through RSA, for instance.

Due to the different nature of RRI, RPA, and MSV, it can be assumed that the strength of the aforementioned other factors varies between their spectra, but the respiration component should be present in all or most of them. Thus, it makes sense to attempt to find significant energy bands or peaks that are present in all of the spectra. What is more, even when the respiratory component is present in all of them, the redundant combination can be used to narrow down the actual respiratory frequency as the resolving power of individual spectra depends on the width/peakedness of the spectral peak that can vary.

In this paper, we approach the issue via spectral domain fusion that strengthens the joint spectral components and diminishes the ones not shared with other spectra. We hypothesize that the fusion will be advantageous in instantaneous breathing rate estimation. Let's assume that the spectrogram of constructed signals can be expressed as:

$$P_{sig} = \begin{pmatrix} p^{sig}_{1,1} & p^{sig}_{1,2} & \cdots & p^{sig}_{1,n} \\ p^{sig}_{2,1} & p^{sig}_{2,2} & \cdots & p^{sig}_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p^{sig}_{m,1} & p^{sig}_{m,2} & \cdots & p^{sig}_{m,n} \end{pmatrix} \quad (1)$$

where *sig* could be RRI, RPA or MSV signal. Every column in the spectrogram corresponds to the PSD of signal in a specific 20-second time window which has one second difference with the consecutive columns of matrix and every row, to the lied energy at specific frequency band. In the fusion matrix, we compute the product of these arrays element-by-element. It should be noted that these PSDs contain the same number of values. Thus, the fusion spectrogram can be stated as:

$$P_{fus} = \begin{pmatrix} p^{rri}_{1,1} \cdot p^{rpa}_{1,1} \cdot p^{msv}_{1,1} & p^{rri}_{1,2} \cdot p^{rpa}_{1,2} \cdot p^{msv}_{1,2} & \cdots & p^{rri}_{1,n} \cdot p^{rpa}_{1,n} \cdot p^{msv}_{1,n} \\ p^{rri}_{2,1} \cdot p^{rpa}_{2,1} \cdot p^{msv}_{2,1} & p^{rri}_{2,2} \cdot p^{rpa}_{2,2} \cdot p^{msv}_{2,2} & \cdots & p^{rri}_{2,n} \cdot p^{rpa}_{2,n} \cdot p^{msv}_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p^{rri}_{m,1} \cdot p^{rpa}_{m,1} \cdot p^{msv}_{m,1} & p^{rri}_{m,2} \cdot p^{rpa}_{m,2} \cdot p^{msv}_{m,2} & \cdots & p^{rri}_{m,n} \cdot p^{rpa}_{m,n} \cdot p^{msv}_{m,n} \end{pmatrix} \quad (2)$$

$P_{fus}$ basically gives a joint spectrogram which considers the agreement between individual spectrogram trends as well as their strength. The frequency bin where the maximum energy is settled at each time instant, is selected as the estimated respiratory frequency. In case there is a correlation between at least two of the signal's PSDs with sufficient strength, the fusion PSD also offers the same trend in the fusion spectrum. However, if there is no correlation between PSDs, the fusion spectrum is affected by the PSD with higher energy. In other words, the fusion method intuitively considers the signal with stronger component as suggested in (Orphanidou et al., 2013) and also decides in a cooperative manner.

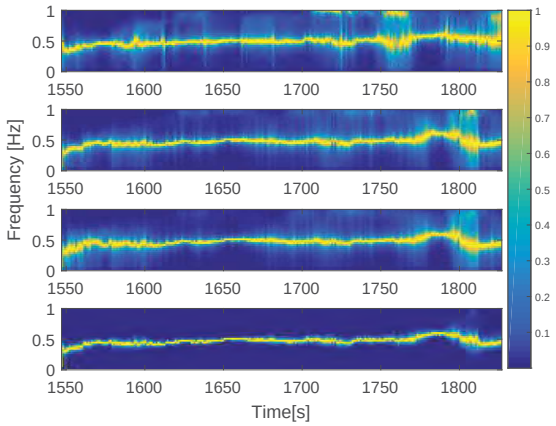



Figure 4: A sample 100-second spectrogram of the constructed signals plus the fusion spectrogram. All the spectrograms are normalized at each time instant for better visualization. The top most sub-figure illustrates RRI, the second top is RPA and the third row shows the MSV normalized spectrogram and the bottom sub-figure is the normalized fusion spectrogram.

It is conceivable that in specific cases and time instants other combinations of spectrograms – such as the joint product of a certain pair of them – may yield better performance than the product triplet. Nevertheless, we expect the presented approach, in which the element-wise multiplication is taken over all the three spectrograms, to perform better on average. Thus, in the following, we consider only the fusion (2) that combines RRI, RPA, and MSV spectrograms.

In Figure 4, the normalized spectrograms of a sample RRI, RPA and MSV signals taken at the same time are depicted. The normalized fusion spectrogram is illustrated in the last sub-figure. Wide distribution of energy in some parts of individual spectrograms is visible, while the fusion spectrogram (bottom sub-figure) earns the common component between the individual spectrums and is considerably narrower.

## 2.4 Database

Since, in this study, we are aiming to evaluate the performance of our instantaneous respiratory frequency estimation methods during uncontrolled ambulatory measurements where the subjects can freely perform their daily activities, we have collected 67 subjects (30 female and 37 male) aged from 18 to 60 years old during household and sport activities. The overall general physiological characteristics of subjects are stated in Table 1.

Table 1: General characteristics of subjects participated in this experiment.

| Characteristic | mean | min | max |
|---|---|---|---|
| Height ($cm$) | 175 | 160 | 195 |
| Weight ($Kg$) | 75.4 | 45.6 | 122.8 |
| Age ($Years$) | 37.9 | 18 | 60 |
| BMI ($Kg/m^2$) | 24.51 | 14.72 | 35.5 |

The ECG signal is recorded using an available commercial electrode belt in the market and upsampled to 1 kHz for HRV analysis and the spirometer data is collected at the rate of one sample per second.

The household activities are comprised of four minutes of floor sweeping (FS) followed by four minutes of table cleaning (TC). The sports activity part of the protocol consists of 10 minutes of cycling (CY) on an ergometer, followed by four minutes of Tennis playing (TN) in a gym hall. During these activities, both spirometer and single-channel ECG data are collected. The relative intensity level of four activity protocols is given in Table 2 as the overall percentage of maximal heart rate ($HR_{max}$) of subjects in that specific activity.

Table 2: Overall intensity of activity protocols as a percentage of $HR_{max}$.

| Activity Protocol | mean | min | max |
|---|---|---|---|
| FS | 52 | 35 | 73 |
| TC | 50 | 34 | 77 |
| CY | 66 | 48 | 83 |
| TN | 81 | 63 | 90 |

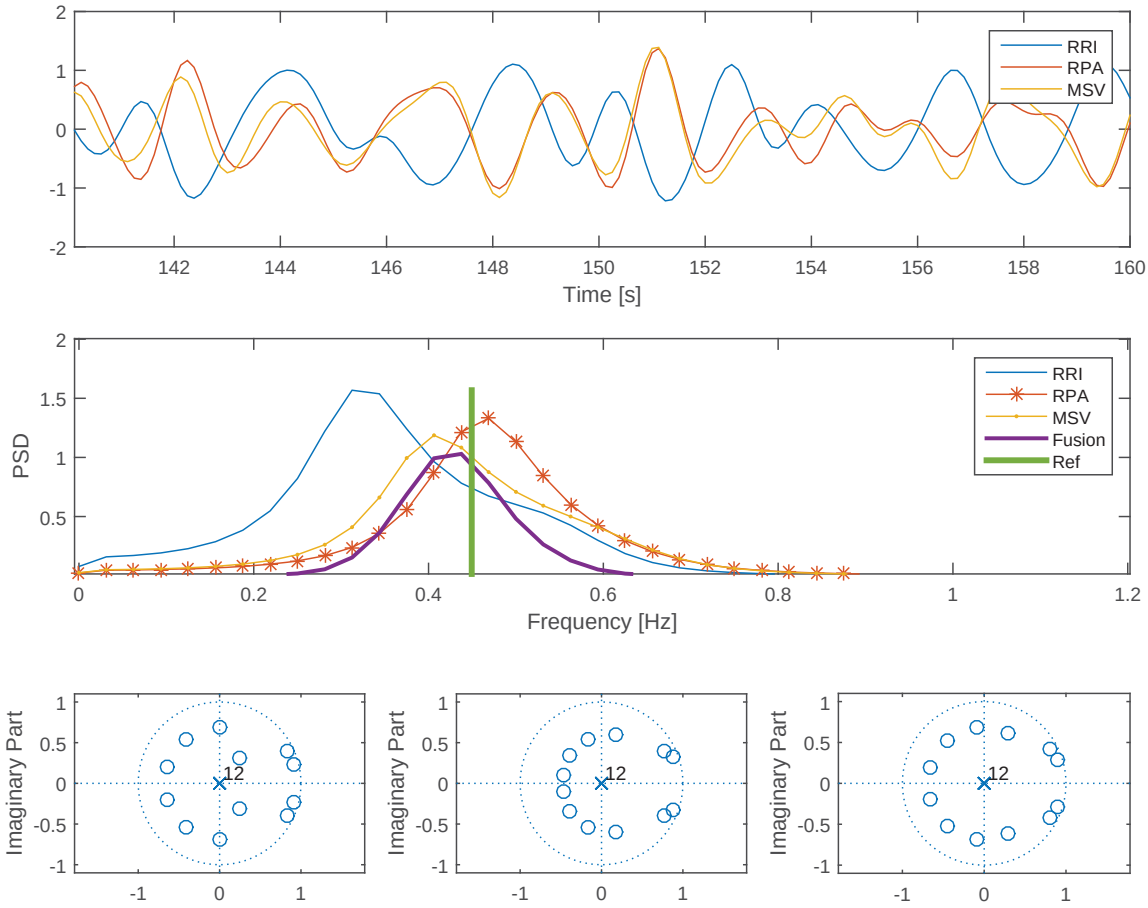



Figure 5: A 20-second sample household activity data including interpolated and filtered RRI, RPA and MSV signal as well as their PSD in the middle sub-figure. The last sub-figure shows RRI, RPA and MSV *z*-plane, respectively, from left to right.

# 3 RESULTS

## 3.1 Performance Measures

Let's assume that our estimated breathing frequency is expressed as $x$ and the original respiratory frequency as $y$ and $n$ is the number of samples, in order to assess our estimation results, the following metrics are computed:

- Root Mean Square Error:

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j - y_j)^2} \tag{3}$$

- Mean Absolute Percentage Error: MAPE weights frequency of respiration. Basically, it considers a larger error margin in higher breathing rate estimation.

$$MAPE = \frac{1}{n}\sum_{j=1}^{n}\left|\frac{x_j - y_j}{y_j}\right| \tag{4}$$

- Concordance Correlation Coefficient: It is a reliable measure to evaluate the agreement between two sets of signals. It can be computed using

$$R_c = \frac{2S_{xy}}{S_x^2 + S_y^2 + (\hat{y} - \hat{x})^2} \tag{5}$$

where $\hat{x}$ and $\hat{y}$ are the average frequency of estimated and original signals and

$$S_{xy} = \frac{1}{n}\sum_{j=1}^{n}(x_j - \hat{x})(y_j - \hat{y}) \tag{6}$$

and $S_x$ and $S_y$ are the standard deviation of $x$ and $y$, respectively.

### 3.2 Quantitative Results

Figure 5 illustrates a 20-second sample data. It shows the signals, their frequency components distributions and specification of poles. In the middle sub-figure, three PSDs as well as computed fusion PSD are depicted. Reference respiratory frequency is also expressed by a straight green line. Instant breathing rate estimation using PSD is corresponded to the frequency where maximum energy is settled.

Table 3 summarizes the results of the methods used in four activity protocols. The data shows that the fusion method outperforms the individual methods in all the four protocols considering different metrics. The performance of EDR derived from RRI signal (RSA-based breathing frequency estimation) is the weakest compared to other two individual signals particularly in sport activities. It might be due to the reason that RRI signal is more vulnerable to CLC or movement artifacts during high intensity exercise.

Table 3: Acquired overall results in four different activity protocols.

| Spect | Metric | Activity | | | |
|---|---|---|---|---|---|
| | | FS | TC | CY | TN |
| RRI | RMSE | 6.2 | 5.4 | 5.3 | 8.3 |
| | MAPE | 19.0 | 18.0 | 16.0 | 18.0 |
| | Rc | 0.23 | 0.19 | 0.39 | 0.33 |
| RPA | RMSE | 5.0 | 4.5 | 3.4 | 6.9 |
| | MAPE | 16.0 | 15.0 | 10.0 | 15.0 |
| | Rc | 0.2 | 0.18 | 0.57 | 0.39 |
| MSV | RMSE | 4.7 | 4.4 | 3.7 | 6.4 |
| | MAPE | 15.0 | 14.0 | 11.0 | 13.0 |
| | Rc | 0.25 | 0.19 | 0.5 | 0.43 |
| Fusion | RMSE | 4.6 | 4.1 | 2.9 | 6.4 |
| | MAPE | 14.0 | 13.0 | 8.8 | 13.0 |
| | Rc | 0.28 | 0.24 | 0.57 | 0.45 |

## 4 CONCLUSION

Ambulatory measurement of instantaneous respiratory frequency can be achieved via ECG surrogate signal processing. However, the performance of breathing rate estimation during uncontrolled condition when the subject is free to move and perform his/her daily activities is in question and not well-studied. This paper proposed a spectral fusion technique which combines the information from individual sources of EDRs, such as RSA-based (RRI signal) and morphological-based (RPA and MSV signals), to boost the performance of estimation using computationally-efficient methods. In essence, the presented method considers the agreement between the individual estimators and their joint spectral power. Overall, our fusion method outperforms the individual methods considering all the metrics and experimented activity protocols.

## REFERENCES


Correa, L. S., Laciar, E., Torres, A., and Jane, R. (2008). Performance evaluation of three methods for respiratory signal estimation from the electrocardiogram. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4760–4763. IEEE.

Cysarz, D., Zerm, R., Bettermann, H., Frühwirth, M., Moser, M., and Kröz, M. (2008). Comparison of respiratory rates derived from heart rate variability, ecg amplitude, and nasal/oral airflow. *Annals of biomedical engineering*, 36(12):2085–2094.

Moody, G. B., Mark, R. G., Bump, M. A., Weinstein, J. S., Berman, A. D., Mietus, J. E., and Goldberger, A. L. (1986). Clinical validation of the ecg-derived respiration (edr) technique. *Group*, 1(3).

Moody, G. B., Mark, R. G., Zoccola, A., and Mantero, S. (1985). Derivation of respiratory signals from multi-lead ecgs. *Computers in cardiology*, 12(1985):113–116.

Nomura, K., Takei, Y., and Yanagida, Y. (2003). Comparison of cardio-locomotor synchronization during running and cycling. *European journal of applied physiology*, 89(3-4):221–229.

Noponen, K., Tiinanen, S., and Seppänen, T. (2012). Deriving respiration from the electrocardiogram by serial comparison with statistical mean shape. In *2012 Computing in Cardiology*, pages 809–812. IEEE.

Novak, V., Hu, K., Vyas, M., and Lipsitz, L. A. (2007). Cardiolocomotor coupling in young and elderly people. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 62(1):86–92.

Orphanidou, C., Fleming, S., Shah, S. A., and Tarassenko, L. (2013). Data fusion for estimating respiratory rate from a single-lead ecg. *Biomedical Signal Processing and Control*, 8(1):98–105.

Schäfer, A. and Kratky, K. W. (2008). Estimation of breathing rate from respiratory sinus arrhythmia: comparison of various methods. *Annals of Biomedical Engineering*, 36(3):476–485.

Thayer, J. F., Sollers III, J. J., Ruiz-Padial, E., and Vila, J. (2002). Estimating respiratory frequency from autoregressive spectral analysis of heart period. *IEEE Engineering in Medicine and Biology*, 21(4):41–45.

Tiinanen, S., Noponen, K., Tulppo, M., Kiviniemi, A., and Seppänen, T. (2015). Ecg-derived respiration methods: Adapted ica and pca. *Medical engineering & physics*, 37(5):512–517.

Tiinanen, S., Tulppo, M., and Seppänen, T. (2009). Rsa component extraction from heart rate signal by independent component analysis. In *2009 36th Annual Computers in Cardiology Conference (CinC)*, pages 161–164. IEEE.

Widjaja, D., Varon, C., Dorado, A., Suykens, J. A., and Van Huffel, S. (2012). Application of kernel principal component analysis for single-lead-ecg-derived respiration. *IEEE Transactions on Biomedical Engineering*, 59(4):1169–1176.

# Technological Approach for Behavior Change Detection toward Better Adaptation of Services for Elderly People

Firas Kaddachi[1], Hamdi Aloulou[1], Bessam Abdulrazak[1,2], Joaquim Bellmunt[3], Romain Endelin[1], Mounir Mokhtari[3,4] and Philippe Fraisse[1]

[1]*Montpellier Laboratory of Informatics, Robotics and Microelectronics (LIRMM), Montpellier, France*

[2]*University of Sherbrooke, Sherbrooke, Canada*

[3]*Image and Pervasive Access Lab (IPAL), Singapore, Singapore*

[4]*Institut Mines-Telecom (IMT), Paris, France*

*name.surname@lirmm.fr, bessam.abdulrazak@usherbrooke.ca, joaquim.bellmunt@ipal.cnrs.fr, mounir.mokhtari@mines-telecom.fr*

Keywords: Behavior, Behavior Change Detection, Elderly People, Sensors.

Abstract: Aging process is associated with behavior change and continuous decline in physical and cognitive abilities. Therefore, early detection of behavior change is major enabler for providing adapted services to elderly people. Today, different psychogeriatric methods target behavior change detection. However, these methods require presence of caregivers and manual analysis. In this paper, we present our technological approach for early behavior change detection. It consists in monitoring and analyzing individual activities using pervasive sensors, as well as detecting possible changes in early stages of their evolution. We also present a first validation of the approach with real data from nursing home deployment.

## 1 INTRODUCTION

Early detection of behavior change is keystone for service providers to better adapt their services for elderly people. Existing psychogeriatric methods for behavior change detection are inconvenient, as they are time-consuming and require manual analysis work from caregivers.

According to existing definitions (Cao, 2010), behavior change is defined as any continuous modification or transformation in way and manner of behavior execution. Behavior change characterizes possible instabilities, variations, impairments, declines, increases or improvements in behavior performance.

Behavior change has significant impact on quality of life. For example, emergence of orientation problems (Cockrell and Folstein, 2002), eating difficulties (Vellas et al., 1999) and mood impairments (Parmelee and Katz, 1990) leads to serious decline in quality of life. On the other hand, any improvement in managing personal finances (Barberger-Gateau et al., 1992), managing household (Lafont et al., 1999) and mobility (Mathias et al., 1986) has positive influence on quality of life.

Early detection of behavior change is major enabler for more efficient intervention, by taking necessary actions in early stages of behavior change. Autonomy of elderly people is consequently improved, by reducing symptoms and evolution of sensor, motor and cognitive diseases.

In this paper, we propose a technological approach for behavior change detection. Changes are detected at temporal scale; i.e., compared to past habits of one particular person.

Our employed technologies (*e.g.,* movement and contact sensors) do not interfere with monitored behavior. These technologies are ubiquitous. They disappear in the environment, without generating unwanted behavior change, and without affecting individual privacy.

Our approach conducts long-term analysis of behavior for detection of continuous changes. We do not study snapshots of behavior, but we analyze overall behavior over long periods. This enables to differentiate between transient and continuous deviations.

Following, section 2 discusses state of the art of behavior definitions and change detection methods. Sections 3 and 4 present our methodology for behavior change detection and our implementation approach. Section 5 introduces a first validation of the proposed approach through real results from nursing home deployment. Section 6 concludes this paper.



Kaddachi, F., Aloulou, H., Abdulrazak, B., Bellmunt, J., Endelin, R., Mokhtari, M. and Fraisse, P.
Technological Approach for Behavior Change Detection toward Better Adaptation of Services for Elderly People.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 96-105
ISBN: 978-989-758-213-4

# 2 RELATED WORK

Researchers study behavior change from two different perspectives: initiation and maintenance of behavior change, and detection of behavior change.

## 2.1 Behavior Definitions

Cao defines behavior as actions or reactions made by individuals (Cao, 2010). Behavior is a response to internal or external stimuli or inputs.

From sociological point of view, Wilson considers behavior as interactions between individuals (Wilson, 2000). It can be influenced by family structure, work or school environment relationships, health conditions and psychiatric issues.

Economists recognize behavior as processes consumers go through or reactions they have toward purchasing or consuming products or services (Perner, 2008; Szwacka-Mokrzycka, 2015). It is influenced by internal factors, such as attitudes, needs, motives and preferences. External factors have also significant influence, such as marketing activities, social, economical and cultural aspects.

In the medical field, behavior refers to persons' beliefs and actions regarding their health (Miller et al., 2007; Lavikainen et al., 2009). While positive behaviors promote healthy life (*e.g,* maintain moderate alcohol intake, not smoke and avoid snacks), negative behaviors present health risks.

These definitions consider behavior as a response to internal or external factors, such as intentions, desires, social interactions and marketing activities. However, individuals respond differently to these factors; e.g., being hungry stimulates individuals to prepare meals with different duration, frequency and difficulty. Therefore, we define behavior as the way and manner individuals perform actions, inactions and beliefs.

## 2.2 Behavior Change Initiation and Maintenance Models

Numerous models have been proposed to predict the amount of effort individuals require for behavior change initiation and maintenance (Ormrod, 2013). In fact, initiating and maintaining behavior changes are related to individuals' perception of their own ability to perform demanding or challenging tasks. This perception is influenced by individuals' prior success in those tasks or related tasks, their psychological state and outside sources of persuasion.

In the medical field, behavior change refers to abandoning health-compromising behaviors and maintaining health-improving behaviors. Rosenstock suggests that individuals' belief about health problems and perceived benefits of actions plays important role in adopting health-promoting behaviors (Rosenstock, 1974).

Schwarzer considers behavior change as two continuous processes: goal setting and goal pursuit (Schwarzer, 2008). While goal setting is related to factors that motivate behavior change, goal pursuit consists in planning and performing intended change.

Prochaska *et al.* propose a five-step model of behavior change (Prochaska and DiClemente, 2005). In the first step, individuals have not thought about changing their behaviors. Then, individuals begin thinking about changing particular behaviors. Afterwards, they prepare their plans for behavior change. In the fourth step, individuals adopt and perform new behaviors. Finally, they consistently conserve their new behaviors.

While these models target behavior change initiation and maintenance, detection of behavior change enables better fulfillment of both objectives. In deed, methods for behavior change detection allow to make better decisions of when to initiate new behavior changes, and which services to select for behavior change initiation and maintenance.

## 2.3 Behavior Change Detection Methods

In the literature, we distinguish psychogeriatric and technological methods for behavior change detection. While psychogeriatric methods use formal tests and questionnaires, technological solutions are developed to automate detection of anomalies.

### 2.3.1 Psychogeriatric Methods

Psychologists and geriatricians propose several internationally validated methods for behavior change detection (Table 1). Using formal scales and questionnaires, trained clinicians and caregivers request that seniors reply to specific questions and perform specific tasks, such as "How many falls did you have in the last six months?" (Tardieu et al., 2016) and "Could you please get up and walk three meters away!" (Mathias et al., 1986).

Following, we present the psychogeriatric tests of Table 1:

- **Short Emergency Geriatric Assessment (SEGA)** allows to evaluate frailty of elderly people (Tardieu et al., 2016). It considers multiple behavior changes, such as falls, nutrition

Table 1: Examples of Psychogeriatric Tests for Behavior Change Detection.

| | SEGA | MMSE | 4Tests | GDS | IADL | AGGIR | GetUp AndGo | MNA | BEHA VEAD | NPI |
|---|---|---|---|---|---|---|---|---|---|---|
| **ADL** | X | | | | X | X | | | X | X |
| **Mobility** | X | | | | | X | X | | | |
| **Cognition** | X | X | X | | X | X | | X | X | X |
| **Social Life** | X | | | | X | X | | X | X | X |
| **Nutritional Status** | X | | | | | X | | X | | |
| **Biological Status** | X | | | | X | X | | X | | |
| **Mood and Emotions** | X | | | X | | | | | X | X |

problems, mobility impairments and memory troubles.

- **Mini Mental State Examination (MMSE)** targets detection of changes in cognitive abilities, such as orientation problems, attention difficulties and language troubles (Cockrell and Folstein, 2002).
- **Benton, Five Word, Clock and Verbal Fluency Tests (4Tests)** target changes in cognitive functions, such as learning problems, memory troubles and construction difficulties (Neuropsy, 2016).
- **Geriatric Depression Scale (GDS)** investigates changes in mood and emotions (Parmelee and Katz, 1990); e.g., feeling sad and that one's life is empty is associated with possible depression.
- **Instrumental Activities of Daily Living (IADL)** identifies changes in activities of daily living that are associated with autonomy loss, such as using telephone, using means of transport, taking medicines and managing personal finances (Barberger-Gateau et al., 1992).
- **Autonomie Gerontologique et Groupes Iso-Ressources (AGGIR)** investigates changes in autonomy of seniors, such as movement troubles, household difficulties and orientation impairments (Lafont et al., 1999).
- **Get-Up and Go** targets motor behavior changes (Mathias et al., 1986). This test asks elderly people to get up, walk and turn around, in order to analyze task execution and identify possible mobility impairments.
- **Mini Nutritional Assessment (MNA)** investigates changes in nutritional status, such as eating difficulties, weight loss and protein intake insufficiency (Vellas et al., 1999).
- **Behavioral Pathology in Alzheimers Disease (BEHAVE-AD) and Neuropsychiatric Inventory (NPI)** allow to detect possible behavioral impairments for elderly people, such as presence of hallucinations, aggressiveness and anxiety (Reisberg et al., 1997) (Cummings et al., 1994).

Using these tests, clinicians observe task execution and analyze senior behavior, in order to identify cognitive impairments, autonomy problems, rapid mood changes, nutritional and behavioral anomalies.

Certain inconveniences limit the efficiency of psychogeriatric tests. In fact, it is inconvenient for seniors to recall past events with full details at assessment time. It is also often not convenient for elderly people to move to assessment place.

Besides, requesting that individuals reply to given questions and perform determined tasks has potential negative impact on their future behaviors after assessment. For example, anxiety of seniors can increase in case they feel their inability to correctly reply to orientation questions or perform mobility tasks. Furthermore, subjective evaluation of assessment results cause possible assessment inaccuracies.

#### 2.3.2 Technological Methods

Different technological methods target behavior change detection. They employ technologies deployed in the environment (*e.g.,* movement sensors, bed sensors, cameras and microphones) or worn by seniors (*e.g.,* smart phone, smart watch and neurosensors). These methods conduct advanced analysis of acquired data, in order to detect changes in monitored behaviors.

Allin *et al.* propose technological method for social behavior change detection (Allin et al., 2003). This method detects emergence of physically and verbally aggressive interactions. It employs cameras and microphones for continuous collection of video and audio recordings. Using hidden markov models, complex analysis of these recordings allows to build typical movement patterns for anomaly detection. However, employed cameras and microphones affect privacy of individuals.

Avvenuti *et al.* target detection of wandering and falls from bed during sleep (Avvenuti et al., 2010).

They study correlation between brain activity and body movement, in order to define rules and derive threshold values for anomaly detection. This method employs modern neurosensors placed on person's scalp to record brain activities. Yet, neurosensors limit individual movements indoors and outdoors.

Another technological method studies mental state changes that lead to possible depression (Magill and Blum, 2012). It provides objective feedback to patients using body and environmental sensors, in addition to subjective questionnaire-based records for their health. Content and timing of questionnaires are personalized for individuals and altered over time as individual's mental health changes. However, it has been reported that technical trial of developed system reveals acceptability issues from participants regarding questionnaires.

Kaye *et al.* investigate changes in computer use (Kaye et al., 2014). Based on statistical analysis of mouse events, they compare frequency and duration of computer use between two aging populations with or without mild cognitive impairments (MCI). They conclude that MCI patients use computers less than regular persons.

Hayes *et al.* target detection of medication adherence changes (Hayes et al., 2009). Using electronic pillbox, subjects take medicines twice per day at specific times. This method monitors number of days when subjects take both medicines, and verifies whether volunteers adhere to given times. It compares two senior groups with low or high performance in given cognitive tests, and concludes that lower performing group has risk of non-adherence.

Another technological study investigates motor behavior changes (Hayes et al., 2008). Careful in-series placement of wireless infrared sensors at home identifies how quickly and frequently seniors pass through sensor lines per day. Comparing two aging populations with or without MCI, MCI patients show a coefficient of variation in median walking speed as twice as high compared to regular subjects.

Petersen *et al.* propose further solution to detect changes in telephone use (Petersen et al., 2014). Employed land-line phone monitors record phone events, such as dialed numbers and ring rate. These recordings allow to have a picture on size and contact frequency of friend, family and acquaintance network. Results show that seniors with high cognitive abilities receive significantly more phone calls.

These last four studies detect behavior changes between different individuals. However, they do not target detection of behavior changes that affect one particular individual.

# 3 BEHAVIOR CHANGE DETECTION METHODOLOGY

We target behavior change detection at temporal scale. Over long periods, we analyze behavior of elderly people, in order to identify changes compared to past habits. Our technologies do not interfere with monitored behavior and do not affect individual privacy.

Our behavior analysis identifies indicators of behavior change, such as activities of daily living, mobility and social life (Figure 1). These indicators are associated with changes in physical and cognitive abilities of elderly people.

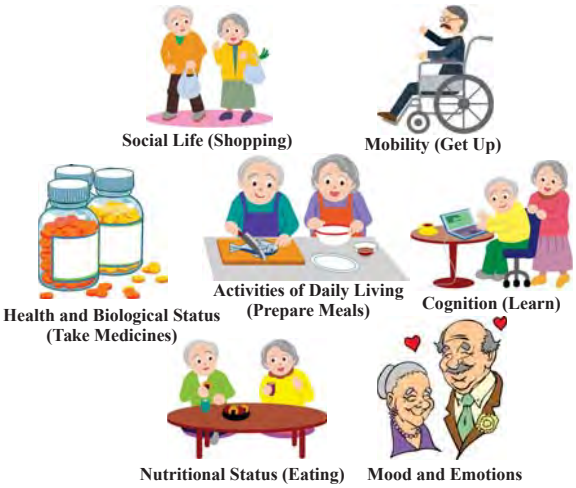



Figure 1: Examples of Behavior Change Indicators.

We also analyze these indicators considering different dimensions, such as quantity, duration, time and location. These dimensions are metrics that quantify collected data and allow to apply algorithms on these data for change detection.

Furthermore, we correlate identified changes with global context, such as weather conditions, family status and house architecture. Considering these factors provides better understanding of detected changes; e.g., senior stays at home for seven days due to heavy snow and not due to eventual social isolation.

## 3.1 Behavior Change Indicators

We have considered different validated psychogeriatric scales (*e.g.,* SEGA, AGGIR, MNA and NPI) to identify indicators of behavior change. That can be captured via ambient technologies (Table 2). Analyzing these indicators allows to detect significant changes in physical and cognitive abilities. Figure 1 shows following examples of indicators:

- **Activities of Daily Living** are essential complex tasks of daily living that demand important physical and cognitive capacities, such as performing household, preparing meals, dressing, hygiene, and urinary and fecal elimination.
- **Mobility** refers to motor behaviors, such as moving indoors and outdoors, getting up, turning around and walking.
- **Cognition** includes essential cognitive tasks such as learning, language and managing financial situation. These complex tasks are associated with temporal orientation, spatial orientation, attention, calculation and construction.
- **Social Life** refers to social behaviors, such as communicating with others, using means of transport, shopping and participating in collective free time activities.
- **Nutritional Status** is related to serving oneself and eating.
- **Health and Biological Status** targets health behaviors that indicate vision, audition and vital sign impairments, such as irregularities in taking medicines and increased hospitalization number.
- **Mood and Emotions** correlate with physical and cognitive functions and are significant depression and stress indicators.

Table 2: Examples of Ambient Technologies for Behavior Change Indicator Monitoring.

| Environment | Technologies |
|---|---|
| **Indoor** | Movement, contact, proximity, vibration and pressure sensors |
| **Outdoor** | Smart phone and smart watch with beacons |

## 3.2 Metrics

We analyze selected behavior change indicators regarding different dimensions. These dimensions are metrics that quantify way and manner of performing these indicators, and allow to apply algorithms on collected data for change detection. Following, we discuss four significant metrics:

- **Quantity** refers to number and amount of behavior execution; e.g., number of friend visits decreases due to social isolation, number of movements decreases due to walk impairments, number of sport center visits increases thanks to raised interest in physical exercise and number of hospitalizations decreases thanks to health status improvement.
- **Duration** is related to length of behavior execution; e.g., duration of preparing meals increases due to cognitive impairments, duration of stair climbing increases due to walk impairments, time spent out of home increases thanks to raised interest in social interactions and time spent in free time activities considerably increases thanks to raised interest in active aging.
- **Time** refers to start and end times of behavior execution; e.g., sleep hours are irregular due to sleep troubles, eating meal hours are inappropriate due to nutritional problems, going out hours are changing thanks to raised interest in social activities and taking medicine hours are adhered thanks to cognitive status improvement.
- **Place** describes where behavior is executed; e.g., detected falls outdoors become more frequent due to fear of going outside, visiting senior activity center becomes less usual due to social isolation and visiting city park becomes more frequent thanks to raised interest in physical exercise.

## 3.3 Global Context

Analyzing behavior in correlation with global context enables better understanding of behavior change. Following, we discuss influence of general personal information, general context information and specific temporary information on behavior change.

General personal information are general descriptors of persons; e.g., age over 85, health care history including more than three physical and mental diseases, and inconvenient family status increase the probability of behavior changes (Tardieu et al., 2016).

General context information describe the environment of behavior execution; e.g., changing one's house affects activities of daily living, moving television in room not easily accessible by elderly people reduces watching television frequency, opening smart city subways adapted for elderly people has positive influence on outdoor activities and building senior activity centers raises interest in social interactions.

Specific temporary information refer to short-term events, such as several consecutive days of heavy snow that obligate senior to stay at home, recent hospitalization of husband that raises wife's anxiety, and recent friend visits that improve emotional state.

# 4 IMPLEMENTATION APPROACH

We perform a first implementation of our behavior change detection methodology in our ambient assisted living platform UbiSMART (Aloulou, 2013; Aloulou et al., 2013). This platform uses environmental sensor data for activity recognition, detection of abnormal activity change and provision of personalized services for elderly people (Figure 2).

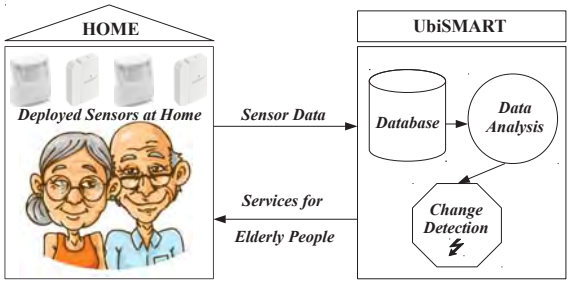



Figure 2: Overview on our Ambient Assisted Living Platform.

Our implementation approach considers following stages:

- **Deployment** consists in installing our hardware infrastructure. This includes environmental sensors (*e.g.,* movement and contact sensors), gateways, receivers and internet access points.
- **Data Acquisition** is essential to build our database. Via internet, data are transmitted to our dedicated server for permanent storage.
- **Data Pre-processing** allows to discard inaccurate and erroneous data for better analysis quality.
- **Data Analysis** quantifies data by considering different metrics, such as daily number and duration of shopping activity. Afterwards, we apply algorithms on these data to detect possible changes at temporal scale; e.g., these algorithms identify decrease in shopping activity periods.

## 4.1 Algorithms

We select statistical algorithms for our data analysis, as they differentiate between transient and continuous deviations; e.g., these statistical algorithms ignore occasional decreases in going out frequency, and consider only continuous decreases as significant changes in going out frequency.

We can distinguish offline and online algorithms for change detection in the literature (Basseville et al., 1993; Liu et al., 2013). Offline algorithms require fully available data as input, such as full history of free time activity number and duration. However, online algorithms iteratively operate on data one by one, such as number and duration of free time activity day by day.

Existing online algorithms use probabilistic models (Takeuchi and Yamanishi, 2006), singular spectrum analysis (Moskvina and Zhigljavsky, 2003) and cumulative sum control charts (Mesnil and Petitgas, 2009). Existing offline algorithms apply relative density-ratio estimation (Liu et al., 2013) and cumulative sum control charts with binary segmentation (Andersson, 2014; Cho, 2015) or bootstrapping (Taylor, 2000).

In order to detect changes as early as possible, we select online algorithms. Following, we discuss two algorithms investigated in our research; i.e., cusum-based (Page, 1954) and window-based (Bland and Altman, 1995) algorithms. These algorithms apply different filters on detected deviations and identify changes with different granularity.

### 4.1.1 CUSUM-based Algorithm

Page proposes Cumulative Sum Control Chart (CUSUM) algorithm for change detection in time series (Page, 1954). This algorithm considers two phases: reference phase and analysis phase.

In reference phase, initial data allow to compute parameters that will condition change detection:

- **M** refers to mean of reference data.
- **SD** is standard deviation of reference data.
- **SHIFT** is related to shift of interest, that determines smallest deviation we target to detect. Ledolter *et al.* set SHIFT to 1x SD (Ledolter and Kardon, 2013).
- **K** refers to allowance parameter, that is related to shift of interest. Mesnil *et al.* set K to 0,5x SHIFT (Mesnil and Petitgas, 2009).
- **H** is decision parameter that determines whether change occurs or not. In the literature, researchers define H with published tables, specific software or set it to 5x SD (Mesnil and Petitgas, 2009; Kibria, 2016).

In the analysis phase, mean and standard deviation parameters allow to standardize data by applying formula 1:

$$data[i] = (data[i] - M)/SD \quad (1)$$

For each datum, cumulative sums recursively accumulate positive and negative deviations, using formula 2 and 3:

$$S_{HIGH}[i] = max(0, S_{HIGH}[i-1] + data[i] - K) \quad (2)$$

$$S_{LOW}[i] = min(0, S_{LOW}[i-1] + data[i] + K) \quad (3)$$

In case $S_{HIGH}$ is higher then +H or $S_{LOW}$ is lower than -H, positive or negative change occurs.

#### 4.1.2 Window-based Algorithm

Based on Bland-Altman analysis, window-based algorithm applies moving window on input data to distinguish between transient deviations and continuous change (Bland and Altman, 1995). Only in case selected number of deviations are consecutively detected without interruption, change occurs.

Positive or negative deviations are data values that are higher or lower than $M \pm SD$, where M and SD correspond respectively to mean and standard deviation of all previously observed data including currently observed datum.

Window length (N) depends on analyzed behavior; e.g., seven consecutive days of staying at home correspond to change in going out frequency or three consecutive months of loosing weight indicate change in nutritional status. Positive or negative changes are detected in case N consecutive positive or negative deviations occur.

## 5 VALIDATION

Following, we present a first validation of our approach through real data from nursing home deployment. Considering mobility as indicator of behavior change, collected data allow to analyze movements of patients inside their individual rooms.

### 5.1 Data Collection

Table 3: Patient Gender, Age and Monitoring Period.

| Patient | Gender | Age | Period(months) |
|---|---|---|---|
| A | M | 90 | 6 |
| B | M | 89 | 5 |
| C | M | 81 | 2 |
| D | F | 84 | 11 |
| E | F | 95 | 2 |
| F | F | 85 | 13 |
| G | F | 87 | 13 |
| H | F | 92 | 9 |
| I | F | 92 | 4 |

Over one year, we deploy movement sensors in bedrooms and bathrooms of 9 patients in a french nursing home in Occagnes (Table 3). Average age of patients is 88 years.

### 5.2 Data Analysis

We use movement sensor data to analyze physical activity periods (PAP) of persons. We simply define a PAP as period of consecutive movements, that are detected with time difference less than 3 minutes.

We do not consider days of inactivity, that correspond to hospitalizations or holidays outside individual rooms. In our analysis, we quantify collected movement sensor data using following metrics:

- **Number** refers to quantity of detected movements and PAPs.
- **Duration** is total length of detected PAPs.
- **Intensity** measures mean number of detected movements per PAP. This corresponds to number of detected movements divided by number of detected PAPs.

### 5.3 Results

Figure 3 shows our analysis results for patient F over 13 months. For each month, we compute average of daily number of movements, PAPs, their duration and intensity. We also study influence of mean ambient temperature on physical activities.

We observe decrease in movement and PAP number in case ambient temperature increases. However, PAP duration grows and PAP intensity is quite stable. This is also observed for other patients. Higher temperature stimulate them to perform less activities with longer total duration inside individual rooms.

For early change detection, we apply cusum-based and window-based algorithms on collected data after each day. In order to validate their results, we also apply an offline algorithm on full months of data.

In the literature, offline algorithms provide more robust results than online algorithms, as they retrospectively analyze longer periods of data (Basseville et al., 1993; Liu et al., 2013).

We select offline algorithm of Change Point Analyzer (CPA) tool (Taylor, 2000). This algorithm implements an iterative combination of cumulative sum control charts and bootstrapping to detect changes. Table 4 shows dates and values of identified changes in movement number data of patient F.

Results of cusum-based and window-based algorithms are compared to those obtained with CPA tool in Figure 4, considering true positive rate (TPR), precision (P), true negative rate (TNR) and accuracy (A).

Cusum-based and window-based (N=5 and N=4) algorithms show true positive rate of 28%, 40% and 45% respectively, as they do not detect all changes. Their precision is 64%, 56% and 33% respectively,

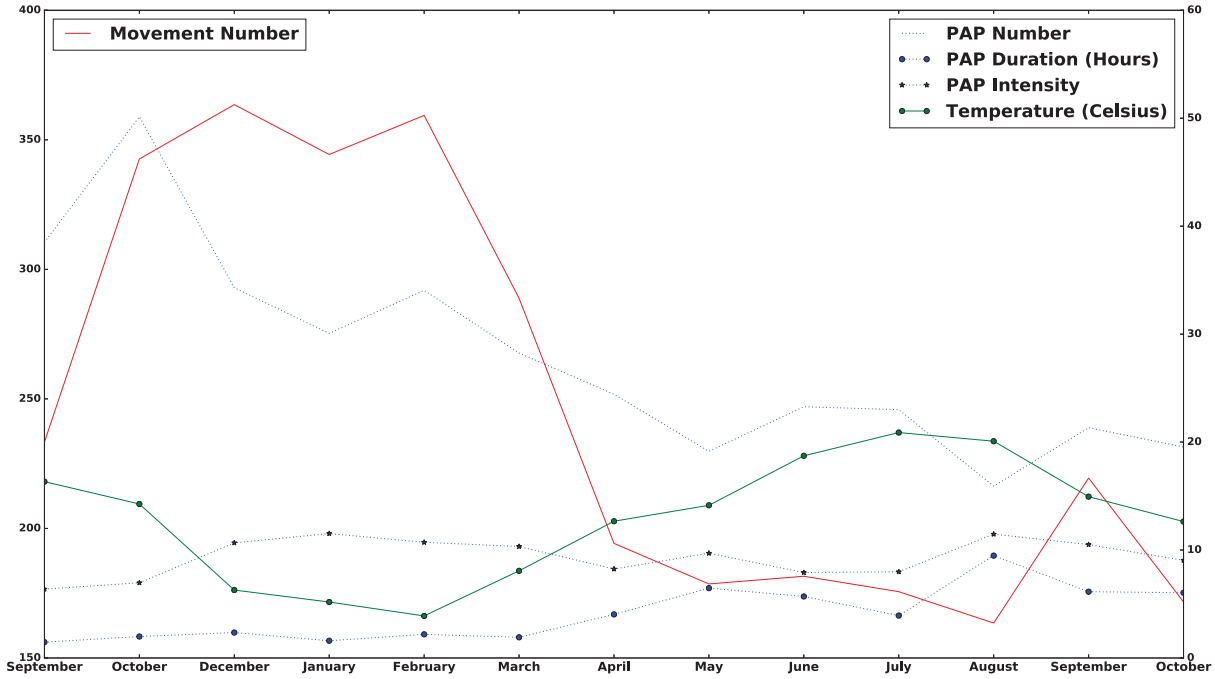



Figure 3: Monthly Average of Movement Number, PAP Number, Duration and Intensity for Patient F.

Table 4: Change Dates and Values of Movement Number for Patient F.

| Change Date | From | To |
|---|---|---|
| 2014, October 3 | 239 | 354 |
| 2015, March 15 | 354 | 253 |
| 2015, April 3 | 253 | 180 |
| 2015, September 13 | 180 | 282 |
| 2015, September 18 | 282 | 176 |

which indicates that not all identified changes are relevant.

However, true negative rate is 99%, 99% and 98% respectively, as they correctly identify almost all normal data. Their accuracy is 97%, 98% and 97% respectively, which corresponds to good overall results.

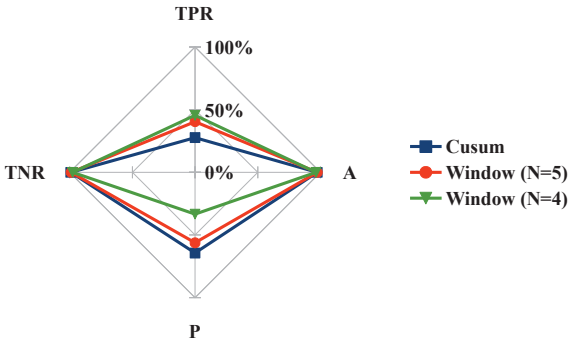



Figure 4: Comparison of Cusum-based and Window-based Algorithms to CPA Tool.

## 6 CONCLUSION

We propose a technological approach for behavior change detection at temporal scale. We analyze overall behavior to identify changes compared to past habits over long periods. Our technologies disappear in the environment, in order to avoid generation of unwanted changes and protect individual privacy.

We also present a first validation of our methodology through real data from nursing home deployment. Over months, employed movement sensors allow to monitor physical activities of patients. Collected data are quantified considering different metrics, such as number and duration. Our selected statistical change detection algorithms provide good overall results.

We are working on improving our behavior change detection in the context of the European project City4Age (City4Age, 2016). The City4Age project target using data generated by technologies deployed in urban areas, in order to provide new adaptable services for elderly people. These services target capturing frailty of elderly people, and provisioning subsequent individualized interventions.

Further technologies are investigated for more diversified analysis of behavior; e.g., bed sensors can be

used for sleep period and vital sign recognition, kinect sensors enable more accurate monitoring of walking activity, and beacon sensors with smart phones allow more precise understanding of outdoor activities.

New reasoning techniques are studied to correlate identified statistical changes with overall changes in behavior toward better adaptation of provided services; e.g., decrease in weight indicates negative nutritional change and triggers sending of personalized notifications to improve nutritional status.

## ACKNOWLEDGEMENT

We give our special thanks to Saint Vincent de Paul nursing home in Occagnes, France. Our deployment in this nursing home is also supported by VHP inter@ctive project and the Quality Of Life chair.

Our work is part of the European project City4Age that received funding from the Horizon 2020 research and innovation program under grant agreement number 689731.

## REFERENCES

Allin, S., Bharucha, A., Zimmerman, J., Wilson, D., Robinson, M., Stevens, S., Wactlar, H., and Atkeson, C. (2003). Toward the automatic assessment of behavioral distrubances of dementia.

Aloulou, H. (2013). *Framework for ambient assistive living: handling dynamism and uncertainty in real time semantic services provisioning*. PhD thesis, Evry, Institut national des télécommunications.

Aloulou, H., Mokhtari, M., Tiberghien, T., Biswas, J., Phua, C., Lin, J. H. K., and Yap, P. (2013). Deployment of assistive living technology in a nursing home environment: methods and lessons learned. *BMC medical informatics and decision making*, 13(1):1.

Andersson, J. (2014). Locating multiple change-points using a combination of methods.

Avvenuti, M., Baker, C., Light, J., Tulpan, D., and Vecchio, A. (2010). Non-intrusive patient monitoring of alzheimers disease subjects using wireless sensor networks. In *2009 World Congress on Privacy, Security, Trust and the Management of e-Business*.

Barberger-Gateau, P., Commenges, D., Gagnon, M., Letenneur, L., Sauvel, C., and Dartigues, J.-F. (1992). Instrumental activities of daily living as a screening tool for cognitive impairment and dementia in elderly community dwellers. *Journal of the American Geriatrics Society*, 40(11):1129–1134.

Basseville, M., Nikiforov, I. V., et al. (1993). *Detection of abrupt changes: theory and application*, volume 104. Prentice Hall Englewood Cliffs.

Bland, J. M. and Altman, D. G. (1995). Comparing methods of measurement: why plotting difference against standard method is misleading. *The lancet*, 346(8982):1085–1087.

Cao, L. (2010). In-depth behavior understanding and use: the behavior informatics approach. *Information Sciences*, 180(17):3067–3085.

Cho, H. (2015). Change-point detection in panel data via double cusum statistic.

City4Age (2016). Elderly-friendly city services for active and healthy aging. *available as http://www.city4ageproject.eu/*.

Cockrell, J. R. and Folstein, M. F. (2002). Mini-mental state examination. *Principles and practice of geriatric psychiatry*, pages 140–141.

Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., and Gornbein, J. (1994). The neuropsychiatric inventory comprehensive assessment of psychopathology in dementia. *Neurology*, 44(12):2308–2308.

Hayes, T. L., Abendroth, F., Adami, A., Pavel, M., Zitzelberger, T. A., and Kaye, J. A. (2008). Unobtrusive assessment of activity patterns associated with mild cognitive impairment. *Alzheimer's & Dementia*, 4(6):395–405.

Hayes, T. L., Larimer, N., Adami, A., and Kaye, J. A. (2009). Medication adherence in healthy elders: small cognitive changes make a big difference. *Journal of aging and health*.

Kaye, J., Mattek, N., Dodge, H. H., Campbell, I., Hayes, T., Austin, D., Hatt, W., Wild, K., Jimison, H., and Pavel, M. (2014). Unobtrusive measurement of daily computer use to detect mild cognitive impairment. *Alzheimer's & Dementia*, 10(1):10–17.

Kibria, G. (2016). Cumulative sum and exponentially weighted moving average control charts. *preprint, available as http://www2.fiu.edu/ kibriag /Stat5666/Handout/Chapter99.pdf*.

Lafont, S., Barberger-Gateau, P., Sourgen, C., and Dartigues, J. (1999). Relation entre performances cognitives globales et dépendance évaluée par la grille aggir. *Revue d'épidémiologie et de santé publique*, 47(1):7–17.

Lavikainen, H. M., Lintonen, T., and Kosunen, E. (2009). Sexual behavior and drinking style among teenagers: a population-based study in finland. *Health promotion international*, 24(2):108–119.

Ledolter, J. and Kardon, R. (2013). Detecting the progression of eye disease: Cusum charts for assessing the visual field and retinal nerve fiber layer thickness. *Translational vision science & technology*, 2(6):2–2.

Liu, S., Yamada, M., Collier, N., and Sugiyama, M. (2013). Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, 43:72–83.

Magill, E. and Blum, J. M. (2012). Personalised ambient monitoring: supporting mental health at home. *Advances in home care technologies: Results of the Match project*, pages 67–85.

Mathias, S., Nayak, U., and Isaacs, B. (1986). Balance in elderly patients: the" get-up and go" test. *Archives of physical medicine and rehabilitation*, 67(6):387–389.

Mesnil, B. and Petitgas, P. (2009). Detection of changes in

time-series of indicators using cusum control charts. *Aquatic Living Resources*, 22(2):187–192.

Miller, J. W., Naimi, T. S., Brewer, R. D., and Jones, S. E. (2007). Binge drinking and associated health risk behaviors among high school students. *Pediatrics*, 119(1):76–85.

Moskvina, V. and Zhigljavsky, A. (2003). An algorithm based on singular spectrum analysis for change-point detection. *Communications in Statistics-Simulation and Computation*, 32(2):319–352.

Neuropsy, R. (2016). Four Tests. http://fmc31200.free.fr/MG-liens/Neurologie/quatre_tests.pdf. [Online; accessed September-2016].

Ormrod, J. E. (2013). *Educational Psychology: Pearson New International Edition: Developing Learners*. Pearson Higher Ed.

Page, E. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.

Parmelee, P. A. and Katz, I. R. (1990). Geriatric depression scale. *Journal of the American Geriatrics Society*.

Perner, L. (2008). Consumer behaviour and marketing strategy.

Petersen, J., Austin, D., Yeargers, J., and Kaye, J. (2014). Unobtrusive phone monitoring as a novel measure of cognitive function. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 10(4):P366–P367.

Prochaska, J. O. and DiClemente, C. C. (2005). The transtheoretical approach. *Handbook of psychotherapy integration*, 2:147–171.

Reisberg, B., Auer, S. R., and Monteiro, I. M. (1997). Behavioral pathology in alzheimer's disease (behave-ad) rating scale. *International Psychogeriatrics*, 8(S3):301–308.

Rosenstock, I. M. (1974). Historical origins of the health belief model. *Health Education & Behavior*, 2(4):328–335.

Schwarzer, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology*, 57(1):1–29.

Szwacka-Mokrzycka, J. (2015). Trends in consumer behaviour changes. overview of concepts. *Acta Scientiarum Polonorum. Oeconomia*, 14(3).

Takeuchi, J.-i. and Yamanishi, K. (2006). A unifying framework for detecting outliers and change points from time series. *IEEE transactions on Knowledge and Data Engineering*, 18(4):482–492.

Tardieu, É., Mahmoudi, R., Novella, J.-L., Oubaya, N., Blanchard, F., Jolly, D., and Drame, M. (2016). External validation of the short emergency geriatric assessment (sega) instrument on the safes cohort. *Geriatrie et psychologie neuropsychiatrie du vieillissement*, 14(1):49–55.

Taylor, W. A. (2000). Change-point analysis: a powerful new tool for detecting changes. *preprint, available as http://www. variation. com/cpa/tech/changepoint. html*.

Vellas, B., Guigoz, Y., Garry, P. J., Nourhashemi, F., Bennahum, D., Lauque, S., and Albarede, J.-L. (1999). The mini nutritional assessment (mna) and its use in grading the nutritional state of elderly patients. *Nutrition*, 15(2):116–122.

Wilson, E. O. (2000). *Sociobiology*. Harvard University Press.

# Identifying Serendipitous Drug Usages in Patient Forum Data
## *A Feasibility Study*

Boshu Ru[1], Charles Warner-Hillard[2], Yong Ge[3] and Lixia Yao[4,1]
[1]*Department of Software and Information Systems, University of North Carolina at Charlotte, Charlotte, NC, U.S.A.*
[2]*Department of Public Health Sciences, University of North Carolina at Charlotte, Charlotte, NC, U.S.A.*
[3]*Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC, U.S.A.*
[4]*Department of Health Sciences Research, Mayo Clinic, Rochester, MN, U.S.A.*
*{bru, cwarnerh, yong.ge}@uncc.edu, yao.lixia@mayo.edu*

Keywords: Social Media, Drug Repositioning, Machine Learning, Patient-Reported Outcomes.

Abstract: Drug repositioning reduces safety risk and development cost, compared to developing new drugs. Computational approaches have examined biological, chemical, literature, and electronic health record data for systematic drug repositioning. In this work, we built an entire computational pipeline to investigate the feasibility of mining a new data source – the fast-growing online patient forum data for identifying and verifying drug-repositioning hypotheses. We curated a gold-standard dataset based on filtered drug reviews from WebMD. Among 15,714 sentences, 447 mentioned novel desirable drug usages that were not listed as known drug indications by WebMD and thus were defined as serendipitous drug usages. We then constructed 347 features using text-mining methods and drug knowledge. Finally we built SVM, random forest and AdaBoost.M1 classifiers and evaluated their classification performance. Our best model achieved an AUC score of 0.937 on the independent test dataset, with precision equal to 0.811 and recall equal to 0.476. It successfully predicted serendipitous drug usages, including metformin and bupropion for obesity, tramadol for depression and ondansetron for irritable bowel syndrome with diarrhea. Machine learning methods make this new data source feasible for studying drug repositioning. Our future efforts include constructing more informative features, developing more effective methods to handle imbalance data, and verifying prediction results using other existing methods.

## 1 INTRODUCTION

Drug repositioning, also known as drug repurposing, is the identification of novel indications for marketed drugs and drugs in the late-stage development (Dudley et al., 2011). A well-known example is sildenafil, which was originally developed to treat angina in clinical trial. However, after failure, it was resurrected to treat erectile dysfunction (Ashburn and Thor, 2004). Another example is the repositioning of duloxetine from depression to stress urinary incontinence, which was irresponsive to many drug therapies at that time (Ashburn and Thor, 2004). These successful stories demonstrated advantages of drug repositioning over new drug discovery and development. Repositioned drugs have a better safety profile than compounds in the early discovery and development stage, as they have already passed several preclinical tests in animal models and safety tests on human volunteers in the Phase I clinical trials. Thus the time and cost of early drug discovery and development can be saved, making repositioned drugs more available to the patients of currently not properly treated diseases and more cost-efficient to pharmaceutical companies (Yao et al., 2011). Despite some potential intellectual property issues, drug repositioning carries the promise of significant societal benefits and has attracted broad interests from the biomedical community in the past decade.

Traditionally, drug-repositioning opportunities were discovered by serendipity. In the case of sildenafil, the clinical team was inspired with the new repositioning idea when they found that some patients enrolled in the original trial for angina were reluctant to return the medicine due to the desirable side effect (Shandrow, 2016). Various computational methods have been developed to systematically explore more drug-repositioning opportunities. One common strategy is to mine chemical, biological, or



Ru, B., Warner-Hillard, C., Ge, Y. and Yao, L.
Identifying Serendipitous Drug Usages in Patient Forum Data - A Feasibility Study.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 106-118
ISBN: 978-989-758-213-4

clinical data for drug similarity, disease comorbidity, or drug-disease associations that imply repositioning opportunities (Dudley et al., 2011, Andronis et al., 2011). For instance, Keiser et al. (2009) compared chemical structure similarities among 3,665 drugs and 1,400 protein targets to discover unanticipated drug-target associations and implicated the potential role of Fabahistin, an allergy drug, in treating Alzheimer's disease. Sanseau et al. (2012) investigated data from genome-wide association studies to systematically identify alternative indications for existing drugs and suggested repositioning denosumab, which was approved for osteoporosis, for Crohn's disease. Hu and Agarwal (2009) created a drug-disease network by mining the gene-expression profiles in GEO datasbase and the Connectivity Map project. By analyzing topological characteristics of this network, they inferred the effects of cancer and AIDS drugs for Huntington's disease. Wren et al. (2004) constructed a network of biomedical entities including genes, diseases/phenotypes, and chemical compounds from MEDLINE (U.S. National Library of Medicine, 2016a), and computationally identified novel relationships between those biomedical entities in scientific publications. One such relationship they found and validated in the rodent model was between chlorpromazine and cardiac hypertrophy. Gottlieb et al. (2011) designed an algorithm called PREDICT, to discover novel drug-disease associations from OMIM, DrugBank, DailyMed, and Drugs.com. Their algorithm predicted 27% of drug-disease associations in clinical trials registered with clinicaltrial.gov. Although these computational methods have demonstrated their promise, they often face the issue of high false positive rates (Dudley et al., 2011, Shim and Liu, 2014). One primary reason is sharing similar chemical structures or co-occurring in the same publication does not always imply medical relevance. Also, ignoring the context (e.g., whether the similarity or validation is observed in experiments on molecular, cell line, or animal models) might impact their capability to be translated to human beings.

More recently, researchers began to verify some drug-repositioning hypotheses using the Electronic Health Record (EHR) data. For example, Khatri et al. (2013) retrospectively analyzed the EHR of 2,515 renal transplant patients at the University Hospitals Leuven to confirm the beneficial effects of atorvastatin on graft survival. Xu et al. (2014) verified that metformin, a common drug for type 2 diabetes, is associated with improved cancer survival rate by analyzing the patients' EHR data from Vanderbilt University Medical Center and Mayo Clinic. These proof-of-concept studies also witnessed several limitations, due to the nature of EHR data: (1) EHR systems do not record the causal relationships between events (e.g., drugs and side effects) as they are mostly designed for clinical operation and patient management instead of research. Whether a statistical association is causal needs to be verified through temporal analysis with a lot of assumptions. Therefore, the models become disease and/or drug specific and remain difficult to generalize and automate in large scale. (2) A significant amount of valuable information, such as the description of medication outcomes, is stored in clinicians' notes in free-text format (Yao et al., 2011). Mining these notes requires advanced natural language processing techniques and presents patient privacy issues. (3) In the US, data from a single provider's EHR system only provide an incomplete piece of patient care (Xu et al., 2014). Integrating EHR data from multiple providers may be a solution, but currently encounters legal and technical challenges, as discussed in depth by Jensen et al. (2012). Due to these limitations, neither EHR, nor any of scientific literature, biological, and chemical data alone appear sufficient for drug repositioning research. We need to identify additional data sources that contain patient medication history and outcomes, as well as develop advanced data integration methods to identify synergistic signals.

In the last decade or so, another type of patient data has increased exponentially in volume with the emergence of smart phones and social media websites. People today not only post their travel pictures but also share and discuss their experiences with diseases and drugs in patient forums and social media websites, such as WebMD, PatientsLikeMe, Twitter, and YouTube (Ru et al., 2015). Such data directly describes drug-disease associations in real human patients and bypasses the translational hurdle from cell-line or animal model to human, thus has led to increased research interests. For example, Yang et al. (2012) detected adverse drug reaction (ADR) signals from drug related discussions in the MedHelp forum by using an ADR lexicon created from the Consumer Health Vocabulary. Yates and Goharian (2013) extracted ADR in the breast cancer drug reviews on askpatient.com, drugs.com, and drugratingz.com using a ADR synonym list generated from the United Medical Language System (UMLS) specifically for breast cancer. Rather than collecting existing social media discussions, Knezevic et al. (2011) created a Facebook group for people to report their ADR outcomes and found social media a highly sensitive

instrument for ADR reporting . Powell et al. (2016) investigated the MedDRA Preferred Terms that appeared on Twitter and Facebook and found 26% of the posts contained useful information for post-marketing drug safety surveillance.

In this work, we expand current social media mining research that is primarily ADR focused to the discovery of serendipitous drug usages, which can suggest potentially new drug repositioning hypotheses. We build a computational pipeline based on machine learning methods to capture the serendipitous drug usages on the patient forum published by WebMD, which was reported in a previous study (Ru et al., 2015) to have high-quality patient reported medication outcomes data. However, this is an extremely difficult machine learning task because: (1) User comments on patient forum are unstructured and informal human language prevalent with typographic errors and chat slangs. It is unclear how to construct meaningful features with prediction power; (2) the mentioning of serendipitous drug usages by nature is very rare. Based on our experience with the drug reviews on WebMD, the chance of finding a serendipitous drug usage in user posts is less than 3% (See Methods). Therefore, we caution the audience that our objective in this work is not to build a perfect pipeline or a high performance classifier, but to perform a feasibility check and identify major technical hurdles in the entire workflow. We plan to direct our systems engineering efforts towards improving the performance of those bottleneck modules as the next step.

# 2 METHODS

In this feasibility study, we built the entire computational pipeline using standard tools and applications, to identify serendipitous drug usages in patient forum data, which includes data collection, data filtering, human annotation, feature construction and selection, data preprocessing, machine learning model training and evaluation, as illustrated in Figure 1. Each module is further described below.

## 2.1 Data Collection

We started by collecting drug reviews posted by anonymous users on the patient forum hosted by WebMD. WebMD is a reputable health care website that exchanges disease and treatment information among patients and healthcare providers. In its patient forum, after filling the basic demographic information including gender and age group, users are allowed to rate drugs in terms of effectiveness, ease of use, overall satisfaction, and post additional comments about their medication experience (See Figure 2). We chose it based on two considerations: (1) With over 13 years' history of operation and on average over 150 million unique visits per month, WebMD contains a large volume of drug reviews that is highly desirable for conducting systematic studies. (2) The quality of drug reviews was reported to be superior to many other social media platforms in a previous study (Ru et al., 2015). Spam reviews, commercial advertisements, or information irrelevant to drugs or diseases are rare, probably thanks to their forum modulators. We downloaded a total number of 197,883 user reviews on 5,351 drugs by the date of March 29, 2015. Then, we used Stanford CoreNLP (Manning et al., 2014) to break down each free-text comment into sentences, which is the standard unit for natural language processing and text mining analysis.

## 2.2 Gold Standard Dataset for Serendipitous Drug Usages

In machine learning and statistics, gold standard, or accurately classified ground truth data is highly desirable, but always difficult to obtain for

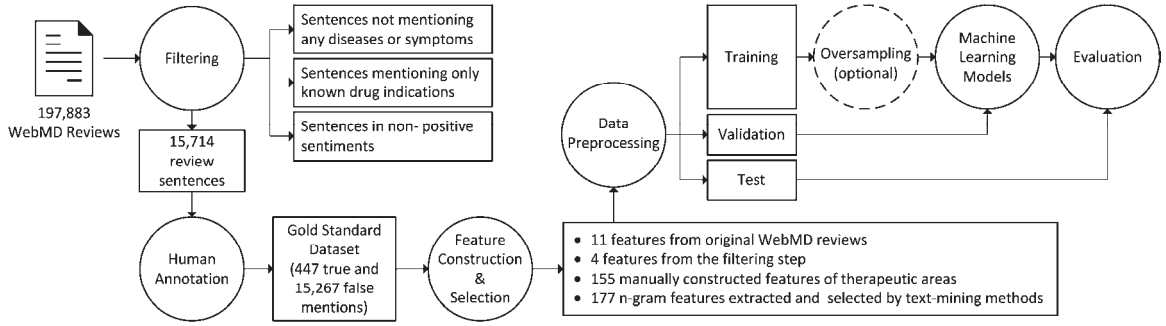



Figure 1: A workflow to identify serendipitous drug usages in patient forum data.

supervised learning tasks. For identifying serendipitous drug usages, it would be ideal if a database of drug usages approved globally or customarily used off-label were readily available as the benchmark for known drug usages. The professional team at WebMD has published monographs to introduce each drug, including information on drug use, side effects, interactions, overdose, etc. We thus used such data as the benchmark for known drug usages in this work. We assume a drug use is serendipitous if the user mentioned improvement of his or her condition or symptom that was not listed in the drug's known indications according to WebMD (See the examples in Figure 2). Otherwise, we set the mentioned drug use to be non-serendipitous. Below we explain in more details how we applied this principal to semi-automatically prepare our gold standard dataset for serendipitous drug usages.

## 2.3 Data Filtering

Three filters were designed to reduce the number of drug review sentences to a number more manageable for human annotation. Firstly, we identified and removed review sentences that did not mention any disease or symptom at all, because these sentences have no chance to be related to serendipitous drug usages. To do this, we selected the UMLS concepts in English and with the semantic types equal to *Disease or Syndrome*, *Finding*, *Injury or Poisoning*, *Mental or Behavioral Dysfunction*, *Neoplastic Process*, or *Sign or Symptom* and used them to approximate medical concepts that could be related to serendipitous drug usages. We then used MetaMap (Aronson and Lang, 2010) to identify these medical concepts in each review sentence. Next, for sentences that did mention any of those concepts, we used SNOMED CT (U.S. National Library of Medicine, 2016b) to determine whether the mentioned concept is semantically identical or similar to the drug's known indications listed on WebMD. Mathematically SNOMED CT is a directed acrylic graph model for medical terminology. Medical concepts are connected by defined relationships, such as *is-a*, *associated with*, and *due to*. The semantic similarity between two concepts was usually measured by the length of the shortest path between them in the graph (Pedersen et al., 2007, Shah and Musen, 2008). If the medical concept mentioned in a review sentence was more than three steps away from the known indications of the drug, we assumed the mentioned medical concept was more likely to be an unanticipated outcome for the drug and kept the sentence in the dataset for the third filter. Otherwise, we excluded the sentence from further evaluation, as it was more likely to be related to the drug's known usage rather than serendipitous usage we were looking for. In the third step, we used the sentiment analysis tool, Deeply Moving (Socher et al., 2013) offered by the Stanford Natural Language Processing Group to assess the sentiment of each sentence where unanticipated medical concept occurred. We filtered out all sentences with *Very Negative*, *Negative*, or *Neutral* sentiment and only kept those with *Positive* or *Very Positive* sentiments because serendipitous drug usages are unexpected but desirable outcomes to patients. Negative sentiment is more likely to be associated with undesirable side effects or potential drug safety concerns. After these three filtering steps, 15,714 drug review sentences remained for further human annotation.

## 2.4 Human Annotation

One public health professional and one health informatics professional with master degrees, independently reviewed the 15,714 sentences and annotated whether each sentence was a true mention of serendipitous drug usage based on the benchmark


Figure 2: Examples of serendipitous drug usage mention on WebMD. In the example on the left, a patient reported that his irritable bowel syndrome (IBS) symptoms were alleviated when taking sulfasalazine to treat rheumatoid arthritis. In the example on the right, an asthma patient taking prednisone reported the improvement of her eczema.

dataset of known drug usages defined by WebMD. That is, they labeled a drug use to be serendipitous if the user mentioned an improved condition or symptom that was not listed in the drug's known indications according to WebMD. Otherwise, they assigned the mentioned drug use to be non-serendipitous. In case that the annotators did not agree with each other, they discussed and assigned a final label together. Six months later, the two professionals reviewed their annotation again to avoid possible human errors. In total, 447 or 2.8% of sentences were annotated to contain true serendipitous drug usage mentions, covering 97 drugs and 183 serendipitous drug usages. The rest 15,267 sentences were annotated to contain no serendipitous drug usage mentions. This dataset was used throughout the study as the gold standard dataset to train and evaluate various machine learning models.

## 2.5 Feature Construction and Selection

Feature construction and selection is an important part of data mining analysis, in which the data is processed and presented in a way understandable by machine learning algorithms. The original drug reviews downloaded from WebMD website come with 11 features, including patients' ratings of drug effectiveness, ease of use, overall satisfaction, and the number of people who thought the review is helpful (See Table 1).

In the data-filtering step, we created four more features, which are (1) whether the sentence contains negation, (2) the UMLS semantic types of mentioned medical concepts; (3) the SNOMED CT-based semantic distance between a drug's known indication and the medical concept the user mentioned in a review sentence; (4) the sentiment score of the review sentence.

Prior knowledge in drug discovery and development also tells that some therapeutic areas, such as neurological disorders, bacteria infection, and cancers are more likely to have "dirty" drugs, which bind to many different molecular targets in human body, and tend to have a wide range of effects (Yao and Rzhetsky, 2008, Frantz, 2005, Pleyer and Greil, 2015). Therefore, drugs used in those therapeutic areas have higher chance to be repositioned. We manually selected 155 drug usages from those therapeutic areas and used them as binary features, which hopefully capture useful information and improve machine learning predictions of serendipitous drug usages.

We also adopted a commonly used text-mining

Table 1: List of the features constructed for the annotated datasets.

| Name | Data Type | Source |
|---|---|---|
| **Original Features obtained from the Patient Forum** | | |
| User rating of effectiveness | Numerical | WebMD |
| User rating of ease of use | Numerical | WebMD |
| User rating of overall satisfaction | Numerical | WebMD |
| Number of users who felt the review was helpful | Numerical | WebMD |
| Number of reviews for the drug | Numerical | WebMD |
| The day of review | Categorical | WebMD |
| The hour of review | Categorical | WebMD |
| User's role (e.g., Patient, Caregiver) | Categorical | WebMD |
| User's gender | Categorical | WebMD |
| User's age group | Categorical | WebMD |
| The time on the drug (e.g., less than 1 month, 1 to 6 months, 6 months to 1 year) | Categorical | WebMD |
| **Additional Features** | | |
| Whether the sentence contains negation | Binary | MetaMap |
| Semantic types of medical concepts mentioned in the sentence | Categorical | MetaMap |
| Semantic distance between the mentioned medical concept and the drug's known indications in SNOMED CT | Numerical | SNOMED |
| Sentiment score | Numerical | Deeply Moving |
| Therapeutic areas (155) | Binary | Self-constructed |
| N-grams extracted from drug review sentences (177) | Binary | Self-constructed |

method, n-gram (Fürnkranz, 1998), to generate more textual features. An *n*-gram is a contiguous sequence of *n* words from a given text and it captures the pattern about how people use word combination in their communication. We used the *tm* package in R (Feinerer and Hornik, 2012) to do this. After the steps of punctuation and stop words removal, word stemming, and rare words pruning, we extracted 3,264 unigrams, 10,064 bigrams, and 5,058 trigrams. For each *n*-gram, we calculated the information gain (Michalski et al., 2013) to assess its differentiating power between true and false classes in Weka (Hall et al., 2009). We excluded n-grams whose information gain equaled zero and kept 177 n-grams with positive information gain (namely 64 unigrams, 73 bigrams, and 40 trigrams) as additional textual features. In total, 347 features were constructed for the machine learning classification, as summarized in Table 1.

## 2.6 Data Preprocessing

We normalized the data by linearly re-scaling all numerical features to the range of [-1, 1]. Such processing is necessary for support vector machine (SVM) to ensure no features dominate the classification just because of their order of magnitude, as SVM calculates the Euclidean distances between support vectors and the separation hyperplane in high-dimensional space (Ali and Smith-Miles, 2006). Then we split the 15,714 annotated sentences into training, validation, and test datasets, according to their post dates. Sixty percent of them, or 9,429 sentences posted between September 18, 2007 and December 07, 2010, were used as the training dataset to build machine learning models. Twenty percent of the data, or 3,142 sentences posted between December 08, 2010 and October 11, 2012 were used as the validation dataset to tune the model parameters. The remaining 20% of data, or 3,143 sentences that were posted between October 12, 2012 and March 26, 2015, were held as the independent test dataset. The proportion of serendipitous drug usages in the three datasets was between 2.0% and 3.2%. This arrangement is essential to pick up the models that could generalize on future and unseen data and minimize the bias led by overfitting, as the validation and test datasets occur temporally after the training dataset.

## 2.7 Machine Learning Models

We selected three state-of-art machine learning algorithms, namely SVM (Cortes and Vapnik, 1995), random forest (Breiman, 2001) and AdaBoost.M1 (Freund and Schapire, 1996) to build the prediction models. The implementation was based on Weka (version 3.7) (Hall et al., 2009) and LibSVM library (Chang and Lin, 2011). For SVM, we used the radial basis function (RBF) kernel and conducted grid search to find the optimal parameters including $C$ and gamma ($\gamma$). LibSVM is able to produce both probability estimates (Wu et al., 2004) and class labels as output. For random forest, we empirically set the number of trees to be 500 and iteratively searched for the optimal value for number of features. By default the prediction gives a probability estimate for each class. For AdaBoost.M1, we selected the decision tree built by C4.5 algorithm (Quinlan, 2014) as the weak learner and obtained the optimal value for number of iterations through iterative search. The Weka implementation of AdaBoost.M1 only provides class labels as prediction results. Our evaluation therefore is based on class label predictions from all three algorithms, without considering the probability estimates from SVM and random forest.

As the chance of finding a serendipitous drug usage (positive class) is rare and the vast majority of the drug reviews posted by users do not mention any serendipitous usages (negative class), we were facing an imbalanced dataset problem. Therefore, we used the oversampling technique (He and Garcia, 2009, Batuwita and Palade, 2010, Kotsiantis et al., 2006) to generate another training dataset where the proportion of positive class was increased from 2.8% to 20%. Afterward, we tried the same machine learning algorithms on the oversampled training dataset, and compared the prediction results side-by-side with those from the original, imbalanced training dataset.

## 2.8 Evaluation

We were cautious about choosing appropriate performance evaluation metrics because of the imbalanced dataset problem. Of commonly used metrics, accuracy is most vulnerable to imbalanced dataset since a model could achieve high accuracy simply by assigning all instances into the majority class. Instead we used a combination of three commonly used metrics, namely precision, recall, and area under the receiver operating characteristic curve (also known as AUC score) (Caruana and Niculescu-Mizil, 2004), to evaluate the performance of various prediction models on the independent test dataset. We also conducted 10-fold cross validation

by combining training, validation and testing datasets together, in order to compare our results directly with some other drug-repositioning studies.

In addition, we manually reviewed 10% of instances in the test dataset that were predicted to be serendipitous drug usages and searched through the scientific literature to check if these predictions based purely on machine learning methods can replicate the discoveries from biomedical scientific community, as another verification on whether machine learning methods alone can potentially predict completely new serendipitous drug usages.

All our data and scripts from this work will be made available to academic users upon request.

# 3 RESULTS

## 3.1 Parameter Tuning

We used AUC score to tune the model parameters on the validation dataset. In case that the AUC scores of two models were really close, we chose the parameter and model that yielded higher precision. This is because end users (e.g., pharmaceutical scientist) are more sensitive to cases that were predicted to be the under-presented, rare events, which are serendipitous drug usages in this work, when they evaluate the performance of any kind of machine learning based predictive models. For SVM models, the optimal value of gamma ( $\gamma$ ), the width of RBF kernel was 0.001 without oversampling and 0.1 with oversampling. The optimal value of *C*, which controls the trade-off between model complexity and ratio of misclassified instances, was equal to 380 without oversampling and 0.1 with oversampling. For random forest models, the number of features decides the maximum number of features used by each decision tree in the forest, which was found to be 243 without oversampling and 84 with oversampling at the best performance on validation dataset. For AdaBoost.M1, the number of iterations specifies how many times the weak learner will be trained to minimize the training error. Its optimal value equaled 36 without oversampling and 58 with oversampling.

## 3.2 Performance Metrics

We evaluated the performance of six prediction models, namely SVM, random forest and AdaBoost.M1 with and without oversampling, on independent test dataset. The results were summarized in Table 2. The highest AUC score (0.937) was achieved from the AdaBoost.M1 model, whereas the lowest score (0.893) was from the SVM with oversampling. On the whole, AUC scores for all models were higher than 0.89, demonstrating the promise of machine learning models for identifying serendipitous drug usages from patient forums.

The precision of random forest and AdaBoost.M1 models with and without oversampling, and the SVM model without oversampling were between 0.758 and 0.857, with the highest precision achieved on the random forest model without oversampling. However, the precision for the SVM model with oversampling was 0.474, which was significantly lower than the other models. The recall of all models was less than 0.50. This means more than 50% of serendipitous usages were not identified. Obtaining either low recall or low precision remains a common challenge for making predictions from extremely imbalanced datasets like ours (He and Garcia, 2009). In many cases, it becomes a compromise depending on the application and the users' need. In our experiment, after we increased the proportion of the positive class to 20% by oversampling, the recall of SVM and random forest models increased slightly; but the precision and the AUC score decreased. Oversampling seemed ineffective on AdaBoost.M1 models. The AUC score, precision and recall for AdaBoost.M1 with oversampling all decreased,

Table 2: Model performance in terms of precision, recall and AUC score.

| Model | Test dataset | | | 10-fold cross validation | | |
|---|---|---|---|---|---|---|
| | AUC | Precision | Recall | AUC | Precision | Recall |
| SVM | 0.900 | 0.758 | 0.397 | 0.926 | 0.817 | 0.539 |
| SVM - Oversampling | 0.893 | 0.474 | 0.429 | 0.932 | 0.470 | 0.620 |
| Random Forest | 0.926 | 0.857 | 0.381 | 0.935 | 0.840 | 0.506 |
| Random Forest - Oversampling | 0.915 | 0.781 | 0.397 | 0.944 | 0.866 | 0.530 |
| AdaBoost.M1 | 0.937 | 0.811 | 0.476 | 0.949 | 0.791 | 0.575 |
| AdaBoost.M1 - Oversampling | 0.934 | 0.800 | 0.444 | 0.950 | 0.769 | 0.559 |

compared to the metrics on AdaBoost.M1 models without oversampling. In the 10-fold cross validation experiment, both recall and AUC scores seemed to be better than what were observed on the independent test set. Our AUC scores were close to the same scores reported by the drug-repositioning algorithm of PREDICT (Gottlieb et al., 2011), which were also from a 10-fold cross validation.

## 3.3 Prediction Review

For the 10% of instances in the test dataset that were predicted to be serendipitous drug usages, we conducted a literature and clinical trial search to provide a closer verification of our prediction models. Table 3 summarizes the analysis. We also presented the condensed evidences in literature and/or clinical trial below, for each instance.

### 3.3.1 Metformin and Obesity

A patient reported weight loss while taking metformin, a type 2 diabetes drug. Actually in the past two decades, metformin's effectiveness and safety for treating obesity in adult and child patients have been clinically examined in dozens of clinical trials and meta-analyses studies with promising results (Igel et al., 2016, Desilets et al., 2008, Paolisso et al., 1998, Peirson et al., 2014, McDonagh et al., 2014). According to the literature review by Igel et al. (2016), one possible explanation is that metformin could increase the body's insulin sensitivity, which helps obese patients (who typically develop resistance to insulin) to reduce their craving for carbohydrates and to reduce the glucose stored in their adipose tissue. Other explanations include that metformin may enhance energy metabolism by accelerating the phosphorylation of the AMP-activated protein kinase system, or it may cause appetite loss by correcting the sensitivity and resistance of leptin.

### 3.3.2 Painkiller and Depression

When tramadol was taken for back pain, a patient found it also helpful with his depression and anxiety. Tramadol is an opioid medication, which have been long used for the psychotherapeutic benefits (Tenore, 2008). Tetsunaga et al. (2015) have demonstrated tramadol's efficacy in reducing depression levels among lower back pain patients with depression in an 8-week clinical trial. The self-reported depression scale of patients in the tramadol group was 6.5 points lower than the control group. Similarly the combinatory therapy of acetaminophen and oxycodone, another painkiller, was reported by Stoll and Rueter (1999) to have antidepressant effect too.

### 3.3.3 Bupropion and Obesity

In the specific comment, the patient reported that Bupropion, an anti-depressant, helped him to lose weight. The weight loss effect of bupropion might be attributed to increased dopamine concentration in the brain, which leads to suppressed appetite and reduced food intake (Greenway et al., 2010). This serendipitous drug usage was also supported by several clinical trials (Gadde et al., 2001, Anderson et al., 2002, Jain et al., 2002).

### 3.3.4 Ondansetron and Irritable Bowel Syndrome with Diarrhea

Ondansetron is a medication for nausea and vomiting. Sometimes it causes the side effect of constipation in patients. Interestingly, this patient also had irritable bowel syndrome with diarrhea and thus ondansetron helped to regulate that. This serendipitous usage actually highlights the justification of personalized medicine and has been tested in a recent clinical trial (Garsed et al., 2014).

### 3.3.5 Desvenlafaxin and Lack of Energy

In the last case, anti-depressant desvenlafaxine was reported to boost energy. Strictly speaking, lack of energy is not a disease but a symptom. With limited information on the patient's physical and psychological conditions before and after medication, it remains unclear whether the energy boost effect was due to changes in the neural system or was purely a natural reflection of more positive moods after the patient took the anti-depressant medicine. We did not find any scientific literature discussing the energy boost effect of desvenlafaxine. So this case could represent either a new serendipitous drug use or a promiscuous drug usage.

Table 3: Examples of serendipitous drug usages predicted by the models.

| **True positive examples** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Drug** | **Known indications** | **Serendipitous usage** | **Example** | **SVM** | **SVM-Oversampling** | **RF*** | **RF-Oversampling*** | **Ada*** | **Ada-Oversampling*** | **Literature evidence** |
| Metformin | Type 2 Diabetes Mellitus, Polycystic Ovary Syndrome, etc. | Obesity | I feel AWFUL most of the day, but the ***weight loss*** is great. | x | x | x | x | x | x | Igel et al. (2016), Desilets et al. (2008), Paolisso et al. (1998) |
| Tramadol | Pain | Depression, anxiety | It also has helped with my ***depression*** and ***anxiety***. | x | x | | | x | x | Tetsunaga et al. (2015) |
| Acetaminophen & oxycodone | Pain | Depression | While taking for pain I have also found it relieves my major ***depression*** and actually gives me the energy and a clear mind to do things. | x | x | x | | x | | Stoll and Rueter (1999) |
| Bupropion | Depression, attention deficit & hyperactivity disorder | Obesity | I had energy and experienced needed ***weight loss*** and was very pleased, as I did not do well on SSRI or SNRIs. | x | x | | x | x | x | Greenway et al. (2010), Gadde et al. (2001), Anderson et al. (2002), Jain et al. (2002) |
| Ondansetron | Vomiting | Irritable bowel syndrome with diarrhea | A lot of people have trouble with the constipation that comes with it, but since I have ***IBS-D*** (irritable bowel syndrome with diarrhea), it has actually regulated me . | | | | | x | x | Garsed et al. (2014) |
| Desvenlafaxine | Depression | Lack of energy | I have had a very positive mood and ***energy*** change, while also experiencing much less anxiety. | x | x | x | x | x | | |
| **False positive examples** | | | | | | | | | | |
| 5-HTP | Anxiety, depression | Thyroid Diseases, Obesity | i have ***Hoshimitos thyroid disease***** and keeping stress levels down is extremely important for many reasons but also for ***weight loss***. | | x | | x | | | |
| Cyclobenzaprine | Muscle spasm | Pain | While taking this medication for neck stiffness and ***pain***; I discovered it also helped with other muscle spasms. | | x | | | | | |

*RF stands for random forest. Ada stands for AdaBoost.M1. "x" indicates the model recognized the example as a serendipitous usage. **Hoshimitos thyroid disease was a typo. The correct spelling should be Hashimoto's Thyroiditis.

### 3.3.6 False Positive Predictions

Besides the true positive examples, we also found two cases where some of our models made false positive predictions due to difficult language expression and terminology flaw. The first example is 5-HTP, an over-the-counter drug for anxiety and depression. One patient commented that stress relief brought by this drug was important to her Hashimito's thyroid disease and weight loss. Although Hashimoto's disease and weight loss were mentioned, the patient did not imply the 5-HTP can treat Hashimoto's disease or control weight. But SVM and random forest models with over-sampling became confused by the subtle semantic difference. In the second case, a patient taking cyclobenzaprine for neck stiffness and pain said the drug also helped with other muscle spasms. Pain, neck stiffness and muscle spasms are really close medical concepts. We found that this false positive prediction was actually due to imperfect terminology mapping.

## 4 DISCUSSION

In this very first effort to identify serendipitous drug usages from online patient forum, we designed an entire computational pipeline. This feasibility study enabled us to thoroughly examine the technical hurdles in the entire workflow and answer the question if patient-reported medication outcome data on social media is worthwhile to explore for drug repositioning research. The best-performing model was built from AdaBoost.M1 method without oversampling, which had precision equal to 0.811, recall equal to 0.476 and AUC score equal to 0.937 on independent test data. The 10-fold cross validation results are also comparable to existing drug-repositioning method (Gottlieb et al., 2011). Therefore our confidence in applying machine learning methods to identify serendipitous drug usages from online patient forum data is increased. More specifically we have addressed the following tasks in this work:

Previously, there was no curated social media dataset available for the purpose of identifying serendipitous drug usages. We spent a considerable amount of time and effort to collect, filter and annotate 15,714 drug review sentences from the WebMD patient forum site. Two health professionals at master level annotated all the sentences independently and discussed on cases when disagreement occurred. They repeated this process six months later. If more resource available, we would like to recruit a larger group of professionals to curate a larger and more reliable gold standard dataset. But the current annotated dataset is comprehensive enough for this work, as it covers not only easy instances, but also challenging ones for machine learning prediction, as shown in Table 3.

In addition, the drug reviews posted on patient forum are unstructured and informal human language prevalent with typographic errors and chat slangs, which need to be transformed to a representation of feature vectors before machine learning algorithms could comprehend. We used patients' demographic information, ratings of drug effectiveness, ease of use, and overall satisfaction from the patient forum. We calculated negation, semantic similarity between the unexpected medication outcome mentioned in a review sentence and the known drug indications based on SNOMED CT, and sentiment score of the review sentence. We also leveraged our known knowledge on dirty drugs, and extracted informative n-gram features based on information gain. The results from this feasibility study showed that these features are useful to predict serendipitous drug usages. For example, dirty drugs for neurological conditions did show up predominantly in the results. But these features seemed not sufficient to predict all serendipitous drug usages correctly. As shown in the false positive examples of Table 3, the n-grams such as *also*, *also help*, and *also for* were often associated with true serendipitous drug usages, but could occur in false positive cases too. Current medical terminology mapping tools (i.e., MetaMap) could be the performance-limiting step in cases like *pain* and *muscle spasm*, despite the close connection of these two concepts from the perspective of medicine. We will explore more sophisticated methods such as DNorm (Leaman et al., 2013), as well as additional methods of semantic similarity calculation as shown in (Pedersen et al., 2007, Sánchez et al., 2012) in future.

Thirdly, the data are extremely imbalanced between two classes (2.8% vs. 97.2%) because serendipitous drug usages are rare events by nature. Such imbalance inevitably impedes the performance of machine learning algorithms. We tried to increase the proportion of serendipitous usages in the training dataset to 20%, using the random oversampling method (He and Garcia, 2009). We have also tried two other methods, namely synthetic minority over-sampling technique (Chawla et al., 2002) and under-sampling (Kotsiantis et al., 2006), but their performance was inferior to that of random

oversampling and not shown here. More robust machine learning algorithms that are less sensitive to imbalanced data or robust sampling methods will be desirable to further improve serendipitous drug usage predictions.

Last but not least, we acknowledge that as an emerging data source, online patient forums have limitations too. Many patients who write drug reviews online lack of basic medical knowledge. Their description of the medication experience can be ambiguous, hyperbolic or inaccurate. Also important contextual information, such as co-prescribed drugs, may be missed in the review. Without a comparison between an experiment group and a control group, serendipitous drug usages extracted from patient forums need to be further verified for drug repositioning opportunities by integrating with existing data sources, such as EHR and scientific literature.

## 5 CONCLUSIONS

Drug repositioning is an important but not yet fully utilized strategy to improve the cost-effectiveness of medicine and to reduce the development time. The dawn of social media brings large volumes of patient-reported medication outcome data, and thus creates an urgent need to examine it for the purpose of drug repositioning. In this work, we collected, filtered, and annotated drug review comments posted on WebMD patient forum. We built an entire computational pipeline based state-of-art machine learning and text mining methods to mine serendipitous drug usages. Our models achieved AUC scores that are comparable to existing drug repositioning methods. Most instances that were predicted to be serendipitous drug usages are also supported by scientific literature. So machine learning approaches seem feasible to address this problem of looking for a needle in the haystack. More of our future efforts will be directed to develop more informative features, improve disease mapping accuracy, handle imbalanced data, and integrate findings from social media with other data sources, in order to build really functional drug-repositioning applications.

## REFERENCES


Ali, S. & Smith-Miles, K. A. Improved support vector machine generalization using normalized input space. *In: Proceedings of the 19th Australasian Joint Conference on Artificial Intelligence*, 2006 Hobart, Australia. Springer, 362-371.

Anderson, J. W., Greenway, F. L., Fujioka, K., Gadde, K. M., Mckenney, J. & O'neil, P. M. 2002. Bupropion SR Enhances Weight Loss: A 48-Week Double-Blind, Placebo-Controlled Trial. *Obesity Research,* 10**,** 633-641.

Andronis, C., Sharma, A., Virvilis, V., Deftereos, S. & Persidis, A. 2011. Literature mining, ontologies and information visualization for drug repurposing. *Briefings in Bioinformatics,* 12**,** 357-368.

Aronson, A. R. & Lang, F.-M. 2010. An overview of MetaMap: historical perspective and recent advances. *Journal of the American Medical Informatics Association,* 17**,** 229-236.

Ashburn, T. T. & Thor, K. B. 2004. Drug repositioning: identifying and developing new uses for existing drugs. *Nature Review Drug Discovery,* 3**,** 673-683.

Batuwita, R. & Palade, V. Efficient resampling methods for training support vector machines with imbalanced datasets. *In: Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 2010 Barcelona, Spain. IEEE, 1-8.

Breiman, L. 2001. Random forests. *Machine Learning,* 45**,** 5-32.

Caruana, R. & Niculescu-Mizil, A. Data mining in metric space: an empirical analysis of supervised learning performance criteria. *In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004 Seattle, WA, USA. ACM, 69-78.

Chang, C. & Lin, C. 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2**,** 27.

Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research,* 16**,** 321-357.

Cortes, C. & Vapnik, V. 1995. Support-vector networks. *Machine Learning,* 20**,** 273-297.

Desilets, A. R., Dhakal-Karki, S. & Dunican, K. C. 2008. Role of metformin for weight management in patients without type 2 diabetes. *Annals of Pharmacotherapy,* 42**,** 817-826.

Dudley, J. T., Deshpande, T. & Butte, A. J. 2011. Exploiting drug–disease relationships for computational drug repositioning. *Briefings in Bioinformatics,* 12**,** 303-311.

Feinerer, I. & Hornik, K. 2012. tm: text mining package. *R package version 0.5-7.1*.

Frantz, S. 2005. Drug discovery: playing dirty. *Nature,* 437**,** 942-943.

Freund, Y. & Schapire, R. E. Experiments with a new boosting algorithm. *In: Proceedings of the 13th International Conference on Machine Learning*, 1996 Bari, Italy. 148-156.

Fürnkranz, J. 1998. A study using n-gram features for text categorization. *Austrian Research Institute for Artifical Intelligence,* 3**,** 1-10.

Gadde, K. M., Parker, C. B., Maner, L. G., Wagner, H. R., Logue, E. J., Drezner, M. K. & Krishnan, K. R. R. 2001. Bupropion for weight loss: an investigation of efficacy and tolerability in overweight and obese women. *Obesity Research,* 9**,** 544-551.

Garsed, K., Chernova, J., Hastings, M., Lam, C., Marciani, L., Singh, G., Henry, A., Hall, I., Whorwell, P. & Spiller, R. 2014. A randomised trial of ondansetron for the treatment of irritable bowel syndrome with diarrhoea. *Gut,* 63**,** 1617-1625.

Gottlieb, A., Stein, G. Y., Ruppin, E. & Sharan, R. 2011. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Molecular Systems Biology,* 7**,** 496.

Greenway, F. L., Fujioka, K., Plodkowski, R. A., Mudaliar, S., Guttadauria, M., Erickson, J., Kim, D. D., Dunayevich, E. & Group, C.-I. S. 2010. Effect of naltrexone plus bupropion on weight loss in overweight and obese adults (COR-I): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet,* 376**,** 595-605.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H. 2009. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter,* 11**,** 10-18.

He, H. & Garcia, E. A. 2009. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering,* 21**,** 1263-1284.

Hu, G. & Agarwal, P. 2009. Human disease-drug network based on genomic expression profiles. *PLoS ONE,* 4**,** e6536.

Igel, L. I., Sinha, A., Saunders, K. H., Apovian, C. M., Vojta, D. & Aronne, L. J. 2016. Metformin: an old therapy that deserves a new indication for the treatment of obesity. *Current Atherosclerosis Reports,* 18**,** 1-8.

Jain, A. K., Kaplan, R. A., Gadde, K. M., Wadden, T. A., Allison, D. B., Brewer, E. R., Leadbetter, R. A., Richard, N., Haight, B. & Jamerson, B. D. 2002. Bupropion SR vs. placebo for weight loss in obese patients with depressive symptoms. *Obesity Research,* 10**,** 1049-1056.

Jensen, P. B., Jensen, L. J. & Brunak, S. 2012. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics,* 13**,** 395-405.

Keiser, M. J., Setola, V., Irwin, J. J., Laggner, C., Abbas, A., Hufeisen, S. J., Jensen, N. H., Kuijer, M. B., Matos, R. C., Tran, T. B., Whaley, R., Glennon, R. A., Hert, J., Thomas, K. L. H., Edwards, D. D., Shoichet, B. K. & Roth, B. L. 2009. Predicting new molecular targets for known drugs. *Nature,* 462**,** 175-181.

Khatri, P., Roedder, S., Kimura, N., De Vusser, K., Morgan, A. A., Gong, Y., Fischbein, M. P., Robbins, R. C., Naesens, M., Butte, A. J. & Sarwal, M. M. 2013. A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation. *The Journal of Experimental Medicine,* 210**,** 2205-2221.

Knezevic, M. Z., Bivolarevic, I. C., Peric, T. S. & Jankovic, S. M. 2011. Using Facebook to increase spontaneous reporting of adverse drug reactions. *Drug Safety,* 34**,** 351-352.

Kotsiantis, S., Kanellopoulos, D. & Pintelas, P. 2006. Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering,* 30**,** 25-36.

Leaman, R., Doğan, R. I. & Lu, Z. 2013. DNorm: disease name normalization with pairwise learning to rank. *Bioinformatics,* 29**,** 2909-2917.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S. & Mcclosky, D. The Stanford CoreNLP natural language processing toolkit. *In: The 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2014 Baltimore, MD, USA. 55-60.

Mcdonagh, M. S., Selph, S., Ozpinar, A. & Foley, C. 2014. Systematic review of the benefits and risks of metformin in treating obesity in children aged 18 years and younger. *JAMA Pediatrics,* 168**,** 178-184.

Michalski, R. S., Carbonell, J. G. & Mitchell, T. M. 2013. *Machine learning: An artificial intelligence approach*, Springer Science & Business Media.

Paolisso, G., Amato, L., Eccellente, R., Gambardella, A., Tagliamonte, M. R., Varricchio, G., Carella, C., Giugliano, D. & D'onofrio, F. 1998. Effect of metformin on food intake in obese subjects. *European Journal of Clinical Investigation,* 28**,** 441-446.

Pedersen, T., Pakhomov, S. V. S., Patwardhan, S. & Chute, C. G. 2007. Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, 40**,** 288-299.

Peirson, L., Douketis, J., Ciliska, D., Fitzpatrick-Lewis, D., Ali, M. U. & Raina, P. 2014. Treatment for overweight and obesity in adult populations: a systematic review and meta-analysis. *CMAJ Open,* 2**,** E306-E317.

Pleyer, L. & Greil, R. 2015. Digging deep into “dirty” drugs–modulation of the methylation machinery. *Drug Metabolism Reviews,* 47**,** 252-279.

Powell, G. E., Seifert, H. A., Reblin, T., Burstein, P. J., Blowers, J., Menius, J. A., Painter, J. L., Thomas, M., Pierce, C. E., Rodriguez, H. W., Brownstein, J. S., Freifeld, C. C., Bell, H. G. & Dasgupta, N. 2016. Social media listening for routine post-marketing safety surveillance. *Drug Safety,* 39**,** 443-454.

Quinlan, J. R. 2014. *C4.5: programs for machine learning*, Elsevier.

Ru, B., Harris, K. & Yao, L. A Content Analysis of Patient-Reported Medication Outcomes on Social Media. *In: Proceedings of IEEE 15th International Conference on Data Mining Workshops*, 2015 Atlantic City, NJ, USA. IEEE, 472-479.

Sánchez, D., Batet, M., Isern, D. & Valls, A. 2012. Ontology-based semantic similarity: A new feature-based approach. *Expert Systems with Applications,* 39**,** 7718-7728.

Sanseau, P., Agarwal, P., Barnes, M. R., Pastinen, T., Richards, J. B., Cardon, L. R. & Mooser, V. 2012. Use

of genome-wide association studies for drug repositioning. *Nature Biotechnology,* 30**,** 317-320.

Shah, N. H. & Musen, M. A. UMLS-Query: a perl module for querying the UMLS. *In: AMIA Annual Symposium,* 2008 Washington, DC, USA. 652-656.

Shandrow, K. L. 2016. *The Hard Truth: What Viagra Was Really Intended For* [Online]. Entrepreneur.com. Available: http://www.entrepreneur.com/article/254908 [Accessed 02/22/2016].

Shim, J. S. & Liu, J. O. 2014. Recent advances in drug repositioning for the discovery of new anticancer drugs. *International Journal of Biological Sciences,* 10**,** 654-63.

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y. & Potts, C. Recursive deep models for semantic compositionality over a sentiment treebank. *In: Proceedings of the Conference on Empirical Methods in Natural Language Processing,* 2013 Seattle, WA, USA. Citeseer, 1631-1642.

Stoll, A. L. & Rueter, S. 1999. Treatment augmentation with opiates in severe and refractory major depression. *American Journal of Psychiatry,* 156**,** 2017.

Tenore, P. L. 2008. Psychotherapeutic benefits of opioid agonist therapy. *Journal of Addictive Diseases,* 27**,** 49-65.

Tetsunaga, T., Tetsunaga, T., Tanaka, M. & Ozaki, T. 2015. Efficacy of tramadol–acetaminophen tablets in low back pain patients with depression. *Journal of Orthopaedic Science,* 20**,** 281-286.

U.S. National Library of Medicine. 2016a. *MEDLINE Fact Sheet* [Online]. Available: https://www.nlm.nih.gov/pubs/factsheets/medline.html [Accessed 09/29/2016].

U.S. National Library of Medicine. 2016b. *SNOMED CT* [Online]. Available: http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html [Accessed 08/03/2015].

Wren, J. D., Bekeredjian, R., Stewart, J. A., Shohet, R. V. & Garner, H. R. 2004. Knowledge discovery by automated identification and ranking of implicit relationships. *Bioinformatics,* 20**,** 389-398.

Wu, T.-F., Lin, C.-J. & Weng, R. C. 2004. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research,* 5**,** 975-1005.

Xu, H., Aldrich, M. C., Chen, Q., Liu, H., Peterson, N. B., Dai, Q., Levy, M., Shah, A., Han, X., Ruan, X., Jiang, M., Li, Y., Julien, J. S., Warner, J., Friedman, C., Roden, D. M. & Denny, J. C. 2014. Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. *Journal of the American Medical Informatics Association,* 22**,** 179-191.

Yang, C. C., Yang, H., Jiang, L. & Zhang, M. Social media mining for drug safety signal detection. *In: Proceedings of the 2012 International Workshop on Smart Health and Wellbeing,* 2012 Maui, HI, USA. ACM, 33-40.

Yao, L. & Rzhetsky, A. 2008. Quantitative systems-level determinants of human genes targeted by successful drugs. *Genome Research,* 18**,** 206-213.

Yao, L., Zhang, Y., Li, Y., Sanseau, P. & Agarwal, P. 2011. Electronic health records: Implications for drug discovery. *Drug Discovery Today,* 16**,** 594-599.

Yates, A. & Goharian, N. ADRTrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites. *In: The 35th European Conference on Information Retrieval,* 2013 Moscow, Russia. Springer-Verlag, 816-819.

# Data-driven Web-based Intelligent Decision Support System for Infection Management at Point-Of-Care: Case-Based Reasoning Benefits and Limitations

Bernard Hernandez[1], Pau Herrero[1], Timothy M. Rawson[2], Luke S. P. Moore[2], Esmita Charani[2], Alison H. Holmes[2] and Pantelis Georgiou[1]

[1]*Centre for Bio-Inspired Technology, Imperial College London, U.K.*

[2]*Health Protection Unit in Healthcare Associated Infections and AMR, Imperial College London, U.K.*

*b.hernandez-perez@imperial.ac.uk*

Keywords: Antimicrobial Resistance, Infection Diseases, Antibiotics, Decision Support System, Case-Based Reasoning, Machine Learning, User Interface, Point of Care.

Abstract: Antimicrobial Resistance (AMR) is a major patient safety issue. Attempts have been made to palliate its growth. Misuse of antibiotics to treat human infections is a main concern and therefore prescription behaviour needs to be studied and modified appropriately. A common approach relies on designing software tools to improve data visualization, promote knowledge transfer and provide decision-making support. This paper explains the design of a Decision Support System (DSS) for clinical environments to provide personalized, accurate and effective diagnostics at point-of-care (POC), improving continuity, interpersonal communication, education and knowledge transfer. Demographics, biochemical and susceptibility laboratory tests and individualized diagnostic/therapeutic advice are presented to clinicians in a handheld device. Case-Based Reasoning (CBR) is used as main reasoning engine to decision support for infection management at POC. A web-based CBR-inspired interface design focused on usability principles has also been developed. The proposed DSS is perceived as useful for patient monitoring and outcome review at POC by expert clinicians. The DSS was rated with a System Usability Scale (SUS) score of 68.5 which indicates good usability. Furthermore, three areas of improvement were identified from the feedback provided by clinicians: thorough guidance requirements for junior clinicians, reduction in time consumption and integration with prescription workflow.

## 1 INTRODUCTION

Antimicrobials are drugs that kill or stop the growth of microbes (e.g. bacteria or viruses), thereby are commonly used to treat infections. Recently, Antimicrobial Resistance (AMR) has been reported to be a leading public health and safety problem (Wise et al., 1998; ONeill, 2014) with the inappropriate use of antibiotics in humans identified as a leading driver (Holmes et al., 2016). Microbes are continuously evolving and unnecessary antibiotic prescription, particularly within infection diseases, are a common concern in critical care and infection management, which are observing and suffering the consequences of an increased rate of AMR. In addition, failure to recognize and respond to the early stage infections is considered a major cause of avoidable mortality. Thus, it is needed to develop guidelines and software tools that facilitate healthcare professionals to treat their patients at the patient bedside by collecting and visualizing laboratory test results while providing a support system to assist in decision-making.

Antibiotic resistance is most likely to develop in areas with a considerable concentration of sick patients and high risk of infection where antimicrobials are used extensively. Henceforth, the Intensive Care Unit (ICU), where proportion of inappropriate antibiotic prescription ranges from 41% to 66%, is targeted in our preliminary studies. Handheld Decision Support Systems (DSSs) including local antibiotic guidelines have proved to reduce antibiotic prescribing in the ICU (Sintchenko et al., 2005). Despite their benefits, factors as hardware availability or interface design (Tsopra et al., 2014) obstruct the acceptance of DSSs in clinical environments. To promote their use, it is necessary to determine the best way to present the information (Moxey et al., 2010). The Healthcare Information and Management Systems Society (HIMMS) stressed the benefits of designs based on usability principles (Belden et al., 2009). In addi-



Hernandez, B., Herrero, P., Rawson, T., Moore, L., Charani, E., Holmes, A. and Georgiou, P.
Data-driven Web-based Intelligent Decision Support System for Infection Management at Point-Of-Care: Case-Based Reasoning Benefits and Limitations.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 119-127
ISBN: 978-989-758-213-4

tion, a well designed DSS sustains the advanced algorithms implemented in its core with a fully comprehensible representation to support confidence in doctors. Since medical knowledge is voluminous, it has to be focused on data and decision making while providing access to electronic health records (HER) and personal health information (PHI). Many studies show that young clinicians engage better with the use of mobile applications, displaying great potential to improve learning and augmenting traditional training (Boulos et al., 2014).

In this paper we postulated that an appropriately designed clinical information technology system could improve reliability and consistency of collecting vital signs; their visualization, interpretation and analysis; and the delivery of a more sophisticated DSS. Therefore, health care professionals and biomedical engineers from Imperial College of London have designed a prototype system accessible at the point-of-care (POC) with the specific objectives of improving three main areas: personalization and therefore outcomes of infection management; continuity through POC support for interpersonal communication; and education during interactions between clinicians and infection specialists.

## 2 BACKGROUND

Critical care, infection management and antimicrobial stewardship is predominantly multidisciplinary with involvement of infection specialists being crucial. In practice, antimicrobial prescribing frequently occurs out of hours, and when advice is dispensed by infection specialists, uptake can be variable. Current patient management systems rarely integrate DSSs to assist with this, or if they do, this is very basic. Therefore, there is an evident need for an intelligent clinical DSS.

### 2.1 Decision Support Systems

A clinical DSS can be defined as a computer program that is designed to analyse data to help health care professionals make clinical decisions. They are meant to increase quality of care, enhance health outcomes and reduce human errors while improving efficiency, cost-benefit ratio and patient satisfaction. Most basic systems include assessment, monitoring and informative tools in the form of computerized alerts, reminders and electronic clinical guidelines. For example, therapy and vital signs monitoring (McGregor et al., 2006) or susceptibility test results visualization (Flottorp et al., 2002). More advanced diagnosis and advisory tools usually rely in statistics, machine learning and artificial intelligent techniques to provide a higher level of data extraction. For example, diagnose and therapy advisers (Paul et al., 2006) or infection risk assessment (Mullett et al., 2004).

Different approaches have been used to design intelligent DSS, each one with their own benefits and drawbacks. Decision trees are popular for their simplicity to understand and construct from logical rules and have been applied in dengue fever diagnosis (Tanner et al., 2008) and antibiotic selection (William Miller, 2013). The amount of computing time required for large datasets is still reasonable. However, they do not tend to work well if decision boundaries are smooth (Quinlan, 1986) and are not optimal for uncorrelated variables. As a result of the greedy strategy applied, they also present high variance and are often unstable, tending to over-fit.

Probability-based approaches are emerging due to its capacity to represent and handle uncertainties (Pearl, 1988). Bayesian Networks (BN) are probabilistic networks that represent a set of variables (nodes) and their dependencies (arcs) using a graph. Such causal dependencies, influences or correlations are defined based on the experience of clinicians. Hence, it can be associated with a rule-based system, which uses data to refine the previously defined relationships. They have been widely exploited in health-care (Lucas et al., 2004). Particularly, Causal Probabilistic Networks (CPN) have been used to develop DSS in diagnosis of cancer (Kahn et al., 1997), ventilator-associated pneumonia (Lucas et al., 2003) and sepsis (Tsoukalas et al., 2015). Bayesian Networks offer a natural way of representing uncertainties, however an insufficient understanding of their formal meaning may give rise to modelling flaws. In particular, Causal Probabilistic Networks are best suited to tackle very specific situations as bloodstream infection (Paul et al., 2006). Unfortunately, treatment recommendation is poor since they usually prescribe broad-spectrum antibiotics (Kofoed et al., 2009). Furthermore, there is a lack of guidance to report and interpret their results by non experts.

The Case-Based Reasoning (CBR) methodology (Aamodt and Plaza, 1994) has been used to tackle problems in antibiotic therapy (Heindl et al., 1997) and molecular biology (Jurisica and Glasgow, 2004). The aim is to use previous experience in form of cases to understand and solve new problems.

### 2.2 Case-Based Reasoning

Case-based reasoning is a widely used approach to solve new problems based on previous experience in

form of cases. It is considered as a methodology to follow rather than an algorithm in itself as shown in Figure 1. The CBR cycle is divided in four different phases. The first phase retrieves from the database those cases that are relevant based on a predefined similarity measure (e.g. euclidean distance). In the second phase, advice is commonly given by adapting or combining the solutions from the retrieved cases (i.e. antibiotic therapies). The proposed solution is incorporated to the case and saved in the database. The third phase monitors the treatment evolution to assess its outcome (e.g. success or failure). Finally, a decision to whether retain or not the case based on its reusability is made.

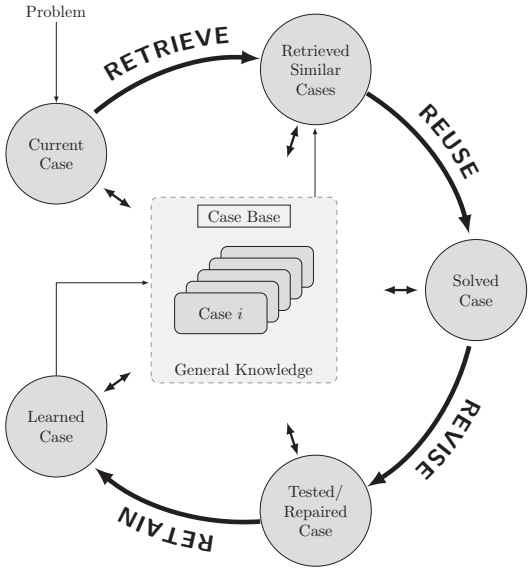



Figure 1: Diagram showing the different phases for a cycle within the Case-Based Reasoning methodology as outlined in (Aamodt and Plaza, 1994).

This methodology is very generic and can be particularized to tackle many different problems. Nevertheless, the most important property that makes CBR appropriate to be used in clinical environments is the straightforward relation that can be found between cases in the CBR methodology and cases as interpreted by clinical staff. Due to this nexus between the clinical and the scientific environments, CBR methodology has been selected to be incorporated in the decision support system and strongly influenced the design of the user interface.

# 3 METHODOLOGY

## 3.1 EPIC IMPOC

Enhanced, Personalized and Integrated Care for Infection Management at Point Of Care (EPIC IMPOC) is a decision support system designed to record a complete set of vital signs at the patients bedside on handheld computing devices while providing instant bedside decision-making assistance to clinical staff. It also pulls in data from the hospital patient administration system, laboratory results and other clinical information stored electronically. It can be used anywhere in the hospital by staff with appropriate access rights, using mobile devices or desktop computers linked to the hospital intranet. EPIC IMPOC has been preliminarily trialled at critical care antimicrobial prescribing, a known reservoir for antimicrobial resistance, and it is being extended to secondary care.

The system architecture is shown in Figure 2 were two parts are clearly differentiated: server and client sides. The server side processes queries, interacts with the permanent storage and serves web pages to the client side. The latter displays information to users. The modules constituting the server side are: *a)* CBR for history review and case comparison. *b)* Probabilistic Inference (PI) aims to provide step-wise guidance fitting the decision pathway followed by clinicians for infection management. *c)* Patient engagement module. *d)* Personalized antibiotic dosing. *e)* Visualization of Antimicrobial Resistance related information. This paper focuses exclusively on the CBR module.

## 3.2 Server Side

The server side has been implemented in Java and uses an object-relational mapping java library (Hibernate ORM) to map an object-oriented domain model to a traditional relational database (SQL). The Lightweight Directory Access Protocol (LDAP) accomplishes the authorization and authentication of users and it is provided in all hospitals at Imperial College Healthcare National Health Service Trust. The server side follows the REST (Representational State Transfer) architectural design, which suggests a group of guidelines to create scalable, more performant and maintainable web services.

The core CBR module implementation is based on the JColibri framework (Díaz-Agudo et al., 2007) including some improvements to achieve better generalization and performance. It is used to retrieve cases from the database based on a similarity measure. A case is defined by a compendium of parameters that

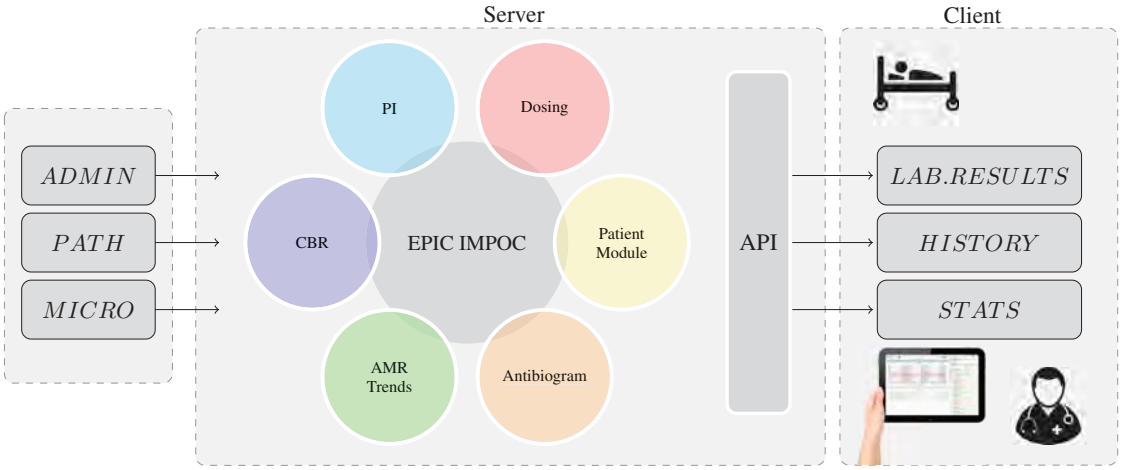



Figure 2: High-level diagram describing the main components of the DSS. The external databases that are currently being accessed are patient administration system (ADMIN), pathology laboratory tests (PATHO) and microbiology results (MICRO). The server side has the following independent modules: Case-Based Reasoning (CBR), Probabilistic Inference (PI), Patient module, Dosing module, antibiogram and AMR trends. All the information is accessed through an API and presented on a handheld device to clinicians.

can be grouped in five different sets: metadata, description, solution, justification and result. However, only those in the description container are used to compute the similarity scores. Some examples of attributes used to define the case are: demographics (age, gender or weight), existing diseases (allergies, HIV or diabetes), respiratory system (ventilation support or oxygen requirements), abdomen (abdominal examination, renal support or catheter), biochemical markers (creatinine or bilirubin) and microbiology (infectious organisms).

## 3.3 Client Side

The client side is a web-based application implemented using HTML, CSS and Javascript which is accessible through the browser. It follows a responsive design approach to render a consistent interface across different devices, from desktop computers to mobile phones and tablets with different screen sizes.

An efficient DSS user interface should present all the information relevant to clinicians neatly, combining different resources of patient-related information (i.e. demographic, pathology and microbiology data). Since some data might be missing or not available, it is also desired to enable clinicians to manually input data or comments for further consideration. In addition, infections evolve with time and so do treatments. Thus, inspection of previous symptoms, treatments and outcomes is desired. Since there is an straightforward relation between cases as interpreted in clinical environments and cases in the CBR methodology, a case is considered as main unit of patient related information to be presented in the interface. A single case is formed by several components, mostly regarding the type and source of the data, and has been divided in different sections (tabs in Figure 3) for visualization purposes. The *Resume* section is read-only and displays the most relevant information (e.g. infectious micro-organisms or organs infected) while *Description* shows additional information and allows the insertion/modification of parameters within the case. *Solution* contains the antibiotic therapy prescribed (including frequency, via and doses) and a section to collect feedback from users.

Six routinely requested biochemical markers were selected as main indicators of infection and patient status after reviewing the scientific literature and discussion with clinicians and infectious disease experts. The temporal evolution of such biochemical markers is shown in *Pathology* (see Figure 3) as time-series where coloured background indicates the normal reference range. Additionally, it is possible to hide/show time-series to improve visualization.

Susceptibility testing is used to determine which antimicrobials will inhibit the growth of micro-organisms causing a infection and is generally performed in vitro. The *Microbiology* section displays the result of all the susceptibility tests requested for the patient. The outcomes of the tests are provided for individual pairs (infectious organism and antibiotic) and are categorised as resistant, intermediate and sensitive. They are also presented during antibiotic therapy selection for further guidance.

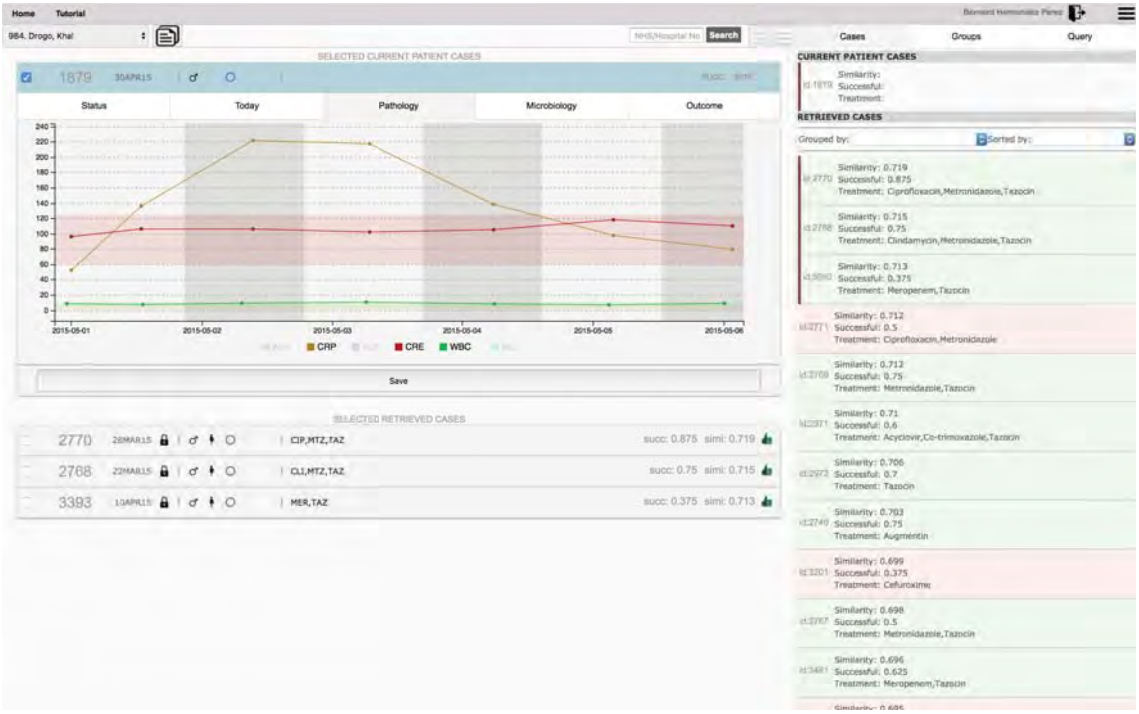



Figure 3: EPIC IMPOC web-based decision support system overview. The main unit of information is the case and its content is displayed among five different tabs (Resume, Description, Pathology, Sensitivity and Solution). The user interface is divided in three main areas: patient selection, dashboard with current patient (top) and retrieved cases (bottom), history review and a side bar to add/remove cases to/from such dashboard.

# 4 RESULTS

A working prototype of EPIC IMPOC incorporating Case-Based Reasoning methodology as decision-support engine was preliminarily trialled in the Intensive Care Unit at Hammersmith Hospital in London for a month. The predefined case base contained 80 cases and information retrieval was performed through handheld computer devices (i.e. ipads) at the patient bed side by clinicians under the supervision of infection specialists. From such study the following conclusions were extracted:

- The system has potential to promote and facilitate communication between nurses, clinicians and infection specialists as shown by the interaction among them during the trial.
- The system improves homogeneous collection of vital signs. Such improvement comes from the introduction of a form in the description of the case to input missing symptoms easily. The form is filled automatically for those symptoms available in external databases (e.g. electronic health records).
- The system facilitates data visualization at POC and simplifies comparison with previous similar cases and outcomes. In addition, biochemical markers evolution, susceptibility tests and history review for the hospitalized patient are easily accessible and found to be very helpful at point of care.
- The system is capable of mimicking clinicians prescription practices in the intensive care unit. In such trial clinicians were under the supervision of infection specialists. As a result, therapies prescribed by clinicians and therapies retrieved by the CBR algorithm matched approximately 90% of the times. It is especially visible in ICU where wide-spectrum antibiotics are commonly used.
- It increases and facilitates the interaction between clinicians and patients. Therefore it helps engaging with patients and opens the possibility to educate population on antibiotic misuse and its consequences (Rawson et al., 2016b).

## 4.1 System Usability Scale Survey

A survey to evaluate the usability of the decision support system interface was performed. The System Usability Scale (SUS) (Brooke et al., 1996) is composed

Table 1: The original SUS statements (Brooke et al., 1996), average agreement and SUS contribution.

| SUS statements | Avg. rating | SUS contribution |
|---|---|---|
| I think that I would like to use this system frequently. | 2.8 | 1.8 |
| I found the system unnecessarily complex. | 1.4 | 3.6 |
| I thought the system was easy to use. | 2.0 | 1.0 |
| I think that I would need the support of a technical person to be able to use this system. | 1.6 | 3.4 |
| I found that the various functions in this system were well integrated. | 3.0 | 2.0 |
| I thought that there was too much inconsistency in this system. | 0.2 | 4.8 |
| I would imagine that most people would learn to use this system very quickly. | 2.8 | 1.8 |
| I found the system very cumbersome to use. | 2.2 | 2.8 |
| I felt very confident using the system. | 3.0 | 2.0 |
| I needed to learn a lot of things before I could get going with this system. | 0.8 | 4.2 |

of 10 statements to which participants indicate their agreement from 1 to 5, where 5 indicates strongly agree. Predefined rules for positive and negative statements are used to obtain the SUS contributions. The SUS contribution for each statement ranges from 1 to 4 where higher scores indicate better usability (see Table 1) and their sum is multiplied by 2.5 to calculate the final SUS score. It ranges from 0 to 100 where poor and great product usability are indicated for SUS scores under 50 and over 80 respectively. This survey is technology agnostic, quick, provides a single score and is non proprietary. A free-text box was added for additional comments and suggestions.

The SUS survey was completed by 10 different participants (83% males) from 27 to 51 years old where technical training in the use of the system was not provided. The profile of those participants was infection specialist (two), clinician (three), nurse (four) and other staff (one). The SUS contribution for each statement is presented in the right column in Table 1. The final SUS score obtained is 68.5 which indicates good product usability with margin to improve. Additionally, a variety of comments were provided by participants and have been synthesized in the following bullet points:

- There is a common concern among experienced clinicians and infection specialists in the use junior doctors would do of such large amount of data displayed in the interface. The decision support system has potential to help training junior doctors and improve their prescription practices, but it needs to narrow the presented information providing specific guidance.
- Clinicians consider the user interface intuitive and helpful for patient long-term monitoring and management, however it might sometimes be time consuming. Additionally, it does not entirely fit with the work-flow followed to prescribe antibiotic therapies.
- They suggested the possibility of recording further parameters, not necessarily directly related with infections.

From the preliminary trial performed by infection specialists and the feedback obtained from the surveys, it is possible to conclude that the CBR algorithm is able to mimic the prescription practices of users. However, that is not enough to promote change in antibiotic prescription practices. Initially, as a quick solution infection specialists were keen in creating an "ideal" case base; that is, a set of cases with optimal antibiotic therapies according to infection guidelines and expert prescriptions. Such optimal therapies would then be suggested by the decision support system to further users. Hence, the knowledge would be transferred from infection specialists to other clinical staff (e.g. nurses and clinicians).

Unfortunately, this approach presents several drawbacks. Creating a complete case base that covers the whole spectrum of possibilities is nearly impossible and time consuming. In addition, infections are often acquired in hospitals by contagious as a consequence of treating more severe diseases which diminish the immune system (e.g. surgeries and cancer). Therefore, future therapies prescribed by clinicians and recorded in the system might not agree with the infection guidelines reshaping the case base and therefore altering CBR recommendations. A case base with strictly guideline oriented therapies on the long-term is unrealistic and limits the scope and usability of the system.

After discussion with a multi-professional team including physicians, nurses, pharmacists and non-medical researchers, an study to map the pathway followed by clinicians to prescribe antibiotics therapies was performed (Rawson et al., 2016a). The reported infection management pathway was defined as a stepwise Bayesian model of estimating probabilities in which each step adds systematically information to allow optimisation of decisions on diagnosis and management of infection. Initially, clinicians estimate the risk of infection and attempt to localize its

source by looking at patient's physiological parameters. Once clinicians construct a picture of the severity of the infection, whether or not to initiate antimicrobial therapy is decided. In this step, local microbiology guidance provided within hospitals was the most commonly cited factor. Finally, they review and refine the treatment accordingly.

This new approach enables to produce very specific step-by-step probability-like decision support. This would improve the guidance provided to junior clinicians and facilitates the validation of decision-making for each individual step. Additionally, since reviewing of previous cases is not necessary it would greatly reduce usability time. Since the methodology integrates with the infection management pathway, it will likely influence prescription practices to a higher degree than CBR.

# 5 DISCUSSION

DSSs are being exploited in several areas such as business or economics but their acceptance by clinical staff is obstructing its use in hospitals and other clinical environments (Kawamoto et al., 2005). Designing a DSS based on usability principles and simply providing the clinical information does not guarantee acceptance (Moxey et al., 2010). Other factors as accessibility, availability, easy of use, time requirements and integration into the clinical workflow are important and need to be considered (Tsopra et al., 2014). Taking previous knowledge in consideration, a DSS to support prescription of antibiotic therapies and patient monitoring at POC exploiting the CBR methodology was implemented.

In many circumstances, as complicated cases, providers prefer to consult their colleges or more specialised clinicians as infection specialists. This consultation among different members of the clinical staff was facilitated by the DSS. In addition, it was believed to enhance decision making and homogeneous collection of vital signs among clinicians resulting in better prescribing practices. Furthermore, re-entering patient data to generate advices is a deterrent to use (Moxey et al., 2010) and integration into existing programs (e.g. electronic medical records) was a clear facilitator.

The usability measured through the SUS survey was 68.5 which is about average and shows potential margin for improvement. A similar strength of agreement was shown by participants for the third and eight statements which had an average rating of 2.0 and 2.2 respectively. The SUS contribution for each statement was 1.0 and 2.8 respectively, indicating that the system is usable but not necessarily easy. Note that some wording used by the original SUS was suggested to be poorly understood by participants affecting the results (Bangor et al., 2008). As an example, the sixth statement which contributed the most contains the wording "too much" which might be unclear. Additionally, users may not have a good sense of the required complexity of such systems since there are no commonly known competing solutions and the wording "unnecessarily" in the second statement might have led to a higher contribution of 3.6. Therefore, the final SUS score is possibly slightly less that the one presented.

There is still much to be done to make these system work in routine clinical practice. Measuring absolute usability using a single metric is very challenging since many external factors influence the results (e.g. technical training or accessible technology). The feedback provided greatly helped to identify areas of improvement and the results of the survey are a useful source for future comparison to asses the benefits of including new components to tackle the identified weaknesses.

# 6 CONCLUSIONS

CBR methodology was incorporated in a DSS as decision-making engine providing similar antibiotic therapies to those prescribed by expert clinicians in the majority of cases in our preliminary trials. In addition, it was widely accepted for information retrieval and long-term patient monitoring and management. Three main areas of improvement were identified from the feedback provided by expert clinicians: specific guidance requirements for junior clinicians, a need to reduce time consumption using the DSS and a better integration into clinical workflow.

The reported infection management pathway was defined as a multi-step Bayesian-like approach (Rawson et al., 2016a) which inherently tackles most of the weaknesses identified: specific guidance, time constraints and integration in the workflow. Therefore, it is more suitable to support and modify prescription practices.

The combination of both approaches into one single decision support system is a very elegant solution that could lead to an increase in acceptability among clinicians. To validate its feasibility, infection risk and source of infection inference from biochemical markers are considered as primary steps. The final integration of such inference in the user interface has potential to reduce misuse of antibiotics and its evaluation forms the basis for future work.

## ACKNOWLEDGEMENT

This report is independent research funded by the National Institute for Health Research Invention for Innovation (i4i) programme, Enhanced, Personalized and Integrated Care for Infection Management at Point of Care (EPIC IMPOC), II-LA-0214-20008.

The authors would like to thank members of Imperial College NHS Healthcare Trust who participated in the study. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the UK Department of Health.

## REFERENCES

Aamodt, A. and Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI communications*, 7(1):39–59.

Bangor, A., Kortum, P. T., and Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, 24(6):574–594.

Belden, J. L., Grayson, R., and Barnes, J. (2009). Defining and testing emr usability: Principles and proposed methods of emr usability evaluation and rating. Technical report, Healthcare Information and Management Systems Society (HIMSS).

Boulos, M. N. K., Brewer, A. C., Karimkhani, C., Buller, D. B., and Dellavalle, R. P. (2014). Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online journal of public health informatics*, 5(3).

Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

Díaz-Agudo, B., González-Calero, P. A., Recio-García, J. A., and Sánchez-Ruiz-Granados, A. A. (2007). Building cbr systems with jcolibri. *Science of Computer Programming*, 69(1):68–75.

Flottorp, S., Oxman, A. D., Håvelsrud, K., Treweek, S., and Herrin, J. (2002). Cluster randomised controlled trial of tailored interventions to improve the management of urinary tract infections in women and sore throat. *Bmj*, 325(7360):367.

Heindl, B., Schmidt, R., Schmid, G., Haller, M., Pfaller, P., Gierl, L., and Pollwein, B. (1997). A case-based consiliarius for therapy recommendation (icons): computer-based advice for calculated antibiotic therapy in intensive care medicine. *Computer Methods and Programs in Biomedicine*, 52(2):117–127.

Holmes, A. H., Moore, L. S., Sundsfjord, A., Steinbakk, M., Regmi, S., Karkey, A., Guerin, P. J., and Piddock, L. J. (2016). Understanding the mechanisms and drivers of antimicrobial resistance. *The Lancet*, 387(10014):176–187.

Jurisica, I. and Glasgow, J. (2004). Applications of case-based reasoning in molecular biology. *Ai Magazine*, 25(1):85.

Kahn, C. E., Roberts, L. M., Shaffer, K. A., and Haddawy, P. (1997). Construction of a bayesian network for mammographic diagnosis of breast cancer. *Computers in biology and medicine*, 27(1):19–29.

Kawamoto, K., Houlihan, C. A., Balas, E. A., and Lobach, D. F. (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*, 330(7494):765.

Kofoed, K., Zalounina, A., Andersen, O., Lisby, G., Paul, M., Leibovici, L., and Andreassen, S. (2009). Performance of the treat decision support system in an environment with a low prevalence of resistant pathogens. *Journal of antimicrobial chemotherapy*, 63(2):400–404.

Lucas, P., Schurink, C., Hoepelman, I., and Bonten, M. (2003). A model-based approach to improved prescription of antibiotics.

Lucas, P. J., van der Gaag, L. C., and Abu-Hanna, A. (2004). Bayesian networks in biomedicine and health-care. *Artificial intelligence in medicine*, 30(3):201–214.

McGregor, J. C., Weekes, E., Forrest, G. N., Standiford, H. C., Perencevich, E. N., Furuno, J. P., and Harris, A. D. (2006). Impact of a computerized clinical decision support system on reducing inappropriate antimicrobial use: a randomized controlled trial. *Journal of the American Medical Informatics Association*, 13(4):378–384.

Moxey, A., Robertson, J., Newby, D., Hains, I., Williamson, M., and Pearson, S.-A. (2010). Computerized clinical decision support for prescribing: provision does not guarantee uptake. *Journal of the American Medical Informatics Association*, 17(1):25–33.

Mullett, C. J., Thomas, J. G., Smith, C. L., Sarwari, A. R., and Khakoo, R. A. (2004). Computerized antimicrobial decision support: an offline evaluation of a database-driven empiric antimicrobial guidance program in hospitalized patients with a bloodstream infection. *International journal of medical informatics*, 73(5):455–460.

ONeill, J. (2014). Antimicrobial resistance: Tackling a crisis for the health and wealth of nations.

Paul, M., Andreassen, S., Nielsen, A. D., Tacconelli, E., Almanasreh, N., Fraser, A., Yahav, D., Ram, R., Leibovici, L., Group, T. S., et al. (2006). Prediction of bacteremia using treat, a computerized decision-support system. *Clinical Infectious Diseases*, 42(9):1274–1282.

Pearl, J. (1988). Probabilistic inference in intelligent systems.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1):81–106.

Rawson, T., Charani, E., Moore, L. S., Hernandez, B., Castro-Sanchez, E., Herrero, P., Georgiou, P., and Holmes, A. H. (2016a). Mapping the decision pathways of acute infection management in secondary care among uk medical physicians: A qualitative study. *BMC medicine [In Press]*.

Rawson, T. M., Moore, L. S., Hernandez, B., Castro-Sanchez, E., Charani, E., Georgiou, P., Ahmad, R., and Holmes, A. H. (2016b). Patient engagement with

infection management in secondary care: a qualitative investigation of current experiences. *BMJ open*, 6(10):e011040.

Sintchenko, V., Iredell, J. R., Gilbert, G. L., and Coiera, E. (2005). Handheld computer-based decision support reduces patient length of stay and antibiotic prescribing in critical care. *Journal of the American Medical Informatics Association*, 12(4):398–402.

Tanner, L., Schreiber, M., Low, J. G., Ong, A., Tolfvenstam, T., Lai, Y. L., Ng, L. C., Leo, Y. S., Puong, L. T., Vasudevan, S. G., et al. (2008). Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. *PLoS Negl Trop Dis*, 2(3):e196.

Tsopra, R., Jais, J.-P., Venot, A., and Duclos, C. (2014). Comparison of two kinds of interface, based on guided navigation or usability principles, for improving the adoption of computerized decision support systems: application to the prescription of antibiotics. *Journal of the American Medical Informatics Association*, 21(e1):e107–e116.

Tsoukalas, A., Albertson, T., and Tagkopoulos, I. (2015). From data to optimal decision making: a data-driven, probabilistic machine learning approach to decision support for patients with sepsis. *JMIR medical informatics*, 3(1):e11.

William Miller, O. (2013). A decision tree: Proper antibiotic selection and use.

Wise, R., Hart, T., Cars, O., Streulens, M., Helmuth, R., Huovinen, P., and Sprenger, M. (1998). Antimicrobial resistance is a major threat to public health. *British medical journal*, 317(7159):609–611.

# Healthcare-Oriented Characterisation of Human Movements by Means of Impulse-Radar Sensors and by Means of Accelerometric Sensors

Paweł Mazurek[1], Jakub Wagner[1], Andrzej Miękina[1], Roman Z. Morawski[1] and Frode Fadnes Jacobsen[2]

[1]*Institute of Radioelectronics and Multimedia Technology, Faculty of Electronics and Information Technology, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland*

[2]*Bergen University College, Center for Care Research, Møllendalsveien 6-8, 5020 Bergen, Norway*

*{p.mazurek, j.wagner, a.miekina, r.morawski}@ire.pw.edu.pl, frode.fadnes.jacobsen@hib.no*

Keywords: Healthcare, Impulse-Radar Sensor, Accelerometer, Measurement Data Processing, Uncertainty Estimation.

Abstract: This paper is devoted to the healthcare-oriented characterisation of the human movements by means of the accelerometric and impulse-radar sensors – the sensors that may be employed in care services for monitoring of elderly and disabled persons. Characterisation of the movements in terms of the so-called self-selected walking velocity can be used by the medical and healthcare personnel to assess the overall health status of a monitored person. The quality of the characterisation, based on the measurement data from accelerometric and impulse-radar sensors, has been assessed in a series of real-world experiments which involved the estimation of the instantaneous and mean walking velocity of a person moving according to predefined patterns. Some indicators of uncertainty of the velocity estimation, determined with respect to assumed predefined velocity values, have been used for comparison of the performance of both types of sensors. The experiments have shown that impulse-radar sensors enable one to estimate the mean walking velocity more accurately than the accelerometric sensors: the estimates obtained on the basis of data from the latter sensors are affected by larger bias and are more widely spread around their mean values.

## 1 INTRODUCTION

The life expectancy has been growing in Europe for many years, while the healthy life expectancy has been slightly diminishing since the last decade of the XXth century (*cf. http://www.healthy-life-years.eu/*). Hence the growing importance of research on new technologies that could be employed in monitoring systems supporting care services for elderly and disabled persons. The capability of those systems to detect dangerous events, such as person's fall, is of key importance (Hamm et al., 2016). However, those systems are expected not only to detect dangerous events, but also to predict those events on the basis of acquired data. The analysis of gait, as well as of the itinerary and timing of activities of the monitored persons, may thus contribute to the prevention (Baldewijns et al., 2016a). The relevance of features related to gait analysis in monitoring of elderly persons, and in particular – in fall prevention, has been emphasised in several recent papers (Buracchio et al., 2010, Studenski et al., 2011, Lusardi, 2012, Egerton et al., 2014, Stone et al., 2015, Thingstad et al., 2015, Baldewijns et al., 2016b).

So far, the most popular monitoring technique, already applied in healthcare practice, is based on wearable devices (Bulling et al., 2014, Cola et al., 2014, Luque et al., 2014, Brodie et al., 2015). Those devices do not require a pre-built infrastructure and thus may be used outdoor. The signals from movement sensors (mainly accelerometers and gyroscopes), worn by a monitored person, are transmitted via radio links to a computer and analysed. This solution makes also possible the acquisition of physiological data (such as values of blood pressure, ECG data or EEG data).

Recently, numerous attempts have been made to apply various radar techniques for monitoring of elderly and disabled persons (Cuddihy et al., 2012, Liu et al., 2012, Tomii and Ohtsuki, 2012, Jian et al., 2014, Su et al., 2015, Miękina et al., 2016b). Those attempts are mainly motivated by the conviction that



Mazurek, P., Wagner, J., Miękina, A., Morawski, R. and Jacobsen, F.
Healthcare-Oriented Characterisation of Human Movements by Means of Impulse-Radar Sensors and by Means of Accelerometric Sensors.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 128-138
ISBN: 978-989-758-213-4

those techniques may be less intrusive than vision-based solutions (*e.g.* digital cameras), less cumbersome than wearable solutions (*e.g.* accelerometers and gyroscopes), and less invasive with respect to the home environment than environmental solutions (*e.g.* pressure sensors).

This paper is devoted to the assessment of the uncertainty of the estimation of the walking velocity, on the basis of data acquired by means of impulse-radar sensors and by means of accelerometric sensors. As suggested in the literature, *e.g.* (Fritz and Lusardi, 2009), the walking velocity is highly informative for healthcare experts; for example:

- the velocity lower than 0.6 m/s enables them to predict an increase in the risk of falls and hospitalisation of a monitored person;
- an improvement in walking velocity of at least 0.1 m/s is a useful predictor for well-being;
- a decrease of the same amount is correlated with deterioration of the health status or advancement of disability.

The comparative study, reported in this paper, is based on an extensive set of real-world experiments which comprise:

- simultaneous recording of measurement data from both types of sensors, representative of the gait characteristics of a person moving according to predefined patterns;
- statistical analysis of those data, aimed at determination of certain indicators of uncertainty of the velocity estimation.

Due to the operation principle of both types of sensors, one may expect that the position of a monitored person can be better estimated on the basis of the data from impulse-radar sensors (hereinafter called *radar data* for brevity), and its acceleration – on the basis of data from the accelerometric sensors (hereinafter called *accelerometric data* for brevity). Therefore, despite the fact that both the position and the acceleration may also be of interest for the healthcare personnel, this study is confined to the uncertainty of the estimation of the velocity, which requires similar degree of the measurement data preprocessing for both types of sensors.

# 2 METHODOLOGY OF EXPERIMENTATION

## 2.1 Data Acquisition

The raw measurement data for experimentation have been acquired by means of the APDM Opal accelerometric sensor (*cf. http://www.apdm.com/wearable-sensors/*) attached to the waist of a monitored person, and by means of a pair of synchronised impulse-radar sensors – *cf.* (Morawski et al., 2014) – whose location is shown in figure 1. A monitored person has moved at the distance of *ca.* 1–6.5 m from each of them.

The walking velocity has been assessed on the basis of real-world data acquired when an experimenter has been walking at various constant velocities, ranging from 0.5 m/s to 1.0 m/s, forth and back along a straight line – $R = 20$ times along the *x*-axis, between points (0,3) and (4,3), and $R$ times along the *y*-axis, between points (2,1) and (2,5) (*cf.* figure 1). In order to assure a known constant walking velocity, a metronome has been used.

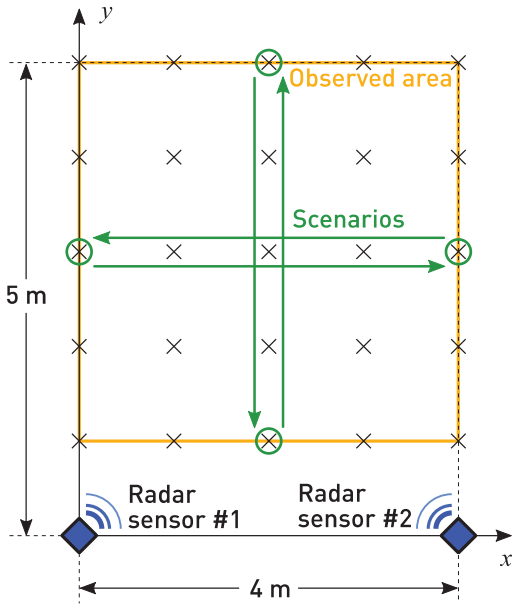



Figure 1: Experimental setup; the crosses indicate the reference points, *i.e.* the points where marks have been placed on the floor.

## 2.2 Data Preprocessing

### 2.2.1 Radar Data

The measurement data from a pair of impulse-radar sensors – after preliminary preprocessing, as described in (Miękina et al., 2016a) – take on the form of a sequence of numbers representative of the *x*-*y* coordinates of a monitored person.

A sequence of the estimates of the instantaneous walking velocity may be obtained by numerical differentiation of the sequence of the position estimates, *e.g.* by means of the central-difference

method (Wagner et al., 2015), defined by the formula:

$$\hat{d}_n^{(1)} \equiv \frac{d_{n+1} - d_{n-1}}{\Delta t_n} \quad \text{for} \quad n = 1, \ldots, N-1 \qquad (1)$$

where $\{d_n\}$ is a sequence of data to be differentiated, and $\Delta t_n \equiv t_{n+1} - t_{n-1}$ are the differentiation steps, with $t_n$ denoting the time moments at which the data have been acquired. That method is, however, very sensitive to errors corrupting the data used for derivative estimation; therefore, it should be regularised through, *e.g.*, optimisation of the differentiation step. The total velocity magnitude has been calculated according to the formula:

$$\hat{v}_n = \sqrt{\left(\hat{x}_n^{(1)}\right)^2 + \left(\hat{y}_n^{(1)}\right)^2} \quad \text{for} \quad n = 1, \ldots, N \qquad (2)$$

where $\hat{x}_n^{(1)}$ and $\hat{y}_n^{(1)}$ are estimates of the first derivatives, computed on the basis of the estimates of the *x*- and *y*-data sequences.

### 2.2.2 Accelerometric Data

An accelerometric sensor – composed of an accelerometer, magnetometer and gyroscope – provides a sequence of data representative of the monitored person's instantaneous acceleration in three directions, *viz.* magnetic north, magnetic west, and vertical. A sequence of the estimates of the instantaneous velocities in these directions can be obtained by numerical integration of the sequences of the acceleration values. It must be, however, taken into account that – since both systematic and random errors corrupting accelerometric data propagate through the integration process (Thong et al., 2004) – the velocity estimates may be subject to a growing-with-time drift and random errors whose standard deviation is also growing with time. As a consequence, non-zero estimates may appear even when a monitored person is standing still; therefore, the velocity estimates have to be corrected by means of a so-called zero-velocity compensation procedure (Bang et al., 2003). It can be applied to a velocity trajectory whose first and last values are known to be zero. In the research reported here, the following correction formula has been used:

$$\hat{v}_n \Leftarrow \hat{v}_n - \delta \frac{n - n_1}{n_2 - n_1} \quad \text{for} \quad n_1 < n < n_2 \qquad (3)$$

where $\delta = \hat{v}_{n_2} - \hat{v}_{n_1}$, $n_1$ and $n_2$ are the indices of the first and last time instants of the movement, respectively; the latter parameters have been determined experimentally. The corrected velocity trajectories in the magnetic north and west directions (denoted with $\hat{v}_n^N$ and $\hat{v}_n^W$, respectively) have been used for computing the total velocity magnitude according to the formula:

$$\hat{v}_n = \sqrt{\left(\hat{v}_n^N\right)^2 + \left(\hat{v}_n^W\right)^2} \quad \text{for} \quad n = 1, \ldots, N \qquad (4)$$

## 2.3 Criteria of Performance Evaluation

In each experiment, $R$ sequences of the instantaneous total velocity estimates have been computed using equations 2 and 4 on the basis of both radar data and accelerometric data:

$$\left\{\hat{v}_n^{(r)} \mid n = 1, ..., N\right\} \quad \text{for} \quad r = 1, \ldots, R \qquad (5)$$

Prior to the evaluation of the uncertainty of the estimation, some outlying sequences have been removed to prevent the misinterpretation of the results. The outlying sequences have been identified as those whose mean value:

$$\hat{\mu}^{(r)} = \frac{1}{N} \sum_{n=1}^{N} \hat{v}_n^{(r)} \qquad (6)$$

deviated from the group mean value:

$$\hat{\mu} = \frac{1}{R} \sum_{r=1}^{R} \hat{\mu}^{(r)} \qquad (7)$$

by more than three standard deviations:

$$\hat{\sigma} = \sqrt{\frac{1}{R-1} \sum_{r=1}^{R} \left(\hat{\mu}^{(r)} - \hat{\mu}\right)^2} \qquad (8)$$

Next, the qualitative assessment of the uncertainty of the estimates has been performed. It has been based on the inspection of the estimates of the mean:

$$\hat{\mu}_n = \frac{1}{R'} \sum_{r=1}^{R'} \hat{v}_n^{(r)} \qquad (9)$$

and standard deviation:

$$\hat{\sigma}_n = \sqrt{\frac{1}{R'-1} \sum_{r=1}^{R'} \left(\hat{v}_n^{(r)} - \hat{\mu}_n\right)^2} \qquad (10)$$

of each element of the sequence of the instantaneous velocity estimates; $R'$ denotes the number of sequences in a set under consideration after removing the outlying sequences.

Finally, the quantitative assessment of the uncertainty of the estimates of the mean walking velocity has been done using the following indicators:

- the absolute discrepancy between the mean value of the estimates of the velocity and the

predefined value of that velocity,
- the absolute root-mean-square discrepancy of the estimates with respect to the predefined value,
- the lower and upper bounds of the absolute discrepancy between the estimates and the predefined value.

The above indicators have been calculated separately for each set of $R'$ estimates of mean walking velocity, obtained in each experiment by averaging its $N$ samples.

## 3 RESULTS AND DISCUSSION

In figures 2–5, the mean sequences of instantaneous velocity estimates of a moving person, obtained on the basis of the radar data and accelerometric data – for both directions of movement (*i.e.* along *x*-axis and along *y*-axis) and for all predefined velocity values – are presented.

It is worth being noticed that the uncertainty of estimation, based on radar data, is direction dependent: for the movement along the *x*-axis and predefined velocity values from 0.5 m/s to 0.7 m/s, the estimated mean value of the velocity oscillates around the predefined value during the movement. This cannot be observed for the movement along the *y*-axis in the same range of velocity values. Moreover, it may be seen that the standard deviation of the velocity is greater for the movement along the *x*-axis.

Those differences are caused by the fact, that the calculation of the position of the moving person is easier when the distance between the person and each of the radars is equal (*i.e.* when each radar *sees* the same side of the human body).

On the other hand, it may be noticed that the uncertainty of estimation, based on accelerometric data, is direction independent.

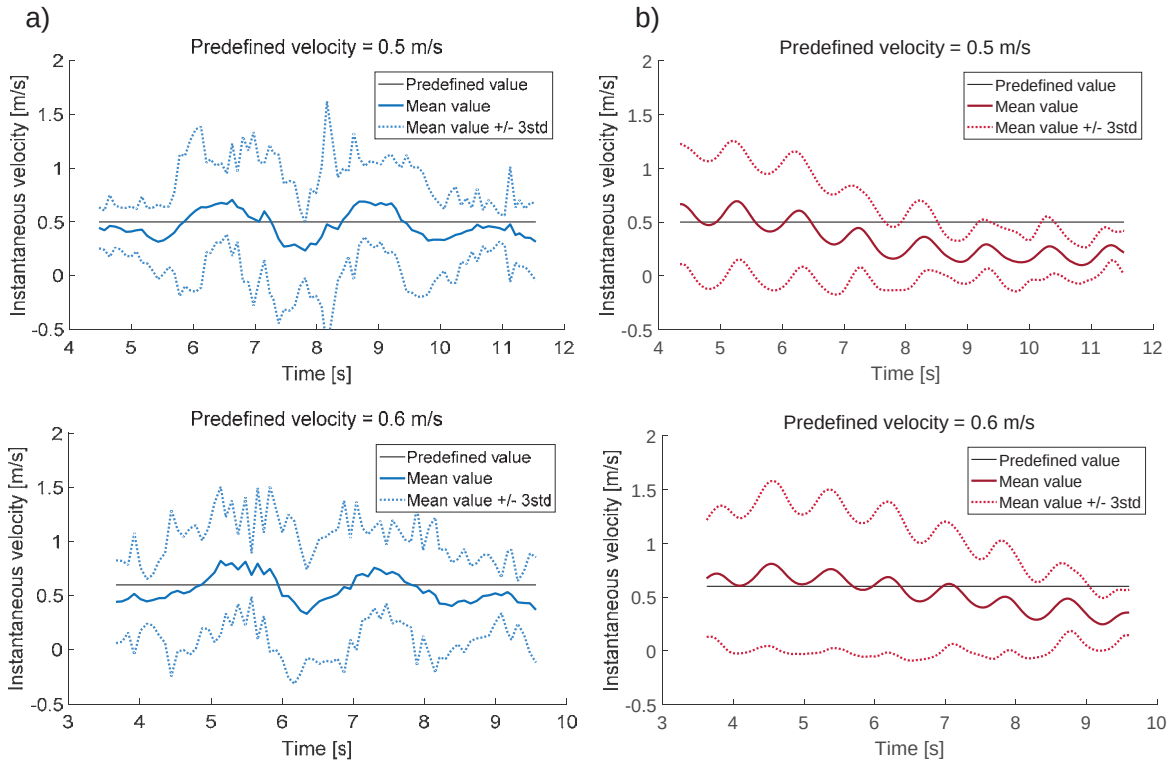



Figure 2: Uncertainty indicators determined for estimates of the velocity of a moving person, obtained on the basis of the radar data (a) and accelerometric data (b), for the movement along *x*-axis with the velocity values ranging from 0.5 m/s to 0.6 m/s. In all sub-figures: the thick solid line denotes the sequence of mean values, while the dotted lines – the sequences of values that are three standard deviations away from the mean sequence.

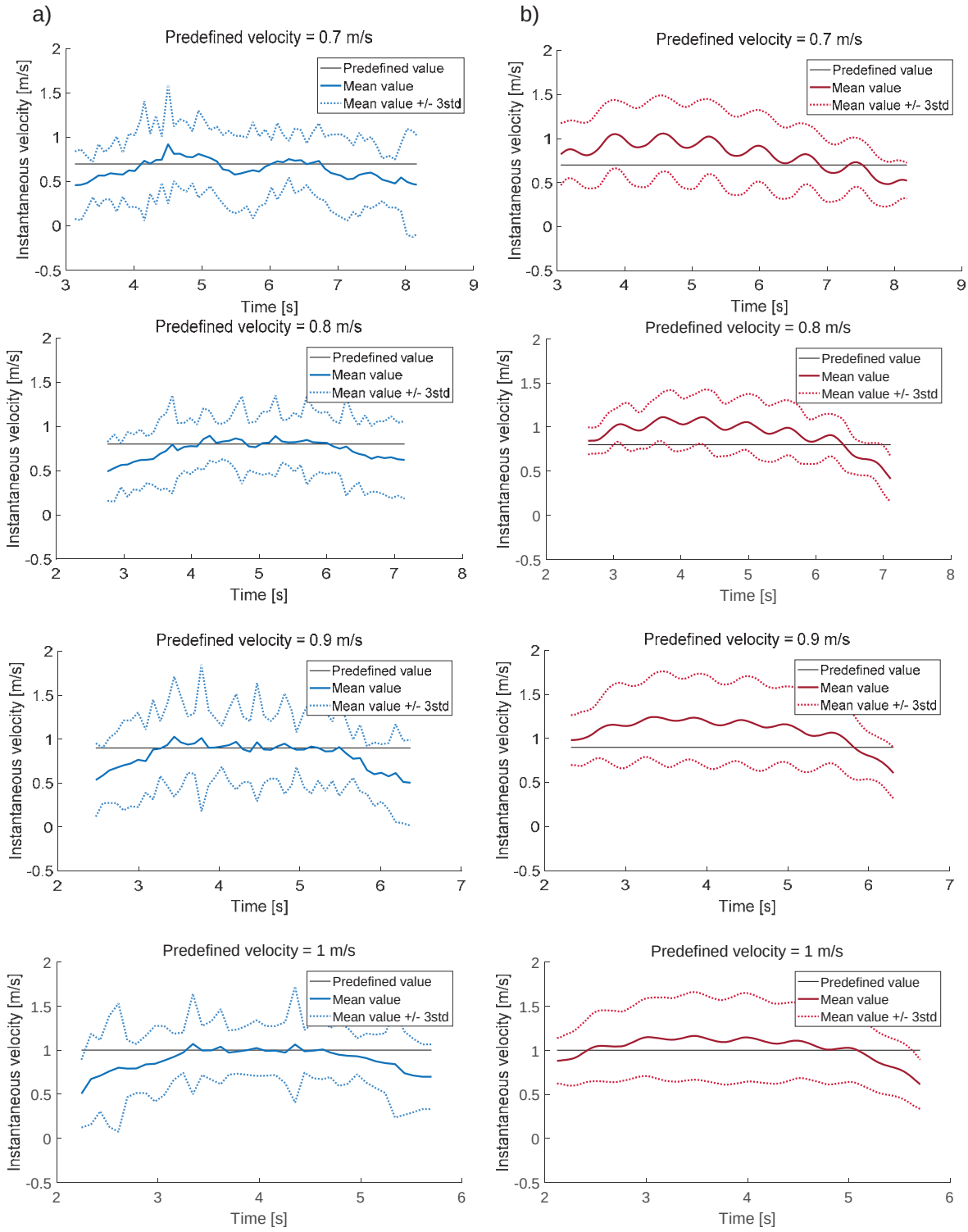



Figure 3: Uncertainty indicators determined for estimates of the velocity of a moving person, obtained on the basis of the radar data (a) and accelerometric data (b), for the movement along *x*-axis with the velocity values ranging from 0.7 m/s to 1.0 m/s. In all sub-figures: the thick solid line denotes the sequence of mean values, while the dotted lines – the sequences of values that are three standard deviations away from the mean sequence.

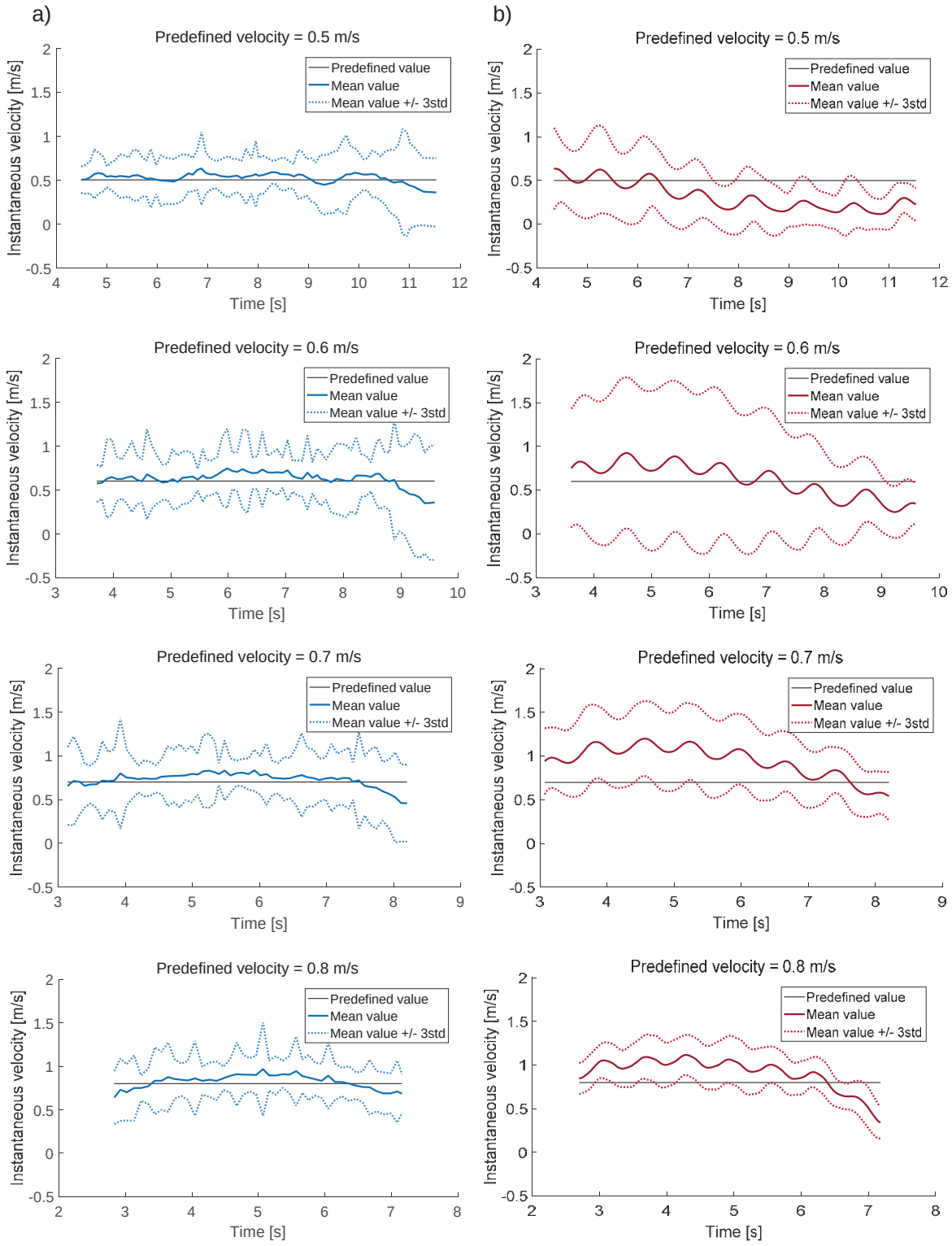



Figure 4: Uncertainty indicators determined for estimates of the velocity of a moving person, obtained on the basis of the radar data (a) and accelerometric data (b), for the movement along *y*-axis with the velocity values ranging from 0.5 m/s to 0.8 m/s. In all sub-figures: the thick solid line denotes the sequence of mean values, while the dotted lines – the sequences of values that are three standard deviations away from the mean sequence.

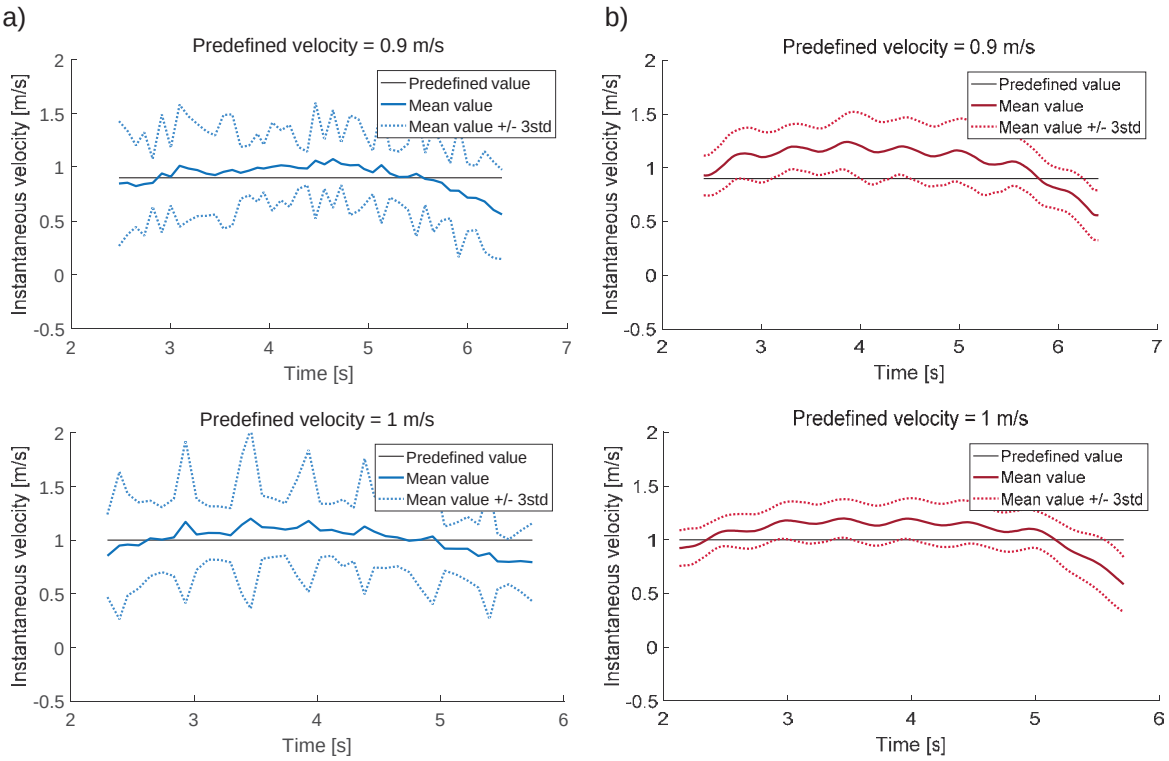



Figure 5: Uncertainty indicators determined for estimates of the velocity of a moving person, obtained on the basis of the radar data (a) and accelerometric data (b), for the movement along *y*-axis with the velocity values ranging from 0.9 m/s to 1.0 m/s. In all sub-figures: the thick solid line denotes the sequence of mean values, while the dotted lines – the sequences of values that are three standard deviations away from the mean sequence.

In figure 6, the so-called box plots representing the aggregated uncertainty of the estimation of the mean walking velocity, performed on the basis of the radar data and accelerometric data, for each investigated value of the walking velocity, are presented. Each box plot indicates:

- the median value;
- the interquartile range (IQR), *i.e.* range between the first and third quartile;
- the smallest value still within 1.5 IQR from the first quartile, and the largest value still within 1.5 IQR from the third quartile;
- the values lying outside 1.5 IQR from the first quartile and 1.5 IQR from the third quartile (marked with crosses).

In table 1 and table 2, the numerical results of all experiments – performed for various walking velocities – are collected.

The results presented in tables 1 and 2 show that the estimates of the mean walking velocity, obtained on the basis of the radar data, are far more accurate than those obtained on the basis of the accelerometric data. For the estimation of the velocity based on the radar data the mean discrepancy, *i.e.* the difference between estimated mean value and a predefined value of the velocity, varies from –0.12 to 0.03 m/s, while it varies from –0.18 to 0.24 m/s for the estimation based on accelerometric data. Moreover, it can be observed that the estimates obtained on the basis of the radar data are more concentrated around their mean values – the root-mean-square discrepancy of the radar-data-based velocity estimates varies from 0.02 to 0.12 m/s, while it varies from 0.08 to 0.27 m/s for the accelerometric-data-based estimates.

It can also be noticed that the estimates of the mean walking velocity, obtained on the basis of the radar data, tend to be underrated with respect to the predefined walking velocity for the movements along the *x*-axis, and very accurate for the

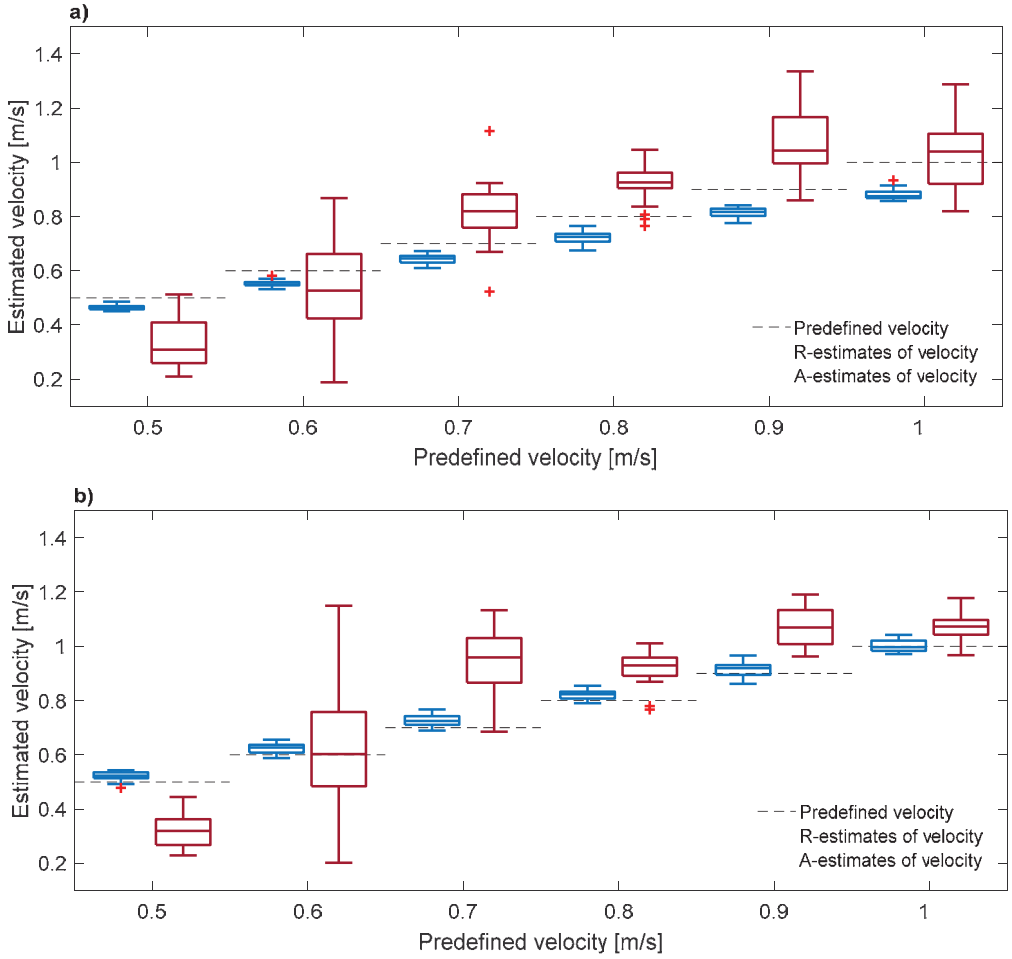



Figure 6: Box plots representing the uncertainty of the estimation of the walking velocity, based on the radar data (R-estimates) and accelerometric data (A-estimates); a) *x*-axis movement, b) *y*-axis movement.

movements along the *y*-axis. On the other hand, the estimates of the mean walking velocity, obtained on the basis of the accelerometric data, seem to be underrated for lower walking velocities and overrated for faster movements.

Lastly, it should be noted that the impact of the imperfections of the movements of the experimenter, reproducing the predefined patterns, are the same for both sensors; so, not changing the result of comparison.

## 4 CONCLUSIONS

The novelty of the study, whose results are presented in this paper, consists in systematic comparison of two monitoring techniques, *viz.* impulse-radar sensors and accelerometric sensors, when applied for healthcare-oriented characterisation of the human movements.

The performance of both types of sensors has been compared on the basis of data acquired by means of them in a series of real-world experiments which involved tracking of a person moving according to predefined patterns. The indicators of uncertainty of the velocity estimation have been determined with respect to the assumed predefined values of velocity.

Prior to the evaluation of the uncertainty, the measurement data from both types of sensors have to be adequately processed. The velocity estimates, obtained on the basis of the accelerometric data, are determined by numerical integration of the sequences of the acceleration estimates and corrected by means of a zero-velocity compensation procedure. The velocity estimates, obtained on the basis of the radar data, are determined using the regularised numerical differentiation of the sequence of the position estimates.

Table 1: Uncertainty of mean velocity estimation for the movement along *x*-axis.

| Uncertainty indicators characterising estimates of mean velocity | Predefined walking velocity [m/s] | | | | | |
|---|---|---|---|---|---|---|
| | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 1.00 |
| Impulse-radar sensors | | | | | | |
| Mean discrepancy [m/s] | –0.03 | –0.04 | –0.06 | –0.08 | –0.08 | –0.12 |
| Root-mean-square discrepancy [m/s] | 0.04 | 0.05 | 0.06 | 0.08 | 0.09 | 0.12 |
| Upper bound of the discrepancy [m/s] | –0.01 | 0.01 | –0.03 | –0.03 | –0.06 | –0.07 |
| Lower bound of the discrepancy [m/s] | –0.05 | –0.07 | –0.09 | –0.13 | –0.12 | –0.14 |
| Accelerometric sensors | | | | | | |
| Mean discrepancy [m/s] | –0.17 | –0.06 | 0.12 | 0.12 | 0.18 | 0.03 |
| Root-mean-square discrepancy [m/s] | 0.19 | 0.17 | 0.17 | 0.14 | 0.23 | 0.13 |
| Upper bound of the discrepancy [m/s] | 0.01 | 0.27 | 0.42 | 0.25 | 0.44 | 0.29 |
| Lower bound of the discrepancy [m/s] | –0.29 | –0.41 | –0.18 | –0.03 | –0.04 | –0.18 |

Table 2: Uncertainty of mean velocity estimation for the movement along *y*-axis.

| Uncertainty indicators characterising estimates of mean velocity | Predefined walking velocity [m/s] | | | | | |
|---|---|---|---|---|---|---|
| | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 1.00 |
| Impulse-radar sensors | | | | | | |
| Mean discrepancy[m/s] | 0.02 | 0.03 | 0.03 | 0.02 | 0.01 | 0.00 |
| Root-mean-square discrepancy [m/s] | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.02 |
| Upper bound of the discrepancy [m/s] | 0.04 | 0.06 | 0.07 | 0.05 | 0.07 | 0.04 |
| Lower bound of the discrepancy [m/s] | –0.02 | –0.01 | –0.01 | –0.01 | –0.04 | –0.03 |
| Accelerometric sensors | | | | | | |
| Mean discrepancy [m/s] | –0.18 | 0.02 | 0.24 | 0.12 | 0.17 | 0.07 |
| Root-mean-square discrepancy [m/s] | 0.19 | 0.21 | 0.27 | 0.13 | 0.18 | 0.08 |
| Upper bound of the discrepancy [m/s] | –0.06 | 0.55 | 0.43 | 0.21 | 0.29 | 0.18 |
| Lower bound of the discrepancy [m/s] | –0.27 | –0.40 | –0.01 | –0.03 | 0.06 | –0.03 |

The experiments performed have demonstrated that impulse-radar sensors enable one to estimate the walking velocity more accurately than the accelerometric sensors. The estimates obtained on the basis of data from the latter sensors are affected by larger bias and are more widely spread around their mean values.

Since falls among elderly persons are the main cause of their admission and long-term stay in hospitals (Abbate et al., 2010), the systems for monitoring of elderly and disabled persons are expected to perform some functions related to fall prevention and/or fall detection. The functions related to fall prevention are implemented to overcome fall risk factors, implied by natural aging-related physical disabilities, and promptly indicate the increasing risk of falling; the functions related to fall detection are to reliably detect falls, when they occur, and minimise the potential injuries. Sensors used for fall prevention are expected to be accurate enough to enable the monitoring system to identify changes in the monitored person's health status on the basis of relatively slow and subtle changes in his/her gait characteristics, *e.g.* changes of the mean

walking velocity. Sensors used for fall detection should be selected and optimised with respect to their sensitivity as to enable the monitoring system to detect short abrupt changes in person's velocity or acceleration.

In light of the results presented in this paper, the impulse-radar sensors seem to be promising means for reliable fall prevention since they enable the through-the-wall monitoring of persons (as the electromagnetic waves propagate through non-metal objects) and highly accurate estimation of their velocity; those sensors are, however, less appropriate for fall detection because of the relatively low rate of data acquisition. On the other hand, the accelerometric sensors appear to be not well-suited for the long-term monitoring of the person's gait characteristics, but better satisfy the requirements related to fall detection, due to their higher sensitivity, significantly higher rate of data acquisition, and suitability for outdoor use.

One may thus conclude that both types of sensors studied in this paper, *viz.* impulse-radar sensors and accelerometric sensors, are in some way complementary, and therefore the combined use of both of them may contribute to the increase in the reliability of the monitoring of elderly and disabled persons.

## ACKNOWLEDGEMENTS

This work has been initiated within the project PL12-0001 financially supported by EEA Grants – Norway Grants (*http://eeagrants.org/project-portal/project/PL12-0001*), and finished within the statutory project supported by the Institute of Radioelectronics and Multimedia Technology, Faculty of Electronics and Information Technology, Warsaw University of Technology.

## REFERENCES

Abbate, S., Avvenuti, M., Corsini, P., Light, J. and Vecchio, A. 2010. Monitoring of Human Movements for Fall Detection and Activities Recognition in Elderly Care Using Wireless Sensor Network: a Survey. In Merrett, G. V. & Yen Kheng Tan (eds.) *Wireless Sensor Networks: Application – Centric Design.* Intech.

Baldewijns, G., Debard, G., Van Den Broeck, B., Mertens, M., Karsmakers, P., Croonenborghs, T. and Vanrumste, B. 2016a. Fall prevention and detection. In Florez-Revuelta, F. & Chaaraoui, A. A. (eds.) *Active and Assisted Living: Technologies and Applications.* Herts, UK: IET.

Baldewijns, G., Luca, S., Vanrumste, B. and Croonenborghs, T., 2016b. Developing a system that can automatically detect health changes using transfer times of older adults. *BMC Medical Research Methodology*, vol.**,** pp. 16-23.

Bang, W.-C., Chang, W., Kang, K.-H., Choi, E.-S., Potanin, A. and Kim, D.-Y., 2003. Self-contained Spatial Input Device for Wearable Computers. In *Proc. 7th IEEE International Symposium on Wearable Computers* (White Plains, NY, USA), pp. 26-34.

Brodie, M. A., Lord, S. R., Coppens, M. J., Annegarn, J. and Delbaere, K., 2015. Eight-Week Remote Monitoring Using a Freely Worn Device Reveals Unstable Gait Patterns in Older Fallers. *IEEE Transactions on Biomedical Engineering*, vol. 62**,** pp. 2588-2594.

Bulling, A., Blanke, U. and Schiele, B., 2014. A tutorial on human activity recognition using body-worn inertial sensors. *Computing Surveys*, vol. 46**,** pp. 33:1-33.

Buracchio, T., Dodge, H. H., Howieson, D., Wasserman, D. and Kaye, J., 2010. The trajectory of gait speed preceding mild cognitive impairment. *Archives of neurology*, vol. 67**,** pp. 980-986.

Cola, G., Vecchio, A. and Avvenuti, M., 2014. Improving the performance of fall detection systems through walk recognition. *Journal of Ambient Intelligence and Humanized Computing*, vol. 5**,** pp. 843-855.

Cuddihy, P. E., Yardibi, T., Legenzoff, Z. J., Liu, L., Phillips, C. E., Abbott, C., Galambos, C., Keller, J., Popescu, M. and Back, J., 2012. Radar walking speed measurements of seniors in their apartments: Technology for fall prevention. In *Proc. 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (San Diego, CA, USA), pp. 260-263.

Egerton, T., Thingstad, P. and Helbostad, J. L., 2014. Comparison of programs for determining temporal-spatial gait variables from instrumented walkway data: PKmas versus GAITRite. *BMC research notes*, vol. 7**,** pp. 1-7.

Fritz, S. and Lusardi, M., 2009. White Paper: "Walking Speed: the Sixth Vital Sign. *Journal of Geriatric Physical Therapy*, vol. 32**,** pp. 2–5.

Hamm, J., Money, A. G., Atwal, A. and Paraskevopoulos, I., 2016. Fall prevention intervention technologies: A conceptual framework and survey of the state of the art. *Journal of Biomedical Informatics*, vol. 59**,** pp. 319–345.

Jian, Q., Yang, J., Yu, Y., Björkholm, P. and Mckelvey, T., 2014. Detection of Breathing and Heartbeat by Using a Simple UWB Radar System. In *Proc. 8th European Conference on Antennas and Propagation* (The Hague, The Netherlands), pp. 3078-3081.

Liu, L., Popescu, M., Ho, K. C., Skubic, M. and Rantz, M., 2012. Doppler Radar Sensor Positioning in a Fall Detection System. *34th Annual International*

*Conference of the IEEE EMBS (San Diego, California USA, 28 Aug. – 1 Sep., 2012).*

Luque, R., Casilari, E., Morón, M.-J. and Redondo, G., 2014. Comparison and Characterization of Android-Based Fall Detection Systems. *Sensors*, vol. 14**,** pp. 18543-18574.

Lusardi, M., 2012. Is Walking Speed a Vital Sign? *Topics in Geriatric Rehabilitation*, vol. 28**,** pp. 67–76.

Miękina, A., Wagner, J., Mazurek, P. and Morawski, R. Z., 2016a. Selected algorithms for measurement data processing in impulse-radar-based system for monitoring of human movements. In *Proc. IMEKO TC1-TC7-TC13 Joint Symposium* (Berkeley, CA, USA), pp. 1-6.

Miękina, A., Wagner, J., Mazurek, P., Morawski, R. Z., Sudmann, T. T., Børsheim, I. T., Øvsthus, K., Jacobsen, F. F., Ciamulski, T. and Winiecki, W., 2016b. Development of software application dedicated to impulse-radar-based system for monitoring of human movements. In *Proc. IMEKO TC1-TC7-TC13 Joint Symposium* (Berkeley, CA, USA), pp. 1-6.

Morawski, R. Z., Yashchyshyn, Y., Brzyski, R., Jacobsen, F. and Winiecki, W., 2014. On applicability of impulse-radar sensors for monitoring of human movements. In *Proc. IMEKO TC-4 International Symposium* (Benevento, Italy), pp. 786–791.

Stone, E., Skubic, M., Rantz, M., Abbott, C. and Miller, S., 2015. Average in-home gait speed: Investigation of a new metric for mobility and fall risk assessment of elders. *Gait & posture*, vol. 41**,** pp. 57-62.

Studenski, S., Perera, S., Patel, K., Rosano, C., Faulkner, K., Inzitari, M., Brach, J., Chandler, J., Cawthon, P., Barrett Connor, E., Nevitt, M., Visser, M., Kritchevsky, S., Badinelli, S., Harris, T., Newman, A. B., Cauley, J., Ferrucci, L. and Guralnik, J., 2011. Gait Speed and Survival in Older Adults. *Journal of the American Medical Association* vol. 305**,** pp. 50-58.

Su, B. Y., Ho, K. C., Rantz, M. and Skubic, M., 2015. Doppler Radar Fall Activity Detection Using The Wavelet Transform. *IEEE Transactions on Biomedical Engineering*, vol. 62**,** pp. 865-875.

Thingstad, P., Egerton, T., Ihlen, E. F., Taraldsen, K., Moe-Nilssen, R. and Helbostad, J. L., 2015. Identification of gait domains and key gait variables following hip fracture. *BMC geriatrics*, vol. 15**,** pp. 1-7.

Thong, Y. K., Woolfson, M. S., Crowe, J. A., Hayes-Gill, B. R. and Jones, D. A., 2004. Numerical double integration of acceleration measurements in noise. *Measurement*, vol. 36**,** pp. 73–92.

Tomii, S. and Ohtsuki, T., 2012. Falling Detection Using Multiple Doppler Sensors. In *Proc. 14th IEEE International Conference on e-Health Networking, Applications and Services* (Beijing, China), pp. 196–201.

Wagner, J., Mazurek, P. and Morawski, R. Z., 2015. Regularised Differentiation of Measurement Data. In *Proc. XXI IMEKO World Congress "Measurement in Research and Industry"* (Prague, Czech Republic), pp. 1-6.

# Fear of Missing out, Social Media Engagement, Smartphone Addiction and Distraction: Moderating Role of Self-Help Mobile Apps-based Interventions in the Youth

Bobby Swar[1] and Tahir Hameed[2]
[1]*Concordia University of Edmonton, Edmonton, Canada*
[2]*SolBridge International School of Business, Daejeon, Republic of Korea*
*bobby.swar@concordia.ab.ca, tahir@solbridge.ac.kr*

Keywords: Basic Psychological Need Satisfaction, Fear of Missing out, Social Media Engagement, Smartphone Addiction, Smartphone Distraction, Self-Help Intervention.

Abstract: Smartphones offer high mobility and internet connectivity at the same time which has led to a substantial increase in the number of active social media users on the move, especially the 'Millennials'. The excessive use of smartphone has been linked with several issues including mental well-being. Recently, different mobile applications have emerged to help users track their excessive use of smartphones and protect them from potential risks to mental health. This paper uses self-determination theory to examine the moderating role of such mobile applications (or self-help interventions) on inter-relationships between social media engagement, smartphone addiction and smartphone distractions. Survey responses from 284 college students reveal that mobile applications could prove to be quite effective self-help interventions that can help the young people in self-regulating their smartphone use. These results have substantial implications for designing effective mobile app-based interventions to save young people from potential risks to their mental health, productivity, and safety in performing their daily tasks. Future research directions have also been pointed out.

## 1 INTRODUCTION

Social media users have grown exponentially in the past decade. There are now 2.34 billion social media users around the globe (*Number of social media users worldwide 2010-2020 | Statistic*, no date). Compared to the general population students attending colleges these days are the heaviest users of the social media (Alt, 2015). The high rate of smartphone penetration is believed to be one of the dominant driving force behind such increase in active social media users.

The advanced functionalities of the smartphone provide smartphone users with ubiquitous accessibility to the Internet transcending the limits of time and place, therefore enabling them to check social media updates in real time (Kim, Chun and Lee, 2014). The smartphone has become the first thing that people look at when they wake up in the morning and the last thing that they look at before going to sleep (Lee *et al.*, 2014). Smartphones have now become central to people's everyday lives (Gill, Kamath and Gill, 2012).

While there are several benefits of smartphones, they do not come without their issues. It has both positive and negative effects in people's daily routine, habits, social behaviors, emancipative values, family relations and social interactions (Samaha and Hawi, 2016). The excessive use of smartphones has also been linked to lot of negative outcomes like health and well-being (Lee et al. 2014; Park and Lee 2012; Samaha and Hawi 2016; Wang et al. 2015), student's academic performances (Duncan et al. 2012; Hawi and Samaha 2016; Li et al. 2015), distracted driving (Rocco and Sampaio 2016; Terry and Terry 2016) and smartphone addictions (Aljomaa et al. 2016; Chiu 2014; Gökçearslan et al. 2016; Haug et al. 2015). Among all the applications that smartphones provide, the use of social media is found to be a stronger predictor of smartphone addiction (Jeong *et al.*, 2016).

In order to prevent excessive smartphone usage its potential negative outcomes, a variety of mobile applications are available nowadays. These



Swar, B. and Hameed, T.
Fear of Missing out, Social Media Engagement, Smartphone Addiction and Distraction: Moderating Role of Self-Help Mobile Apps-based Interventions in the Youth.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 139-146
ISBN: 978-989-758-213-4

applications come with features like tracking the time spent on each of the mobile applications, turning off or restricting the (push-up) distracting notifications from certain applications and also restricting smartphone usage by locking or even turning it off after a specified time period. Most of the applications also provide reports and charts on the smartphone usage behavior. Using such applications is referred to as self-help intervention which helps the smartphone users control their excessive smartphone use and therefore protects them from its potential negative effects.

Research so far have identified and examined some negative effects of excessive social media and smartphone use. However, there are no studies, to the best of our knowledge, that examine whether mobile applications-based self-help interventions actually help in preventing excessive smartphone use, smartphone addiction and smartphone distractions. In this context, the objective of this study is to (1) examine the moderating role of self-help mobile applications-based interventions in preventing excessive smartphone use, smartphone addiction and smartphone distraction, and (2) examine the relationship of social media engagement on smartphone with smartphone addiction and smartphone distraction. To investigate these research objectives, a research model has been proposed and tested using the data collected from a large sample of university students from several countries.

# 2 THEORETICAL BACKGROUND AND HYPOTHESIS DEVELOPMENT

## 2.1 Basic Psychological Need Satisfaction

Self-determination theory (SDT), a macro-theory of human motivation, development, and wellness (Deci and Ryan, 2008) helps to explain the basic psychological need satisfaction of human beings. According to SDT, effective self-regulation and psychological health of human beings are based on satisfaction of their basic psychological needs for competence, autonomy and relatedness (Przybylski *et al.*, 2013). Competence refers to the individuals desire to feel effective in their interactions with the environment (Roca and Gagné, 2008). Autonomy refers to the individual desire to self-initiate and self-regulate own behavior (Sørebø *et al.*, 2009). Relatedness refers to the individual desire to feel connected and supported by others (Sørebø *et al.*, 2009).

Przybylski et al. (2013) suggested that individuals with low basic need satisfaction for competence, autonomy and relatedness have higher levels of Fear of Missing Out (FoMO). On this pretext, the following hypothesis is proposed.

H1: Basic psychological need satisfaction is positively associated with fear of missing out.

## 2.2 Fear of Missing out (FoMO)

FoMO phenomenon has been defined as a "pervasive apprehension that others might be having rewarding experiences from which one is absent, FoMO is characterized by the desire to stay continually connected with what others are doing" (Przybylski et al. 2013, p. 1841).

Research has explored the prevalence of FoMO and its relation to social media (JWT, 2012; Abel, Buff and Burr, 2016). Przybylski et al. (2013) is the first study that operationalizes the FoMO construct by collecting a diverse international sample of participants. A recent study by Alt (2015) in academic arena shows that there is a positive link between social media engagement and two motivational factors: Extrinsic and amotivation for learning are more likely to be mediated by FoMO.

FoMO plays an important role in individuals engaging in social media. Yin et al. (2015) indicate that FoMO and enjoyment are positively related to continuance intention of using social media. According to Przybylski *et al.*, (2013) individual with high levels of FoMO relate to higher levels of social media engagement.

The use of social media has been associated with greater levels of emotional support from close friends (Putnam 2000 as cited in Alt 2015). People with low basic need satisfaction generally perceive social media as a platform to connect with others in order to develop social competence, and an opportunity to deepen social ties (Przybylski *et al.*, 2013). This context leads to the proposal of the following hypothesis.

H2: Fear of missing out is positively associated with social media engagement.

## 2.3 Social Media Engagement

Social media refers to the websites and online tools that facilitate interactions between users by providing them opportunities to share information,

opinions, and interests (Khan, Swar and Lee, 2014). Social media engagement simply refers to the extent of an individual immersing into social media activities. People basically engage into social media by sharing personal or social information with close actors in social networks, such as family and friends (Alt, 2015).

Jeong et al. (2016) suggested that people who use smartphones for social media, games, and entertainment were more likely to be addicted to smartphones. The use of social networking mobile applications is a significant predictor of mobile addiction (Salehan and Negahban, 2013). These days, it has been much easier for individuals to engage in social media activities due to ubiquitous accessibility to the Internet through smartphones. This phenomenon leads to the formulation of the following hypothesis.

H3: Social media engagement is positively associated with smartphone addiction.

Social media engagement is also believed to play an important role in smartphone distraction. A user preoccupied with social media activities tends to be distracted from other primary tasks. This phenomenon of distraction can be explained with the concept of multitasking. Research in cognitive science shows that individuals performance decreases while multitasking (Junco, 2012).

Studies in an academic setting have also shown the negative relationship between the use of social media and academic performance. For example, according to Rosen, Mark Carrier and Cheever, (2013) those who use Facebook and text applications while studying had lower GPAs compared with the students who did not. This is clear evidence that excessive engagement in social media is associated with smartphone distraction and it leads to the proposal of the following hypothesis.

H4: Social media engagement is positively associated with smartphone distraction.

## 2.4 Smartphone Addiction

Smartphone addiction can be defined as the excessive use of smartphones in a way that is difficult to control and its influence extends to other areas of life in a negative way (Park and Lee, 2012).

Smartphones are a significant source of distraction for decision-based activities such as driving, classroom learning, and work-related tasks (Gill, Kamath and Gill, 2012). In academic settings, studies have linked the excessive smartphone use with the poor academic performance of the smartphone users by distracting them from their primary tasks. Duncan, Hoekstra and Wilcox (2012) have found a significant negative correlation between in-class phone use and final grades of the students. Similarly, Hawi and Samaha (2016) have reported that students who were at a high risk of smartphone addiction were less likely to achieve cumulative GPAs suitable for distinction or higher.

According to Gill, Kamath and Gill, (2012) smartphones are known to be detrimental to cognitive performance and the use of smartphones increases reaction time, reduces focus (attention), and lower performance of task needing mental concentration and decision making. Based on the observation the following hypothesis is proposed.

H5: Smartphone addiction is positively associated with smartphone distraction.

## 2.5 Self-help Intervention

Installation and use of mobile applications that could help the users in regulating their excessive smartphone use and therefore help in preventing smartphone addictions and smartphone distractions is referred to as self-help intervention in this study. There are a variety of such mobile applications available these days.

The mobile applications come with a variety of features such as blocking the websites one would want to avoid and periodic, customize setting on social media sites to give time-fixed updates (Andreassen, 2015), tracking the time spent on specific mobile applications, turning off or restricting the distracting (push-up) notifications from certain applications, and limiting the overall smartphone usage time by locking or even turning off the smartphone device.

Generally, self-motivated people use such mobile applications and with a belief in their ability and intention to prevent excessive smartphone use. This argument leads to the formulation of the following hypothesis.

H6: Self-help intervention has a moderating effect on the relationship between social media engagement, smartphone addiction and smartphone distraction.

# 3 RESEARCH MODEL

To examine the relationship between the use of social media through smartphone and its impact on smartphone addiction and distraction, and to

examine the moderating role of self-help intervention, a research model is proposed in Figure 1. The model shows that basic psychological need is a direct antecedent of fear of missing out which affects social media engagement, smartphone addiction, and smartphone distraction. Smartphone distraction is also directly influenced by social media engagement and smartphone addiction.

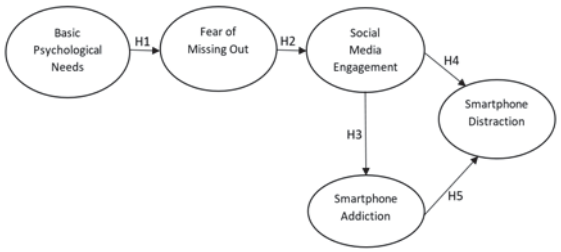



Figure 1: Research Model.

# 4 RESEARCH METHOD

The proposed research model was tested using the data collected through an online survey. The survey instruments were adapted from the existing measures to this research context. Each of the items was measured on a seven-point Likert-type scale.

The data was collected from university students from several countries. As shown in Table 1 majority of the respondents are 18 to 27 years old. According to Pew Research Center, ages 18 to 29 have always been the most likely users of social media by a considerable margin. Total of 284 useful responses were identified for the analysis in this study. The sample size of 284 should be adequate to test the research model against the required 50 samples for five paths in the research model according to Hair et al. (2011).

The respondents were categorized into two different groups based on whether or not they use self-help mobile applications in their smartphone to monitor and control their smartphone usage behavior. The respondents that use such kind of mobile applications are identified as “Self-help” group and the respondents that do not use such kind of self-help mobile application are categorized as “No-help intervention” group in this study.

Table 1: Sample demographics.

| Category | | Frequency | Percent (%) |
|---|---|---|---|
| Gender | Male | 130 | 45.77 |
| | Female | 154 | 54.23 |
| Age | <17 | 0 | 0 |
| | 18 - 22 | 205 | 72.18 |
| | 23 - 27 | 72 | 25.35 |
| | 28 - 32 | 5 | 1.76 |
| | >33 | 2 | 0.70 |
| Nationality | Bangladesh | 10 | 3.52 |
| | China | 71 | 25 |
| | Kazakhstan | 25 | 8.80 |
| | Russia | 17 | 5.98 |
| | South Korea | 113 | 39.79 |
| | Uzbekistan | 23 | 8.10 |
| | Others | 25 | 8.80 |
| Groups | Self-Help Intervention | 182 | 64.08 |
| | No Intervention | 102 | 35.92 |

# 5 RESULTS

## 5.1 Assessment of Measurement Model

This study utilizes structural equation modeling (SEM) supported by partial least squares (PLS) method to examine the research model and its hypotheses. The study in particular uses SmartPLS 3 software package for data analysis (Ringle, Wende and Becker, 2015).

Table 2 shows the assessment results of the measurement model for both the groups. Internal consistency reliability is investigated by using composite reliability. The constructs in the proposed model are above the 0.7 thresholds indicating a high reliability of items used for each construct.

Convergent validity is assessed by evaluating the average variance extracted (AVE) from the measures. The AVE is above the threshold value of 0.5, meeting the criteria of convergent validity.

Discriminant validity is assessed by examining the square root of AVE as recommended by (Fornell and Bookstein, 1982). Table 3 and 4 shows Fornell-Larcker tests of discriminant validity for no-help and self-help group respectively. Table 3 and 4 shows that the square root of AVE of each construct is greater than the correlations between it and all other constructs. Moreover, all the constructs are found to have a stronger correlation with their own measures than to those of others. This shows the proper assessment of discriminant validity.

Table 2: Assessment of the measurement model.

| Variables | Average Variance Extracted (AVE) | | Composite Reliability | |
|---|---|---|---|---|
| | No Intervention | Self-Help Intervention | No Intervention | Self-Help Intervention |
| Basic Psychological Needs | 0.65 | 0.68 | 0.96 | 0.96 |
| Fear of Missing Out | 0.55 | 0.55 | 0.90 | 0.89 |
| Social Media Engagement | 0.61 | 0.71 | 0.89 | 0.92 |
| Smartphone Addiction | 0.69 | 0.69 | 0.90 | 0.90 |
| Smartphone Distraction | 0.58 | 0.63 | 0.88 | 0.89 |

Table 3: Fornell-Larcker tests of discriminant validity for no-help group.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| (1) Basic Psychological Needs | **0.81** | | | | |
| (2) Fear of Missing Out | 0.40 | **0.74** | | | |
| (3) Social Media Engagement | 0.51 | 0.54 | **0.78** | | |
| (4) Smartphone Addiction | 0.14 | 0.48 | 0.42 | **0.83** | |
| (5) Smartphone Distraction | 0.34 | 0.50 | 0.51 | 0.66 | **0.76** |

Table 4: Fornell-Larcker test of discriminant validity for self-help group.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| (1) Basic Psychological Needs | **0.83** | | | | |
| (2) Fear of Missing Out | 0.52 | **0.74** | | | |
| (3) Social Media Engagement | 0.56 | 0.59 | **0.84** | | |
| (4) Smartphone Addiction | 0.37 | 0.52 | 0.54 | **0.83** | |
| (5) Smartphone Distraction | 0.46 | 0.48 | 0.47 | 0.56 | **0.79** |

Note: In Table 3 and table 4 the diagonal elements (in bold) represent the square root of AVE.

## 5.2 Testing the Model

Table 5 shows the results obtained from the PLS analysis. The coefficient of determination, $R^2$, is 0.499 for "No-help group" and 0.355 for "Self-help intervention" group respectively. The results show that the model explains a substantial amount of variance for smartphone distraction in both groups. As shown in Table 5 all the hypotheses are for both the groups are statistically significant.

Table 5: Results of the structural model with path coefficients.

| Path | Groups | |
|---|---|---|
| | No-help ($R^2$=0.499) | Self-help ($R^2$=0.355) |
| Basic Psychological Needs - Fear of Missing Out (β1) | 0.4 (4.66)** | 0.523 (8.823)** |
| Fear of Missing Out - Social Media Engagement (β2) | 0.544 (8.237)** | 0.594 (11.843)** |
| Social Media Engagement - Smartphone Addictions (β3) | 0.423 (4.811)** | 0.541 (9.057)** |
| Social Media Engagement - Smartphone Distractions (β4) | 0.278 (3.255)** | 0.231 (2.765)* |
| Smartphone Addictions - Smartphone Distractions (β5) | 0.542 (7.488)** | 0.438 (5.178)** |

Note: Associated t-statistics are in parentheses and *p<0.05, **p<0.01

Table 6 summarize the results of the multi-group analysis with their path coefficients. All the relationships (path coefficients) differ significantly across the two groups of smartphone users. These findings make intuitive sense considering that self-motivated people use self-help mobile applications to control their levels of social media engagement, smartphone addictions and smartphone distractions.

Table 6: Results of multi-group analysis with path coefficients.

| Paths | Path coefficients differences | P Value |
|---|---|---|
| Basic Psychological Needs - Fear of Missing Out (β1) | 0.119 | 0.873** |
| Fear of Missing Out - Social Media Engagement (β2) | 0.050 | 0.726** |
| Social Media Engagement - Smartphone Addictions (β3) | 0.119 | 0.873** |
| Social Media Engagement - Smartphone Distractions (β4) | 0.047 | 0.346** |
| Smartphone Addictions - Smartphone Distractions (β5) | 0.104 | 0.178* |

Note: *p<0.05, **p<0.01

# 6 DISCUSSION AND CONCLUSION

The primary objective of this study was to examine the moderating role of self-help mobile applications in regulating social media engagement, smartphone addictions and distractions. Additionally, it would examine the relationship between social media engagement on smartphone and smartphone addictions or smartphone distractions. To achieve the objective, this study established a path model and tested six hypotheses.

Table 7: Summary of results.

| Paths | Hypotheses | Results |
|---|---|---|
| Basic Psychological Needs - Fear of Missing Out | H1 | Supported |
| Fear of Missing Out - Social Media Engagement | H2 | Supported |
| Social Media Engagement - Smartphone Addictions | H3 | Supported |
| Social Media Engagement - Smartphone Distractions | H4 | Supported |
| Smartphone Addictions - Smartphone Distractions | H5 | Supported |
| Moderating role of self-help mobile applications | H6 | Supported |

The summary of the result is shown in Table 7. Empirical analysis of the research model provides several key findings, which are discussed below. The findings clearly show that self-help mobile applications can actually help in regulating social media engagement, smartphone addictions and distractions. This finding is very crucial as smartphone addictions and distractions have become somewhat a new illness in today's society. People needing help for their smartphone usage behavior can use self-help mobile applications to protect themselves from the negative effects of the smartphones. In academic settings, educators can recommend students needing help to use such self-help mobile applications.

This study also shows that people with low psychological needs have a high fear of missing out which leads to higher levels of social media engagement. The high levels of social media engagement then lead to smartphone addictions and smartphone distractions. The findings are in line with the emerging recent research (e.g., Przybylski *et al.*, 2013; Alt, 2015). The findings contribute to the existing literature by illustrating the mediating role of social media engagement in explaining smartphone addiction and smartphone distractions.

Nevertheless, there are some limitations of this study and the others that can provide opportunities for future research direction. This study does not take into considerations the types of (and underlying techniques) of mobile applications used as self-help interventions. These days such mobile applications come with a variety of features and capabilities that may have several different types of impact. Future research may take into consideration the types of self-help mobile apps-based interventions and their impact on smartphone use.

This study only deals with the smartphone users that are already using self-help mobile applications. It can be assumed that this group of users are self-motivated and are already less likely to get affected by the negative effects of the smartphones. Future research could ask smartphone users to install such self-help mobile applications and then examine the same effect after some period of time in a regulated environment.

This study has used the data from university students with the majority of respondents ranging from 17 to 27 years old. This group of people are among the heaviest users of social media and smartphones. Therefore, caution is needed while generalizing the results of this study. Different age groups can show different results. Future research is also recommended for other age group and settings.

## REFERENCES


Abel, J. P., Buff, C. L. and Burr, S. A. (2016) 'Social Media and the Fear of Missign Out: Scale Development and Assessment', *Journal of Business & Economics Research*, 14(1).

Aljomaa, S. S., Al.Qudah, M. F., Albursan, I. S., Bakhiet, S. F. and Abduljabbar, A. S. (2016) 'Smartphone addiction among university students in the light of some variables', *Computers in Human Behavior*, 61, pp. 155–164. doi: 10.1016/j.chb.2016.03.041.

Alt, D. (2015) 'College students' academic motivation, media engagement and fear of missing out', *Computers in Human Behavior*. Elsevier Ltd, 49, pp. 111–119. doi: 10.1016/j.chb.2015.02.057.

Andreassen, C. S. (2015) 'Online social network site addiction: A comprehensive review', *Current Addiction Reports*, pp. 175–184. doi: 10.1007/s40429-015-0056-9.

Chiu, S. I. (2014) 'The relationship between life stress and smartphone addiction on taiwanese university student: A mediation model of learning self-Efficacy and social self-Efficacy', *Computers in Human Behavior*. Elsevier Ltd, 34, pp. 49–57. doi: 10.1016/j.chb.2014.01.024.

Deci, E. L. and Ryan, R. M. (2008) 'Self-determination theory: A macrotheory of human motivation, development, and health.', *Canadian Psychology/Psychologie canadienne*, 49(3), pp. 182–185. doi: 10.1037/a0012801.

Duncan, D. K., Hoekstra, A. R. and Wilcox, B. R. (2012) 'Digital Devices, Distraction, and Student Performanc - Does In-Class Cell Phone Use Reduce Learning?', *Astronomy Education Review*, 11(1), pp. 1–4. doi: 10.1119/1.3317469.

Fornell, C. and Bookstein, F. L. (1982) 'Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory', *Journal of Marketing Research*, 19(4), p. 440. doi: 10.2307/3151718.

Gill, P. S., Kamath, A. and Gill, T. S. (2012) 'Distraction: An assessment of smartphone usage in health care work settings', *Risk Management and Healthcare Policy*, 5, pp. 105–114. doi: 10.2147/RMHP.S34813.

Gökçearslan, Ş., Mumcu, F. K., Haşlaman, T. and Çevik, Y. D. (2016) 'Modelling smartphone addiction: The role of smartphone usage, self-regulation, general self-efficacy and cyberloafing in university students', *Computers in Human Behavior*, 63, pp. 639–649. doi: 10.1016/j.chb.2016.05.091.

Hair, J. F., Ringle, C. M. and Sarstedt, M. (2011) 'PLS-SEM: Indeed a silver bullet', *Journal of Marketing Theory and Practice*. Taylor & Francis, 19(2), pp. 139–152.

Haug, S., Castro, R. P., Kwon, M., Filler, A., Kowatsch, T. and Schaub, M. P. (2015) 'Smartphone use and smartphone addiction among young people in Switzerland', *Journal of Behavioral Addictions*. Akademiai Kiado, 4(4), pp. 299–307. doi: 10.1556/2006.4.2015.037.

Hawi, N. S. and Samaha, M. (2016) 'To excel or not to excel: Strong evidence on the adverse effect of smartphone addiction on academic performance', *Computers & Education*, 98, pp. 81–89. doi: 10.1016/j.compedu.2016.03.007.

Hawi, N. S. and Samaha, M. (2016) 'To excel or not to excel: Strong evidence on the adverse effect of smartphone addiction on academic performance'. doi: 10.1016/j.compedu.2016.03.007.

Jeong, S.-H., Kim, H., Yum, J.-Y. and Hwang, Y. (2016) 'What type of content are smartphone users addicted to?: SNS vs. games', *Computers in Human Behavior*, 54, pp. 10–17. doi: 10.1016/j.chb.2015.07.035.

Junco, R. (2012) 'In-class multitasking and academic performance', *Computers in Human Behavior*, 28(6), pp. 2236–2243. doi: 10.1016/j.chb.2012.06.031.

JWT (2012) *Fear Of Missing Out (FOMO), March 2012 Update*. Available at: http://www.jwtintelligence.com/wp-content/uploads/2012/03/F _JWT_FOMO- update _3.21.12.pdf.

Khan, G. F., Swar, B. and Lee, S. K. (2014) 'Social Media Risks and Benefits: A Public Sector Perspective', *Social Science Computer Review*, 32(5), pp. 1–22. doi: 10.1177/0894439314524701.

Kim, D., Chun, H. and Lee, H. (2014) 'Determining the factors that influence college students' adoption of smartphones', *Journal of the Association for Information Science and Technology*, 65(3), pp. 578–588. doi: 10.1002/asi.22987.

Lee, Y. K., Chang, C. T., Lin, Y. and Cheng, Z. H. (2014) 'The dark side of smartphone usage: Psychological traits, compulsive behavior and technostress', *Computers in Human Behavior*, 31(1), pp. 373–383. doi: 10.1016/j.chb.2013.10.047.

Li, J., Lepp, A. and Barkley, J. E. (2015) 'Locus of control and cell phone use: Implications for sleep quality, academic performance, and subjective well-being', *Computers in Human Behavior*, 52, pp. 450–457. doi: 10.1016/j.chb.2015.06.021.

*Number of social media users worldwide 2010-2020 | Statistic* (no date). Available at: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/.

Park, N. and Lee, H. (2012) 'Social Implications of Smartphone Use: Korean College Students' Smartphone Use and Psychological Well-Being', *Cyberpsychology, Behavior, and Social Networking*. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA , 15(9), pp. 491–497. doi: 10.1089/cyber.2011.0580.

Przybylski, A. K., Murayama, K., Dehaan, C. R. and Gladwell, V. (2013) 'Motivational, emotional, and behavioral correlates of fear of missing out', *Computers in Human Behavior*. Elsevier Ltd, 29(4), pp. 1841–1848. doi: 10.1016/j.chb.2013.02.014.

Putnam, R. D. (2000) *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.

Ringle, C. M., Wende, S. and Becker, J.-M. (2015) *SmartPLS 3, Hamburg: SmartPLS*. Available at: http://www.smartpls.com.

Roca, J. C. and Gagné, M. (2008) 'Understanding e-learning continuance intention in the workplace: A self-determination theory perspective', *Computers in Human Behavior*, 24(4), pp. 1585–1604. doi: 10.1016/j.chb.2007.06.001.

Rocco, L. and Sampaio, B. (2016) 'Are handheld cell phone and texting bans really effective in reducing fatalities?', *Empirical Economics*. Springer Berlin Heidelberg, 51(2), pp. 853–876. doi: 10.1007/s00181-015-1018-8.

Rosen, L. D., Mark Carrier, L. and Cheever, N. A. (2013) 'Facebook and texting made me do it: Media-induced task-switching while studying', *Computers in Human Behavior*, 29(3), pp. 948–958. doi: 10.1016/j.chb.2012.12.001.

Salehan, M. and Negahban, A. (2013) 'Social networking on smartphones: When mobile phones become addictive', *Computers in Human Behavior*. Elsevier Ltd, 29(6), pp. 2632–2639. doi: 10.1016/j.chb.2013.07.003.

Samaha, M. and Hawi, N. S. (2016) 'Relationships among smartphone addiction, stress, academic performance, and satisfaction with life', *Computers in Human Behavior*, 57, pp. 321–325. doi: 10.1016/j.chb.2015.12.045.

Sørebø, Ø., Halvari, H., Gulli, V. F. and Kristiansen, R. (2009) 'The role of self-determination theory in explaining teachers' motivation to continue to use e-learning technology', *Computers and Education*, 53(4), pp. 1177–1187. doi: 10.1016/j.compedu.2009.06.001.

Terry, C. P. and Terry, D. L. (2016) 'Distracted Driving Among College Students: Perceived Risk Versus Reality', *Current Psychology*. Springer US, 35(1), pp. 115–120. doi: 10.1007/s12144-015-9373-3.

Wang, J.-L., Wang, H.-Z., Gaskin, J. and Wang, L.-H. (2015) 'The role of stress and motivation in problematic smartphone use among college students', *Computers in Human Behavior*, 53, pp. 181–188. doi: 10.1016/j.chb.2015.07.005.

Yin, F. S., Liu, M. L. and Lin, C. P. (2015) 'Forecasting the continuance intention of social networking sites: Assessing privacy risk and usefulness of technology', *Technological Forecasting and Social Change*. Elsevier Inc., 99, pp. 267–272. doi: 10.1016/j.techfore.2015.07.019.

# A Classification of Healthcare Social Network Analysis Applications

Lamia Benhiba, Asmae Loutfi and Mohammed Abdou Janati Idrissi
*TIME team, ENSIAS, Mohamed V University, Rabat, Morocco*
*{lamia.benhiba, a.janati}@um5s.net.ma, loutfi.asmae@gmail.com*

Keywords: Network Dynamics, Structural Analysis, Social Network Analysis, Healthcare Organization, E-health, Healthcare SNA Applications.

Abstract: As the web, social networks and the internet of things permeated our daily life; a new perspective for understanding the complexity of our interconnectedness has become necessary. One approach that has predominantly proven useful in discovering hidden relationships, connections and trends of complex systems through mathematical and graphical techniques is Social Network Analysis (SNA). This approach has become increasingly appeling for Healthcare in particular as many of this domain's problems examine systems with dynamic actors that interact with each other and exhibit emergent complex behaviors. However, due to their multiplicity, the application of SNA methodologies proves to be a complex and confusing endeavor. In an attempt to support the effort of applying SNA methodologies on Healthcare research problems, this paper offers firstly a categorization of SNA methodologies (structural and dynamic analysis), then inventories Healthcare SNA applications and classifies them into organizational and e-health related problems. The resulting categorization helps identify the Healthcare research problems most auspicious for SNA methodologies and should thus provide a guiding material of adequate SNA methodologies for a given Healthcare research problem.

## 1 INTRODUCTION

With the emergence of the web, online social networks, the internet of things etc., we are increasingly aware of our interconnectedness and its quantifiability. There is thus a growing realization that the behavior of a system is shaped by the interactions among its discrete components (Bullmore, 2009). Thereby, the study of the underlying network has become a stepping stone into understanding complex systems.

Social network analysis (SNA) has gained a lot of attention from both academia and practitioners of various domains (from social science (Lewis, 2008), economics (Krempel, 2002), politics (Klofstad, 2003), fight against crime and terrorism (Paulo, 2013), to neuroscience (Rubinov, 2010) and epidemiology (Chen, 2007)). SNA offers a new perspective for analysis and prediction as it focuses on the interconnectedness between the various constituents of the system and not on their inherent characteristics. It relies on Graph theory to express complex systems as a set of nodes (e.g. persons, organizations etc.) interconnected through social relationships (e.g. friendship, collaboration, transfer of funds, co-occurrence etc.). SNA aims to model, map, characterize and quantify topological properties of the network, identify patterns of relations and recognize the roles of sub-groups and nodes within it.

With the increasing availability of data and the advent and development of methods used to (a) collect, store and (b) visualize network data (Abraham, 2010), the interest in SNA has grown massively. Healthcare is among the chief domains where this particular approach is increasingly appealing. Many healthcare research problems examine systems with dynamic actors that interact with each other and exhibit emergent complex behavior. This makes these problems an auspicious application of SNA's methodologies and techniques.

The rest of the paper is organized as follows: Section II will introduce SNA and its underlying principles. It will also present the classification of the different SNA methodologies used throughout the literature into two main categories: structural and dynamics analysis. Section III will particularly focus on organizational healthcare and e-health SNA applications and then match them with the two SNA categories of section II. Section IV will summarize the results and enumerate different opportunities and challenges of the application of SNA in the



Benhiba, L., Loutfi, A. and Idrissi, M.
A Classification of Healthcare Social Network Analysis Applications.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 147-158
ISBN: 978-989-758-213-4

healthcare domains. The last section will conclude the paper with providing hints on future work.

# 2 SOCIAL NETWORK ANALYSIS METHODOLOGIES

SNA is an interdisciplinary descriptive, empirical discipline that studies networks as a mathematical representation of complex systems by expressing them in terms of relationships among actors. SNA has four features: 1) It is motivated by a structural intuition based on ties linking social actors, 2) It is grounded in systematic empirical data, 3) It draws heavily on graphic imagery, and 4) It relies on the use of mathematical and/or computational models (Freeman, 2004). The body of research has used SNA methodologies in various domains to help validate theories made about the structure or the behavior of a social construct or complex system. These methodologies can be categorized in several ways. No matter how limited and flawed the effort, doing so is useful because it guides the first steps when attempting to answer a specific research question.

We propose a categorization based on the purpose of the SNA analysis. A review of seminal works on SNA {(Wasserman, 1994), (Albert, 2002), (Barabási, 2002), (Newman, 2003), (Watts, 2004), (Christakis, 2011), (Scott, 2012), (Blonder, 2012), (Barabási, 2016)} has rendered two distinct purposes of network analysis:

- **Structural Analysis:** describes in discrete time snapshots the topology of the network, the roles of particular nodes, communities and subgroups within the network etc.
- **Dynamics Analysis:** studies the changes of the network's topology through time (its evolution and growth, the removal and adding of nodes and edges, the change in link weight etc.) and examines the diffusion of processes within the network.

## 2.1 Structural Analysis

Structural analysis aims to examine the topology of the network in order to uncover the overall properties of the network and its constituents' characteristics. It offers two perspectives: a micro-view and a macro-view. The micro or Ego-centric view focuses on a select actor (ego) and examines its neighbors (nodes that are connected to it), their neighbors and so forth. It studies the features of personal networks. The macro or Socio-centric view, on the other hand, provides a bird's eye perspective of the network and helps examine the structural patterns of the interactions among nodes with the aim to explain and potentially generalize an outcome. Studying the structure of a network relies on a number of measures. Because of their ability to give an indication on the topology of the network (random, small world or scale free), the most studied concepts in contemporary network research are: degree distribution, clustering and Assortativity.

The *degree of a node* is the number of links it has in the network and thus reflects the size of a node's neighborhood. The average degree has been used to gauge the cohesion (Kratzer, 2005) or connectedness on the network level (Shrader, 1989). The degree distribution is often plotted, using histograms, to obtain insight into the overall structure of the network and detect potential heavy-tailed distributions.

The *clustering coefficient* represents the tendency of nodes to form tightly knit groups within the network. It is measured on the node level and on the network level (Watts, 1998). The local Clustering coefficient of a node is used to quantify the level of transitivity within the network, i.e. the chance that a node u is connected to w, when u is connected to v and v is connected to w (uvw form a triangle). The Network Clustering coefficient on the other hand is defined as the average of the local clustering coefficients of all the nodes.

*Assortativity* detects the level of *homophily* in a network and measures the similarity of connections in the graph with respect to the node degree (McPherson, 2001). Assortativity can hint to the existence of a core-periphery structure where a set of closely knit nodes constitute the core of a network and low degree nodes are left on the periphery.

Along these core concepts, many studies have focused on community detection where algorithms are applied to uncover locally dense connected subgraphs (barabasi, 2016). Community detection allows a deeper understanding of the network's structure and hidden connectivity patterns.

## 2.2 Dynamics Analysis

The study of network dynamics refers to two distinct phenomena. We borrow the classification given by (Blonder, 2012) in which they distinguish the dynamics of the network from the dynamics on the network. The first examines the growth of the network, the factors behind the creation or dissolution of new nodes and edges and the

evolution of link strength through time. The second studies propagation phenomena and the transfer, throughout the network, of cascades such as information, trust, opinion, behavior, money, goods or pathogens etc.

### 2.2.1 Dynamics of the Network

The study of the dynamics of networks stems from the need to understand the rules of networks' growth in order to predict their evolution. Networks evolve by adding or removing nodes or links over time. Research on the evolution of networks focuses on the various dynamical processes that affect the change of the network's structure. The most popular evolving networks' models are Barabàsi and Albert's Preferential attachment (Barabàsi, 1999) and (McPherson, 2001)'s homophily model.

- In the preferential attachment model, nodes present a bias to connect to popular nodes that have a large number of connections. These hubs gain more connectivity as the network grows, following a rich-gets-richer model (Bollobás, 2003).
- Homophily represents the likeliness of nodes to connect to nodes that resemble them and which are generally the neighbors in the network. Nodes' connections are thus based on a conscious action with embedded bias (It's more likely for example to connect to a friend of a friend or an individual with common interests than it is to a random person).

While the main goal of these models is to predict the probability of link formation, enabling thus Link recommendations, nodes and links dissolution is another aspect of network evolution that is increasingly drawing interest. The goal here is to predict links that are more likely to be dropped from the network and to understand how it would affect the structure of the network.

### 2.2.2 Dynamics on the Network

In an attempt to understand the dynamic effect of network properties on diffusion, various studies relied on mathematical models originally used in fields such as epidemiology, sociology and economics. Louni et al. (Louni, 2014) classified the most popular information diffusion models into three categories:

- *Contagion Models*: these models build on the idea that a cascade flows in a network in the same way a contagious disease spreads through a population. The most widely used models for studying contagion are usceptible-infected (SI), susceptible-infected-susceptible (SIS) and susceptible-infected-recovered (SIR). The models consider cascades to spread from adopters (infected) to susceptible nodes and consider the possibility of retracting the cascade for recovered nodes.
- *Social Influence Models*: These models assume that the social influence between nodes affects the diffusion of cascades (opinions or behaviors for instance). The most widely studied and used social influence models are the Linear Threshold (LT) (Granovetter, 1978) and the Independent Cascade (IC) (Goldenberg, 2001).
- *Social Learning Models:* In contrast with previous models which ignore the actions and decision making of actors, the nodes in social learning models are considered rational agents who observe outcomes of prior behaviors and decide accordingly. The decision of a user to forward information is modeled using game theory concepts where the user maximizes some utility for himself (Jackson, 2008).

Studying the spread of cascades within a network offers theoretical and empirical tools to not only quantify the propagation process, but to forecast it as well.

## 3 A CATEGORIZATION OF HEALTHCARE SNA APPLICATIONS

Healthcare's purpose is to ensure the well-being of people by taking both proactive and active actions. Healthcare organizations take preventive actions like sharing information about healthy life styles, the vaccines in the market etc., providing psychological council, or conducting research for improving health services and the health life of people. They also take reactive actions by administrating drugs, doing surgery helping people with chronic illness etc.

Healthcare research covers a lot of areas such as clinical, biomedical, health systems and services and social, cultural, environmental and population health research. Healthcare research is undertaken to establish the foundation for developing effective therapeutic interventions to expose to individuals and communities, to support enhancing and understanding illness and health and safeguard and enhance the health of persons and populations (Steinwachs, 2008). Due to the complexity of the healthcare system, a methodological approach is needed to analyze, monitor and ensure the

effectiveness of its endeavors. SNA is thus introduced as a powerful new way to discover valuable hidden connections, relationships, trends and insights.

## 3.1 Methods and Materials

The purpose of this work is to establish a matching between SNA methodologies, described in section II, and healthcare application domains in order to uncover trends of SNA applications in the healthcare field. To accomplish this, we classified SNA applications in healthcare according to their functional domain and finally assigned SNA methodologies to each healthcare domain.

To identify the SNA applications in healthcare, we scanned three databases (Scopus, Science Direct and IEEExplore Digital Library) for the last 10 years, using various research terms related to: (social network analysis OR graph analytics) AND (healthcare, e-health, health organization, behavioral OR epidemiology) in Title, keywords and abstract. The search was restricted to English scientific literatures that are in peer-reviewed venues and duplicated works were eliminated. A paper is selected when the algorithm and methodology of social network analysis in e-health or healthcare organizations. In this paper, the 16 works listed in Table 1 will be considered. During the data extraction process, we included information about the title of the article, the year of publication, the authors, the country, the application of healthcare research, the data sources, the applied methodology and the type of the modeled graphs.

## 3.2 Categories of Healthcare SNA Applications

There are many areas of healthcare that can apply SNA. In this paper, we focus on the areas that drew the most attention: healthcare organization and e-health.

### 3.2.1 Healthcare Organization

A report of the Institute of Medicine suggested six aspects for improvement of the healthcare system. It needs to be: Safe (healthcare services to patients should be secure and not cause any injuries), Effective (care services based on scientific evidence for increasing healthy outcomes), Patient-centered (present to the patient the care service that respect their needs, values and preferences), Timely (provide care assistance early on before any complications occur), Efficient (make healthcare services available with minimum costs and without waste), Equitable (people should have the same access to healthcare services).

To achieve these purposes, healthcare organizations need to collaborate to share information about their operational and research works, establish policies for more effective and safe treatments and manage their waste by detecting fraud of healthcare providers.

#### i. Health policy

The World Health Organization (WHO) defined health policy as “the decisions, plans, and actions that are undertaken to achieve specific health care goals within a society”. The specification of rules that healthcare stakeholders should follow in terms of defining characters for differing groups, making a reference for treatments and actions undertaken by healthcare practitioners and sharing this information with people, are the various things attained by a health policy institution.

The study proposed by (Millard, 2015) adresses WHO's Essential Medicine List (EML). EML is a list of medicines that assists countries on selecting the treatments of each priority requirement. In this article, SNA is used to inspect the social, political and economic areas for adding the encouragement for Misoprostol’s use for preventing and treating a postpartum hemorrhage, especially in low income countries, according to the WHO's EML in 2011. A study the chronology of WHO misoprostol applications and evolution of related social networks are applied to evaluate the relation of health policy and this social area.

In (Takahashi, 2016), a descriptive analysis of duplicative prescription practices is performed. When patients take orders for the same state from two or more sources, we talk about duplicative prescription practices. This practice is the origin of medical waste. Some patients resell drugs for extra cash and can also cause adverse effects. The descriptive analysis was conducted by using the measurement of SNA and describing the prevalence (the rate of persons with an illness or characteristic) of duplicative prescription through ages. The study also calculated the density of the medical facilities and patients network for each class of drugs defined by their prevalence.

In (Bramhachari, 2016), the authors conducted a qualitative ego-network analysis to understand dominance of Rural Medical Practitioners (RMPs) in West Bengal, India. They inspected the genesis of RPMs’ social links with various actors in the health

system and showed the operators donating their subsistence over the years, by using SNA. By identifying the ties in RMPs' network that are formal healthcare providers, the healthcare market and the community, we can comprehend the dynamics of the healthcare market.

Guo et al. use healthcare claims data of the Medical Insurance Association of Anhui Province to look into details of referrals social network. They design a referral social network where the nodes are hospitals and ties are patient-transferred between hospitals. The authors conduct a structural analysis to measure the degree and centrality to describe the relationship between this variable and other patients and hospital variables. Finally, they explore rules between the variables of the referral social network and variables of quality of the healthcare to help healthcare providers minimize cost and length of stay in the hospital and increase the efficiency of medical resources (Guo, 2015).

## ii. Healthcare organizational collaboration

Healthcare organizations need to collaborate with each other in order to improve the quality of care to the patients, in term of efficient research, cost decrease, good management of resources etc.

**- Intra-organizational Collaboration (actors)**

Soulakis et al. made use of patients' Electronic Health Records (EHR) with heart failure to explore the collaboration between healthcare providers and patients in The Northwestern Memorial Hospital (NMH). The access to EHR provides a large amount of data about interactions between providers and patients (Soulakis, 2015). A structural SNA methodology is used to describe the collaboration between patients and providers through a bipartite network (the source node was a provider, the target node was a patient and the edge represented the patient record accessed by this provider), and a provider collaboration network which is a network of the common access of patient records by providers (the node represent providers and the edges are established when two providers have access to more than 10 common patient records). Data is extracted from the Enterprise Data Warehouse (EDW) of NMH. The network is afterwards visualized and clique formation is analyzed. A graph database is used to process queries and answer questions about care and provider-patient collaboration.

**- Inter-organizational Collaboration (institutions)**

Caniato et al. conducted a case study on management of healthcare waste in a region with specific characteristics: Gaza Strip (Caniato, 2015). They employ an SNA and stakeholder analysis to explore and comprehend the effects of a range of logistical and socio-economic factors on the effectiveness of stakeholder networks in the region. Caniato et al. applied a structural analysis of interaction frequency and information exchanged among stakeholders that are public authorities, health providers, supporting actors and others.

The study performed by (Schoen, 2014) used SNA to confirm the suggestion that when we take funding to concentrate on multi-sector collaboration in Social Innovation for Missouri (SIM) program, a public health program interventions to prevent obesity and stop tobacco, develop various partnership structures than other grantees. The authors explore different variables as the level of collaboration and frequency of contact by applying SNA. They measure the network descriptors such as average degree, density, betweeness centralization and degree centralization to evaluate the network of contacts and the collaboration of different stakeholders.

Dianas et al. studied an excellence program in low and middle-income countries provided by the National Heart, Lung, and Blood Institute-UnitedHealth to fund 11 centers of excellence (Dianis, 2016). In order to prove the effect of collaboration with a federal support, they used SNA. They created a network of the program's stakeholders by considering links as collaborations on administrative support and research projects. They later compared the resulting network before the development of the Centers of Excellence Program and after.

Kawonga et al. presented a case study of HIV monitoring and evaluation to examine and understand the way that Disease Control Program(DCP) and General Health Services(GHS) managers communicate when they make a health reform to make an administrative integration in South Africa (Kawonga, 2015). For this purpose, they described the entire network by using density, degree and betweenness centrality. They also used density and a measure of homophily to analyze sub-groups networks. A block-model analysis was used to identify the connections between management committees and manager groups.

The paper presented by Khosla et al. introduced a study of collaboration between HIV agencies in Baltimore (Khosla, 2016). SNA and relation coordination were used to analyze the quality of coordination between HIV agencies when they accessed resources like information, around seven

dimensions such as accuracy of communication, knowledge of agencies' work, frequency, problem-solving communication, timeliness, shared goals and mutual respect. Density and centrality of the network of agencies collaboration were calculated as part of an SNA structural analysis. For the study of relation coordination, a questionnaire was used among these seven dimensions about communication and relationships between HIV agencies. SNA measures were used to describe the whole network: density and degree centralization and to describe a position of an actor in the network: degree, indegree, centrality, degree centrality, weighted degree centrality, betweenness centrality and closeness centrality.

Wang et al. apply SNA to explore the collaboration between surgeons, assistants and anesthetist working at different hospitals by using data from Private Health Insurance (PHI) claims in Australia. They also studied their impact on quality and cost of care (Wang, 2014). SNA is used to analyze the collaboration among the three healthcare providers, study the topologies of the network to see how doctors work while treating patients and examine the effect of these topologies on quality and cost of care for patients. The effect of network structure on quality and cost is analyzed around efficiency metrics that are Length of Stay (LoS), Medical costs and Complication rate. They thereafter designed two kind of networks: one for collaboration between surgeons, assistant surgeons and anesthetists; and the second centered on a surgeon collaboration network to study the connections of each surgeon. The measures of SNA used in this paper are: the size of node (charged number of this provider), tie strength (total of common admission between two providers), centrality to have an idea on the influence of a vertex in the network) and density.

**- Research collaboration**

Collaboration research is important to enhance the quality of research by determining the leaders in a subject and affording reasonable proposals and scientific evidence to make a finance of specific area of research policy (Wu, 2015).

Bien et al. presented a case study of the use of SNA in the context of biomedical research grants collaboration at the University of Arkansas for Medical Sciences (UAMS) (Bian, 2013). The objective of this study was to evaluate the research collaboration networks (RCNs) for both level inter- and intra-institution in the community of the Clinical Translational Science Award (CTSA) and examine the effectiveness of CTSA funded at UAMS and their influence on environment of research collaboration in an institution. For categorizing the network, the authors calculated the network's path length and its clustering coefficient. They also measured the structural characteristics such as centrality to identify the important (the influencer or contributor) node in the research community. They examined the structural characteristics and the network dynamics of the RCNs.

Wu et al. (Wu, 2015) performed a study on the scientific research collaboration in the specialty of psychiatry. This work applied SNA to analyze the structure of scientific collaboration in psychiatry by using the notion of co-authorship, which can determine the authors, institutions and countries involved in the scientific collaboration network. In each level of authors, institutions and countries, the author characterized psychiatry research collaborative behaviors, K-plex analysis and Core-periphery are the methods used in this paper to describe the collaborative connections. The authors measure centrality to detect the central, the core position and actor with control and possession of valuable research resources in each collaboration network.

### 3.2.2 E-health

WHO defines e-Health as "the use of information and communication technologies (ICT) for health"(WHO, 2016). By using ICT in this area, we can assist patients for treatments; share information about healthy life styles, follow people with diseases etc.

The study presented by Chomutare et al. adresses weight loss performance by monitoring online interaction behaviors for forcast them (Chomutare, 2014). The authors captured data from a sub-forum of two online communities concerned with obesity. The first was for people older than 50 years and the second was for people that needed surgical interventions as they interacted before and after the intervention for weight loss performance. Structural SNA is used to create a classification of people who lose significant weight (performers) and the others (non-performers) to predict weight loss. Authors remarked that the top performers were connected at different sub-community and were more active online.

Pachucki et al. measure objectively the social interaction between 6th-grade students at a private K-8 School in the State of California by using accelometers and RFID technology. The purpose of this paper is to study the relations between social interaction and mental health behaviors such as self-

esteem and depressive symptoms of early adolescences. Due the focus on health behaviors, health status and changes in network structure; the authors measure the characteristics of social environment, health behavior, social interaction network and mental health to analyze them using bivariate associations. They use a stochastic actor-based modeling (SABM) framework to join the dynamic co-evolution of social ties and self-esteem or depressive symptoms (Pachucki, 2015).

Goodall et al. explored the importance of ICT for searching information behavior by older migrants with Culturally And Linguistically Diversity (CALD) (Goodall, 2014). They determined factors that leverage the use of ICT to locate information. These factors can be education, migration, socio-economic status, ethnicity and English proficiency of older migrants. The study undertook by Goodall et al. focused on the search of cancer-related information by the group. The authors used SNA and a constructivist grounded theory method to analyze the data captured in the interviews, and then they studied the preferences and uses of traditional information sources compared to modern ones (PC, Internet and mobiles).

Table 1: A methodological classification of each healthcare application represented in this paper.

| **Reference of paper and the country** | **The SNA application** | **Healthcare categorization** | **Methodological categorization** | **Dataset/Size of the network** | **Algorithms/Metrics** |
|---|---|---|---|---|---|
| (Millard, 2015), UK | They establish a chronology of WHO misoprostol applications and they examine the evolution of related social networks and the nested subset network of the WHO EML misoprostol applications | Healthcare organization: Health policy | Dynamics of the network | 238 organizations and individuals | Chronological approach combined with SNA (evolution of social network) : density, geodesic distance, diameter, centrality, nesting, clique formation |
| (Takahashi, 2016), Japan | They conduct a descriptive analysis of medical waste by studying duplicative prescription practices | Healthcare organization: Health policy | Structural | Data are from health insurance claims database 1,243,058 insured people and their dependents | Statistical analyses: correlation by using scatter plots and the Pearson correlation coefficient SNA: bipartite networks, density |
| (Bramhachari, 2016), India | They use a qualitative ego-network method to understand the RMP network | Healthcare organization: Health policy | Structural | 35 participants | Qualitative Ego-network method |
| (Guo, 2015), China | They analyze the healthcare claims data of the Medical Insurance Association of Anhui Province to design a referral social network. | Healthcare organization: Health policy | Structural | 72 hospitals and 8856 patients in the claim data from Medical Insurance Bureau | Community detection: spinglass, edge betweenneess, label propagation, optimal, walktrap. Simple linear regression: Los, Medical cost, Degree, closeness centrality, betweenness centrality, eigenvector centrality, rank of Hospital. Rules exploration: Decision tree. |

Table 1: A methodological classification of each healthcare application represented in this paper (cont.).

| | | | | | |
|---|---|---|---|---|---|
| (Soulakis, 2015), USA | They make a bipartite network of providers accessing patients' records and a provider collaboration network to describe collaboration between patients and providers. | Healthcare collaboration intra-organization | Structural | Collaborative electronic health record (HER) 1504 nodes and 83 998 edges | Bipartite network Module and clique identification: heuristic community detection algorithm, kCliques algorithm |
| (Caniato, 2015), Italy | They conduct a Stakeholder analysis and structural SNA of interaction frequency and information exchanged between stakeholders | Healthcare collaboration inter-organization | Structural | Dataset constructed from 16 structured and two semi-structured interviews | SNA and stakeholder analysis |
| (Schoen, 2014), USA | They apply structural SNA to both contact and collaboration networks | Healthcare collaboration inter-organization | Structural | 23 Missouri communities in early 2012 | SNA: average degree, density, degree centralization, and betweenness centralization |
| (Dianis, 2016),USA | They conduct structural SNA on the network of all stakeholders in an excellence program | Healthcare collaboration inter-organization | Structural | 11 contracts in 10 countries 128 nodes | SNA: density, average distance |
| (Kawonga, 2015), South Africa | They apply structural and dynamic methodologies on the communication network of GHS and DCP managers | Healthcare collaboration inter-organization | Dynamics of the network | 51 managers in two provinces during 2010-2011 Dataset: HIV data collation and HIV data use | SNA: density, degree, betweenness centrality and E-Index (measure of homophily) Block modelling |
| (Khosla, 2016), USA | They combine SNA and relational coordination to measure the quality of coordination among HIV agencies | Healthcare collaboration inter-organization | Structural | 57 agencies | SNA: density, degree centralization, weighted degree centralization, closeness centrality, betweenness centrality Relational coordination: frequency, timeliness and accuracy of communication, problem-solving communication, knowledge of agencies' work, mutual respect and shared goals |
| (Wang, 2014), Australia | They use SNA to explore the collaborative network and surgeon centric collaboration network to analyze the impact of collaboration on the quality and cost of care | Healthcare collaboration inter-organization | Structural | Health insurance claims: 59256 admissions performed by 870 surgeons | SNA: degree centrality, closeness centrality, betweenness centrality , density, clustering coefficient, number of triangles |

Table 1: A methodological classification of each healthcare application represented in this paper (cont.).

| | | | | | |
|---|---|---|---|---|---|
| (Bian, 2013), Arkansas, United States | They apply structural and dynamic methodologies on the network to identify leaders and influencers in a research collaboration network | Healthcare research collaboration | Dynamics of the network | The Automated Research Information Administrator (ARIA) and the Translational Research Institute (TRI) | SNA: measures of centrality, mean path length, clustering coefficient, characteristic path length, diversity Temporal evolution: average number of new edges, centrality leaders |
| (Wu, 2015), China | They use the measure of centrality to conduct an SNA on authors, institutions and countries collaborating on psychiatric research | Healthcare research collaboration | Structural | 36557 papers about psychiatry from Science Ciation Index Expanded (SCI-Expanded) in web of science | SNA: centrality, K-plex analysis, Core periphery Hierarchical clustering |
| (Goodall, 2014), Australia | They use Grounded theory and a qualitative SNA on the egocentric network of individuals and their sources of information, and compare the resulting networks | E-health: Information technology access | Structural | Interview with 54 participants aged 63–94 years | Constructivist grounded theory method (CGTM) SNA: egocentric network |
| (Chomutare, 2014), Norway | They use structural SNA to classify people according to their ability to lose weight significantly | E-health: Social influence and behavior analysis | Structural | | Binomial classification: Bayes and decision tree method SNA: bipartite graph , degree centrality, betweenness centrality Expansion-reduction method Community detection: hierarchical clustering |
| (Pachucki, 2015) USA | They measure the association between social interactions and depressive symptoms and self-esteem of early adolescences at a private K-8 School in the State of California | E-health: Behavioral analysis | Dynamics on the network | 40 students of sixth-graders at a private K-8 school | Measures of Social environment, Health behaviors ( physical activity and food choice), Social interaction networks, dependent variables ( self-esteem and depressive symptoms) to make bivariate associations and SNA ( size of personal networks, transitivity and closeness centrality) |

## 4 SYNTHESIS AND DISCUSSION

The following table represents the matching between the categorization of healthcare SNA applications and the SNA methodologies.

Figure 1 shows that the highest number of the

included research works studied healthcare collaboration and focused on collaboration between institutions. This might be due to the accessibility of inter-collaboration data compared to intra-collaboration data (insurance claims vs. EHRs). E-health is a new area of SNA application and can thus present new opportunities to researchers, although the lack of data especially in low-income countries may be problematic.

Structural SNA methodologies are the most used, whereas dynamic methodologies are mainly used in problems related to healthcare organization (cf. Table 2). The prevalence of structural analyses could be due to the complexity of dynamic methodologies compared to structural ones. Associations between SNA structural metrics and domain-specific metrics are however rarely examined; which constitutes a research question that deserves further attention. There is also a pressing need to move beyond the static view of the network, visualized in snapshots, to a visualization that captures more accurately the dynamic processes that reshape the network (a movie-like visualization for instance). Another research opportunity relates to the application of dynamics on the network methodologies such as propagation and diffusion models. While these methodologies have been exclusively used in behavioral analysis, they have the potential to examine the propagation of information in social networks and uncover hidden processes shaping collaboration or policy making endeavors.

With respect to data collection, a third of the included studies gathers data from questionnaires. This raises data completion issues and inaccuracies arising from informant bias and stresses the pertinence of alternative data collection methods relying on RFID technologies, accelerometers or EHRs etc. Online social networks (OSN) such as Facebook, Twitter etc. are widely used nowadays and can help perform social behavioral analyses. We can comprehend e-health tools and design future IPC to promote effective interaction behaviors in OSNs by correlating interaction behaviors and a specific disease (Chomutare, 2014). However, many challenges could face such studies. When SNA methodologies are used on OSNs' social data, it is difficult to distinguish between the effects of Homophily and those of peer influence. Questions relating to the sufficiency of collected data and its representativeness of a given behavior to infer conclusions remain unanswered. There are also the pressing issues of privacy and ethics regarding data collection and which are consequences of the inherent processes of the social graph's construction and design. Anonymization, consent and privacy are among the issues that need further attention.

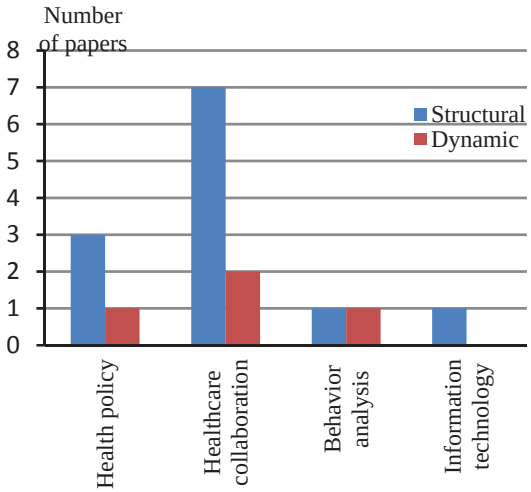



Figure 1: Number of papers of each healthcare application and methodological classification (Dynamic, Structural).

Table 2: The SNA methodology used in each healthcare domain.

| Healthcare categorization | Functional Sub-categorizing | Methodological categorization |
|---|---|---|
| Healthcare organization | Health policy | Dynamics of the network, Structural |
| | Healthcare collaboration intra-organization | Structural |
| | Healthcare collaboration inter-organization | Structural, Dynamics of the network |
| | Healthcare research collaboration | Dynamics of the network, Structural |
| E-heath | Information technology access | Structural |
| | Social influence and behavior analysis | Structural, Dynamics on the network |

Furthermore, social media provides a large amount of data. The resulting networks are thus very large and new tools are needed to process them. Big data network analysis is increasingly drawing the attention of researchers. However, due to the complexity of healthcare research problems, further

research is needed in order to produce domain-specific tools.

## 5 CONCLUSIONS

The purpose of this paper was to propose a classification of healthcare SNA applications based on a review of papers that used structural and dynamic SNA methodologies to answer healthcare-related research problems. We classified these research works into two categories: One concerning healthcare organizations and pertaining to policy making, communication, and collaboration and a patient-oriented category which concerns patients' behaviors, social influence and healthcare information access.

The proposed classification of healthcare SNA applications is preliminary and requires further enrichment through the inclusion of other research works. The level of adequacy of a chosen SNA methodology to a given Healthcare research problem is yet to be examined. Experimental studies will have to be conducted to establish comparative analyses between variations of a given methodology for a particular problem. For instance, different subsets of metrics can be used and compared for structural SNA methodologies, various propagation models can be simultaneously tried for dynamic SNA methodologies.

## REFERENCES


Bullmore, E., & Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience,10(*3), 186-198.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N., 2008. Tastes, ties, and time: A new social network dataset using Facebook. com.*Social networks, 30(4*), 330-342.

Krempel, L., & Plümper, T., 2003. Exploring the dynamics of international trade by combining the comparative advantages of multivariate statistics and network visualizations. *Journal of social structure, 4(1),* 1-22.

Klofstad, C. A., Sokhey, A. E., & McClurg, S. D., 2013. Disagreeing about disagreement: How conflict in social networks affects political behavior.American *Journal of Political Science, 57(1),* 120-134.

Paulo, D., Fischl, B., Markow, T., Martin, M., & Shakarian, P., 2013. Social network intelligence analysis to combat street gang violence. *arXiv preprint arXiv:1306.6834.*

Rubinov, M., & Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage, 52(3),* 1059-1069.

Chen, Y. D., Tseng, C., King, C. C., Wu, T. S. J., & Chen, H., 2007. Incorporating geographical contacts into social network analysis for contact tracing in epidemiology: A study on Taiwan SARS data. In *Intelligence and Security Informatics: Biosurveillance (pp. 23-36).* Springer Berlin Heidelberg.

Abraham, A., Hassanien, A. E., & Snâaésel, V. (Eds.)., 2010. Computational social network analysis: trends, tools and research advances. *Springer*.

Freeman, L. C., 2004. The development of social network analysis. *Vancouver: Empirical Press*.

Wasserman, S., 1994. Social network analysis: Methods and applications (Vol. 8). *Cambridge university press*.

Albert, Réka, and Albert-László Barabási, 2002. "Statistical mechanics of complex networks." *Reviews of modern physics 74.1 : 47*.

A.-L. Barabási. *Linked: The New Science of Networks*. Perseus Books Group, 2002.

Newman, Mark EJ, 2003. "The structure and function of complex networks." *SIAM review 45.2* : 167-256.

D. Watts. *Six Degrees: The Science of a Connected Age*. Norton, 2004.

N. Christakis and J. Fowler. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives.* Back Bay Books, 2011. (Scott, 2012).

Scott, John. *Social network analysis*. Sage, 2012.

Blonder B, Wey T, Dornhaus A, James R, Sih A. 2012. Te mporal dynamics and network analysis. *Meth Ecol Evol*. doi: 10.1111/j.2041-210X.2012.00236.x.

Barabási, Albert-László. "Network science" *Cambridge University Press, Cambridge* (August 5, 2016).

Kratzer, J., Leenders, R. T. A., & Van Engelen, J. M., 2005. Informal contacts and performance in innovation teams. *International Journal of Manpower*, 26(6), 513-528.

Shrader, C. B., Dellva, W. L., & McElroy, J. C., 1989. Social networks and group performance – A partial test of the Pearce and David hypotheses. *International Journal of Small Group Research, 5(2),* 145-162.

Watts, Duncan J.; Strogatz, Steven H., June 1998. "Collective dynamics of 'small-world' networks". *Nature 393 (6684):* 440–442.

McPherson, M., Smith-Lovin, L., & Cook, J. M., 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415-444.

Barabási, A. L., & Albert, R., 1999. Emergence of scaling in random *networks.science, 286(5439),* 509-512.

Bollobás, B., & Riordan, O., 2003. Mathematical results on scale-free random graphs. *Handbook of graphs and networks*, 1, 34.

Louni, Alireza, and K. P. Subbalakshmi, 2014. "Diffusion of information in social networks." *Social Networking. Springer International Publishing.* 1-22.

Bollobás, B., & Riordan, O., 2003. Mathematical results on scale-free random graphs. *Handbook of graphs and networks*, 1, 34.

Louni, Alireza, and K. P. Subbalakshmi, 2014. "Diffusion of information in social networks." *Social Networking. Springer International Publishing*. 1-22.

Granovetter, Mark, 1978. "Threshold models of collective behavior." *American journal of sociology*: 1420-1443.

Goldenberg, J., Libai, B., Muller, E, 2001.: *Talk of the network: a complex systems look at the underlying process of word-of-mouth.* Mark. Lett. 12(3), 211–223.

Jackson,M.O, 2008.: Social and Economic Networks. *Princeton University Press.*

Steinwachs DM, Hughes RG, 2008. *Patient Safety and Quality: An Evidence-Based Handbook for Nurses*, Hughes RG, editor.

Millard, Colin, Petra Brhlikova, and Allyson Pollock. 2015. 'Social networks and health policy: The case of misoprostol and the WHO model essential medicine list', *Social Science & Medicine*, 132: 190-96.

Takahashi, Yoshimitsu, Tatsuro Ishizaki, Takeo Nakayama, and Ichiro Kawachi. 2016. 'Social network analysis of duplicative prescriptions: One-month analysis of medical facilities in Japan', *Health Policy*, 120: 334-41.

Bramhachari, Rittika, and Sabyasachi Mandal. 2016. 'a social network analysis of rural medical practitioners in the sundarbans, west bengal', *BMJ Global Health*, 1: A41-A42.

Guo H., Wei F., Cheng S., Jiang F., 2015. Find Referral Social Networks, *International Symposium on Security and Privacy in Social Networks and Big Data*, 58-63.

Soulakis, Nicholas D., Matthew B. Carson, Young Ji Lee, Daniel H. Schneider, Connor T. Skeehan, and Denise M. Scholtens. 2015. 'Visualizing collaborative electronic health record usage for hospitalized patients with heart failure', *Journal of the American Medical Informatics Association : JAMIA*, 22: 299-311.

Caniato, M., T. Tudor, and M. Vaccari. 2015. 'Understanding the perceptions, roles and interactions of stakeholder networks managing health-care waste: A case study of the Gaza Strip', *Waste Management*, 35: 255-64.

Schoen, Martin W., Sarah Moreland-Russell, Kim Prewitt, and Bobbi J. Carothers. 2014. 'Social network analysis of public health programs to measure partnership', *Social Science & Medicine*, 123: 90-95.

Dianis, Nancy L., Tracy L. Wolbach, and Maura Spiegelman. 2016. 'The NHLBI–UnitedHealth Global Health Centers of Excellence Program: Assessment of Impact of Federal Funding Through a Social Network Analysis', *Global Heart*, 11: 145-48.e1.

Kawonga, Mary, Duane Blaauw, and Sharon Fonn. 2015. 'Exploring the use of social network analysis to measure communication between disease programme and district managers at sub-national level in South Africa', *Social Science & Medicine*, 135: 1-14.

Khosla, Nidhi, Jill Ann Marsteller, Yea Jen Hsu, and David L. Elliott. 2016. 'Analysing collaboration among HIV agencies through combining network theory and relational coordination', *Social Science & Medicine*, 150: 85-94.

Wang, F., U. Srinivasan, S. Uddin, and S. Chawla. 2014. "Application of network analysis on healthcare." *In Advances in Social Networks Analysis and Mining (ASONAM)*, 596-603.

Wu, Y., and Z. Duan, 2015. 'Social network analysis of international scientific collaboration on psychiatry research', *International Journal of Mental Health Systems.*

Bian, J., M. Xie, U. Topaloglu, T. Hudson, and W. Hogan. 2013. "Understanding biomedicai research collaborations through social network analysis: A case study." *In Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference*, 9-16.

World health organization 2016. Health topics, e-health, url: http://www.who.int/topics/ehealth/en/

Chomutare, T., A. Xu, and M. S. Iyengar. 2014. "Social Network Analysis to Delineate Interaction Patterns That Predict Weight Loss Performance." *In 2014 IEEE 27th International Symposium on Computer-Based Medical Systems*, 271-76.

Pachucki M. C., Ozer E. J., Barrat A., Cattuto C., 2015. Mental health and social networks in early adolescence: A dynamic study of objectively-measured social interaction behaviors. *Social Science and Medecine 125*, 40-50.

Goodall, K. T., L. A. Newman, and P. R. Ward. 2014. 'Improving access to health information for older migrants by using grounded theory and social network analysis to understand their information behaviour and digital technology use', *European Journal of Cancer Care*, 23: 728-38.

# Understanding Jump Landing as an Oscillating System: A Model-based Approach of Balance and Strength Analyses

Sandra Hellmers[1], Sebastian Fudickar[1], Lena Dasenbrock[2], Andrea Heinks[2], Jürgen M. Bauer [3] and Andreas Hein[1]

[1]*Assistive Systems and Medical Technologies, Carl von Ossietzky University Oldenburg, Oldenburg, Germany*

[2]*Department of Geriatric Medicine, Carl von Ossietzky University Oldenburg, Oldenburg, Germany*

[3]*Chair of Geriatric Medicine, Heidelberg University, Agaplesion Bethanien Hospital Heidelberg, Heidelberg, Germany*

{*sandra.hellmers, sebastian.fudickar*}*@uni-oldenburg.de*

Keywords: Force Platform, Counter Movement Jump, Vertical Jump, Balance, Strength, Geriatric Assessment, Jump Landing, DPSI.

Abstract: Counter movement jumps (CMJ) are well-suited to measure the muscle power and balance. Since it has been clarified that well accepted CMJ amplification-based balance measures (such as TTS or CoP) are significantly influenced by algorithmic, and measurement settings and thus, measurement results have limited meaningfulness among force platforms, we introduce a new model-based approach measuring the postural stability. In this, during the landing and recovery phases after vertical jumps, the lower extremities can be represented by an oscillating system and the corresponding transfer function is described by a second-order delay (PT2) element.

In an initial prospective study with 20 subjects aged over 70 years, we observed an inverse relationship between the calculated parameter *w* and the jump height and could also identify an influence of sex, and body weight on the jump height. Furthermore, we also found a relation between the parameter *w* and the dynamic postural stability index (DPSI), even though these results must be ensured statistically using a larger cohort, due to the current limited number of subjects.

Nevertheless, we could confirm the general applicability of the Systems and Control Technology perspective on describing human movements in a potentially more robust manner than current amplification based approaches. Further investigations on our model and the oscillating behavior in the phase of landing are needed to improve our system and to interpret the calculated parameters in a technical and physiological point of view.

## 1 INTRODUCTION

Geriatric assessments are well-established instruments to identify early changes associated with functional and cognitive decline, as they can occur in common geriatric syndromes, such as frailty or sarcopenia (Clegg et al., 2013; Cruz-Jentoft et al., 2010; Elsawy and Higgins, 2011). Thus, the assessments gain increasing relevance with the ongoing age-related demographic shift. Therefore, it exists a strong research interest to identify degrading abilities very early in geriatric assessments or with technical monitoring systems (Hein et al., 2010; Fudickar et al., 2012), like for example with systems in domestic environments to identify changes in the user behavior (Steen et al., 2013), to trigger preventive measures.

Nevertheless, for a self-determined healthy life and low fall risk, functional abilities and physical fitness are fundamental for healthy aging. Muscular strength of the lower extremities, balance, and endurance are essential factors (Granacher et al., 2013) for the fall risk, frailty, and sarcopenia.

Due to the relevance of muscular strength of the lower extremities, postural stability, and endurance, these factors are covered by various standardized assessments and tests (see Table 1). Most of these assessments and tests consist of several assessment items. For example, the Short Physical Performance Battery (SPPB) consists of a walk test, a static balance test, and the chair rising test and can cover strength and balance only in the combination of the assessment items. Consequently, among the common assessments, only the Counter Movement Jump (CMJ) is well-suited to test both components, strength, and balance within a single item. In detail, the CMJ allows to measure postural stability (balance) (Granacher et al.,



Hellmers, S., Fudickar, S., Dasenbrock, L., Heinks, A., Bauer, J. and Hein, A.
Understanding Jump Landing as an Oscillating System: A Model-based Approach of Balance and Strength Analyses.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 159-168
ISBN: 978-989-758-213-4

Table 1: Selection of assessment items in our geriatric study (Hellmers et al., 2016), their test duration, and classification regarding the components of physical fitness (- none, + to +++ increasing significance). The test durations are based on literature and estimated on own experiences (*) in a study with 250 participants. The values in brackets are the durations with introduction and instructions or a test jump.

| Geriatric Assessments | Balance | Strength | Endurance | Test Duration |
|---|---|---|---|---|
| de Morton Mobility Index | ++ | - | - | 9 min |
| -static balance | ++ | - | - | |
| -dynamic balance | ++ | - | - | |
| Short Physical Performance Battery | ++ | ++ | - | 15 min |
| -static balance | ++ | - | - | |
| -chair rise test | - | ++ | - | |
| Frailty Index | - | ++ | - | 10-17 min* |
| -grip strength | - | ++ | - | 3-5 min |
| Stair Climb Power Test | - | ++ | - | 2-(5) min* |
| 6 Minute Walk Test | - | + | ++ | 6 min |
| Counter Movement Jump | ++ | ++ | - | 5-(6) min* |

2013) via the time to stability (TTS) during the landing and stabilization phase and muscle performance of the lower extremities (strength) via muscle power ahead of jumps (Buehring et al., 2015; Rittweger et al., 2004; Dietzel et al., 2015; Kalyani et al., 2014).

The assessment of patients' functional status for balance and muscle strength through a single test item instead of several tests lowers costs and personal efforts (supporting an increased assessment density) and can reduce stress and potential fatigue for patients, which hold the risk that assessments results lose significance (Siglinsky et al., 2015). Consequently, the CMJ is a well-suited assessment item to cover both muscle strength of the lower extremities and postural stability.

However, since relying just on the force amplification, current CMJ-based balance measures (such as COP and TTS) have been shown to be significantly prone to algorithmic and technical variations thus have limited viability regarding the generalization of classification and measurements.

Thus, we propose a robust approach to measure postural stability based on the natural frequencies during the landing and recovery phase of CMJs and evaluate its practicability for 20 subjects with an age of 71 to 82 years.

# 2 COUNTER MOVEMENT JUMP

Within this section, the characteristics of CMJs are discussed and are followed by a description of the technically supported measurement of ground reaction forces.

## 2.1 Biomechanical Characteristics

Counter movement jumps (CMJs) are vertical jumps that are performed from standing, and according to (Palma et al., 2008) consist of the following phases (as shown in Figure 1): In the first phase (a) the participant is standing. Phase (b) is characterized by the preparation (b) with a downward movement by the flexion of the knees and hips, followed by an immediate and impulsive extension of the knees and hips again to jump vertically up and take-off (c) and flight (d). At the end of the jump, a stage of landing (e) with the absorption of the forces of the impact, and a stage of recovery (f) of the balance can be identified, followed by a standing phase after compensation of the forces (a).

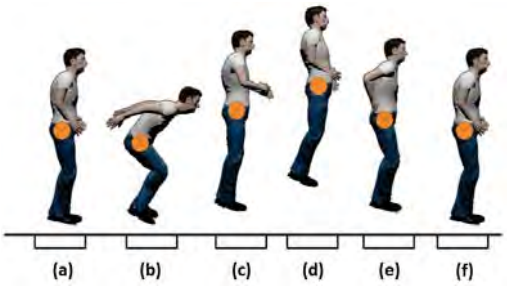



Figure 1: Counter movement jump and its separate phases: standing (a), preparation (b), take-off (c), flight (d), landing (e), and recovery (f). The marks indicate the participant's center of mass in each jump phase.

## 2.2 Common Technical Measurements via Force Platforms

Since pure observational evaluations are difficult, due to the fast progress of a jump, analyses are typi-

cally technology supported via force platforms, contact mats or optical systems (Bui et al., 2015). Force platforms measure the ground reaction force intensity and distribution.

Due to the specific force distribution characteristics, the phases of a CMJ can be identified by variations of the ground force (if measured by ground force reaction platforms or similar devices), as shown exemplarily in Figure 2 and discussed in the following: The transition from standing or rest (a) to preparation (b) can be clearly recognized by changes in the amplitudes of the force, which was nearly constant during the standing phase. Take-off (c) is characterized by the decrease of force affecting on the platform. During the flight phase (d), the force amounts to zero. In the moment of landing (e), the force increases to a maximum. In the phase of recovery (f), the subject tries to compensate the forces, in order to enter the phase of standing or resting (a) again.

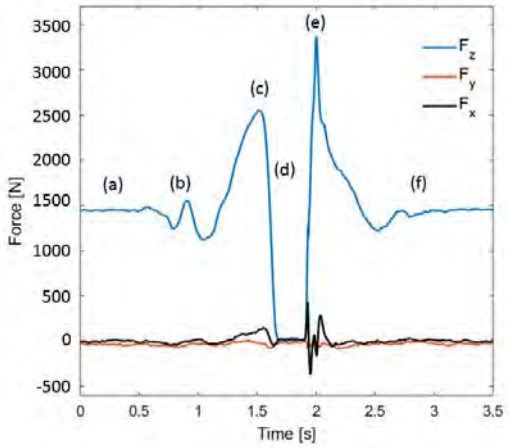



Figure 2: Variations over time of force intensities per axis during a counter movement jump. The force was measured by a force platform. The phases of the counter movement jump are marked in the graph: standing (a), preparation (b), take-off (c), flight (d), landing (e), and recovery (f).

Figure 3 shows the coordinate orientation of the force platform. It is clear that the mean force acts perpendicular to the force plate in the vertical (z) direction. But there are also reaction forces in the mediolateral (y) and especially in the anteroposterior (x) direction. In the anteroposterior direction is a peak during the take-off phase pushing the feet off the ground, and during the landing while the toes and heels strike the ground and compensate the movements.

Force platforms can estimate power, velocity and the related jump height (Samozino et al., 2008). The physical relationships between the measured force and the power, as well as the jump height, is shown in Section 3. The peak force measurements during the preparation and take-off phases e.g. of CMJs are, as long as being considered relative to the body mass, significantly related to muscle strength (Nuzzo et al., 2008; Markovic et al., 2014). The strength can be analyzed by the take-off phase of the jump, and the balance was shown to be estimated based on the force-intensities and distributions during the landing and recovery phase of jumps. For example, the time to stabilization (TTS) is the time it takes for an individual to return to a stable state following a jump or hop landing, and it is a used factor for balance analyses. Thereby, a longer TTS indicates more difficulty controlling the posture of landing and might indicate impaired neuromuscular control (Fransz et al., 2015).

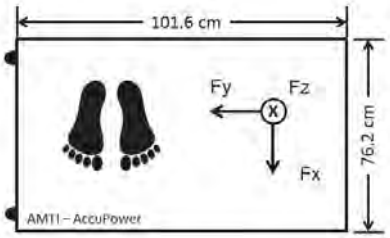



Figure 3: The dimensions of the used AccuPower ground reaction force platform of AMTI with its coordinate orientation. The footprint symbolizes the orientation of the subjects during the jumps.

## 3 STATE OF THE ART

The jump power, as an indicator of muscle strength, can be identified by the force measurements during the vertical jump, especially in the phase of the take-off.

Force plates measure the force acting on the plate. According to Newton's second law, the force $F$ is equal to the mass $m$ of an object times its acceleration $a$.

$$F = m \cdot a \tag{1}$$

In the example of Figure 2, we can clearly see the influence of the mass of the jumper on the force $F_z$ acting perpendicular to the surface of the plate. The offset at rest amounts about $1500N$, which corresponds approximately to a mass of $150kg$.

The power $P$ is defined by the force $F$ times the velocity $v$:

$$P = F \cdot v \tag{2}$$

In many studies, the maximum jump power is observed and seems to be a sensitive indicator of the muscle performance and the strength (Dietzel et al., 2015; Kalyani et al., 2014).

A further important parameter for jump analyses is the jump height. The jump height $h$ can be estimated by the following equation:

$$h = (v_t \cdot t) - \frac{1}{2} \cdot g \cdot t^2, \tag{3}$$

where $v_t$ is the vertical velocity of the center of mass of the jumper at take-off, $t$ is the time to peak flight and $g$ the gravity.

Besides measuring muscle strength, force platforms can be utilized to measure dynamic postural stability, which has been shown as related to balance and ankle stabilities. Therefore, functional deficits such as chronic functional ankle instability (FAI) (Hertel, 2002), can be indicated based on the recorded vertical, anteroposterior or mediolateral reaction forces, which enable the calculation of time to stabilization (TTS) and variations over time of the center of pressure (COP), range of motion (ROM), and the dynamic postural stability index (DPSI) as accepted measures for postural stability and FAI. The DPSI is at least as accurate and precise as TTS but provides a comprehensive measurement of dynamic postural stability that is sensitive to change in 3 directions.

DPSI provides a comprehensive measurement of dynamic postural stability that is sensitive to change in all directions since combining three (vertical, anteroposterior and mediolateral) stability indexes and considers as well the subject's weight for the vertical stability and thus has been shown to be a reliable measure (Wikstrom et al., 2005b; Meardon et al., 2016). While COP and ROM have shown mixed correlations to FAI stabilities, TTS is a well-accepted measure to quantify performance. Typically, the force is considered in order to measure the TTS, as a measure of the ability to stabilize posture (which is applied within numerous studies). TTS typically ranges from 0 to 7s. By investigating 20 TTS calculation methods (as identified via a structured literature review), Fransz et al. (Fransz et al., 2015) have shown that all use threshold-based approaches based on the ground force and 90% can be described based on four aspects: (1) the input signal, (2) signal processing, (3) the stable state (threshold), and (4) the definition of when the (processed) signal is considered stable.

Wikstrom et al. identified a significant variability among TTS measurements due to differences between the TTS calculation methods used in various studies (Wikstrom et al., 2005a). By evaluating the influences of parameter variations, Fransz et al. (Fransz et al., 2015) have indicated that the TTS measure does produce non-standardized parameters if estimated via ground forces reaction parameter. They indicated variations of the TTS of up to 56% for sample rate (100 to 1000 Hz), 37% for filter settings (no filter, 40, 15 or 10 Hz), 28-282% for trial lengths (20, 14, 10, 7, 5 and 3 s), as well as calculation methods. Thereby they clarified the difficulties to compare TTS results recorded among different systems based on the power measure.

While these analyses are performed based on single jump measurements for 25 healthy younger adults (20-53 years), its insights will generally apply due to the indicated computational differences and the drastic effect sizes.

Consequently, alternative measures are desired, which are more robust regarding measurement-variations such as sample rates.

Ideally, these measures should be equally applicable to rather mobile measurement devices such as inertial measurement units (IMU), which will be increasingly applied due to their lower price and the higher grade of mobility (Choukou et al., 2014; Elvin et al., 2007; Milosevic and Farella, 2015).

# 4 SYSTEMS AND CONTROL TECHNOLOGY

Considering the situation, that a system is stimulated by an action (input signal). Usually, the system reacts on this stimulation in any manner (output signal). Now we want to describe this system to predict the reaction of the system to an action. In the systems and control technology the relation of an input and an output function, and therefore, the system can be described by a transfer function (see Figure 4).

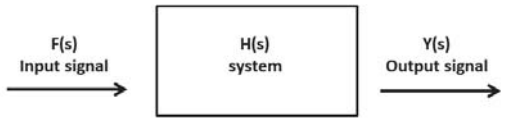



Figure 4: Relation between the output function $Y(s)$, the transfer function $H(s)$ and $F(s)$ the input function.

The mathematical relation is given by:

$$Y(s) = H(s)F(s), \tag{4}$$

where $Y(s)$ is the output function, $H(s)$ the transfer function and $F(s)$ the input function. If assuming, that the landing and recovery phase of a vertical jump (Figure 2 (e)) is an oscillating system, the transfer function is described by a second-order delay element (PT2-element).

$$H(s) = \frac{Y(s)}{F(s)} = \frac{a}{cs^2 + bs + 1} \tag{5}$$

Considering the general second-order system of an oscillator $H(s)$ can be described by

$$H(s) = \frac{K\omega_0}{s^2 + 2D\omega_0 s + \omega_0^2}, \tag{6}$$

where $\omega_0$ is the natural frequency, $K$ the DC gain of the system and $D$ the damping ratio. It is assumed, that the error resulting from the use of the time-continuous model is small, because of a relatively high sampling ratio of 200 Hz. The natural frequency determines how fast the system oscillates during the response. The damping ratio determines how much the system oscillates as the response decays toward a steady state. These parameters can be deduced from equation 6, after transferring it in the form of equation 5:

$$H(s) = \frac{k}{(\frac{s^2}{\omega_0^2} + \frac{2Ds}{\omega_0} + 1)} \tag{7}$$

$$\omega_0 = \sqrt{\frac{1}{c}} \qquad \text{and} \qquad D = \frac{b\omega_0}{2} \tag{8}$$

Therefore, the natural frequency and the damping ratio can characterize the landing phase, the absorption of the impact, and the restoring of the balance and stability. As we will see in the next section, these parameters might be an alternative possibility to characterize the balance ability, the muscle strength, and allow conclusions to postural stabilization and neuromuscular control.

# 5 MODEL-BASED APPROACH

We propose the use of the oscillation behavior as an alternative approach to drawing conclusions about muscle strength, balance ability, postural stability, and neuromuscular control instead of using the DPSI, TTS, COP or ROM. The advantage of the model-based approach of the oscillation behavior (during landing and recovery phase) over existing amplification-based methods might be its potentially lower dependability on sample rates, and trial lengths.

In detail, we aim to model (as schematically illustrated in Figure 5) the human's lower extremities as a spring that oscillates during the landing and recovery phase. During free fall the spring is slack and will be compressed at the impact on the floor and the landing phase and depresses during the recovery phase to the steady state in one or more oscillations.

From a physical point of view, this system can be described by

$$F = -kx, \tag{9}$$

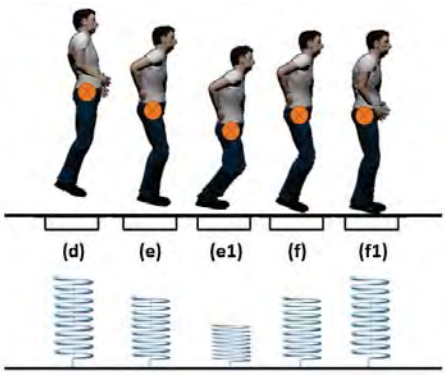



Figure 5: Comparing of the human's lower extremities with a spring during the landing and the recovery phase of a jump. The spring will be compressed during the landing (e-e1) and depresses during the recovery phase (f-f1) to the steady state in one or more oscillations.

with the force $F$, the displacement $x$ and the spring constant $k$. The frequency can be estimated with

$$\omega_0 = \sqrt{k/m}. \tag{10}$$

This equation shows that the frequency correlates to the spring constant $k$.

In our model, the spring is characterized by the spring constant. Consequently, if comparing the spring with the muscles of the humans' lower extremities, the spring constant characterizes the stiffness of the muscles in a first approximation. Therefore, the natural frequency $\omega_0$ of our system describes the ability to absorb the impact at the landing and characterizes the muscles of the lower extremities.

The damping ratio $D$ indicates the influence within or upon an oscillatory system that has the effect of reducing its oscillations and might also be a relevant parameter for the characterization of the balance ability and the postural stability.

# 6 EVALUATION

## 6.1 Study Design

Each of the 20 considered healthy older adults of our study (12 subjects are female (60%) and 8 male (40%)) has performed three sequential CMJs with a rest of 1 min between the jumps to avoid signs of fatigue. Further characteristics of the subjects are listed in Table 2. The group covers a typical range of age, weight, and height for the group of pre-frail elderlies. The test procedures were approved by the local ethics committee (ethical vote: Hannover Medical School No. 6948) and conducted in accordance with the Declaration of Helsinki.

Table 2: Population characteristics of our study with the minimum (min.), maximum (max.), mean values and standard deviation (SD).

| n=20 | min. | max. | mean | SD |
|---|---|---|---|---|
| age [years] | 71 | 82 | 75.1 | 3.02 |
| weight [kg] | 51.6 | 97.25 | 74.87 | 12.49 |
| height [cm] | 154.1 | 189.1 | 167.25 | 9.92 |

The jumps have been performed on an AMTI AccuPower ground reaction force plate, which is specified for jumping and power analyses and is an accepted gold standard. Figure 3 shows the coordinate orientation and the dimensions of the plate. The sampling rate amounts $200\,Hz$. The AccuPower sensitivity is based on a $8900\,N$ full-scale $F_z$ capacity and a 12 bit internal AD ($\pm 2048\,bit$) or about 4.3 N/bit.

The transfer functions and FPE (as describing the transfer functions fit) have been estimated for each (of the 3x20) performed CMJs with Mathworks' MATLAB (version R2015a) using the System Identification Toolbox (version 9.2).

Per subject, the jump with the best fit estimation was taken into account in our analyses, whereby fits below 70% are rejected. Thereby, the function for the fit corresponds to the form of equation 5. An impulse function with the height of the impact force at landing will be assumed for the input signal. Figure 6 shows a typical input function.

Figure 7 shows a characteristic phase of landing and recovery and is assumed as an output function. To take the force in all directions into account, the absolute values for each axis are summed up. In the step of reprocessing the means are removed. In accordance with the theoretical considerations of Section 4, the number of poles is set to 2 and the number of zeros to 0 for the model of the transfer function, to describe an oscillating system (see equation 6). A discrete-time spectrum with T=0.005 s is chosen, because of the sampling rate of 200 Hz of the force plate.

## 6.2 Natural Frequency and Damping Ratio

The natural frequency $\omega_0$ and damping ratio $D$ were determined as described in section 4. The results for the natural frequency are in the range of about $1\,Hz$. Due to the literature, we expected higher frequencies (Wakeling et al., 2001). The analyze of the poles of the poles of the determined transfer functions show that there are only two real poles, which indicate a non-oscillating function instead of complex pole pairs, which lead to exponentially modulated oscillations. Consequently, the calculated parameter (called

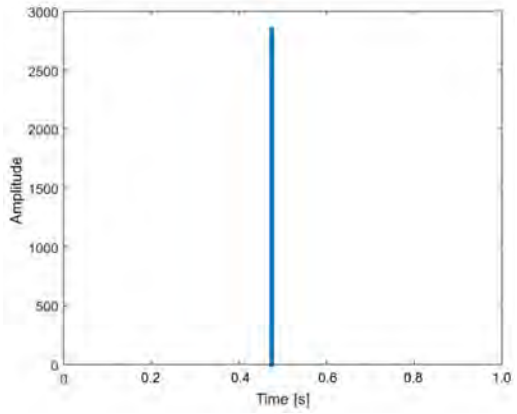



Figure 6: Typical impulse function as an input function.

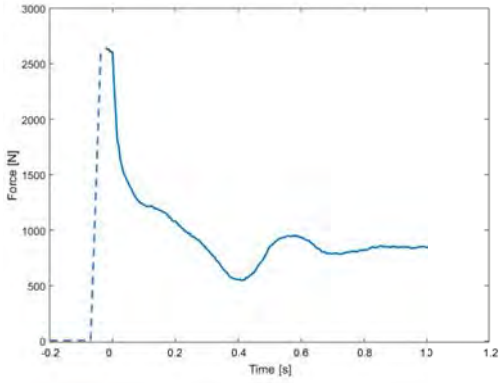



Figure 7: Typical phase of landing and recovery as an output function. The dashed line indicates the previous progress of the force during the flight phase.

$w$) is not equivalent to the natural frequency $\omega_0$ and describes rather a damping factor. Nevertheless, the parameter $w$ seems to be significant: Figure 8 shows the parameter $w$ in relation to the jump height $h$. A linear regression results in

$$d = -0.0013 \cdot h + 1 \tag{11}$$

The relation shows an inverse relationship between these parameters. Therefore, the parameter $w$ decreases with an increasing jump height.

Figure 9 shows the relation between jumping height and the calculated parameter $d$ for the damping ratio $D$. Within the study group, the parameter $d$'s effect size is small, since varying only in a small range. As an approximative assumption, the damping is linearized. Hence, an inverse relationship between the jumping height and the damping ratio was also recognized. The relation is not significant, due to the small effect size. It needs further investigations via data of both larger cohorts and heterogeneous groups, and additionally if the parameter $d$ is equivalent to the damping ratio $D$.

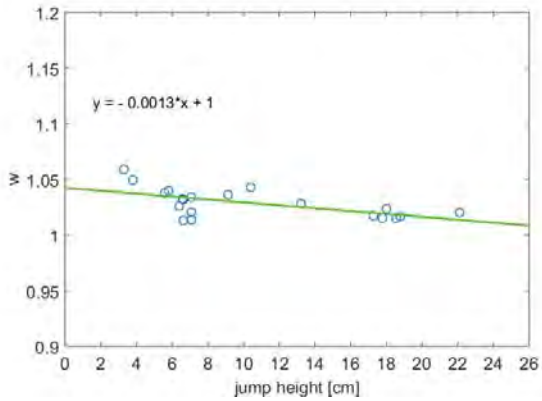



Figure 8: The parameter $w$ is shown in relation to the jump height $h$. There is an inverse relationship between this parameter and the jump height.

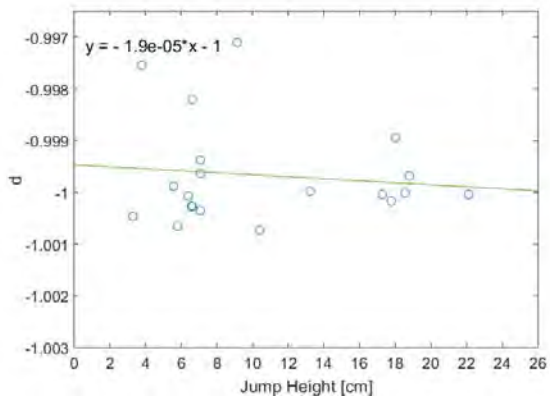



Figure 9: The estimated parameter $d$ is shown in relation to the jump height $h$. The variation of this parameter lies only in a small range.

## 6.3 Major Factors of Influence

Due to the fact, that several factors can influence the parameters $w$ and $d$, we analyzed the influence of the age, the sex, the body weight, the body height, $w$ and $d$ as a function of the jump height.

Therefore we used a generalized regression model. Due to the left skewed distribution of jump height (see Figure (10)) we estimated a gamma-distribution. Considering the jump height as clearly observable element, we performed a stepwise model selection by the Akaike information criterion (AIC).

The results of the regression analyses are listed in Table 3. As mentioned in Section 6.2 the parameter $d$ varies only in a small range and is not significant. Also, the body height has no influence on the jumping height. But we can see a small influence of the body weight and an influence of the sex and the parameter $w$.

Table 3 shows the changes of the jump height for a

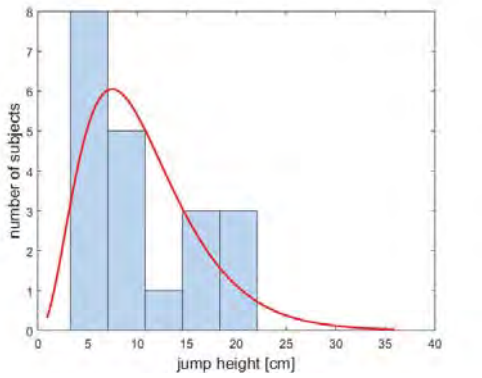



Figure 10: Distribution of the jump height and the estimated gamma-distribution (red line).

one unit change of the listed parameters. For example, the sex has an influence of about 7 $cm$. The change of 0.01 unit of $w$ results in a mean change of 1.5 cm of jump height.

Table 3: Influence of weight, sex, and $w$ depending on the jump height.

| | Estimation | SD | p($>|t|$) |
|---|---|---|---|
| $w$ | -152.20 | 43.02 | <0.01 |
| weight | -0.073 | 0.04 | 0.08 |
| sex | -7.34 | 1.95 | <0.01 |

Figure 11 illustrates the differences between male and female subjects in jump height as a function of $w$.

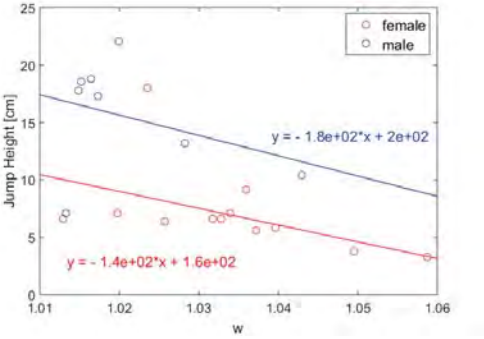



Figure 11: Comparison of the jump height as a function of $w$ of male and female subjects.

Considering equation 10 in Section 5 and the simplified comparison of a human as a spring, the natural frequency increases with increasing spring constant and therefore increasing stiffness. On the assumption that the jump height $h$ corresponds with the muscle strength $S$, there seems to be a relationship between jump height, the paramter $w$, muscle strength, and spring constant $k$:

$$h \sim S \sim \frac{1}{w} \sim \frac{1}{\omega_0} \sim \frac{1}{k}. \quad (12)$$

Thus, our model explains the inverse relation between jump height and the parameter $w$.

In order to analyze the postural stability, the dynamic postural stability index (DPSI) was determined in accordance with the approach of (Wikstrom et al., 2005b) by equation 13.

$$DPSI = \sqrt{\frac{\sum(0-F_x)^2 + \sum(0-F_y)^2 + \sum(m-F_z)^2}{n}}, \quad (13)$$

where $F_x$, $F_y$, $F_z$ are the forces in anteroposterior (x), mediolateral (y) and vertical(z) direction, $m$ the body weight, and $n$ the number of data points. Figure 12 shows the resulting distribution of the DPSI over the considered CMJs of our study population. Next to the normal distribution of the DPSI values and the distribution-range (in relation to other age-related groups such as in (Wikstrom et al., 2005b; Meardon et al., 2016)) confirms the suitability of the DPSI for the considered group.

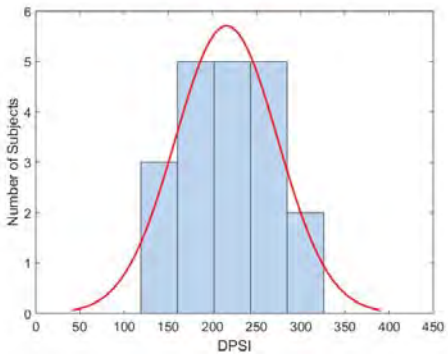



Figure 12: Distribution of the DPSI over our study population with a normal distribution fit.

In order to investigate the relation between $w$ and postural stability (represented by the DPSI), we also analyzed the influence of the factors age, body weight, and $w$ as a function of the DPSI. Table 4 lists the results.

Table 4: Influence of weight, sex and $w$ depending on the DPSI.

| | Estimation | SD | p($>|t|$) |
|---|---|---|---|
| $w$ | -1244.73 | 791.75 | 0.14 |
| weight | 2.25 | 0.816 | 0.01 |
| age | 7.71 | 3.37 | 0.04 |

We found an influence of these factors on the DPSI: For example, a change of 0.01 units in $w$ results in a change of 12 of the DPSI-value. A change in the age of 10 years causes a change of 77 of the value of DPSI. While a significant correlation of the weight and age with the DPSI was found, the strong influence of $w$ (indicated by an estimate of 1244.73, which is 161 times stronger than for age, the next strongest estimate) could not yet be statistically ensured due to the low number of subjects ($n$=20). Consequently, we are looking forward to confirming the expected significant correlation in an upcoming analysis with a larger study group.

## 7 CONCLUSION

While the CMJ is well-suited to measure muscle power and strength within a single assessment item, common traditional amplification-based balance measures for CMJ (such as TTS or CoP) have been shown to be significantly influenced by measurement settings including trial length, sample rate, and filter settings. Thus, a reliable alternative approach to detect balance ability for vertical jumps is required.

As an alternative approach, we propose to model the human body during the landing after a jump from a Systems and Control Technology perspective. Therefore, we used an impulse function of the maximum force at the impact on the ground as input function and the landing and recovery phase of a jump as output function. This phase is characterized by the balance and muscle strength of a subject.

Since the landing and recovery phase of vertical jumps can be represented by an oscillating system, the transfer function is described by a second-order delay element (PT2-element), where the natural frequency determines the systems oscillation frequency, and the damping ratio determines the system oscillation intensity as the response decays towards a steady state.

In an initial prospective study with 20 elderly probands, we could not observe the expected oscillating behavior in the phase of landing. Nevertheless, an inverse relationship between the calculated parameter $w$ and the jump height and an inverse relation to the muscle strength could be determined. We could identify an influence of sex, weight, and $w$ on the jump height.

Furthermore, a potential correlation between DPSI (as a common standard-index for balance) and $w$ was seen but could not be clearly clarified due to the limited group size. Thus, we will investigate these effects in a further larger study.

Moreover, we considered in our model only one dynamic mode (one modal mass and one frequency). Using two-dimensional models of the mus-

culoskeletal system (Blache and Monteil, 2013) or even an anatomically realistic three-dimensional musculoskeletal model (Farahani et al., 2016) could open perspectives for more robust models.

## ACKNOWLEDGEMENTS

The study is funded by the German Federal Ministry of Education and Research (Project No. 01EL1422D). We would also like to thank Dr. Fabian Sobotka for his support.

## REFERENCES

Blache, Y. and Monteil, K. (2013). Effect of arm swing on effective energy during vertical jumping: experimental and simulation study. *Scandinavian Journal of Medicine & Science in Sports*, 23(2):e121–e129.

Buehring, B., Krueger, D., Fidler, E., Gangnon, R., Heiderscheit, B., and Binkley, N. (2015). Reproducibility of jumping mechanography and traditional measures of physical and muscle function in older adults. *Osteoporosis International*, 26(2):819–825.

Bui, H. T., Farinas, M.-I., Fortin, A.-M., Comtois, A.-S., and Leone, M. (2015). Comparison and analysis of three different methods to evaluate vertical jump height. *Clinical Physiology and Functional Imaging*, 35(3):203–209.

Choukou, M.-A., Laffaye, G., and Taiar, R. (2014). Reliability and validity of an accelerometric system for assessing vertical jumping performance. *Biology of Sport*, 31(1):55.

Clegg, A., Young, J., Iliffe, S., Rikkert, M. O., and Rockwood, K. (2013). Frailty in elderly people. *The Lancet*, 381(9868):752–762.

Cruz-Jentoft, A. J., Baeyens, J. P., Bauer, J. M., Boirie, Y., Cederholm, T., Landi, F., Martin, F. C., Michel, J.-P., Rolland, Y., Schneider, S. M., et al. (2010). Sarcopenia: European consensus on definition and diagnosis report of the european working group on sarcopenia in older people. *Age and Ageing*, 39:412423.

Dietzel, R., Felsenberg, D., and Armbrecht, G. (2015). Mechanography performance tests and their association with sarcopenia, falls and impairment in the activities of daily living-a pilot cross-sectional study in 293 older adults. *Journal of Musculoskeletal and Neuronal Interactions*, 15(3):249–256.

Elsawy, B. and Higgins, K. E. (2011). The geriatric assessment. *American Family Physician*, 83(1):48–56.

Elvin, N. G., Elvin, A. A., and Arnoczky, S. P. (2007). Correlation between ground reaction force and tibial acceleration in vertical jumping. *Journal of Applied Biomechanics*, 23(3):180.

Farahani, S. D., Andersen, M. S., de Zee, M., and Rasmussen, J. (2016). Optimization-based dynamic prediction of kinematic and kinetic patterns for a human vertical jump from a squatting position. *Multibody System Dynamics*, 36(1):37–65.

Fransz, D. P., Huurnink, A., de Boode, V. A., Kingma, I., and van Dieën, J. H. (2015). Time to stabilization in single leg drop jump landings: An examination of calculation methods and assessment of differences in sample rate, filter settings and trial length on outcome values. *Gait & Posture*, 41(1):63–69.

Fudickar, S., Karth, C., Mahr, P., and Schnor, B. (2012). Fall-detection simulator for accelerometers with in-hardware preprocessing. In *Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments*, page 41. ACM.

Granacher, U., Gollhofer, A., Hortobágyi, T., Kressig, R. W., and Muehlbauer, T. (2013). The importance of trunk muscle strength for balance, functional performance, and fall prevention in seniors: A systematic review. *Sports Medicine*, 43(7):627–641.

Hein, A., Winkelbach, S., Martens, B., Wilken, O., Eichelberg, M., Spehr, J., Gietzelt, M., Wolf, K.-H., Büsching, F., Hülsken-Giesler, M., et al. (2010). Monitoring systems for the support of home care. *Informatics for Health and Social Care*, 35(3-4):157–176.

Hellmers, S., Fudickar, S., Büse, C., Dasenbrock, L., Heinks, A., Bauer, J. M., and Hein, A. (2016). Technology supported geriatric assessment. In *Ambient Assisted Living*. Springer. (in press).

Hertel, J. (2002). Functional anatomy, pathomechanics, and pathophysiology of lateral ankle instability. *Journal of Athletic Training*, 37(4):364.

Kalyani, R. R., Corriere, M., and Ferrucci, L. (2014). Age-related and disease-related muscle loss: the effect of diabetes, obesity, and other diseases. *The Lancet Diabetes & Endocrinology*, 2(10):819–829.

Markovic, S., Mirkov, D. M., Nedeljkovic, A., and Jaric, S. (2014). Body size and countermovement depth confound relationship between muscle power output and jumping performance. *Human Movement Science*, 33:203–210.

Meardon, S., Klusendorf, A., and Kernozek, T. (2016). Influence of injury on dynamic postural control in runners. *International Journal of Sports Physical Therapy*, 11(3):366.

Milosevic, B. and Farella, E. (2015). Wearable inertial sensor for jump performance analysis. In *Proceedings of the 2015 Workshop on Wearable Systems and Applications*, pages 15–20.

Nuzzo, J. L., McBride, J. M., Cormie, P., and McCaulley, G. O. (2008). Relationship between countermovement jump performance and multijoint isometric and dynamic tests of strength. *The Journal of Strength & Conditioning Research*, 22(3):699–707.

Palma, S., Silva, H., Gamboa, H., and Mil-Homens, P. (2008). Standing jump loft time measurement-an acceleration based method. In *Biosignals (2)*, pages 393–396.

Rittweger, J., Schiessl, H., Felsenberg, D., and Runge, M. (2004). Reproducibility of the jumping mechanography as a test of mechanical power output in physically competent adult and elderly subjects. *Journal of the American Geriatrics Society*, 52(1):128–131.

Samozino, P., Morin, J.-B., Hintzy, F., and Belli, A. (2008). A simple method for measuring force, velocity and power output during squat jump. *Journal of Biomechanics*, 41(14):2940–2945.

Siglinsky, E., Krueger, D., Ward, R. E., Caserotti, P., Strotmeyer, E. S., Harris, T. B., Binkley, N., and Buehring, B. (2015). Effect of age and sex on jumping mechanography and other measures of muscle mass and function. *Journal of Musculoskeletal & Neuronal Interactions*, 15(4):301.

Steen, E.-E., Frenken, T., Eichelberg, M., Frenken, M., and Hein, A. (2013). Modeling individual healthy behavior using home automation sensor data: Results from a field trial. *Journal of Ambient Intelligence and Smart Environments*, 5(5):503–523.

Wakeling, J. M., Von Tscharner, V., Nigg, B. M., and Stergiou, P. (2001). Muscle activity in the leg is tuned in response to ground reaction forces. *Journal of Applied Physiology*, 91(3):1307–1317.

Wikstrom, E., Tillman, M., and Borsa, P. (2005a). Detection of dynamic stability deficits in subjects with functional ankle instability. *Medicine & Science in Sports & Exercise*, 37(2):169–175.

Wikstrom, E., Tillman, M., Smith, A., and Borsa, P. (2005b). A new force-plate technology measure of dynamic postural stability: the dynamic postural stability index. *Journal of Athletic Training*, 40(4):305–309.

# C-LACE: Computational Model to Predict 30-Day Post-Hospitalization Mortality

Janusz Wojtusiak, Eman Elashkar and Reyhaneh Mogharab Nia
*Health Informatics Program, George Mason University, 4400 University Dr., MSN 1J3, 22030, Fairfax, VA, U.S.A.*
*jwojt@mli.gmu.edu, {eelashka, rmoghara}@gmu.edu*

Keywords: Mortality Prediction, Machine Learning, Online Calculator.

Abstract: This paper describes a machine learning approach to creation of computational model for predicting 30-day post hospital discharge mortality. The Computational Length of stay, Acuity, Comorbidities and Emergency visits (C-LACE) is an attempt to improve accuracy of popular LACE model frequently used in hospital setting. The model has been constructed and tested using MIMIC III data. The model accuracy (AUC) on testing data is 0.74. A simplified, user-oriented version of the model (Minimum C-LACE) based on 20-most important mortality indicators achieves practically identical accuracy to full C-LACE based on 308 variables. The focus of this paper is on detailed analysis of the models and their performance. The model is also available in the form of online calculator.

## 1 INTRODUCTION

Risk Adjusted Mortality Rates are important indicators for care outcome. They are used by administrators, Policy makers and organizations including government agencies, managed care companies and consumer groups (Inouye et al, 1998) to compare effectiveness of care among different facilities and utilize results in quality improvement efforts. Clinicians are mostly interested in accurate and valid mortality prediction models to use as tools for better planning of care, evaluation of medical effectiveness among treatment groups while controlling for patients' baseline risk, and to help clinicians decide if a patient may benefit from intensive care units and when. From patient's family perspective, discussing outcome of critically ill patients is always welcomed and appreciated. (Rocker et al, 2004)

Many illness severity scoring systems that are primarily used to measure prognosis early in the course of critical illness had been widely used to calculate in-hospital mortality. The Simplified Acute Physiology Score (SAPS) and the Mortality Prediction Model (MPM) use data collected within one hour of ICU admission. Sequential Organ Failure Assessment (SOFA) scoring uses data obtained 24 hours after admission and then every 48 hours. Logistic Organ Dysfunction Score (LODS) and Multiple Organ Dysfunction Score (MODS) also had been used to measures severity of illness at time of ICU admission. Acute Physiologic and Chronic Health Evaluation (APACHE) scoring system widely used to predict risk of in-hospital mortality among ICU patients. The score uses the worst physiologic values measured within 24 hours of admission to the ICU and requires a large number of clinical variables including age, diagnosis, some laboratory results, and other clinical variables and run the result on a computer generated logistic regression model to calculate risk of mortality. However, these scoring systems have shown limited accuracy predicting risk of mortality for individual patients.

Most relevant to the presented work, the LACE index, which has been used to predict mortality within 30 days of hospital discharge can use both primary and administrative data. The name LACE explains variables required: length of stay ("L"); acuity of the admission ("A"); comorbidity or diagnoses of the patient (uses Charlson comorbidity score) ("C"); and number of emergency department visits in the six months before admission ("E"). LACE index scoring ranges from 0 (2.0% expected risk) to 19 (43.7% expected risk) (Walraven et al, 2010). However, standard LACE didn't show sufficient accuracy and it is not always possible to obtain data on the 4th item ("E"), as emergency room visits are not necessarily available.



Wojtusiak, J., Elashkar, E. and Nia, R.
C-Lace: Computational Model to Predict 30-Day Post-Hospitalization Mortality.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 169-177
ISBN: 978-989-758-213-4

A recent study added an extension of the LACE (LACE+) which uses the same 4 items of LACE as well as age and items unique to Canadian administrative databases (such as the Canadian Institute for Health Information Case Mix Groupings and number of hospital days awaiting alternate level of care arrangements). LACE+ had shown more accuracy in predicting death within 30 days of hospital discharge (c-statistic 0.77) than LACE index had shown (c-statistic 0.68) (Walraven et al, 2010). However, both instruments didn't show sufficient accuracy, besides it is not always possible to obtain data on the 4th item of LACE ("E"), as emergency room visits are not necessarily recorded in available data.

In the presented work we propose a computational alternative to LACE index, called C-LACE, constructed by application of machine learning methods to data containing information about length of stay, acuity of the admission, and comorbidities present during hospitalization. We decided not to use patients' emergency visits due to possible problems with data availability when applying model.

A number of other models based on machine learning and computational methods have been proposed to predict patient mortality. For example, (Levy et al., 2015) proposed a Multimorbidity Index tuned to predict mortality among nursing home patients. A number of methods have been created for prediction of mortality among specific disease groups such as pneumonia (Cooper et al., 1997), prostate cancer (Ngufor et al, 2014), or sepsis (Taylor et al., 2016).

The main contributions of the presented work are construction of C-LACE model that can be used to predict 30-day post-hospitalization mortality, and more importantly detailed analysis of the model and its behavior on real and simulated data.

# 2 DATA ANALYSIS AND MODEL CONSTRUCTION

## 2.1 MIMIC III Data

In order to construct and test the C-LACE model, we obtained and analyzed Medical Information Mart for Intensive Care III (MIMIC III) data. The data is publically available to researchers who satisfy certain conditions (Goldberg et al, 2000). The MIMIC III data has been collected between 2001 and 2012 in the Beth Israel Deaconess Medical Center. It consists of over 58,000 hospital admissions for more than 40,000 patients. It is structured into 26 tables organized as a relational database (Johnson et al, 2016).

From the MIMIC III data, we selected only admissions for patients at least 65 years old and alive at hospital discharge. This results in selection of 21,651 admissions. The distribution of selected attributes in the data is presented in Tables 1a and 1b. The tables also show likelihood ratios (RL) associated with each of the attributes for predicting mortality. Within the data, the majority of patients were treated in Medical Intensive Care Units (MICU), followed by Cardiac Surgery Recovery Units (SCRU), Cardiac Care Units (CCU), Surgical Intensive Care Units (SICU) and Trauma Surgical Intensive Care Units (TSICU). It can also be noted from the data that the majority of patients were hospitalized only once.

In the presented work, instead of loading to relational database, the data has been analyzed within distributed computing infrastructure designed and implemented as a part of the larger research project conducted in GMU's Machine Learning and Inference Laboratory. The data has been mapped to concepts within the Unified Medical Language System (UMLS) and integrated during analysis based on unique concept identifiers. The mapping process is a combination of manual labor-intensive identification of appropriate concepts which requires strong domain background of the person performing the mapping, with automated search for concepts between different terminologies in UMLS. The latter can be done when original data stored in database are coded using one of standard terminologies, but the final results still need to be verified by human experts. In fact, the presented construction of the model served as a testing application for the developed platform, whose description is out of scope of this paper (Wojtusiak et al., 2016).

Table 1a: Distribution of values in the data.

| Variable | Died in 30 days N = 1425 | | Not died in 30 days N = 20226 | | LR |
|---|---|---|---|---|---|
| Age (mean, SD) | 79.33 years (7.26) | | 76.93 years (7.16) | | |
| **Length of Stay** | | | | | |
| Hospital | 13.73 days (11.33) | | 10.52 days (9.15) | | |
| CCU (mean, SD) | 121.22 days (115.56) | 19.79% | 72.45 days (86.18) | 19.02% | 1.05 |
| CSRU (mean, SD) | 262.05 days (322.26) | 10.74% | 92.67 days (132.29) | 27.16% | 0.32 |
| MICU (mean, SD) | 106.10 days (122.87) | 57.89% | 85.32 days (119.07) | 36.14% | 2.43 |
| SICU (mean, SD) | 143.88 days (222.66) | 17.54% | 111.51 days (170.28) | 16.64% | 1.07 |
| **Admission Location** | | | | | |
| Emergency Room Admit | | 53.75% | | 39.22% | 1.80 |
| Clinic Referral/Premature | | 18.95% | | 19.93% | 0.94 |
| Phys Referral/Normal Deli | | 6.95% | | 21.73% | 0.27 |
| Transfer From Hosp/Extram | | 18.04% | | 18.39% | 0.98 |
| Transfer From Skilled Nur | | 1.75% | | 0.61% | 2.89 |
| Transfer From Other Healt | | 0.49% | | 0.10% | 4.75 |
| Info Not Available | | 0.07% | | 0.00% | 14.20 |

Table 1b: Distribution of values in the data, cont.

| Comorbidities | Died in 30 days N = 1425 | Not died in 30 days N = 20226 | LR |
|---|---|---|---|
| Cardiac dysrhythmias | 42.25% | 36.73% | 1.26 |
| Acute and unspecified renal failure | 37.05% | 21.12% | 2.20 |
| Essential hypertension | 39.16% | 52.57% | 0.58 |
| Respiratory failure; insufficiency; arrest (adult) | 33.40% | 17.88% | 2.30 |
| Congestive heart failure; nonhypertensive | 22.60% | 16.28% | 1.50 |
| Pneumonia (except that caused by TB or STD) | 25.40% | 12.66% | 2.35 |
| Urinary tract infections | 24.70% | 16.20% | 1.70 |
| COPD | 24.84% | 17.75% | 1.53 |
| Diabetes mellitus without complication | 25.47% | 24.55% | 1.05 |
| Deficiency and other anemia | 29.19% | 22.87% | 1.39 |
| Fluid and electrolyte disorders | 27.93% | 20.52% | 1.50 |
| Disorders of lipid metabolism | 26.95% | 39.20% | 0.57 |
| Coronary atherosclerosis and other heart disease | 18.67% | 23.09% | 0.76 |

## 2.2 Model Construction

During the analysis, the data has been randomly split into training set (80%) and testing set (20%). The testing portion of the data has been set aside and the experimental work has been performed on the training set. Only final application of models has been done on the testing set.

The data (diagnoses, ICU stays, lab tests, and medications) has been aggregated on the level of admission, i.e., one example in the final dataset corresponds to hospital admission. Because of specific implementation of machine learning library that was used, all data had to be coded as numeric attributes. Values of nominal attributes were coded as 0, 1, 2, etc.

- Basic demographic information (age, gender, race, etc.) for patient has been retrieved and coded.
- Diagnoses present during hospitalization were coded in the original data as ICD-9-CM codes. They were aggregated to CCS categories that group together similar ICD codes while preserving their clinical meaning (AHRQ, 2016).
- Lab values were coded as normal and abnormal. This coding was created as part of the original MIMIC dataset. Then, if at least one abnormal value for a test was detected, the overall value was coded as abnormal. This corresponds to taking the worst case and is consistent with several other approached to patient modeling. However, this is a significant oversimplification, since the values should be treated as a time series and patient trajectory analyzed accordingly Verduijn et al., 2007; Moskovitch and Shahar, 2015).
- Drugs were coded with a single binary attribute indicating use of immunosuppressant drugs. The drugs were extracted using their LOINC codes.
- Binary output attribute indicating mortality within 30 days after discharge has been calculated using the dates of discharge and death.

The data has been transformed into a single analytic file (or technically corresponding data structures) in order to be used by machine learning software.

A number of supervised machine learning methods have been explored in order to arrive at most accurate and useful set of models. Among the tested methods were logistic regression, random forest, naïve Bayes, and support vector machines. Comparison of the methods is presented in section 3.1, and actual descriptions of the methods is outside of the scope of this paper and can be found in the literature.

## 2.3 Implementation

The presented work has been implemented in Python 3 programming language (Anaconda distribution Python 3.5.2). The main libraries used are Pandas (v. 0.18.1) for data processing and sciencekit-learn (sklearn v. 0.17.1) for machine learning.

All developed source code is open source and available on request. We are in the process of preparing release code that will be available on the project website.

# 3 RESULTS

## 3.1 Method Selection

The first set of results concern selection of the most appropriate method that can handle the data. Table 2 shows comparison of accuracy of six methods applied to training data and testing data. The methods have been executed with multiple parameters and top results are presented.

Table 2: Comparison of Methods applied to complete dataset.

| Method | AUC (training) | AUC (testing) |
|---|---|---|
| Logistic | 0.73 | 0.663 |
| SVM | 1.0 | 0.5 |
| Linear SVM | 0.522 | 0.512 |
| Bayesian | 0.514 | 0.512 |
| Decision Tree | 1.0 | 0.543 |
| Random Forest | 1.0 | 0.743 |

The table clearly indicates that SVM and naïve Bayesian approaches are not performing well on the data. Decision tree is strongly overfit and useless on

testing data. Logistic regression preforms reasonably on both sets. Although its performance on testing data is below desired level.

Random Forest (Breiman, 2001) has consistently shown the highest accuracy on testing data, despite clear overfit. Detailed analysis of the model presented in Section 4 shows that the model is stable and appropriate. Based on the result, the remainder of this paper will focus on using Random Forest as the prediction algorithm. It is a well-studied approach, previously used in healthcare (i.e., Gu et al., 2015), in which large number of shallow decision trees are generated based on subsets of data (both examples and attributes). In our case, the best performance was achieved when generating 1,000 trees.

## 3.2 Use of Administrative and Clinical Data

The primary dataset used to test the research question is MIMIC III (Johnson et al., 2016) which is part of PhysioNet project (Goldberger et al., 2000). The dataset includes a variety of patient and clinical information about hospitalizations, ICU, and patient history. MIMIC III comprises over 58,000 hospital admissions for 38,645 adults and 7,875 neonates. The data spans June 2001 - October 2012. The rationale of using MIMIC III in this project is that it includes much more complex and diverse information than typically found in claims data. One of our goals is to illustrate that learning models from such data using the described method leads to better results than those that can be obtained from claims only data.

In the second set of experiments we tested if addition of clinical data (lab values) to administrative data (coded diagnoses) improves accuracy of prediction of 30-day mortality. Inclusion of lab values is consistent with existing models such as APACHE II.

The results indicate that addition of clinical data makes small difference in the accuracy. The AUC increases from 0.72 to 0.74. The ROC for combined administrative and clinical data is consistently above one for administrative data only, as shown in Figure 1. Interestingly, when applied to Medical Intensive Care Unit (MICU) and Surgical Intensive Care Unit (SICU) patients only, the accuracy worsens. While contradictory to the fact that these are two distinct types of patients and separate modeling should improve accuracy, this discrepancy can be explained by the amount of data available and thus overfitting of models.

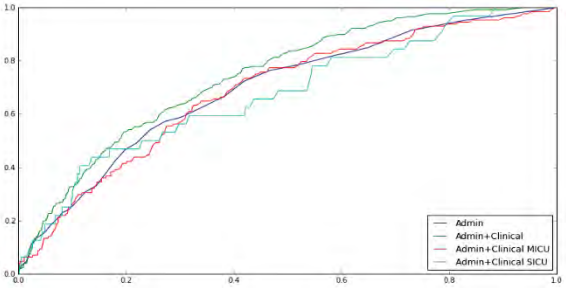



Figure 1: Receiver-operator curves for four variants of C-LACE model learned from administrative data only and administrative and clinical data. Curves for MICU and SICU patients are additionally presented.

## 3.3 Minimum C-LACE Model

Finally, we investigated possibility of reducing number of attributes needed to accurately predict 30-day mortality. Such a reduction is important for simplification of the model and, as described in Section 4, allows for creation of online calculator in which data can be entered manually.

All 308 attributes used in the full model were ranked based on their Mean Decrease Impurity calculated by the Random Forest model. It is a standard measure reported by RF after forests are built. We created a set of models while increasing number of attributes until the accuracy became comparable to one in full model. This resulted in selection of top 20 attributes listed in Table 3 along with their weights. The table also includes counts of patients and likelihood ratio as additional measure of attribute quality.

Table 3: Selected top 20 attributes along with their importance.

| Feature | Importance |
|---|---|
| Age | 0.0452 |
| HOSPITAL_LOS | 0.0346 |
| MICU_LOS | 0.0320 |
| CCU_LOS | 0.0177 |
| CCS 106 | 0.0176 |
| CCS 157 | 0.0169 |
| CCS 98 | 0.0159 |
| ADMISSION_LOCATION | 0.0157 |
| CCS 131 | 0.0152 |
| CCS 108 | 0.0145 |
| CCS 122 | 0.0133 |
| SICU_LOS | 0.0130 |
| CCS 159 | 0.0129 |
| CCS 127 | 0.0127 |
| CCS 49 | 0.0127 |
| CSRU_LOS | 0.0126 |
| CCS 59 | 0.0123 |
| CCS 55 | 0.0123 |
| CCS 53 | 0.0110 |

The AUC of the model based only on age was 0.516 which is basically a random guess based on prior class distribution. Similarly, the AUC of the model based on Age and Length of Hospital Stay was 0.576. Interestingly models based on 5 and 10 top attributes performed very close to each other with AUC values of 0.6961 and 0.697, respectively. Finally, the model based on 20 attributes performed only slightly worse than one based on all 308 attributes (AUCs 0.734 and 0.743 respectively). Figure 2 below illustrates ROC for these models.

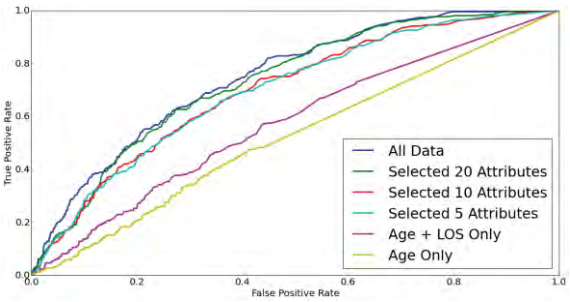



Figure 2: Accuracy of models for different selection of attributes given as ROC.

Additional analysis indicates that in fact predicted probabilities from both models are very close. When applied to training data Mean Squared Error (MSE) between probabilities of 30-day mortality calculated between both models was 0.000439 as illustrated in scatterplot in Figure 3.

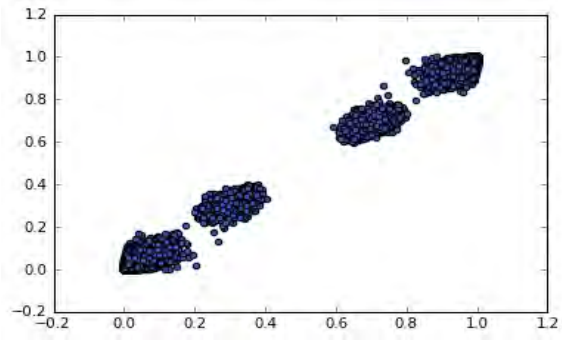


Figure 3: Comparison of probabilities of C-LACE and Minimum C-LACE on training data.

Similarly, when compared on testing data the MSE between the two models was 0.00335 as shown in Figure 4. While there is a slight difference in the predicted probabilities, the data are clearly clustered into two groups that correspond to low and high risks of mortality. Assignment to these groups is virtually identical regardless of models used.

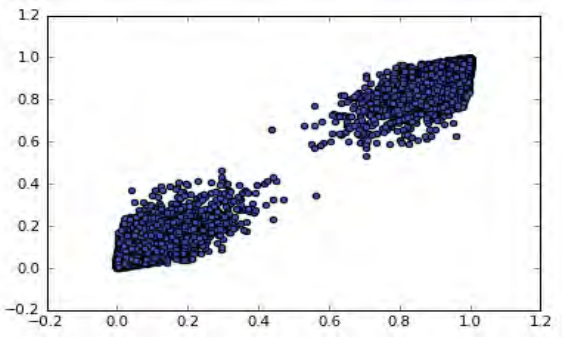


Figure 4: Comparison of probabilities of C-LACE and Minimum C-LACE on testing data.

The above analysis indicates that the two models are almost identical in terms of predictions, thus the simpler of the models should be used.

## 4 MODEL ANALYSIS

In addition to standard testing of the created model presented in the previous section, this section discusses a more detailed analysis of the created Minimum C-LACE model. The goal is to understand the model's behavior and its sensitivity to changes in input attributes.

The first set of experiments was to investigate how probabilities of 30-day mortality depend on changes in single variables. This is particularly important for continuous variables for which model should be "smooth" and not produce sudden changes in output probabilities. This property can be investigated be applying the model to large simulated data and comparing output to distribution of values in real dataset.

First created simulated dataset was *completely random*, that is, each input attribute was assigned uniformly a random value from list of allowed values for that attribute with exception for one attribute being controlled. For example, generation of simulated data to test age attribute followed the procedure:

```
for a = min(age) to max(age):
    Generate 1,000 random examples:
    age = a
    for each attribute x other than age:
        x = random(domain(x))
```

After simulated dataset is generated, C-LACE model is applied to predict mortality probabilities. These probabilities can then be investigated to check model's behavior based on changes in age.

Obviously, accuracy measures are not applicable to this simulated data since no true answer is known. The result is shown in Figure 5, which also includes distribution of average values depending on age in training, testing and complete data.

One can immediately note that the probabilities based on "completely random" simulated data are much higher than those in real data. This is correct, because a completely randomly generated patient is much "sicker" than real patients due to the way data are generated. The data on the plot shows that the model is smooth in regard to changes of probability with age. An interesting fact about model is that probabilities are somewhat higher for the lowest allowed value of age, namely 65.

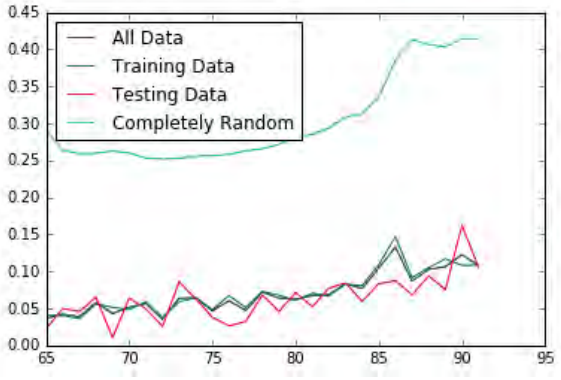



Figure 5: Distribution of predicted probability of 30-day mortality based on patient age for completely random simulated data compared with real data.

The second (*averaged training*) method used to generate simulated data started with original dataset used for training C-LACE model and multiplied the data by copying all examples for each fixed age and applying low probability random distortion to all other attributes.

```
for a = min(age) to max(age):
    For each example in training data:
        Copy the example
        age = a
        for each attribute x other than age:
            distort x
```

One can notice that probabilities of mortality in the simulated data are no longer higher than those of real data. This is due to the fact that all attributes other than age are distributed as in the original dataset (Figure 6). In the plot, one can immediately see that there is a similar "jump" of probability at the age of 65 indicating possible instability of model there.

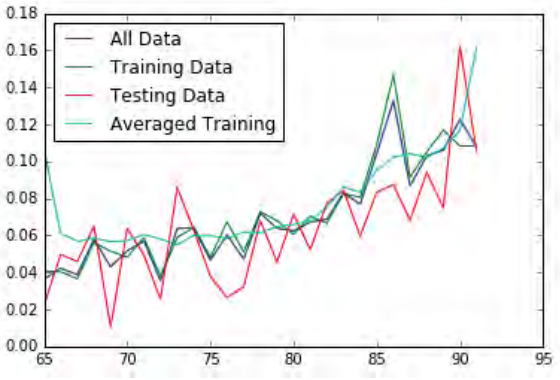



Figure 6: Distribution of predicted probability of 30-day mortality based on patient age for averaged training simulated data compared with real data.

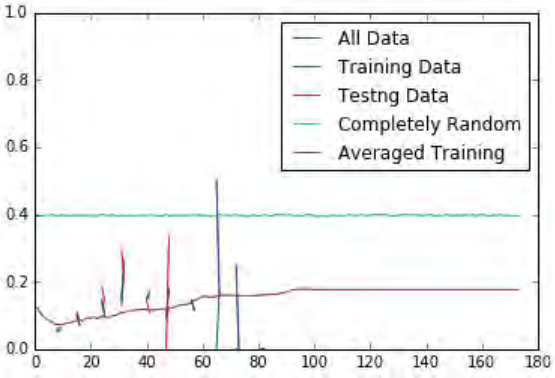



Figure 7: Distribution of predicted probability of 30-day mortality based on hospital length of stay for completely random simulated data compared with averaged training data and actual data.

The same methodology for creating completely random and averaged training simulated data has been applied to other attributes in the data with similar results. One interesting result was obtained when simulating data for fixed hospital length of stay (LOS) shown in Figure 7. When applied to completely random data, LOS has absolutely no effect on predicted probability (straight line on the plot). Interestingly, on simulated averaged training data, LOS shows clear trend. One possible explanation of this fact is that within the model LOS is strongly confounded with other attributes. The visible trend is in fact one of other attributes interacting with LOS to affect predicted mortality indirectly. Finally, a number of colored randomly looking lines in Figure 7 show that in the original data there is no clear pattern of how LOS affects predicted 30-day mortality.

The fact that when working with simulated data probabilities output by the model are smooth,

confirms the hypothesis that the constructed C-LACE model is stable.

## 4.1 Analysis of Errors

An interesting and important question concerns finding cases for which the model makes mistakes. If successful, such analysis may allow for predicting when C-LACE is more likely to make a mistake, and thus preventing it.

As shown in Figures 8 and 9, there is basically no pattern on when the model makes mistakes based on distribution of age and length of hospital stay. In both figures, green dots representing patients who died should be clustered towards the top, and red ones representing alive patients towards the bottom. The distribution errors in the model (how far green dots are from the top) is practically uniform with respect to age. While the distribution of hospital length of stay is clearly positively skewed, there seems to be no pattern in when errors are made (Figure 9).

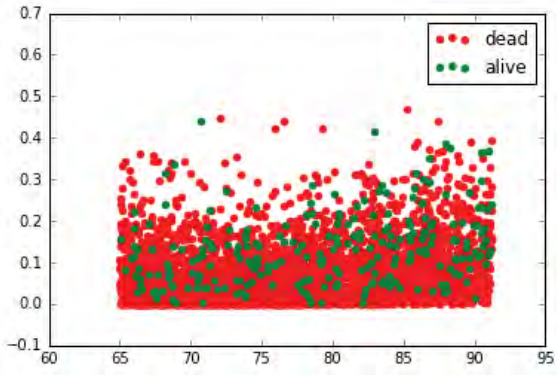



Figure 8: Predicted probabilities of 30-day mortality for training data in relation to patient age. Color of dots represents true class.

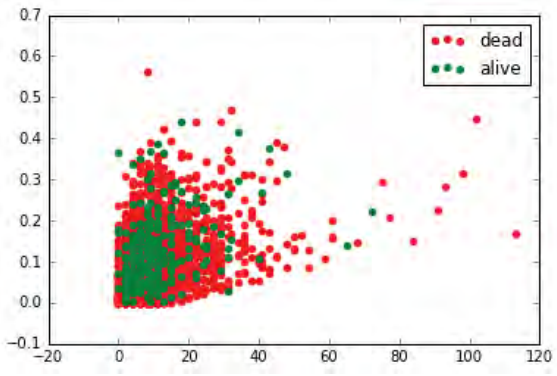



Figure 9: Predicted probabilities of 30-day mortality for training data in relation to hospital length of stay. Color of dots represents true class.

Secondary model was learned from data to predict when C-LACE is likely to misclassify positive mortality examples. Specifically, it was built from data labeled as correct classification/misclassification of testing data used to evaluate C-LACE. The secondary model has been learned using logistic regression. Following the standard procedure the misclassification data was split into training (80%) and testing (20%). When tested, the model showed very high promise of predicting when C-LACE is likely to make mistakes. It achieved AUC 0.867 on misclassification training and AUC 0.858 on misclassification testing data.

The final set of performed tests investigated optimal classification threshold based on precision and recall. Using C-LACE it is possible to achieve any value or recall, precision in general stays very low as shown in Figure 10. The figure indicates that selection of classification threshold for C-LACE around 0.1 may be the most reasonable. More detailed cost-benefit analysis of false positives and false negatives of the model is needed to arrive at final threshold applicable for final use.

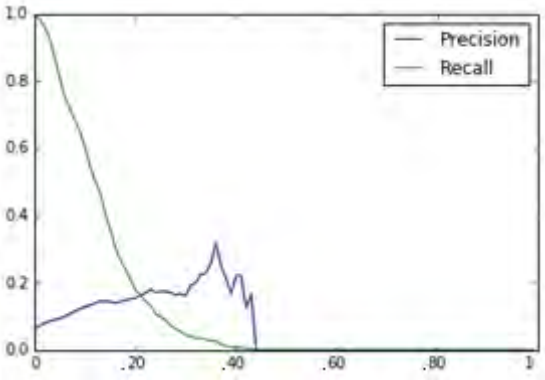



Figure 10: Analysis of Precision and Recall of the Minimum C-LACE model on testing data.

# 5 ONLINE CALCULATOR

In order for other researchers to test the developed mortality prediction models, an online calculator which includes Minimum and Full C-LACE models was created. The minimum model is available through a web form that can be used by entering data, as well as Application Programing Interface (API) for automated use. The full model is available only through an API, since it is unlikely for anyone to answer 308 questions on a web form. At this stage, the online calculator is intended only for research purposes and not for clinical use, since additional

validation is needed. The online calculator is available at the website http://hi.gmu.edu/cgi-bin/calculatros/c-lace/c-lace.cgi.

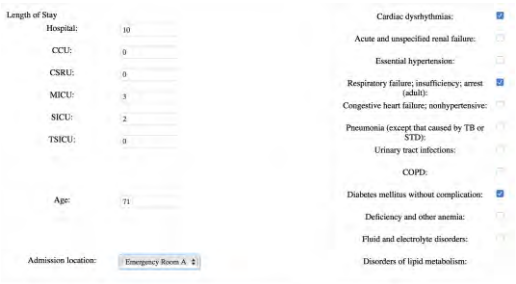



Figure 11: Design of the simple form used to enter patient and hospitalization information.

Simple online form (Figure 11) is used to enter patient and hospitalization characteristics. The entry is split into sections related to length of stay in hospital and specific ICUs, age, admission location and selected conditions most predictive of 30-day mortality. After submitting the form, user is provided with estimated probability of 30-day mortality. Because of the way the data was analyzed, the calculator is intended to be used at the time of hospital discharge.

It is important to note that within the scope of this project it was impossible to completely test the calculator and in particular assess its impact on patient care. Thus, the site contains a disclaimer that the calculator is intended to be used only for research purposes.

# 6 CONCLUSIONS

This paper presented construction and analysis of C-LACE method for predicting probability of 30-day post-hospitalization mortality. The presented solution based on application of Random Forest algorithm gives accuracy comparable to other methods available in the literature and superior to accuracy of the original LACE index. It shows that Minimum C-LACE, a 20-attributes version of the presented method, achieves the same results as one that uses 308 attributes.

Detailed analysis of the constructed model shows that the model is not sensitive to changes in values of key variables and, in fact, smoothens the data (the most visible for length of stay). While the accuracy of the model precludes its use completely independently, it is a reasonable improvement over popular LACE method. The model can be used to inform clinicians when performing patient risk assessment. Analysis has indicated that it may be possible to automatically assess classification errors from the model, though additional work is needed in this area.

The current continuation of research proceeds in two main directions:

- Possible improvement of the model accuracy by using additional clinical variables. There is significant work that remains to be done in the area of incorporating detailed clinical information and patient notes with specific focus on temporal aspect of the data. In the presented Minimum C-LACE model, no clinical attributes were included, which may be result of oversimplification of how the values were coded (see Section 2.2).
- Analysis of how the model should be presented to end-users so they understand predicted probabilities and model limitations. The latter is particularly important to make the presented online calculator useful.

# ACKNOWLEDGEMENTS

This project has been supported by the LMI-Academic partnership program grant. The authors thank Donna Norfleet, Chris Bistline, Brent Auble and all those who participated in our monthly meetings for their support and feedback that helped improve the project.

The authors thank anonymous reviewers whose feedback helped improve the paper, particularly by pointing to relevant previously published literature.

# REFERENCES

AHRQ (2016) https://www.hcup-us.ahrq.gov/tools software/ccs/ccs.jsp

Au, AG, McAlister, FA, Bakal, JA, et al. 2012. Predicting the Risk of Unplanned Readmission or Death within 30 Days of Discharge After a Heart Failure Hospitalization. American Heart Journal. 164(3):365-372.

Breiman, L., 2001. Random forests. Machine learning, 45(1), 5-32.

Cooper, Gregory F., et al., 1997. An evaluation of machine-learning methods for predicting pneumonia mortality. Artificial intelligence in medicine 9.2. 107-138.

Davis, RB, Iezzoni LI, Phillips RS, Reiley P, et al. 1995. Predicting in-hospital mortality: the importance of functional status information. Med Care.33:906-921.

Goldberger AL, et al., 2000. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic

Signals. Circulation 101(23):e215-e220 [Circulation Electronic
Pages; http://circ.ahajournals.org/content/101/23/e215.full

Gu, W., Vieira, A. R., Hoekstra, R. M., Griffin, P. M., & Cole, D. , 2015. Use of random forest to estimate population attributable fractions from a case-control study of Salmonella enterica serotype Enteritidis infections. Epidemiology and Infection, 1-9.

Ho KM, Dobb GJ, Knuiman M, et al. 2006. A comparison of admission and worst 24-hour Acute Physiology and Chronic Health Evaluation II scores in predicting hospital mortality: a retrospective cohort study. Crit Care; 10:R4.

Iezzoni LI. Risk Adjustment for Measuring Health Outcomes Ann Arbor. 1994. Mich: Health Administration Press.

Inouye, S.K., Peduzzi, P., Robison, J., Hughes, J., Horwitz, R., Concato, J, MPH, 1998. Importance of Functional Measures in Predicting Mortality Among Older Hospitalized Patients. JAMA. 1998;279(15):1187-1193. doi:10.1001/jama.279.15.1187.

Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. 2016. MIMIC-III, a freely accessible critical care database. Scientific Data. 10.1038/sdata.2016.35.

Kuzniewicz MW1, Vasilevskis EE, Lane R, Dean ML, Trivedi NG, Rennie DJ, Clay T, Kotler PL, Dudley RA. 2008. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. Chest. 2008;133(6):1319-27. doi: 10.1378/chest.07-3061.

Levy, C., Kheirbek, R,, Alemi, F., Wojtusiak, J., Sutton, B,, Williams, A.R. and Williams, A., 2015. Predictors of six-month mortality among nursing home residents: diagnoses may be more predictive than functional disability. Journal of Palliative Medicine, 18(2), 100-6.

Moreno, RP, Metnitz, P.G., Almeida, E, et al., 2005. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med; 31:1345.

Moskovitch, R. and Shahar, Y., 2015. Classification-driven temporal discretization of multivariate time series. Data Mining and Knowledge Discovery, 29(4), pp.871-913.

Ngufor, C., Wojtusiak, J., Hooker, A., Oz, T. and Hadley, J., 2014. Extreme Logistic Regression: A Large Scale Learning Algorithm with Application to Prostate Cancer Mortality Prediction. Proceedings of the 27th International Florida Artificial Intelligence Research Society Conference.

Rocker, G., Cook, D., Sjokvist, V., Weaver, B., Finfer, S., McDonald, E., Marshall, J., Kirby, A., Levy, M., at al., 2004. Clinician Predictions of Intensive Care Unit Mortality, Crit Care Med. 2004;32 (5)1

Rose, Sherri, 2013. "Mortality risk score prediction in an elderly population using machine learning." American journal of epidemiology 177.5. 443-452.

Taylor, R. Andrew, et al. 2016. Prediction of In- hospital Mortality in Emergency Department Patients with Sepsis: A Local Big Data–Driven, Machine Learning Approach. Academic Emergency Medicine.

Van Walraven, C., Dhalla, IA, Bell, C., et al., 2010. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. CMAJ 2010 Apr 6; 182(6): 551–557. doi: 10.1503/cmaj.091117. PMCID: PMC2845681

Van Walraven C, Wong J, Forster AJ. 2012. LACE+ index: extension of a validated index to predict early death or urgent readmission after hospital discharge using administrative data. Open Med. 2012 Jul 19;6(3):e80-90. PMCID: PMC3659212

Vasilevskis EE, Kuzniewicz MW, Cason BA, Lane RK, Dean ML, Clay T, Rennie DJ, Vittinghoff E, Dudley RA. 2009. Mortality probability model III and simplified acute physiology score II: assessing their value in predicting length of stay and comparison to APACHE IV. Chest. 136(1):89-101. doi: 10.1378/chest.08-2591.

Verduijn, M., Sacchi, L., Peek, N., Bellazzi, R., de Jonge, E. and de Mol, B.A., 2007. Temporal abstraction for feature extraction: A comparative case study in prediction from intensive care monitoring data. Artificial intelligence in medicine, 41(1), pp.1-12.

Wojtusiak, J., Elashkar, E., Mogharab, R., 2016. Integrating Complex Health Data for Analytics. The Machine Learning and Inference Laboratory, electronic circulation. MLI 16-1.

# Predicting Outcome of Ischemic Stroke Patients using Bootstrap Aggregating with M5 Model Trees

Ahmedul Kabir[1], Carolina Ruiz[1], Sergio A. Alvarez[2] and Majaz Moonis[3]
[1]*Dept. of Computer Science, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, U.S.A.*
[2]*Dept. of Computer Science, Boston College, Chestnut Hill, MA 02467, U.S.A.*
[3]*Dept. of Neurology, Univ. Massachusetts Medical School, Worcester, MA 01655, U.S.A.*
*{akabir, ruiz}@cs.wpi.edu, alvarez@bc.edu, majaz.moonis@umassmemorial.org*

Keywords: Ischemic Stroke, mRS Score, M5 Model Tree, Bootstrap Aggregating, Predicting Stroke Outcome.

Abstract: The objective of our study is to predict the clinical outcome of ischemic stroke patients after 90 days of stroke using the modified Rankin Scale (mRS) score. After experimentation with various regression techniques, we discovered that using M5 model trees to predict the score and then using bootstrap aggregating as a meta-learning technique produces the best prediction results. The same regression when followed by classification also performs better than regular multi-class classification. In this paper, we present the methodology used, and compare the results with other standard predictive techniques. We also analyze the results to provide insights on the factors that affect stroke outcomes.

## 1 INTRODUCTION

Stroke is defined as the rapid loss of brain function caused by disturbances in the blood supply to the brain. It is one of the leading causes of death worldwide (Raffeld et al., 2016). Stroke can be broadly classified into two types: Ischemic, which occurs due to lack of blood flow; and hemorrhagic, which is caused by internal bleeding. In this study we deal with data from patients with ischemic stroke which is the more common of the two types, accounting for around 87% of all strokes (Mozaffarian et al., 2016). The data are collected retrospectively from the University of Massachusetts Medical School, Worcester, Massachusetts, USA and comprise demographic information, medical history and treatment records of 439 patients.

The objective of this study is to predict the outcome of a stroke patient in terms of the modified Rankin Scale (mRS) score, an integer value between 0 and 6 measuring the degree of disability or dependence in daily activities of people who have suffered a stroke (Rankin, 1957). There are two approaches one may use to solve this problem. One is to treat the target as a numeric attribute and apply some form of regression. The other approach would be to think of the several different mRS scores as different categories, in which case the problem becomes that of multi-class classification. We have addressed the prediction task from both perspectives.

### 1.1 Scope of this Paper

In this paper, we aim to predict the mRS score of a patient after 90 days of an ischemic stroke based on the data we have about the patient at the time of discharge. Knowledge gained from this prediction task may help medical practitioners manage stroke more effectively and allocate resources more efficiently. The predictive (or independent) attributes in our study consist of demographic information, medical history and treatment records. The target attribute is mRS-90, the mRS score at 90 days following stroke onset (described in Table 1). We treat the target as a numeric attribute first and apply different regression techniques for prediction. Our studies show that M5 model trees used in tandem with bootstrap aggregating (bagging) significantly outperforms other common regression methods such as linear regression. We then treat the target as a multiclass categorical attribute and apply several classification techniques. Classification using the aforementioned regression technique followed by translation of the target to a discrete value performs better than well-known classification methods such as logistic regression and C4.5 decision trees.



Kabir, A., Ruiz, C., Alvarez, S. and Moonis, M.
Predicting Outcome of Ischemic Stroke Patients using Bootstrap Aggregating with M5 Model Trees.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 178-187
ISBN: 978-989-758-213-4

## 1.2 Modified Rankin Scale

The modified Rankin Scale (mRS) measure is the most widely used clinical outcome measure for stroke. It was first introduced by Dr. John Rankin (Rankin, 1957) and later modified to its current form by a group of researchers during the late 1980s (Van Swieten et al., 1988). The mRS score is an integer between 0 and 6 signifying the various degrees of impairment caused by stroke, with 0 being the least amount of impairment and 6 being death. Table 1 presents a summary description of the different mRS scores. The mRS score can be calculated at various stages of stroke. In this study, the mRS scores are recorded in three different stages. The first, mRS before admission, presents the degree of disability the patient had before the onset of stroke. The next is mRS at discharge, which gives the mRS score at the time the patient is discharged from the hospital after initial treatment of stroke. The last one is mRS at 90 days after stroke (mRS-90), the score this study attempts to predict.

Table 1: Different mRS scores and their description (Banks and Marotta, 2007).

| Score | Description |
|---|---|
| 0 | No symptoms |
| 1 | No significant disability |
| 2 | Slight disability |
| 3 | Moderate disability: requires assistance |
| 4 | Moderately severe disability |
| 5 | Severe disability: patient bedridden |
| 6 | Death |

## 1.3 Related Work

The mRS-90 score has been used as a measure of stroke outcome in numerous studies. Most of these studies focus on a particular treatment or condition, the efficacy of which is examined by how it affects the mRS-90 score. In most cases, the mRS-90 score has been dichotomized to convert the task of prediction to that of binary classification. The classification task is performed usually by multivariate logistic regression which allows the authors to comment on the influence of one or more variables on stroke outcome based on the odds ratios computed from the logistic regression model. For example, (Moonis et al., 2005) reported that using statins for treatment of ischemic stroke improved stroke outcome since the statins obtained an odds ratio of 1.57 in a logistic regression model predicting mRS-90 $\leq$ 2. This means that the patients who are administered statins have 1.57 times the probability of attaining mRS-90 $\leq$ 2 than those who are not treated with statins. (Marini et al., 2005) studied the effects of atrial fibrillation in stroke outcome. In (Yong and Kaste, 2008), hyperglycemia is associated with poor outcome, while in (Nogueira et al., 2009) successful revascularization is associated with good outcome. (Henninger et al., 2012) reported that leukoaraiosis is a factor in poor 90-day outcome of stroke. These are only a handful of the studies using mRS-90 prediction as a means of discovering effects of factors in stroke outcome. All of the above studies dichotomized the mRS score to two levels – one consisting of mRS-90 $\leq$ 2 and the other of mRS > 2.

In contrast, there have not been many studies that focused solely on predicting the stroke outcome and employing machine learning models to assist in the prediction task. (Gialanella et al., 2013) aimed to predict stroke outcome using linear regression, but used the functional independence measure (FIM) which is a scale that measures stroke recovery in terms of activities of daily living (Keith et al., 1987). A similar effort was made by (Brown et al., 2015), again focusing on FIM. Neither of these papers considered regression techniques other than linear regression. To the best of our knowledge, there is no study that has methodically explored regression analysis methods to predict the mRS-90 score as a measure of stroke outcome.

## 1.4 Plan of the Paper

In Section 2 of this paper, we present the methodology of our research. That section deals with the steps that are taken to prepare and preprocess the data, and also describes in full details our prediction techniques. Section 3 presents a comparison of different prediction methods, and analyzes the results to gain more insights about the models discovered. Section 4 concludes with a summary of findings and directions for future work.

# 2 METHODOLOGY

## 2.1 Data Collection and Preparation

Our study is conducted on retrospective data obtained from medical records of 439 ischemic stroke patients admitted at the University of Massachusetts Medical School, Worcester, MA,

USA between 2012 and 2015. Information relevant for stroke outcome prediction is extracted into a dataset. Patients who died within 90 days of stroke, therefore having a mRS score of 6, are excluded from this analysis. The reason for this exclusion is that patient death can occur for a combination of several reasons apart from stroke, such as advanced age or other comorbid conditions. Therefore, for stroke outcome prediction, we decide to work only with the patients who survived the stroke after 90 days. Prominent works on this area such as the Copenhagen Stroke Study (Nakayama et al., 1994) have also excluded dead patients in some of their models.

The process of selecting relevant predictive attributes is a combination of domain expertise and empirical knowledge of machine learning procedures. In the first step, one of the authors of this paper, a clinical neurologist and expert on stroke, has helped select a large set of attributes for extraction from the patients' medical records. We then inspect each attribute to see whether they are conducive for machine learning. Attributes with a large amount of missing values, or with almost all instances having the same value are removed. In the end, the chosen set of attributes include demographic information (such as age and gender), medical history (such as diabetes and hypertension), habits history (such as smoking and drinking), subtype of stroke (such as large vessel and cardioembolic) (Adams et al., 1993), prescribed medication (such as anticoagulants), and mRS scores at different stages (before admission, at discharge and at 90 days). A measure of stroke severity determined by the National Institutes of Health Stroke Scale (NIHSS) score (Brott et al., 1989) is also included. Table 2 presents summary statistics of all the attributes of the stroke dataset used in this study.

For the multivalued attribute *stroke subtype*, five binary attributes for the five possible values are created, with each attribute value specifying whether (1) or not (0) the patient has that particular subtype of stroke. This is done since there is no ordinal relationship among the different stroke types; so giving them numeric scores would make the model incorrect.

## 2.2 Regression

In statistics and machine learning, regression is the process of analyzing how a numeric dependent variable changes with regards to changes in one or more independent variables. In this study the regression task is performed by a meta-learning technique called bootstrap aggregating where the base learner is a model tree generated using the M5 algorithm. The machine learning tool Weka (Hall et al., 2009) is used for the experiments.

Table 2: Summary statistics of the attributes of the stroke dataset. The total number of patients is 439. For continuous attributes, the mean and standard deviation are shown in a *Mean* ± *Std. Dev.* format. For categorical attributes the percentages of different values are given. For binary attributes, only the percentages of TRUE values are shown. For mRS scores at different stages, we summarize the overall mean and standard deviation along with the distribution of individual scores.

| **Attribute** | **Distribution of values** |
|---|---|
| Stroke subtype | Small vessel: 12.3%,<br>Large vessel: 23.7%,<br>Cardioembolic: 31.4%<br>Cryptogenic: 23.7%,<br>Others: 8.9% |
| Gender | Male: 57.4%,<br>Female: 42.6% |
| Age | 67.2 ± 14.6<br>Range: 19 - 97 |
| NIHSS score at admission | 7.2 ± 7.1<br>Range: 0 - 32 |
| Hypertension | 74.7% |
| Hyperlipidemia | 58.8% |
| Diabetes | 29.8% |
| Smoking | 29.4% |
| Alcohol problem | 14.6% |
| Previous history of stroke | 19.4% |
| Atrial Fibrilation | 27.7% |
| Carotid Artery Disease | 21.0% |
| Congestive Heart Failure | 8.7% |
| Peripheral Artery Disease | 6.4% |
| Hemorrhagic conversion | 11.2% |
| tPA | 20.5% |
| Statins | 47.4% |
| Antihypertensives | 62.9% |
| Antidiabetics | 20.5% |
| Antiplatelets | 45.3% |
| Anticoagulants | 10.3% |
| Perfusion | 8.7% |
| Neurointervention | 18.7% |
| mRS before admission | 0.41 ± 0.86<br>0: 74.0%, 1: 15.0%<br>2: 5.9%, 3: 2.1%<br>4: 1.4%, 5: 0.5% |
| mRS at discharge | 1.60 ± 1.63<br>0: 35.3%, 1: 13.7%<br>2: 15.3%, 3: 9.8%<br>4: 11.6%, 5: 5.0% |
| mRS at 90 days | 1.28 ± 1.46<br>0: 46.9%, 1: 17.5%<br>2: 14.4%, 3: 11.6%<br>4: 6.2%, 5: 3.4% |

### 2.2.1 M5 Model Trees

A *decision tree* is a tree where each node represents a choice among a number of alternatives, and each leaf represents a decision that can be reached by following a series of choices starting from the root of the tree. Specifically in terms of machine learning, each node of a decision tree specifies a test of some attribute in the dataset while branches emanating from the node correspond to possible values or outputs of the test in the node (Tan et al., 2005). In the more common case, decision trees perform classification where the leaf represents one of the classes the instance is to be categorized to. But a decision tree can be used to perform regression too, in which case the leaf outputs a numeric value of the target attribute instead of a class (Breiman et al., 1984). This type of tree is called a *regression tree*. A *model tree* is a special form of regression tree where the decision in each leaf is a not a value, but is itself a multivariate linear model. The numeric value predicted by the tree for a given test data instance is obtained by evaluating the linear equation in the leaf of the branch where the data instance belongs. (Quinlan, 1992) describes an algorithm, called *M5*, that is used to construct such a tree. Some improvements to the algorithm were made by (Wang and Witten, 1996).

The construction of the model tree is a two-stage process. In the first stage, a decision tree induction algorithm is used which employs a splitting criterion that minimizes the intra-subset variability in the values down from the root through the branch to the node. The variability is measured by the standard deviation of the target values that reach that node. Taking the standard deviation of the values as a measure of error, M5 examines all attributes and possible split points to choose one that maximizes the expected reduction in error. The splitting process stops when the instances reaching a leaf have low variability or when few instances remain (Etemad-Shahidi and Mahjoobi, 2009). In the second stage, the tree is pruned starting from the leaves upward. A linear regression model is computed for every interior node, including only the attributes tested in the sub-tree rooted at that node. As the final model for this node, M5 selects either this linear model or the model subtree built in the first stage, depending on which has the lower estimated error. If the linear model is chosen, pruning takes place and the subtree at this node is converted to a leaf containing this linear model (Quinlan, 1992).

M5 model tree essentially builds a piecewise linear model. The problem space is divided into several subspaces based on the branching decisions of the tree, and separate linear models to fit the data points in each subspace are generated. Figure 1 illustrates this concept.

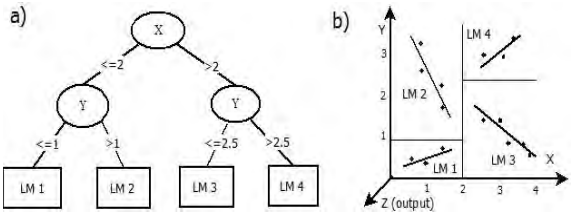



Figure 1: a) An example model tree built with the M5 algorithm with input attributes X and Y. Linear models LM 1 to LM 4 are built in the leaves. b) The corresponding problem space showing separate subspaces defined by the tree and how each linear model fits points in the subspace.

### 2.2.2 Bootstrap Aggregating

*Bootstrap aggregating*, commonly known as "*bagging*", is a meta-learning technique where multiple versions of a predictor are generated and later used to get an aggregated predictor. Each version of the predictor is learned from a bootstrap, which is a sample with replacements of the data instances drawn according to a uniform probability distribution. For the task of predicting a numerical outcome, the aggregation averages over the predictor versions (Breiman, 1996). Bagging improves generalization error by reducing the variance of the individual predictive models. If a base predictor is unstable - if it is not robust to fluctuations - the bagging process helps to stabilize it (Tan et al., 2005).

In the most common case, the size of each bootstrap $B_i$ is *n*, the same as that of the entire dataset. In this case, on average $B_i$ contains approximately 63% of the original training data since each sample has a probability of $1 - (1 - 1/n)^n$ of being picked, which converges to about 0.63 for sufficiently large *n* (Aslam et al., 2007). This is, of course, because of the fact that sampling is done with replacement, resulting in duplicate instances in each bootstrap. Once *k* bootstraps $B_{1, \ldots, } B_k$ are created, one predictor is trained on each of the bootstraps, thus producing *k* predictors. In the prediction step, a given test data instance is fed to the *k* predictors and the final prediction is the average of the values output by the *k* predictors. Figure 2 summarizes the bagging process. For the bagging models reported in this study, the value of *k* is 10.

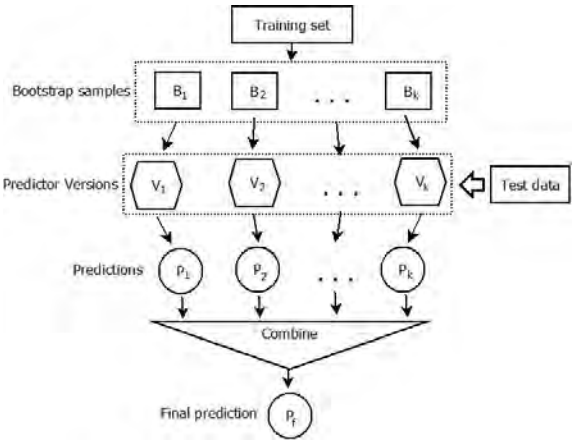



Figure 2: Summary of the process of bagging. From the training set, *k* bootstraps are created. Each bootstrap $B_1$, …, $B_k$ is used to build predictor versions $V_1$, …, $V_k$ which make separate predictions $P_1$, …, $P_k$. The final prediction $P_f$ is a combination (average for regression, majority voting for classification) of all the predictions.

### 2.2.3 Evaluation Criteria

To evaluate the performance of the regression models, we examine the degree of similarity between the actual values of the target attribute, and the predicted values returned by the models. To assess how well the models will generalize to an independent dataset, *10-fold cross validation* is used (Kohavi, 1995). The degree of similarity between the actual and predicted values is checked via three criteria: the Pearson correlation coefficient, mean absolute error and root mean squared error.

The *Pearson correlation coefficient*, R, is a measure of the linear dependence between $X = \{X_1,\ldots,X_n\}$ and $Z = \{Z_1,\ldots,Z_n\}$. It gives a value between -1 and +1 where -1 stands for total negative correlation, 0 for no correlation and +1 for total positive correlation. It can be defined as follows (Rodgers and Nicewander, 1988):

$$R = \frac{\sum(X_i - \bar{X})\,(Z_i - \bar{Z})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Z_i - \bar{Z})^2}} \quad (1)$$

where $\bar{X}$ and $\bar{Z}$ are means of $X$ and $Z$ respectively.

*Mean absolute error* (MAE) and *root mean squared error* (RMSE) are both widely used in prediction tasks to measure the amount of deviation of the predicted values from the actual values. The two are defined in the following way:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|z_i - \hat{z_i}| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(|z_i - \hat{z_i}|)^2} \quad (3)$$

Where n is the number of predictions, $z_1, \ldots, z_n$ are the actual and $\hat{z_1}, \ldots, \hat{z_n}$ are the predicted values respectively (Moore, 2007).

## 2.3 Classification

The different levels of mRS scores can be viewed as different categories and hence predicting the mRS score can be viewed as a multi-class classification problem. We consider three classifiers in this study. Two of them are widely used classification algorithms: logistic regression (McCullagh, 1980) and C4.5 decision tree (Quinlan, 1993). The choice of logistic regression is motivated by the fact that it is the standard classification method used in clinical trial studies. As for decision tree, it gives a good diagrammatic representation of the prediction process as well as proving to be empirically successful in classification tasks.

The other classification method in this study is actually one that uses the results of the regression method involving bagging and model trees. Once a numeric prediction is obtained from the regression method, we round it to the nearest integer and assign the instance to the class corresponding to that integer. We denote this approach here as *classification via regression*.

The evaluation criterion for the classification algorithms used in this study is *accuracy* – the percentage of cases where the actual and the predicted classes are the same. For the prediction of mRS-90 score, however, we may consider a predicted score which is close enough to the actual score to be fairly accurate as well. We therefore define "*near-accuracy*" to be the percentage of cases where the prediction is either fully correct or is incorrect by a margin of just one score, and use it as an additional evaluation metric.

# 3 RESULTS

## 3.1 Regression Models to Predict mRS-90

Supervised regression is performed on the stroke data to predict the patient outcome after 90 days of stroke onset. The target attribute is mRS-90, the mRS score after 90 days, and the predictive

attributes are all the other attributes described in Table 2. We construct an M5 model tree and compare its results with linear regression, the most commonly used method for regression analysis. We then apply bootstrap aggregating (bagging) using M5 model trees and separately linear regression models as respective base predictors. For comparison purposes, we construct also the simple regression model whose prediction is always the average of the values of the dependent variable in the training set.

Parameter optimization is done for both model tree and bagging. For M5 model trees, we experiment with the minimum number of instances to allow in a leaf. It is found that having a minimum of 10 instances in the leaf produces the best performing tree. Increasing this number creates shorter trees that underfit the data while reducing this number creates larger trees that are prone to overfitting. For bagging, we experiment with different number of iterations for bootstrapping (number of bags) and different bootstrap sizes. Our conclusion is that 10 iterations with each bootstrap containing the same number of instances as the training set produces the best results.

Table 3 compares the results of these five methods in terms of correlation coefficient (R), mean absolute error (MAE) and root mean squared error (RMSE). We can observe from the table that bagging used in tandem with M5 model trees performs much better than all the other techniques. An interesting observation is that M5 model tree (without bagging) shows an impressive improvement over linear regression in terms of mean absolute error, but performs only slightly better in terms of root mean squared error. Large errors have a relatively greater influence when the errors are squared. So as the variance associated with the frequency distribution of the error magnitude increases, the difference between MAE and RMSE also increases (Willmott and Matsuura, 2005). It therefore makes sense that a variance-reducing procedure like bagging should reduce RMSE when applied to model trees, as observed in Table 3. Note also that bagging does not have the same kind of effect in improving the performance of linear regression.

Table 3: Comparison of different regression methods on stroke data in terms of R, MAE and RMSE. For R, higher values indicate better model fit, whereas for the MAE and RMSE metrics lower values are better.

| Method | R | MAE | RMSE |
|---|---|---|---|
| Average Prediction | -0.136 | 1.235 | 1.461 |
| Linear regression | 0.779 | 0.654 | 0.916 |
| M5 model tree | 0.785 | 0.577 | 0.905 |
| Bagging with Linear Regression | 0.783 | 0.649 | 0.908 |
| Bagging with M5 model trees | 0.822 | 0.537 | 0.832 |

To see if the improvement is statistically significant, we perform paired t-tests in terms of correlation coefficient on each pair of the four methods considered. The difference between means for each pair are examined at a p-value of 0.05. The results of the tests are presented in Table 4, showing that the bagging method with M5 model trees performs significantly better than the other four methods on the stroke dataset.

Table 4: Results of statistical significance analysis on correlation coefficient with p-value of 0.05. Each cell represents the result of the paired t-test between a pair of algorithms. If the algorithm in the row is significantly better than the one in the column, a '>>' is shown. If it is significantly worse, a '<<' is shown. A '<->' indicates that there is no statistically significant difference.

| | Avg Pred | Lin Reg | M5 tree | Bagging Lin Reg | Bagging M5 trees |
|---|---|---|---|---|---|
| Avg Pred | - | << | << | << | << |
| Lin Reg | >> | - | <-> | <-> | << |
| M5 tree | >> | <-> | - | <-> | << |
| Bagging Lin Reg | >> | <-> | <-> | - | << |
| Bagging M5 trees | >> | >> | >> | >> | - |

### 3.1.1 Observations from the M5 Model Tree

We investigate the model returned by the M5 model tree algorithm to find insights about stroke outcome. Figure 3 shows the model tree where each leaf is a linear equation. The equations appear below. The sign and magnitude of coefficients of each predictive attribute in the equations give an indication of how the output attribute responds to changes in the given input attribute. The continuous variables age and NIHSS at admission are scaled to the range between 0 and 1, so that the magnitudes of all attributes are within the [0,1] range.

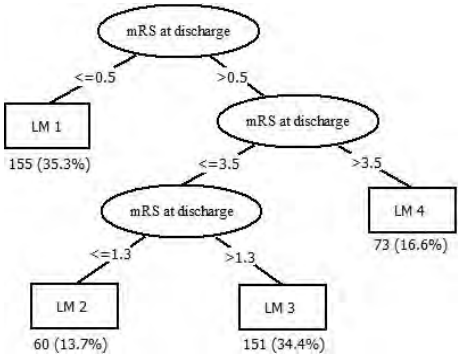



Figure 3: The M5 model tree built on the stroke dataset with minimum 10 instances in each leaf. Each leaf is a linear model predicting the target attribute mRS-90. The numbers under the leaves indicate how many instances are covered under that particular linear model.

LM 1 (here the value of mRS at discharge is 0)
mRS 90 days =
- 0.1309 * Subtype - Large Vessel
- 0.1472 * Subtype - Small Vessel
- 0.1552 * Subtype - Cardio
- 0.0532 * Subtype - Crypto
- 0.1454 * Subtype - other
+ 0.064 * NIHSS at admission
+ 0.0189 * MRS before admission
+ 0.0996 * Age
+ 0.0155 * Diabetes
- 0.0472 * Antiplatelets
+ 0.0534 * mRS at discharge
+ 0.1285

LM 2 (here the value of mRS at discharge is 1)
mRS 90 days =
0.0446 * Subtype - Large vessel
- 0.0837 * Subtype - Small vessel
- 0.4857 * Subtype - Cardio
- 0.6028 * Subtype - Crypto
- 0.0827 * Subtype - other
+ 0.3298 * NIHSS at admission
+ 0.084 * MRS before admission
+ 0.4344 * Age
+ 0.0959 * Diabetes
- 0.0137 * Tobacco
+ 0.2618 * Antihypertensives
- 0.0057 * Antiplatelets
+ 0.1265 * mRS at discharge
+ 0.3596

LM 3 (here the value of mRS at discharge is 2 or 3)
mRS 90 days =
0.3911 * Subtype - Large vessel
- 0.0837 * Subtype - Small vessel
- 0.0882 * Subtype - Cardio
- 0.0832 * Subtype - Crypto
- 0.807 * Subtype - other
+ 1.5475 * NIHSS at admission
+ 0.3333 * MRS before admission
+ 1.5486 * Age
+ 0.4281 * Diabetes
- 0.0137 * Tobacco
- 0.0057 * Antiplatelets
+ 0.0951 * mRS at discharge
- 0.3414

LM 4 (here the value of mRS at discharge is 4 or 5)
mRS 90 days =
- 0.0119 * Subtype - Large vessel
- 0.0837 * Subtype - Small vessel
- 0.0882 * Subtype - Cardio
- 0.0832 * Subtype - Crypto
- 0.0827 * Subtype - other
+ 0.1919 * NIHSS at admission
+ 0.0438 * MRS before admission
+ 0.2979 * Age
+ 0.0567 * Diabetes
- 0.0351 * Tobacco
- 0.0057 * Antiplatelets
- 0.4463 * Neurointervention
+ 1.4419 * mRS discharge
- 3.0914

From the model tree of Figure 3, it is clear that mRS at discharge plays the major role in deciding the mRS score at 90 days. The tree simply first decides what the mRS discharge score is, and then builds linear models to predict mRS-90 for the patients with that score. By following the decision branches of the tree, we can see that the linear models LM 1 and LM 2 corresponds to mRS discharge scores of 0 and 1 respectively. Similarly LM 3 is associated with mRS discharge scores of 2 and 3, and LM 4 with scores of 4 and 5.

Looking at LM 1, we find that the y-intercept is a very small value and there is no other attribute that has a large coefficient that could change the prediction substantially. This means that the prediction for almost all patients reaching this point of the tree will be close to 0. At LM 2, since the mRS discharge score is 1 with a coefficient of 0.1265 and the y-intercept is 0.3596, the baseline prediction for this leaf (if all other conditions are not present) is 0.4861. Older age, higher NIHSS at admission and presence of antihypertensives contribute towards increasing the mRS-90 score. On the other hand, cardioembolic and cryptogenic strokes contribute significantly towards lowering the mRS-90 score. At LM 3, if the mRS discharge score is 2, then the baseline prediction is 2*0.0951 – 0.3414 = - 0.1512. If the mRS discharge = 3, it is 3*0.0951 – 0.3414 = - 0.0561. However, there are some attributes in this model that may have a major impact on the final prediction, notably age, NIHSS at admission, diabetes, large vessel stroke subtype and mRS before admission. Higher values for some or all of the above attributes will result in increased

mRS-90 score. For LM 4, the baseline prediction is either 2.6762 (for mRS discharge = 4) or 4.1181 (for mRS discharge = 5). If a patient reaches this leaf, the output is likely to be quite high, since only neurointervention has a major effect of lowering the mRS-90 score.

## 3.2 Classification Models to Predict mRS-90

We now consider the mRS-90 attribute as discrete (i.e., consisting of individual classes 0, 1, …, 5) instead of a continuous numeric attribute, and construct classification models to predict this discrete attribute. We explore two main approaches to constructing classification models: One is to apply traditional multi-class classification techniques; another one is to use regression followed by classification (i.e., classification via regression). For this experiment we choose two well-known and empirically successful classification algorithms, namely logistic regression and C4.5 decision tree. For classification via regression we use the bagging with M5 model tree method discussed in section 3.1, and convert the predicted mRS-90 numeric value to a discrete class by rounding this value to the nearest integer between 0 and 5.

As a first evaluation metric, we use classification accuracy (the percentage of correct predictions). But since there are six different classes with subtle variations between two adjacent mRS scores, we also consider the case when the classifier makes an error, but by only one mRS score. We define the metric "near-accuracy" to refer to the percentage of cases in which the classifier either makes an accurate prediction or makes a wrong prediction which is either one more or one less than the correct mRS score.

Table 5 shows a comparison of the performance of classification via regression with those of multi-class classification using Logistic regression and C4.5 decision trees. For comparison purposes, we include also that majority class classifier which classifies any test instance with the mRS-90 value that appears most frequently in the training set.

For C4.5 decision trees, the result of the best model after experimentation with pruning is shown. The classification via regression method performs better in terms of both accuracy and near-accuracy. Table 6 shows the confusion matrix obtained by this method. Paired t-tests are performed on the classification accuracy for the three algorithms. The results, given in Table 7, show that classification via regression performs significantly better than logistic regression, but not significantly better than the C4.5 decision tree at a level of p = 0.05.

Table 5: Comparison of logistic regression, C4.5 and classification via regression (bagging with M5 model trees) on the stroke dataset in terms of accuracy and near-accuracy.

| **Method** | **Accuracy** | **Near-accuracy** |
|---|---|---|
| Majority class | 46.9% | 64.4% |
| Logistic Regression | 54.2% | 83.6% |
| C4.5 (with pruning) | 56.7% | 86.8% |
| Classification via regression | 59.7% | 90.0% |

Table 6: Confusion matrix for the method of supervised classification via regression using bagging with M5 model trees. The rows show the actual mRS scores while the columns show the ones predicted by the model. The diagonals (in bold) are the correct predictions. The cells adjacent to the diagonals (in bold and italic) are near-correct predictions missing the actual score by 1.

| Actual | Predicted | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| 0 | **159** | ***36*** | 11 | 0 | 0 | 0 |
| 1 | ***10*** | **40** | ***19*** | 8 | 0 | 0 |
| 2 | 2 | ***15*** | **31** | ***14*** | 1 | 0 |
| 3 | 0 | 8 | ***19*** | **21** | ***3*** | 0 |
| 4 | 0 | 3 | 5 | ***8*** | **10** | ***1*** |
| 5 | 0 | 3 | 1 | 2 | ***8*** | **1** |

Table 7: Results of statistical significance analysis on classification accuracy with p-value of 0.05. Each cell represents the result of the paired t-test between a pair of algorithms. If the algorithm in the row is significantly better than the one in the column, a '>>' is shown. If it is significantly worse, a '<<' is shown. A '<->' indicates that there is no statistically significant difference.

| | Majority class | Logistic Regression | C4.5 tree | Classif via regression |
|---|---|---|---|---|
| Majority class | - | << | << | << |
| Logistic Regression | >> | - | <-> | << |
| C4.5 tree | >> | <-> | - | <-> |
| Classif via regression | >> | >> | <-> | - |

# 4 CONCLUSIONS

This paper has presented the results of predicting the 90-day outcome of stroke patients based on the data consisting of demographics, medical history and treatment records of ischemic stroke patients. The problem of prediction is treated first as the regression task of predicting the numeric score according to the modified Rankin Scale which measures the degree of disability in patients who have suffered a stroke. A meta-learning approach of bootstrap aggregating (bagging) using M5 model trees as the base learner proved to be a very effective regression technique in this case, significantly outperforming other more commonly used regression methods. The same method, after translation of the target output from numeric to nominal, performs better as a multi-class classification scheme than other commonly used classifiers.

The high performance of the M5 model tree can be attributed to the fact that the mRS score at 90 days is highly dependent on one of the attributes - the mRS score at discharge from the hospital. Therefore, a model predicting mRS-90 would do well by dividing the input space into a number of subspaces defined around the value of mRS at discharge, building a separate specialized model for each of the subspaces. A model tree does exactly that. Examination of the M5 model tree that is constructed on the stroke dataset reveals that the tree simply directs the prediction task towards different ranges of values for the mRS score at discharge. A multivariate linear regression model is then built for each of the leaves, which are more specialized for predicting the outcome of those particular patients. The superior performance of bagging in enhancing the prediction results can be explained by the variance in error of the base M5 model trees. By examining the model tree prediction errors for the stroke dataset considered, it is found that the variability of errors is much higher for model trees than for other regression methods such as logistic regression. Since bagging is empirically known to reduce the instability and error variance of its base learners, it shows good performance for this particular dataset.

Further examination of the models reveals interesting insights into how different factors affect stroke outcome. It is found, rather unsurprisingly, that patients who have a low mRS score ($\leq 1$) at discharge tend to maintain a low mRS score at 90 days as well. However, patients who have some minor disability (mRS = 1) at discharge tend to have poorer outcome if they have older age, more severe initial stroke and hypertension, while patients suffering from cardioembolic or cryptogenic types of stroke actually make a better recovery. The patients who have slight or moderate disability at the time of discharge (mRS 2 or 3) may end up in a wide spectrum of outcomes at 90 days based on several factors; older age, more severe initial stroke, presence of diabetes, preexisting disability before stroke and large vessel thrombosis are associated with poorer outcome. For patients who have fairly severe disability at the time of discharge (mRS 4 or 5), only neurointervention performed during the hospital stay has the effect of improving the recovery rate after discharge and within 90 days of stroke.

One limitation of the study is the exclusion of the patients who died within 90 days of stroke. As mentioned before, this is in line with other work in the literature (e.g., the Copenhagen Stroke Study (Nakayama et al., 1994)), but it would be interesting in future work to extend our approach to include these patients. We are also limited by a large amount of missing values in attributes that are not included in this study but which may have been instrumental in stroke outcome prediction. In the future we would like to address these shortcomings to develop better models for prediction. Another future goal is to improve the process of classification via regression by discovering better ways to translate the numeric predictions to discrete classes.

# ACKNOWLEDGEMENTS

The authors would like to thank Theresa Inzerillo and Preston Mueller of Worcester Polytechnic Institute for stimulating discussions around the analysis of stroke data.

# REFERENCES

Adams, H. P., Bendixen, B. H., Kappelle, L. J., Biller, J., Love, B. B., Gordon, D. L. & Marsh, E. E. 1993. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke,* 24**,** 35-41.

Aslam, J. A., Popa, R. A. & Rivest, R. L. 2007. On Estimating the Size and Confidence of a Statistical Audit. *USENIX/ACCURATE Electronic Voting Technology Workshop,* 7**,** 8.

Banks, J. L. & Marotta, C. A. 2007. Outcomes validity and reliability of the modified Rankin scale:

Implications for stroke clinical trials a literature review and synthesis. *Stroke,* 38**,** 1091-1096.

Breiman, L. 1996. Bagging predictors. *Machine learning,* 24**,** 123-140.

Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. 1984. *Classification and regression trees*, CRC press.

Brott, T., Adams, H., Olinger, C. P., Marler, J. R., Barsan, W. G., Biller, J., Spilker, J., Holleran, R., Eberle, R. & Hertzberg, V. 1989. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke,* 20**,** 864-870.

Brown, A. W., Therneau, T. M., Schultz, B. A., Niewczyk, P. M. & Granger, C. V. 2015. Measure of functional independence dominates discharge outcome prediction after inpatient rehabilitation for stroke. *Stroke,* 46**,** 1038-1044.

Etemad-Shahidi, A. & Mahjoobi, J. 2009. Comparison between M5′ model tree and neural networks for prediction of significant wave height in Lake Superior. *Ocean Engineering,* 36**,** 1175-1181.

Gialanella, B., Santoro, R. & Ferlucci, C. 2013. Predicting outcome after stroke: the role of basic activities of daily living predicting outcome after stroke. *European journal of physical and rehabilitation medicine,* 49**,** 629-637.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H. 2009. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter,* 11**,** 10-18.

Henninger, N., Lin, E., Baker, S. P., Wakhloo, A. K., Takhtani, D. & Moonis, M. 2012. Leukoaraiosis predicts poor 90-day outcome after acute large cerebral artery occlusion. *Cerebrovascular Diseases,* 33**,** 525-531.

Keith, R., Granger, C., Hamilton, B. & Sherwin, F. 1987. The functional independence measure. *Adv Clin Rehabil,* 1**,** 6-18.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI.*

Marini, C., De Santis, F., Sacco, S., Russo, T., Olivieri, L., Totaro, R. & Carolei, A. 2005. Contribution of atrial fibrillation to incidence and outcome of ischemic stroke results from a population-based study. *Stroke,* 36**,** 1115-1119.

McCullagh, P. 1980. Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological)***,** 109-142.

Moonis, M., Kane, K., Schwiderski, U., Sandage, B. W. & Fisher, M. 2005. HMG-CoA reductase inhibitors improve acute ischemic stroke outcome. *Stroke,* 36**,** 1298-1300.

Moore, D. S. 2007. *The basic practice of statistics*, New York, WH Freeman

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., Das, S. R., de Ferranti, S., Després, J.-P. & Fullerton, H. J. 2016. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. *Circulation,* 133**,** 447.

Nakayama, H., Jørgensen, H., Raaschou, H. & Olsen, T. 1994. The influence of age on stroke outcome. The Copenhagen Stroke Study. *Stroke,* 25**,** 808-813.

Nogueira, R. G., Liebeskind, D. S., Sung, G., Duckwiler, G., Smith, W. S. & Multi MERCI Writing Committee 2009. Predictors of good clinical outcomes, mortality, and successful revascularization in patients with acute ischemic stroke undergoing thrombectomy pooled analysis of the Mechanical Embolus Removal in Cerebral Ischemia (MERCI) and Multi MERCI Trials. *Stroke,* 40**,** 3777-3783.

Quinlan, J. R. 1992. Learning with continuous classes. *5th Australian joint conference on artificial intelligence.* Singapore.

Quinlan, J. R. 1993. *C4. 5 Programs for Machine Learning*, San Francisco, Morgan Kauffmann.

Raffeld, M. R., Debette, S. & Woo, D. 2016. International Stroke Genetics Consortium Update. *Stroke,* 47**,** 1144-1145.

Rankin, J. 1957. Cerebral vascular accidents in patients over the age of 60. II. Prognosis. *Scottish medical journal,* 2**,** 200.

Rodgers, J. L. & Nicewander, W. A. 1988. Thirteen ways to look at the correlation coefficient. *The American Statistician,* 42**,** 59-66.

Tan, P.-N., Steinbach, M. & Kumar, V. 2005. *Introduction to data mining*, Boston, Addison-Wesley.

Van Swieten, J., Koudstaal, P., Visser, M., Schouten, H. & Van Gijn, J. 1988. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke,* 19**,** 604-607.

Wang, Y. & Witten, I. H. 1996. Induction of model trees for predicting continuous classes. *European Conference on Machine Learning.* University of Economics, Prague.

Willmott, C. J. & Matsuura, K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research,* 30**,** 79-82.

Yong, M. & Kaste, M. 2008. Dynamic of hyperglycemia as a predictor of stroke outcome in the ECASS-II trial. *Stroke,* 39**,** 2749-2755.

# Short Papers

# Sign-Lingo
## *Feasibility of a Serious Game for Involving Parents in the Language Development of Their Deaf or Hearing Impaired Child*

Ivor van der Schalk and Marco Spruit
*Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands*
*{i.l.vanderschalk, m.r.spruit}@uu.nl*

Keywords: Feasibility Study, Serious Game, Sign Language, Kinect, Language Development.

Abstract: Family involvement plays a critical factor in the language development of a deaf or hearing impaired child. Hearing parents often have major difficulties in communicating with their child when it is deaf or hearing impaired. These difficulties often lead to issues in the language development of the child. In this research we investigate the feasibility of a serious game for involving parents in the language development of their deaf or hearing impaired child by using sign language together in a fun and engaging way. From the feasibility analysis we find that such a serious game is feasible and could help deaf and hearing impaired children to improve their language development.

## 1 INTRODUCTION

About 3 out of every 1,000 children are born either deaf or with a detectable level of hearing loss in one or both ears. More than 90% of these children are born to hearing parents (NIDCD, 2016). Language development requires a child to have access to communication (Marschark, 2001). Unlike children with no hearing impairment it is difficult or even impossible for deaf and hearing impaired children to use a spoken language. These children mainly use Sign language as communication for language development.

Hearing parents often have major difficulties in communicating with their child when it is deaf or hearing impaired. This obstacle makes it difficult for parents to be involved in the language development of the child. Previous research has shown that Serious games can be used to improve a child's language development by using natural language (Sørensen and Meyer, 2007). This makes one wonder whether or not a Serious game can also be used to improve the language development of a deaf or hearing impaired child by using Sign language.

The Kinect device offers the possibility to create a fun and engaging Serious game for both the child and its parents. So far only very limited research has been performed on Serious games that are intended to be played together by deaf and hearing impaired children and their parents.

In this research we conduct a feasibility study for a Serious game which helps involving parents of deaf and hearing impaired children to use Sign language together in a fun and engaging way. The main goals of the game, which we call Sign-Lingo, are to improve the child´s language development and to improve the interaction between the child and its parents. Through this research it will become clear how feasible such a Serious game is.

## 2 RESEARCH METHOD

In this research a feasibility study is performed based on the CORETEST feasibility study framework (Meulendijk *et al.*, 2013). In this study the feasibility of all five aspects are explored. These aspects are:

- conceptual feasibility
- organizational feasibility
- economic feasibility
- technological feasibility
- societal feasibility

The feasibility of these aspects is explored by performing a qualitative study. Note that, instead, Aarnoutse *et al.* (2016) take a systematic literature study (SLR) approach to study their application's feasibility.



Schalk, I. and Spruit, M.
Sign-Lingo - Feasibility of a Serious Game for Involving Parents in the Language Development of Their Deaf or Hearing Impaired Child.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 191-198
ISBN: 978-989-758-213-4

In this study the initial approach was to explore the field of Sign language and Serious games. From a literature study it became clear there is a problem in the language development of deaf and hearing impaired children. A feasibility study of a Serious game in which Sign language is used by deaf and hearing impaired children to improve their language development was then conducted. Among the studied literature are research journals, papers and books about Sign language, Serious games and studies about deaf and hearing impaired children.

In the feasibility study several models have been developed using the modeling tool ArchiMate (Beauvoir, 2016).

# 3 CONCEPTUAL FEASIBILITY

## 3.1 Problem Exploration

In their work Kyle and Woll (1988) describe that the development of language can only occur when a child is provided with input it can perceive, and when the adult and the child are joint partners in creating communication. The importance of parent involvement in language learning is supported by the research of Moeller (2000) in which she shows that family involvement plays a critical factor in the language development of deaf and hearing impaired children, especially those with hearing parents.

The majority of deaf and hearing impaired children are born to hearing parents who do not know how to use Sign language to communicate. As a result, many of these children do not have full access to language during the early years of life most critical to language acquisition. Marschark (2001) states that having parents who can sign well and who read regularly with their deaf or hearing impaired child are extremely important factors in the child's development of literacy skills. This statement might explain the findings of Allen (1994) who found that only 15 percent of white deaf children who graduated from high school in the United States read above the sixth-grade level.

Akamatsu, Musselman and Zweibel (2000) found that fewer than half of the children who use Sign language in school also sign when they are with their families, and only a small fraction of those children are able to carry on normal everyday conversations with their parents. Marschark (2001) states that deaf children of hearing parents have fewer signed or spoken labels for things around them than hearing children of hearing parents, or deaf children of deaf parents. Special efforts, therefore, need to be made to expand the vocabularies of deaf children of hearing parents through print, sign and speech.

It is clear there is a problem in the language development of deaf and hearing impaired children with hearing parents. There is evidence that the involvement of parents plays a crucial factor in the

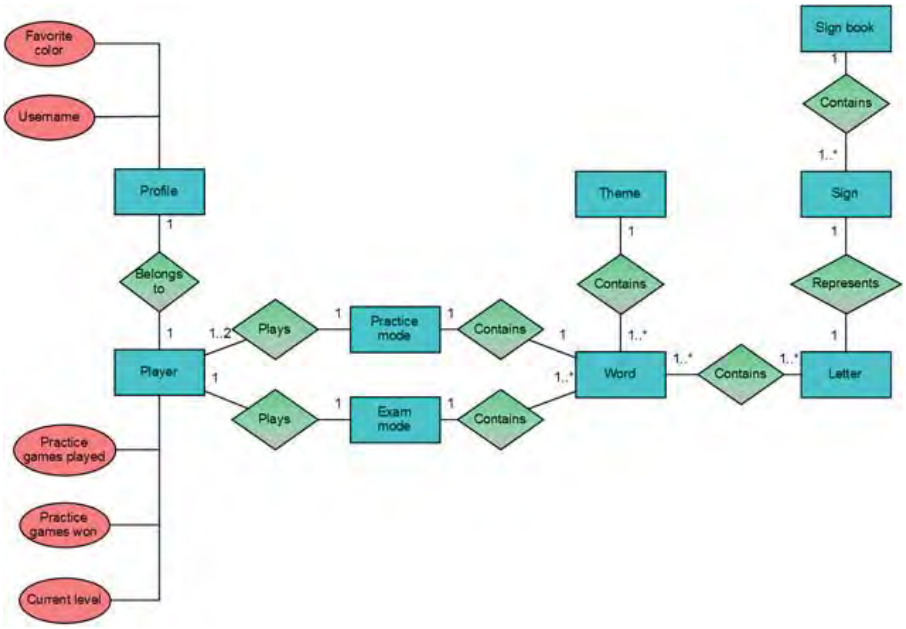



Figure 1: Conceptual model of Sign-Lingo.

language development of their deaf or hearing impaired child but that this involvement is insufficient due to the parents' lack in ability to communicate with their deaf or hearing impaired child.

## 3.2 Concept Modeling

In the previous section it has become clear there is a problem in the language development of deaf and hearing impaired children with hearing parents. In this section a conceptual model is presented of the proposed system. This model visualizes the main concepts of the system. The model is shown in Figure 1.

As can be seen in the model, the system contains nine concepts. These concepts are highlighted in blue rectangles inside the model. Within the model we can also find red circles which represent the attributes of concepts.

Each player has a profile in which general information of the player is stored. This information contains the name of the player and its favorite color.

A player has the option to play practice-mode or exam-mode. In both game modes the player needs to guess words by using Sign language.

Practice-mode is meant to be played together by both the parent and child. The goal of this game mode is to facilitate family involvement in the language development of the child. In this game mode the player who guesses the word first wins the game.

In their work Michael and Chen (2006) mention the importance to measure the progress of the player in a Serious game. The exam-mode is meant be played by the child. The goal of this game mode is for the child to demonstrate its acquired vocabulary knowledge. In this game mode the child needs to guess multiple words correctly to advance to the next level. The difficulty of the words given in the game modes are based on this level.

Tennant and Brown (1998) explain it helps to fingerspell with a partner when contextual clues are used. Therefore, when starting a game mode, the player can select a theme. Examples of themes are animals, sports or family. Based on the theme the game presents words that are related to the selected theme.

From their pilot study Lee *et al.* (2005) found that the flow within a game is very important to keep the child interested in playing. In the Sign-Lingo game we have therefore included a sign book. This book contains examples of all the signs of the Alphabet which can be consulted by the player. This makes it easier for both the parent and child to continue playing even when they do not know how to make a sign.

For every player the number of practice games played, practice games won and the current level are stored. The current level is an indication of the current language level of the child. These statistics can be used to track the progress and target problem areas in the language development of the child.

## 3.3 Process Modeling

Goldin-Meadow and Mayberry (2001) explain that to teach children to read, the first step is to make sure they have a language - any language. In case of a deaf or hearing impaired child this will be a Sign language. After a language has been obtained a child needs to learn the concept of mapping between the language it knows and print. In their research Padden and Ramsey (2000) identified a teaching method that they call "chaining". This method encourages children to see the relation between print and Sign language. Chaining consists out of four steps in which the final step is optional:

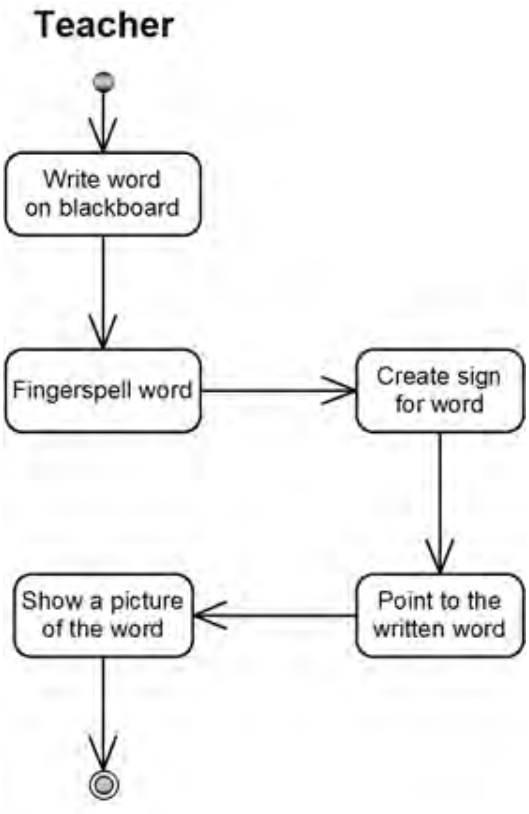



Figure 2: Chaining method.

1. Fingerspell a word.
2. Create the sign that represents the finger spelled word.
3. Point to the written word.
4. Show a picture which represents the word (optional).

Fingerspelling is used to express letters from the Alphabet. Tennant and Brown (1998) explain that fingerspelling is mainly used in Sign language to express names of people, places, brand names and titles. Fingerspelling is performed by creating the sign with the dominant hand. In the chaining method for example a teacher could fingerspell the word 'd-o-g', then create the sign that represents the word 'dog' and finally point to the word 'dog' written down on the blackboard. In addition to that the teacher can show a picture of a dog. This method creates an understanding between the visual spelling of a word (e.g. in English) and the Sign language spelling of the word. This method is presented in Figure 2 and incorporated in the Sign-Lingo game.

# 4 ORGANIZATIONAL FEASIBILITY

## 4.1 Market Modeling

In the previous section we modeled the system's concepts and its interaction with its users. In this section the organizational feasibility is assessed by looking into the market in which the proposed system is intended to participate. Through this assessment it becomes clear who the system's key players, competitors and potential partners are after which an organizational framework can be built.

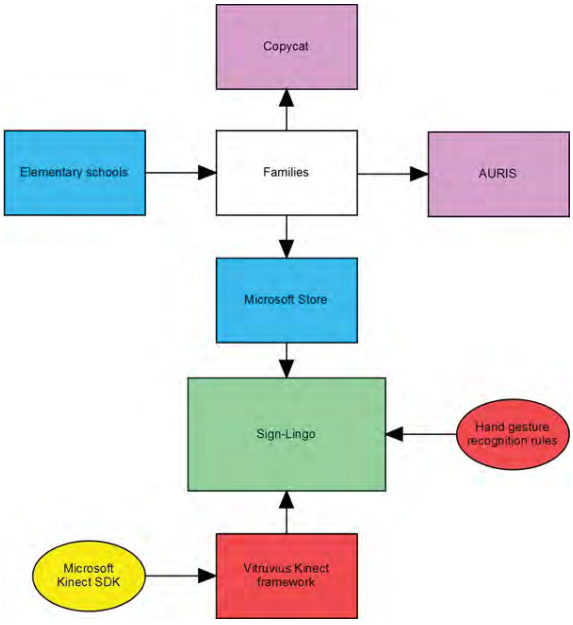



Figure 3: Stakeholder's relationships of Sign-Lingo.

The model presented in Figure 3 shows an overview of Sign-Lingo system's stakeholders. This diagram visualizes the relations between the system and the stakeholders and how the system behaves in relation to third-party software.

In the model the red rounded box contains the rules that need to be incorporated into the system. The rounded yellow box contains third-party software that is required for the system to function properly. The red rectangles in the diagram represent the plugins which are required for the system to function. The blue rectangles represent the partners which are involved in the promotion of the system. The white rectangle represent the system's users. Finally, the purple rectangles represent the competition of other Sign language games related to language development.

The system requires hand gesture recognition rules to recognize letters from the Alphabet in Sign language. The system makes use of the Microsoft Kinect SDK which is designed to help recognize gestures. The system will be built on the Vitruvius Kinect framework which simplifies Kinect development.

The system is used by families with deaf or hearing impaired children. These families can be reached through elementary schools. The system is downloadable from the Microsoft Store.

Existing research has been conducted mainly on the development of Serious games that help a person to learn Sign language (Escudeiro *et al.*, 2015; Fisher *et al.*, 2014; Wilson, 2013), but not specifically on helping a deaf or hearing impaired child in its language development.

When looking at games related to language development for deaf or hearing impaired children we find CopyCat (Lee *et al.*, 2005) and AURIS (Sarmaşik, Serbetcioglu and Kut, 2009) produce interesting results. What differentiates Sign-Lingo from these games is the fact that these games are not based on an existing language development method, do not facilitate family involvement and do not track the progress of the player.

## 4.2 Organization Modeling

In the previous section the key players in the relevant market were identified. In this section the

user roles are identified, and the attitudes towards the newly proposed system are explored.

Through the literature review several methods were identified which are used to help children learn to develop their language. It also became clear parent involvement plays a crucial role in the language development of a child.

Figure 4 presents the organizational model of the Sign-Lingo system. The large rectangles represent the architectural layers on top of which the Sign-Lingo system is designed. The yellow rectangles represent all the stakeholders and their roles.

From the model it becomes clear the Sign-Lingo system is developed using the Microsoft Kinect SDK. The system is used by the parent and child who both have the role of user. It also becomes clear from the model that the developer needs to implement hand gesture recognition rules into the system. In addition to that, these hand gesture recognition rules are tested by a Sign language expert to validate whether the Sign language recognition part of the system performs as expected. Finally, teachers of elementary schools have the role as a promoter. Their task is to introduce the system to families with deaf or hearing impaired children.

# 5 ECONOMIC FEASIBILITY

## 5.1 Development Strategy

In the previous section the market in which the Sign-Lingo system operates was investigated. To reach out to families who could benefit from the Sign-Lingo system there are several options to consider. One option is to promote the system through digital channels such as application platforms like the Microsoft Store. Another option is to promote the system through other channels such as elementary schools for the deaf and hard hearing.

All partners have different objectives regarding Sign-Lingo. For example in case of the Microsoft Store the partner's interests are clearly commercial. In the case of elementary schools the interest is educational. The system will most likely need to have some relationship with the school's curriculum in order for the school to be willing to promote it.

Two different development strategies were created. These strategies can be run independently from each other. Below follows an overview of the two strategies.

1. Digital-channel strategy; the software is promoted on application platforms such as the Microsoft Store. In exchange for this promotion Microsoft wants a percentage of the selling price for every sold product. In addition to that a website is made for the Sign-Lingo system which explains how the system works and where the system can also be bought (*e.g.* Abdat, Spruit and Bos, 2011).
2. Physical-channel strategy; the software is promoted through elementary schools. These schools are contacted initially through

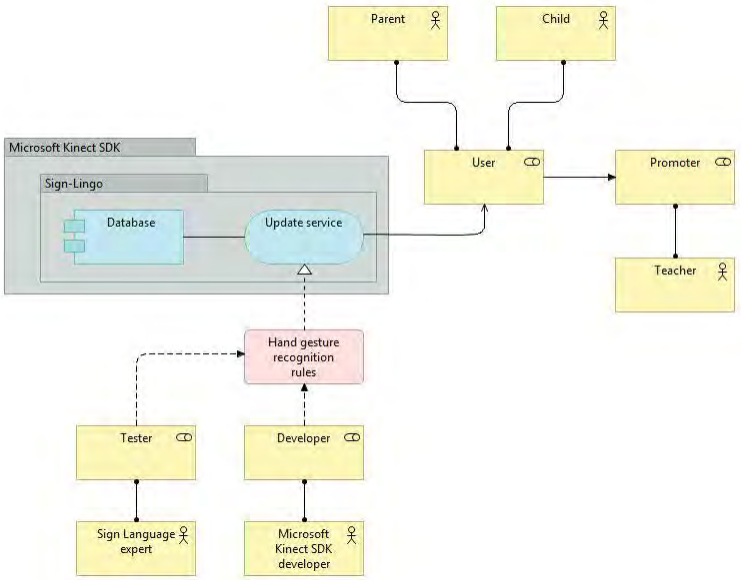



Figure 4: Organizational model of Sign-Lingo.

telephone calls in which the system and the concept behind it are presented to the school's management. Once a school agrees to promote the system teachers demonstrate the system to parents and their children and encourage them to use it.

What both strategies described above share, is the fact they both require partners in order to be successful. Without partnerships the system is unlikely to be used by anyone.

# 6 TECHNOLOGICAL FEASIBILITY

## 6.1 Technological Process Modelling

In this section the activity diagram of the use case 'Play practice-mode' of the Sign-Lingo game is presented. This is the most important and complicated use case in the system. The diagram is shown in Figure 5.

This process contains two actors: the player and

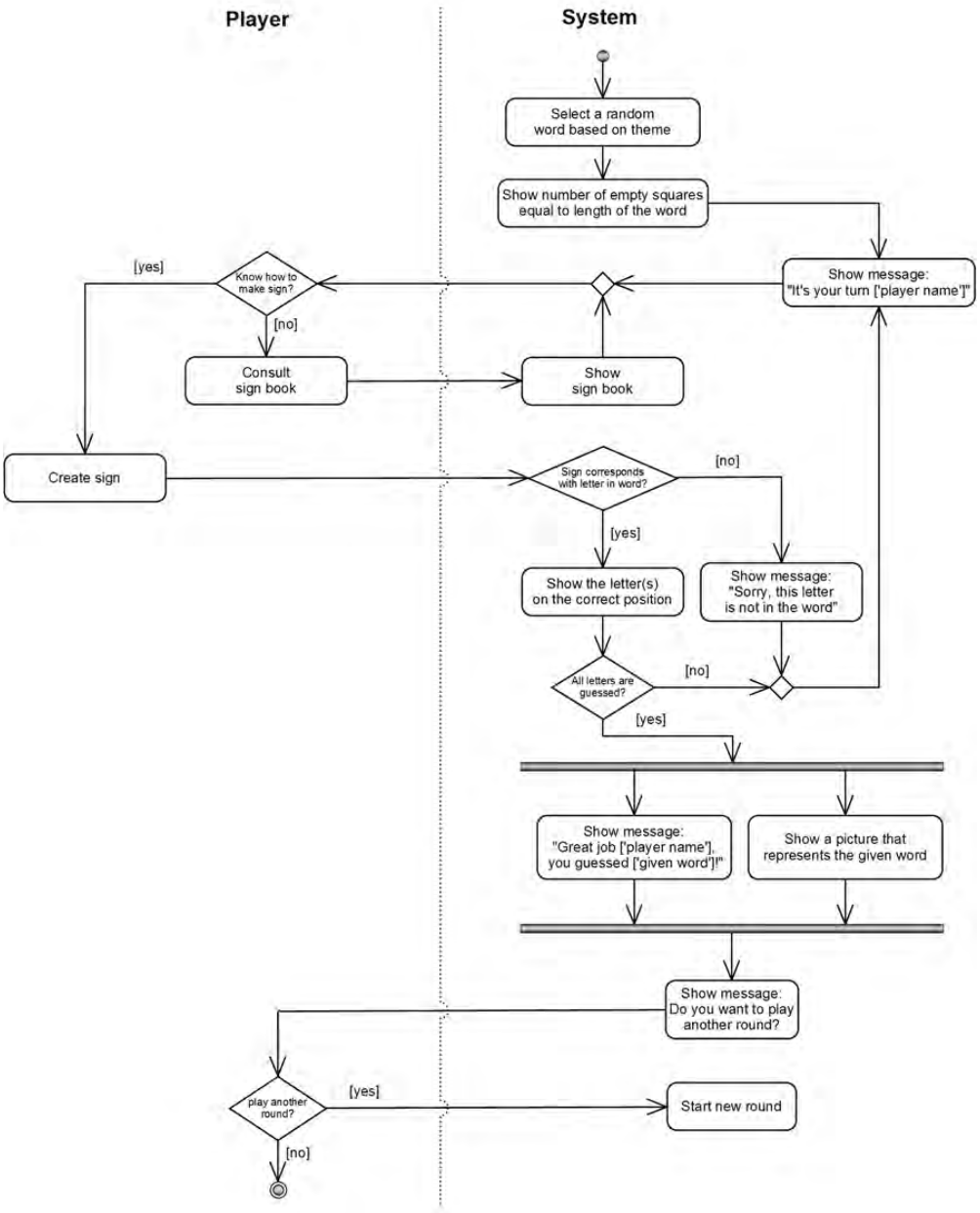



Figure 5: Activity diagram of the Sign-Lingo use case 'Play practice-mode'.

the system. Because this is a multiplayer game this process can be run by two players simultaneously. Before this process can be initiated the player needs to have executed the use case 'View game settings' in which the player has selected the theme and the word length. Based on these two parameters the system selects a random word from the system´s dictionary and displays a lingo screen which contains empty squares equal to the length of the random word. The player can then start to guess the given word.

Each turn, the player needs to guess the word by creating a sign gesture that corresponds with a letter. When the player does not know how to create a specific sign he can consult the 'sign book'. The sign book contains an overview of all the signs of the Alphabet. When the player was successful in making a sign the system will recognize this and check whether the given word contains the signed letter. If the word contains the signed letter the system will show the letter to the player on the correct position in the empty squares. When all the letters of the word are guessed the system shows a message to the player which confirms that the word has been guessed correctly and shows a picture that represents the guessed word. The player can then decide whether he wants to play another round or exit the game.

## 6.2 Wireframe

Figure 6 contains a wireframe of the Sign-Lingo game. In this sketch two players are playing the game in practice-mode and need to guess the word 'DOG'. In their work Tennant and Brown (1998) explain clues such as the length of the word are helpful for a child to improve its receptive skills which makes this game type well suited for language development.

As can be seen in the sketch the players can see each other in the game, this makes it possible for the players to maintain eye contact. Tennant and Brown (1998) explain that maintaining eye contact is an important aspect to understand a signer.

The game keeps track of how many rounds a player has won. The game also shows a message to indicate whose turn it is to guess a letter. Once a letter is guessed correctly it appears on the screen and its background turns green. When one of the players do not know how to make a sign it can consult the sign book which can be found in the bottom left corner.

Over time, we aim to optimize gameplay using machine learning techniques for personalization (*e.g.* Vleugel, Spruit and Daal, 2010), and even investigate the added value of factors such as

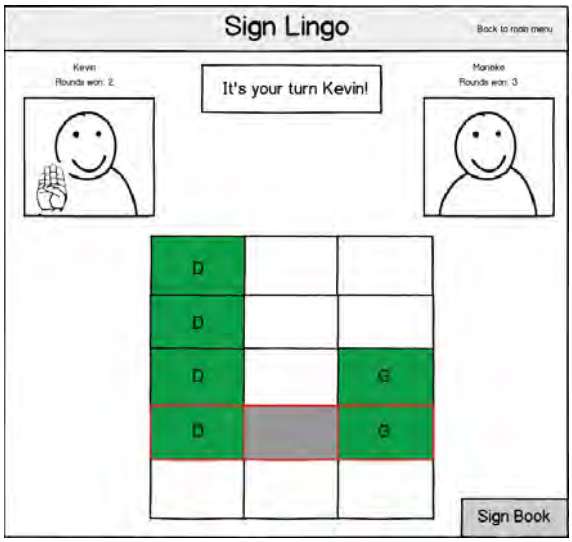



Figure 6: Wireframe of Sign-Lingo user interface.

playroom environment and its geographical location and local population density as possible language variation indicators (Heeringa *et al.*, 2007).

# 7 SOCIETAL FEASIBILITY

In the final step of the feasibility study we focus on how the proposed system can contribute to help deaf and hearing impaired children and their families. The gains can not only be described in monetary terms but in educational terms as well. These gains can be assessed by measuring the language level of children who use the proposed system and compare them with children who do not use the system. The monetization gains can be measured by estimating how much the education associated to the language development of a deaf or hearing impaired child costs and how much of the education can be replaced by the Sign-Lingo system.

# 8 CONCLUSIONS

In this paper we presented a feasibility analysis of a Serious game for involving parents in the language development of their deaf or hearing impaired children by using Sign language together in a fun and engaging way. To assess the feasibility the conceptual, organizational, economic, technological and societal feasibility aspects, known as the CORETEST, have been investigated.

From the CORETEST it became clear there is a problem with the language development of deaf and hearing impaired children and a Serious game could

help these children to improve their language development. The chaining method which is used to help a child to improve its language is well suited to be incorporated in this Serious game.

In future research the proposed system could be implemented and then tested in a case study with a sample group. In this case study the language level of deaf and hearing impaired children who use the system could be compared with the language level of deaf and hearing impaired children who do not use the system to see what the effect of the game is on the language level of a child.

Finally, we aim to incorporate the lessons learned into follow-up research on the feasibility of Serious games for medication adherence in children and quality of care in long-term care (*e.g.* Spruit, Vroon and Batenburg, 2014).

## REFERENCES


Aarnoutse, F., Renes, C., Batenburg, R., & Spruit, M. (2016). *STRIPA: The potential usefulness of a medical app*. In Gasmelseid,T. (Ed.), Advancing Pharmaceutical Processes and Tools for Improved Health Outcomes (pp. 114–135). IGI Global.

Abdat, N., Spruit, M., & Bos, M. (2011). *Software as a Service and the Pricing Strategy for Vendors*. In Strader,T. (Ed.), Digital Product Management, Technology and Practice: Interdisciplinary Perspectives, Advances in E-Business Research (AEBR) Book Series (pp. 154–192). IGI Global.

Akamatsu, C. T., Musselman, C., & Zwiebel, A. (2000). Nature versus nurture in the development of cognition in deaf people. *The deaf child in the family and at school: Essays in honor of Kathryn P. Meadow-Orlans*, 255-273.

Allen, T. E. (1994). Who are the deaf and hard-of-hearing students leaving high school and entering postsecondary education. *A Comprehensive Evaluation of the Postsecondary Educational Opportunities for Students who are Deaf or Hard of Hearing, funded by the US Office of Special Education and Rehabilitative Services. Gallaudet University*.

Beauvoir, P. (2016). *Archi | Free ArchiMate Modelling Tool*. [online] Archimatetool.com. Available at: http://www.archimatetool.com/ [Accessed 29 Nov. 2016].

Escudeiro, P., Escudeiro, N., Reis, R., Baltazar, A. B., Rodrigues, P., Lopes, J., ... & Bidarra, J. (2015). A kinect game in the VirtualSign project: training and learning with gestures. *In ARTECH 2015: 7th International Conference on Digital Arts. Artech International*.

Fisher, J., Ellis, K., Willoughby, L., & Barca, J. (2014).
Taking a User Centred Design Approach for Designing a System to Teach Sign Language. ACIS.

Goldin-Meadow, S. and Mayberry, R. (2001). How Do Profoundly Deaf Children Learn to Read? Learning Disabilities Research and Practice, 16(4), pp.222-229.

Heeringa, W., Nerbonne, J., Bezooijen, R. van, & Spruit, M. (2007). [Geography and inhabitant numbers as clarifying factors of variation in the Dutch dialect region]. Tijdschrift voor Nederlandse taal- en letterkunde, 123(1), Kwantitatieve benaderingen in de taal- en letterkunde, 70–82.

Kyle, J. G., Woll, B., & Pullen, G. (1988). *Sign language: The study of deaf people and their language*. Cambridge University Press.

Lee, S., Henderson, V., Brashear, H., Starner, T., Hamilton, S., & Hamilton, H. (2005). *User-centered development of a gesture-based American Sign Language game*. In Proceedings of instructional technology and education of the deaf symposium. NTID, Rochester, NY.

Marschark, M. (2001). Language Development in Children Who are Deaf: A Research Synthesis. FORUM publications.

Meulendijk, M., Drenth-van Maanen, A., Jansen, P., Brinkkemper,, S., Numans, M. and Spruit, M. (2013). *Introducing the CORETEST feasibility analysis in medical informatics: a case study of a decision-supportive knowledge system in the Dutch primary care sector*. In Miranda, I., Cruz-Cunha, M., & Gonçalves, P. (Eds.), Handbook of Research on ICTs for Healthcare and Social Services: Developments and Applications, pp. 1066–1087, IGI Global.

Michael, D. and Chen, S. (2006). Serious games. Boston, MA: Thomson Course Technology PTR.

Moeller, M. P. (2000). Early intervention and language development in children who are deaf and hard of hearing. *Pediatrics*, *106*(3), e43-e43.

NIDCD (2016). Quick Statistics [NIDCD Health Information]. Retrieved 7 February 2016, from *http://www.nidcd.nih.gov/health/statistics/pages/quick.aspx.*

Padden, C., & Ramsey, C. (2000). American Sign Language and reading ability in deaf children. *Language acquisition by eye*, *1*, 65-89.

Sarmaşik, G., Serbetcioglu, B., & Kut, A. (2009) Computer Aided Education and Training Tool for Hearing Impaired Children: AURIS.

Sørensen, B. H., & Meyer, B. (2007). Serious Games in language learning and teaching–a theoretical perspective. In *Proceedings of the 3rd international conference of the digital games research association* (pp. 559-566).

Spruit, M., Vroon, R., & Batenburg, R. (2014). Towards healthcare business intelligence in long-term care: an explorative case study in the Netherlands. *Computers in Human Behavior*, *30*, 698–707.

Tennant, R. A., & Brown, M. G. (1998). *The American sign language handshape dictionary*. Gallaudet University Press.

Vleugel, A., Spruit, M., & Daal, A. van (2010). Historical data analysis through data mining from an outsourcing perspective: the three-phases method. *International Journal of Business Intelligence Research, 1*(3), 42–65.

Wilson, C. (2013). Portable Game Based Instruction of American Sign Language.

# Time Synchronization in Emergency Response Time Measurement

F. Jordan Srour[1] and Nabil Georges Badr[2]
[1]*Department of Information Technology and Operations Management, Adnan Kassar School of Business, Lebanese American University, Beirut, Lebanon*
[2]*Grenoble Graduate School of Business, Grenoble, France*
*jordan.srour@lau.edu.lb, nabil.badr@alumni.grenoble-em.com*

Keywords: Emergency Medical Dispatch Systems, Computer Aided Dispatch, Emergency Response Times, Time Data, Time Synchronization.

Abstract: Emergency response time reporting requires data provided by multiple systems. Whenever more than one system produces a time stamp, issues of time synchronization across systems manifest. As emergency response time targets are often short (8 minutes or less) and critical to public perceptions of service, errors in reporting these times are unacceptable. This article seeks to quantify the probability and magnitude of such errors through an empirical study of one emergency medical dispatch system.

## 1 INTRODUCTION

One needs only to "google" the expression "ambulance response times" to find a plethora of government agency and ambulance service provider reports on both targets for and reported performance indicators of response times. The most common definition for response time proffered in these reports is: "the period between when a call is recorded at the emergency operations center to when the ambulance arrives at the patient's address.", see eg. (New South Wales Government, 2016; The Capital Region of Denmark, 2016; Manitoba Canada, 2016). Immediately striking in this definition is the implicit recognition that at least two information system components must interface to yield an accurate response time – the operations center call recording system and the mobile, field reporting unit on-board the ambulance specifying arrival to the scene. The potential for time synchronization issues increases with each additional information and communication technology (ICT) that must interface in a system. This paper examines the response time data from a computer aided dispatch system with the aim of quantifying the prevalence of time synchronization data errors in the measurement of emergency response times.

The paper begins with a short review covering emergency response time targets and computing standards on time data recording and synchronization. Section 3 describes the case study setting. Sections 4 and 5 present the results and a conclusion with directives for improved time data management in emergency medical dispatch (EMD) systems.

## 2 BACKGROUND

Emergency response time targets, while highly disputed in the literature (Pons and Markovchick, 2002; Pons et al., 2005; Blackwell et al., 2009; Wilde, 2013), are still considered the standard against which ambulance services are measured. One of the most prevalent targets is an 8-minute response time. The National Health Service of the United Kingdom specifies that 75% of all top priority (Red level) calls should be responded to within 8 minutes; for less urgent calls the standard requires response within 19 minutes for 95% of all calls (Wankhade, 2011). Violations of response time service level standards can, in some locations, lead to the revocation of ambulance operating licenses (Clark County, IN, USA, 2014).

Despite these standards there is still wide variation in ambulance response times. Much of this variance can be attributed to the different systems of garnering and reporting such data. This issue is not particular to the UK, in a June 2015 article on the *EMS World* website, Dr. Bruce Moeller notes that "Time is easy to measure – response times are not." He continues by noting that a study in Florida found nine different definitions of response time in use. Interestingly, Dr. Moeller concludes that Pinellas County overcame this problem for the 19 providers in their



Srour, F. and Badr, N.
Time Synchronization in Emergency Response Time Measurement.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 199-204
ISBN: 978-989-758-213-4

system by using a regional computer-aided dispatch (CAD) system and a fixed definition of response time as the time from agency dispatch until arrival on scene (EMS World, 2015). This implies that the two requisite time stamps recorded by the CAD are considered completely accurate. While there is some basis for this assumption of accuracy as reported in a 1997 article on measuring response intervals in a 911 dispatch system, current time format standards have changed since that time (Campbell et al., 1997).

The International Organization for Standardization's standard ISO 8601 specifies a complete set of date and time formats (ISO, 2004). This standard is based on the Gregorian calendar and is flexible in that it allows one to represent time as both calendar dates and as ordinal dates. The standard also specifies a representation protocol for time zones – again with flexibility as one can specify local time, universal time, or an offset from universal time. Finally, ISO 8601 allows for the representation of time intervals as well. Of course, any standard as flexible as the ISO 8601 can also lead to complexity when used in specific settings. To that end, the Internet Engineering Task Force approved RFC 3339 which simplifies time stamp notation for Internet use limiting the format to yyyy-mm-ddThh:mm:ssZ or +/-hh:mm, where y stands for year, m for month, d for day, T is a character separating date from time, h stands for hour, m for minutes, and s for seconds; the Z is for universal (or Zulu) time or one can specify an offset from universal time (Klyne and Newman, 2002). The World Wide Web Consortium adopts a similar standard for web-based time and date stamps (W3C, 1997).

With regards to EMD, the critical elements to the time stamp standards noted above are the clear use of the yyyy-mm-dd protocol for dates. If one ICT within a system uses the American style yyyy-dd-mm format while another uses the yyyy-mm-dd protocol, then 12 days out of the year would yield reasonable but incorrect time intervals and an additional 11 days worth of data out of every month would yield illogically long intervals. Furthermore, specifying the time zone is also important in emergency medical services as changes between daylight savings time and standard time will influence the calculation of time intervals that span such clock changes. If the components of an EMD system are located in different time zones (eg. a remote server is used to capture and time stamp records), then having the ability to bring all times into one zone for time calculations is critical.

Due to the relative nature of time intervals (eg. a year may be 365 or 366 days depending on a leap year), the ability to accurately perform time related arithmetic is more complex than one might expect. As response time is ultimately the difference between two time stamps, being able to make that subtraction accurately and correctly is important. This calculation is generally performed outside of the ICT or EMD system. Nevertheless, the analytics tools that take possession of the data in order to produce the response time performance indicators should use state of the art time calculation protocols such as those recommended with the `lubridate` package of the software R (R Core Team, 2016; Grolemund and Wickham, 2011).

Regardless of the time stamp format used, the resulting response time calculation will only be as accurate as the time stamps themselves. Even when initially set accurately, real clocks will differ after some amount of time due to clock drift or skew, caused by clocks counting time at slightly different rates. Thus, clock skew in a distributed system must realize the same global time. Clock or time synchronization is a central topic in computer science and engineering that aims to coordinate otherwise independent clocks (Cristian, 1989) especially in distributed systems (Lamport, 1978). The oldest protocol for time synchronization in a network is the Network Time Protocol (NTP) which has been under continual development and updating since 1979 (Mills et al., 2010). The NTP works based on a hierarchical, peer-to-peer structure of computers and servers organized into strata with the top level strata containing a set of high-precision reference clocks. In contrast, the Institute for Electrical and Electronics Engineers also specifies a time synchronization protocol for networked measurement and control systems. Specifically, the Precision Time Protocol, encapsulated in standard, IEEE 1588, operates using a master-slave framework with corrections for both clock offsets and network delays (IEEE, 2008). To ensure the integrity of time stamps within EMD systems, synchronizing all computers to the same clocks using the same synchronization protocols is a must.

The remainder of this paper uses a prototype EMD system to understand the extent of time synchronization related data corruption on response time data when only limited synchronization protocols are implemented.

## 3 CASE STUDY SYSTEM

To achieve time synchronization over the network, our case used the NTP. As a time source, the Global Positioning System (GPS) is used for central clock synchronization. Although GPS time signals are accurate, clock skew still introduces time stamp issues

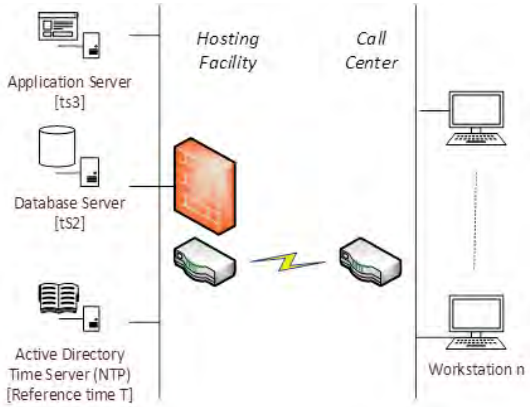



Figure 1: Overview of Case Study EMD.

into database records. The more computers involved in a system, the more likely that skew will be reflected in the data.

Specifically, in the emergency response system that we studied, the call center consists of a few (6-14) workstations taking calls and logging them into a system. The workstations at the call center are physically in a different location than the corresponding servers for data entry. The workstations and data entry servers are separated by a set of firewalls and connected via a Virtual Private Network (VPN), these two end points may incur delay in synchronizing their time. The system is depicted in Figure 1.

In order to follow the process of time stamping and its relationship to the process of time synchronization, Table 1 specifies the component, the time reference to which that component is synched, and a time stamp id associated with the component. As shown in Table 1, at regular intervals, the application and database servers set their clock to the reference time tS1 of the NTP Time Server. When a user/agent logs into a workstation session, the time stamp is reset to the Application Server time, thus at the workstation t = tS3. When opening a new call on the system, a key database record is created with a time stamp of tR1 which is equal to the database clock time at that instance, tS2. In this case, if the server clock is skewed when the record is created, but gets reset subsequently, then a timing error could be introduced. As the call progresses, all statuses and subsequent records are stamped with the workstation time tRn as it is the workstation that initiates the data entry request, tRn = t.

In the case study system, tRn could drift based on the internal workstation clock. If the workstation is not cycled at every shift, to synchronize its internal clock with the system, the absolute difference between tS3 and t will increase inducing errors in time measurements. For example, within 5 minutes a clock could drift 280 milliseconds. If the clock synchronization occurs on a 24 hour basis, then a total of 80 seconds of drift may occur between the system and workstation clock. As a result, time stamp errors arise. These errors could manifest themselves in "events happening in the past" (i.e. tRn $<$ tR1 = tS2) based on the inaccurate time stamp. Furthermore, due to the number of computers used in creating any one record the drift is not confined to that of one computer. As such, measuring and correcting drift from log files is an untenable task.

# 4 RESULTS

The system described in Section 3 was used to monitor the calls made to a nation-wide ambulance response service between 1 January 2016 and 30 June 2016. For each call a time stamp in yyyy-mm-ddThh:mm:ss+hh:mm format was recorded for the time the record was created (eg. call received), the time the mission was scheduled (eg. assigned to an ambulance), the time the ambulance departed, the time the ambulance arrived to the case, the time the ambulance departed the case, the time the ambulance arrived to the transport destination, the time the mission was deemed complete, and the time the ambulance was available for service again. For the purpose of this study, we are interested only in the first four time stamps as these are the time stamps required to calculate ambulance response time following the commonly used definitions of response time – divided into the three intervals: record created to scheduled; call scheduled to ambulance departure; and time from departure to arrival at the case.

Within the six month study period a total of 35,527 calls were recorded. In February, a data extraction error led to the rounding of the time data to exclude the seconds. As such, the data from February were excluded yielding a total of 29,606 records. Of these calls, any record with any of the first four time stamps missing was excluded. Of note is the fact that no records were missing an entry for the created on time stamp; 102 records were missing the mission scheduled time stamp; 1039 were missing the ambulance departure time stamp; and 4450 records were missing the arrival to case time stamp. Each of the three time stamp types with missing entries requires that the ambulance crew report the time back to the dispatch center; when the crew fails to report the time, then the entry is missing. Once the records with these missing entries were culled, the data set had 24,391 records that were considered useable for this study.

Within the 24,391 valid records, time synchro-

Table 1: Time Synchronization and Flow of Time Stamps in Case Study System.

| Component | Time Reference | Time Stamp |
|---|---|---|
| NTP Server | Time Source (GPS) | tS1 |
| Database Server | NTP Server | tS2 |
| Application Server | NTP Server | tS3 |
| Workstation Session Time Stamp | Application Server | t set to tS3; reset at login |
| Database Key Record Time Stamp | Database Server | tR1 = tS2 at initiation of the call; ie. when the user creates the record |
| Time Stamp of Remaining Records Related to the Initial Key Record | Workstation Time | tRn = t* based on workstation's current time |

nization errors were identified on the basis of negative time progressions when moving logically through the data. For example, if the time stamp associated with the mission being scheduled was before the time stamp associated with the creation of the record, then the difference between these two time stamps would yield negative time. Negative time intervals were considered to be a function of time synchronization issues. Had these errors been ignored then the average response time would be incorrectly reduced. Alternatively, if the negative intervals were accommodated in absolute terms, the impact on the average response time would likely show an inconsistent bias.

Of the 24,391 records, the vast majority of negative time intervals (21,716 records) occurred relative to the mission scheduled and created on records; the scheduled to departure interval had only 442 time synchronization errors; and the departure to arrival on scene interval had 187 time synchronization errors. The percent of negative time interval records apportioned by time interval type can be seen in Table 2. While a time synchronization error rate of nearly 90% for the first interval of the response time period may seem shocking, it is logical as the created on time stamp is generated by the call center server while the remaining time stamps are generated on the workstation terminals. It is these ICT component links that are most prone to time synchronization issues.

Having identified the intervals with time synchronization issues, we turn to the magnitude of those errors. Table 3 provides a summary of the mean, standard deviation, minimum, maximum, and first through third quartiles for the time synchronization errors in seconds. The data exhibits an extreme range within all time intervals. The maximum value within the created to scheduled interval reflects an error of 24 hours. This is likely due to poor synchronization between computers at the moment of the change from standard time to daylight savings time. Despite the extreme tail within the distribution of time synchronization errors, 75% of all errors are less than 106, 70, and 70 seconds for the created to scheduled, scheduled to departure, and departure to arrival at case intervals, respectively.

Table 2: Percent of records with negative time intervals by time stamp type.

| Time Interval | Percent of Records with Negative Time |
|---|---|
| Created to Scheduled | 89.04 |
| Scheduled to Departure | 1.81 |
| Departure to Arrival at Case | 0.77 |
| Time on Scene | 0.57 |
| Time to Destination | 0.48 |
| Time at Destination | 0.39 |
| Time from Mission Complete to Unit Available | 1.05 |
| Time from Unit Available to At Station | 0.56 |

# 5 CONCLUSIONS

The issue of proper handling of time related data is significant in the management of an information system. The issue becomes even more significant when the time data is intended for use in time critical settings such as EMD services. This paper serves to highlight the potential magnitude of time synchronization errors within a prototype EMD system.

A straight forward solution to this issue has not yet been devised. Some initiatives have tackled algorithms that may reduce clock skew (Gusella and Zatti, 1989). Other methods impose requirements of minimal delays on the network and are not suitable for distributed VPN based environments (Rentel and Kunz, 2005). "Skewless" clock synchronization is still a favored research subject, but unfortunately without any real field implementations in the context of distributed networks (Mallada et al., 2015). To date, the usage of internet based time synchronization has prevailed (Sherman and Levine, 2016). Ultimately the daunting task of keeping track of the cause-effect relations

Table 3: Descriptive Statistics of Time Synchronization Errors in Seconds.

| Time Interval | n | Mean | Std. Dev. | Min. | Q1 | Median | Q3 | Max. |
|---|---|---|---|---|---|---|---|---|
| Created to Scheduled | 21,716 | 92 | 927 | 1 | 46 | 70 | 106 | 86102 |
| Scheduled to Departure | 442 | 586 | 5819 | 1 | 17 | 37 | 70 | 86400 |
| Departure to Arrival at Case | 187 | 646 | 3645 | 1 | 15 | 30 | 70 | 40534 |

Table 4: Summary of errors found in study – their sources and recommended solutions.

| Source of Error | Error | Recommendation |
|---|---|---|
| Human | Inconsistent time formats | Time formats should be decided on prior to implementation and used consistently in both time stamping and data extraction scripts. |
| Human | Time rounding | Rounding of seconds should never be done in a time critical environment. |
| Human | Adherence to time synchronization protocols | Establish work checklists that ensure regular checking and synchronization of clocks. |
| Design | Large intervals between synchronization lead to significant clock drift | Time synchronization should be invoked every five minutes as opposed to every 24 hours. |
| Design | VPNs over long distances can introduce clock drift as a result of latency | Components of an EMD system should be co-located on a fast local network. |
| System | Misuse of time formats | Ensure that all computers are set to produce time stamps following standard protocols: yyyy-mm-ddThh:mm:ss+/-hh:mm |
| System | Mishandling of time changes due to inter-location differences, daylight savings time regulations, and leap years | Ensure that the offset from universal time is properly set and maintained across all system components. |

among different data manipulations continues to occupy engineers and scientists who seek to develop a way to increase the accuracy of time tracking in distributed systems (Bravo et al., 2015).

At a minimum, a set of primary recommendations for EMD systems emerge from this research. A summary of the time related errors encountered in this research and the corresponding recommendations are listed in Table 4. These recommendations are critical as many response time targets are on the order of seconds, but time synchronization related problems can yield errors that are on the order of minutes.

This study is not without limitations. The time synchronization errors studied in this paper were found by identifying time intervals with "negative" time. Thus, the errors reported here potentially reflect only half of the errors actually present in the data. Unfortunately, it is impossible to identify time interval related errors when the intervals reflect a logical progression within time.

Furthermore, the results are based on a system for which limited time synchronization was performed in order to yield a worst case analysis scenario. Future work includes studying the same system after the recommendations noted in Table 4 have been adopted. Subsequent studies should also seek to sample from multiple dispatch systems in operation with different configurations in order to determine a realistic range of potential time related errors.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Lebanese Red Cross for their interpretation of and comments on the findings in this paper.

# REFERENCES


Blackwell, T. H., Kline, J. A., Willis, J. J., and Hicks, G. M. (2009). Lack of association between prehospital response times and patient outcomes. *Prehospital Emergency Care*, 13(4):444–450.

Bravo, M., Diegues, N., Zeng, J., Romano, P., and Rodrigues, L. E. (2015). On the use of clocks to enforce consistency in the cloud. *IEEE Data Eng. Bull.*, 38(1):18–31.

Campbell, J. P., Gridley, T. S., and Muelleman, R. L. (1997). Measuring response intervals in a system with a 911 primary and an emergency medical services secondary public safety answering point. *Annals of Emergency Medicine*, 29(4):492 – 496.

Clark County, IN, USA (2014). Public safety plan ordinance. http://www.clarkhealth.net/pdf/public_safety_plan_ordinance_adopted_1-2014.pdf.

Cristian, F. (1989). Probabilistic clock synchronization. *Distributed Computing*, 3(3):146–158.

EMS World (2015). Should response time be a performance indicator? http://www.emsworld.com/article/12071652/response-time-expert-opinions.

Grolemund, G. and Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3):1–25.

Gusella, R. and Zatti, S. (1989). The accuracy of the clock synchronization achieved by tempo in berkeley unix 4.3bsd. *IEEE Trans. Softw. Eng.*, 15(7):847–853.

IEEE (2008). 1588-2008 - ieee standard for a precision clock synchronization protocol for networked measurement and control systems. Standard ieee 1588, Institute for Electrical and Electronics Engineers.

ISO (2004). Data elements and interchange formats – information interchange – representation of dates and times. Standard ISO 8601:2004(E) 01.140.03, International Organization for Standardization, Geneva, Switzerland.

Klyne, G. and Newman, C. (2002). Date and time on the internet: Timestamps. Standard rfc 3339, Internet Engineering Taskforce.

Lamport, L. (1978). Time, clocks, and the ordering of events in a distributed system. *Commun. ACM*, 21(7):558–565.

Mallada, E., Meng, X., Hack, M., Zhang, L., and Tang, A. (2015). Skewless network clock synchronization without discontinuity: Convergence and performance. *IEEE/ACM Transactions on Networking*, 23(5):1619–1633.

Manitoba Canada (2016). Statistical and response time information. http://www.gov.mb.ca/health/ems/stats.html.

Mills, D., Martin, J., Burbank, J., and Kasch, W. (2010). Network time protocol version 4: Protocal and algorithms specification. Standard rfc 5905, Internet Engineering Taskforce.

New South Wales Government (2016). Response times. http://www.ambulance.nsw.gov.au/Our-Performance/response-times.html.

Pons, P. T., Haukoos, J. S., Bludworth, W., Cribley, T., Pons, K. A., and Markovchick, V. J. (2005). Paramedic response time: does it affect patient survival? *Academic Emergency Medicine*, 12(7):594–600.

Pons, P. T. and Markovchick, V. J. (2002). Eight minutes or less: does the ambulance response time guideline impact trauma patient outcome? *The Journal of emergency medicine*, 23(1):43–48.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rentel, C. H. and Kunz, T. (2005). A clock-sampling mutual network time-synchronization algorithm for wireless ad hoc networks. In *IEEE Wireless Communications and Networking Conference, 2005*, volume 1, pages 638–644. IEEE.

Sherman, J. A. and Levine, J. (2016). Usage analysis of the nist internet time service. *Journal of Research of the National Institute of Standards and Technology*, 121:33.

The Capital Region of Denmark (2016). Ambulance response times. https://www.regionh.dk/english/Healthcare-Services/Emergency-Medical-Services/Copenhagen-Emergency-medical-services/Pages/Ambulance-response-times.aspx.

W3C (1997). Date and time formats. Standard, World Wide Web Consortium.

Wankhade, P. (2011). Performance measurement and the uk emergency ambulance service: Unintended consequences of the ambulance response time targets. *International Journal of Public Sector Management*, 24(5):384–402.

Wilde, E. T. (2013). Do emergency medical system response times matter for health outcomes? *Health economics*, 22(7):790–806.

# ICT and Ageing in Lebanese Public Hospitals
## *A Resource based View Perspective on Capabilities*

Nabil Georges Badr[1], Charbel Chedrawi[2] and Jamil Arida[2]
[1]*Grenoble Graduate School of Business, Grenoble, France*
[2]*Universite Saint Joseph, Beirut, Lebanon*
*nabil.badr@alumni.grenoble-em.com, {charbel.chedrawi, jamil.arida}@usj.edu.lb*

Keywords: ICT, Ageing, Lebanese Public Hospitals, Hospital Information System.

Abstract: This paper looks into the Lebanese healthcare system and its readiness to care for a growing elderly population and how ICT is used and how it is perceived by the stakeholders. The paper presents concerns on ageing population in Lebanon. It first addresses the status of hospital infrastructure in the country, and then discusses some interviews regarding ICT plans with six general managers of large public hospitals in different regions of Lebanon.

## 1 INTRODUCTION

Population ageing, which entails an increasing share of older persons in the population, is a major global demographic trend; a trend that is expected to intensify during the twenty-first century, driven by remarkable increases in life expectancy and falling fertility rates. By 2050, old people will outnumber children on earth (ESA, 2012).

This paper's setting is the country of Lebanon, a small middle-income developing country with a population estimated at around four million, characterized by unique socio-demographic features that render the ageing of its population a complex challenge (Saxena, 2008). A rapidly ageing society of adults of 65+ years faced by the lack of clear comprehensive government policy, the unavailability of accurate comprehensive database and statistics, limited capabilities of existing institutions complicated by an increased pressure on a resources deficient healthcare system (Sibai, 2014). Findings from the Lebanese National Health Expenditures and Utilization Survey (Ammar, 2009), indicate that, while older adults constitute less than 10 % of the population, they consume over 60 % of the health care resources.

Key questions arise: ***Are Lebanese Public Hospitals ready to face the demands from an ageing population? What role could ICT have in this shift?***

In an attempt to answer these questions, the paper surveys a selection of 6 geographically distributed Lebanese public hospitals serving a mixed demography of patients in order to learn how these hospitals face the demands of an ageing population and identify the role of ICT implementations in addressing this challenge. Grounded in the principles of Resource-Based View theory (RBV), the paper is organized as follows: First, an overview on the relevant literature with respect to ICT implementations for ageing societies. Followed by a look into obstacles faced in developing countries with a focus on the Lebanese healthcare system and related ICT implementations. After the literature review a section presents the methodology used in this research. Then the paper is concluded with the discussion and suggestions for further research.

## 2 BACKGROUND

Similar to the work of Bryson et al. (2007), Rosenberg and Ferlie (2014), and Burton and Malone (2014), this paper uses RBV theory as a springboard in assessing the performance of Lebanese hospitals towards caring for the aging population. A RBV theoretical lens could shed light on the preparedness and capability of Lebanese hospitals to deploy internal resources in order to improve their performance, under severe financing pressure. Therefore, the RBV setting in this context does not assume the existence of competitive market forces but can be a useful assessment of the strategic potential



Badr, N., Chedrawi, C. and Arida, J.
ICT and Ageing in Lebanese Public Hospitals - A Resource based View Perspective on Capabilities.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 205-212
ISBN: 978-989-758-213-4

of Lebanese public hospital organizations. Furthermore, rather than financial performance, healthcare capabilities towards the aging are considered.

Literature on RBV links firm capability and performance to its ability to use its tangible and intangible resources (Eisenhardt and Martin, 2000). Originally developed in private sector firms (Wernerfelt, 1984), RBV theory argues that firms with different resource profiles and capabilities exhibit different performance levels. Thus, the effective use of these resources yields 'core competences'. Core competences are key internal resources, which, when effectively developed and exploited into dynamic capabilities (Teece et al, 1997), allow organizations to perform. These resources in healthcare could be human, such as nursing staff for example, financial and infrastructure such as number of beds, etc. and information technology. Further, the "*level of resource constraint*" in a public hospital setting may mimic "*market volatility*" in the private sector.

## 2.1 ICT for Health in an Ageing Society

New advances in technology make it possible to integrate previously disparate facility systems to form an "*intelligent*" hospital infrastructure (Blumenthal and Glaser, 2007). A plethora of applications generically referred to as '*telemedicine'* (e.g. video chat, mobile devices, and internet connected medical monitors) is expected to extend the provider - patient relationship boundary to remote areas positively impacting healthcare for the ageing in developing countries (Lucas, 2008). Assistive technologies have been proposed to overcome elderly problems such as fall risk, chronic disease, dementia, social isolation, and poor medication management, etc. ICT is one of these technologies along with robotics and gamification (Khosravi and Ghapanchi, 2015).

In general, there is evidence of benefit to society healthcare and patient care from access to Information and Communication Technologies (ICT) infrastructure, such as communication and systems for data interchange (Anwar et al, 2011), or online health information tools (Bolle, et al, 2015) with a keen emphasis on integrated care for the ageing (Merino et al, 2015). Progress in the field of health information systems is rather directly correlated with more quality and efficiency of care, where *''with more efficiency of care''* may in future mean that care will remain affordable (Haux, 2006). Patient access to electronic medical records improved patient communication (Cimino et al, 2002), however might have impacted aspects of physician - patient communication (Makoul et al, 2001). Poissant et al (2005) found that the use of bedside terminals and central station desktops saved nurses around 25% of their overall time spent documenting during a shift. That was encouraging. However, the authors found conflicting evidence supporting undesirable outcomes: bedside or point-of-care systems increased documentation time of physicians by 17.5%. Early empirical data on cost reductions was not very consistent (Hillestad, et al, 2005). This could have been be due to the variability of the health IT systems in their features and implementations (Chaudhry, et al, 2006). Later on, as technology advances were introduced, the benefits of computerized physician order entry systems CPOE were reported (Khajouei et al, 2011). More recent studies show measurable benefits emerging from the adoption of health information technology; these benefits range from efficiency and effectiveness of care (McCarthy, 2009), provider and patient satisfaction, preventive care for chronic diseases (Wildevuur and Simonse, 2015) and patient safety (Buntin et al, 2011) especially in developing countries where health resources are scarce (Marful et al, 2015).

## 2.2 Obstacles to Public Healthcare ICT in Developing Countries

Since the turn of the century, public authorities have been encouraging healthcare organizations to adopt new techniques and systems in order to deliver services of high quality and low cost (Naranjo-Gil, 2009), especially where adoption of innovations tends to be slow and fragmented (Fagerberg et al., 2005). In the public healthcare context, legislation and donor support are fundamental to the rise of propensity and ability to adopt ICT into Healthcare even in its basic capacities (Oak, 2007). In developing countries, studies underscore the possibility of impact on the cost of care as hindering the adoption of ICT into healthcare practice (Panir, 2011), extending beyond initial implementation over the full life-cycle of operating and maintaining these systems (Cohen et al, 2015). Cost of acquisition and maintenance and the lack of ICT skills have been known to present a significant barrier to implementation of ICT into hospitals of developing countries. Lack of skilled resources (Bing-Jonsson, et al, 2015) and the deficiency in ICT infrastructure hinder the capabilities of developing countries to acquire and develop electronic medical records for instance (Anwar et al, 2011). Physicians may perceive a loss

of professional autonomy (Esmaeilzadeh et al., 2015) and English literacy and education levels could curb the intention to use (Hasanain et al., 2015). The increasing familiarity of a new generations of healthcare practitioners is likely to lessen adoption issues (Hennington and Janz, 2007). Thus, organizational barriers to ICT adoption in healthcare have been recognized in the form of structure, policies, incentives and decision processes (Lluch, 2011).

## 2.3 Ageing and the Lebanese Healthcare System

Research on ageing in Lebanon involves three themes: living arrangements, social relations, and health (Abdulrahim et al, 2015). Ageing was related to health in terms of practices and social statuses that encourage good health (Ajrouch et al. 2013), nutrition (Boulos et al, 2014), and discussing threats to good health such as chronic conditions (Waked et al, 2011), even the role of religion (Chaaya et al, 2007). Yet, the literature is scarce when it comes to discussing the contribution of hospitals to the well-being of the elderly in Lebanon. The Lebanese healthcare system is described by a wide network of public (28) and private (136) hospitals and counts 26 beds per 1000 population making this one of the highest ratios in the Middle East; however, only 17% of these hospitals and 16.6% of the beds are in the public sector. There are few geriatricians practicing in the country; few hospitals and health centers, both private and public, have geriatric units. In 2011 the Ministry of Public Health (MOPH) report (Table 1) shows that there were 377,470 of elderly individuals in Lebanon, 65% of them aged between 65 and 74. The number of hospitals and the number of beds don't reflect the effective need (6 hospitals and 470 beds in Mount Lebanon for 145,558 elder versus 5 Hospitals and 365 beds in Nabatieh for 26,033 elderly). Hospitalization rates (days per year spent at hospitals) among older people exceed 28 % which is almost two-fold the national average (Kronfol, 2006). Older persons in Lebanon continue to rely on their relatives for healthcare. The transition from large extended families to small nuclear ones, accompanied with high rates of emigration among young Lebanese, an increased entry of women into the labor force have created a relative shortage in family members available for the provision of care (Sibai and Kronfol 2011). Concerns are growing about providing the elderly with a better quality of life (Silbermann et al., 2015).

## 2.4 Development of ICT in Lebanese Hospitals

In the Lebanon, the ICT sector witnessed significant growth over the period stretching from 2009 to 2014, growing by an average annual rate of 7.9% to reach a market size of USD 381 million in 2014 (MOPH). Though Lebanon has the required capacity, innovation, and skills required to improve its ICT sector, the country's lack of adequate infrastructure and regulations have so far slowed the development of ICT in hospitals. Furthermore, budgetary constraints and the lack of ICT competence in physicians present a challenge in rolling out ICT applications and services (Nicolian et al, 2015). Thus, ICT empirical research in Lebanese hospitals is limited and health data statistics present a daunting task in a fragmented health information system. Studies treating the use of technology tackled the effect of total quality management implementation on innovation skills of hospital staff (Aoun and Hasnan, 2015). Extant studies range from the discussion about Healthcare IS for data mining (Shahin et al, 2014) to a review job satisfaction of nurses is related to the level of information system use in their work (Maamari and Chaanine, 2013). Little research has focused directly on ICT in the care of the aging.

Table 1: Public hospitals capacity overview in relationship with the elderly in each region. Source: CAS, 2007 (The National Survey of Household Living Conditions), and MOPH 2011.

| Region | Beirut | Bekaa | Mount Lebanon | Nabatieh | North Lebanon | South Lebanon | Total |
|---|---|---|---|---|---|---|---|
| Number of public hospitals | 2 | 5 | 6 | 5 | 7 | 3 | 28 |
| Number of beds in public Hospitals | 595 | 430 | 470 | 365 | 455 | 235 | 2,550 |
| Elderly population (% of population in region) | 36,156 (10%) | 50,311 (13%) | 145,558 (39%) | 26,033 (7 %) | 77,281 (21%) | 42,131 (11%) | 377,4 70 |

# 3 METHODOLOGY

Our research employed a qualitative methodology in an exploratory approach (Eisenhardt, 1989). Data were collected by means of semi-structured interviews conducted between May 2015 and May 2016. In this study we interviewed general managers of six public hospitals, in six different regions in Lebanon covering a cross section of the Lebanese demography. Largest ranked by number of beds (by MOPH, 2011), these hospitals have stated that they have implemented or are in the process of implementing components of a Hospital Information System (HIS) as part of their ICT strategy. HIS provides the underpinnings for health-related decision-making that affect health outcomes of the ageing (mortality, morbidity, disease outbreaks, health status, disability, wellbeing). These hospitals were coded (H1…H6) to maintain the desired anonymity of the participants. Data consolidation was carried out by means of the software N*VIVO. Secondary data from documentation provided by the hospitals showed (1) occupancy ratios and capacity; (2) nursing staff information; (3) financing and government support information; and (4) infrastructure details mapped to the stages of the electronic medical record adoption model suggested by the Healthcare Information and Management Systems Society (HIMMS).

# 4 DISCUSSION

## 4.1 Are Lebanese Public Hospitals Ready to Face the Needs of an Aging Population?

The study shows that the most likely answer is "*No*" (Table 2). Major challenges were reported by the informants in this study; mainly high occupancy ratios and low capacity in terms of number of beds, short-staffing on nurses, lack of adequate infrastructure due to a reliance on self-financing, and little government support.

(1) Hospitals reported challenges of high occupancy ratios and low capacity in terms of number of beds; "*The hospital has 167 beds and still we face a daily problem. Sometimes occupancy is above 90% and the hospital cannot respond to the patients' needs as they sit in the emergency waiting (H6)".*

(2) The number of nurses is a major concern for all the hospitals in this study, as they agreed that they were adequately staffed by physicians and understaffed on nurses. This issue is well summarized by the informant of Hospital 4: "*The number of nurses is a major problem […] but the number of doctors is adequate. The hospital performs 2000 surgeries yearly and 70% of these patients are aged more than 65 years"*

(3) The hospitals interviewed are still waiting for the Ministry of Health to define a strategy for quality health care. The General Manager of H6 explains: "*The government finances each hospital for its expenses in proportion to the number of beds with a plus related to the hospital needs. Public hospitals cannot reject patients even if the hospital cost exceeds the amount given by the government. The ministry of health gives our hospital 6 billion Lebanese pounds yearly (4 Million USD). This number is small if we compare it to the needs of care for 167 beds. This places a stress on our operating capability (H6)".*

(4) All hospitals concurred on the lack of infrastructure to support the growth in demand and conveyed that most of their operational budgets is self-funded or financed through donations. "*The hospital upgrades its infrastructure from its own budget"(H2) and the expansion plans are covered by auto financing" (H3, H4, H5, H6)"Otherwise, most of the expansions financing sources in all the public hospitals in Lebanon come from donations and contributions from Kuwait, Islamic Banks, Emirates and Qatar (H1).* This confirms the work of Saxena (2008) and Sibai (2014).

Table 2: Summary of empirical data (last column indicates hospital's stated readiness to face the needs of an aging population).

| Hospital | Patients 65+ | Occupancy ratios | Major ailments treated | Ready? |
|---|---|---|---|---|
| H1 | 50 % | 90 + % | Heart failure | Yes |
| H2 | 55 % | 85 + % | Heart failure, chronic disease (Diabetes, etc.) | No |
| H3 | 70 % | 90 + % | Heart failure and obesity | No |
| H4 | 65 % | 80 + % | Heart failure | Yes |
| H5 | 70 % | 80 + % | Heart failure | No |
| H6 | 50 % | 90 + % | Prostate issues, broken hips, obesity, chronic diseases | No |

## 4.2 ICT's Role in the Enablement of Care for the Ageing

The general manager of H6 gives a pertinent summary on the perceived role of ICT in Lebanese Public hospitals: "*Today there is a greater need than ever to leverage technology to improve health, quality of life, and social connectivity for older adults, and assist in clinical care and monitoring of more impaired older adults. A variety of technology solutions (such as Web-based applications, remote sensing technologies, electronic health records and other devices) support patient engagement. The impact of ICT is indirect and is contingent on the redesign of practices and structures also outside health care. Improvements will only be realized if all parties involved can coordinate their efforts to take advantage of new technology*."

Elderly care is demanding, hospitals must have the resources to act quickly and effectively with a solid decision support system in order to minimize errors and offer quality of care for the elderly (Smith et al, 2006). In addition to a trained and experienced staff in the needs of the ageing, "*a solid database to follow up each case is required in order to find the appropriate solution*" (H2). The GM of H4 indicated that they have installed "*24/7 monitoring systems of health and activities, intelligent devices and appliances, internet enabled services, predictive behaviour models, and so on*" in order to provide a better quality of care for their elderly patients". Such innovations in stationary and mobile solutions would allow practitioners to stay in continuous contact, whether at the patient's bedside, in examination rooms, or in emergency treatment centers, to effectively develop and deliver patient assessments, and make more informed care decisions based on collaborative treatment plans. This findings aligns with the works of Lucas (2008) and Anwar et al. (2011).

It is expected that such ICT technologies would (1) *Reduce the time of treatments* (H1); (2) *provide better access to data for enhanced decision making, preventive care and disease management* (H1, H2, H3, H4, & H5); and (3) *improve interdepartmental collaboration / Emergency services* (H6).

For all 6 hospitals surveyed, the adoption of the information system is of top importance. "*The adoption of the information system is prominent practice for today's hospital, in the use of equipment, archive, and for more efficient medical results*" states the GM of H1 at the first stages of the interview. All hospitals have described their adoption of the Health IS system in advanced stage (Table 3). This is quite an achievement considering that all these public hospitals stated that they rely on self-funded initiatives. Nurses and doctors document patient's vital signs in the system, physician prescription history are tracked by the system, and PACS systems are connected to the online medical history. (PACS is the acronym for picture archive and communication system, a computer network for digitized radiologic images and reports).

Ultimately, addressing medical research advancement benefits, the GM of H1 stressed that "*better access to data that can be studied and help to get better solutions to face the ageing society*". ICT has become necessary to manage routine function up close, and monitor vital signs measurements and control the administration and recording of medications (type, dose & time). Agreeing, three other hospitals (H3, H4, H5) pointed out that "*ICT in the form of an integrated information database aims at disease management to help improve the awareness and preventive care in the elderly patient population*". Furthermore, the case summaries emphasized the role of the hospital senior management in encouraging adoption. "*It was my responsibility to oversee the implementation of the new system and to follow it up step by step with the different actors in order to achieve the results needed*." Adds the GM of H3.

The Lebanese government did not engage with any of the 6 hospitals in the decision to adopt their HIS, such decision was taken by the board of directors of each hospital independently.

Table 3: Level of HIS use in the hospital (Y= in use at the moment; Y*=Incomplete, but in process).

| Level of Use (*based on the HIMMS Model for EMR adoption maturity*) | H1 | H2 | H3 | H4 | H5 | H6 |
|---|---|---|---|---|---|---|
| Nurses / doctors enter patient's vital signs in the system | Y | Y | Y | Y | Y | Y |
| Physician prescription history tracked by the system | Y | Y | Y | Y* | Y | Y |
| PACS systems connected to online medical history | Y | Y* | Y | Y | Y | Y |
| Online medical history | Y | Y | Y* | Y | Y | Y |
| Automated pharmacy, laboratory, and radiology | Y* | Y | Y | Y* | Y | Y* |

Three of the hospital surveyed (H6, H3 and H1) expressed a relatively low barrier to adoption: The general manager of H6 disclosed that for the relatively new hospital, HIS was part of the hospital build out plans. Likewise, H3 pride themselves with a continual development effort of the HIS: "*We had an integrated information system since the start […] we develop it continuously based on the growing and changing needs of the hospital",* affirms the GM. In the case of H1, their GM shared that "*there were no obstacles in the implementation because all the staff started using the system since the establishment of the hospital. No transition was required since the staff was recruited with an IT IS experience. […] When we developed the implementation plan, we defined the different roles of the major actors. I was following each step to make sure that the plan is well executed. This was time consuming but effective*". In contrast the HIS in H5 is older and with limited with disparate data stores that are not integrated. Obstacles such as the qualification of the existing staff and the difficulty of recruiting new talent into the public sector were indicated. Additionally, budgets that need to be approved by governmental authorities have not yet been allocated to this hospital for the refresh of their system. Budgetary concerns were raised by most hospitals (H2, H3, H4, and H5), "*The process was time and energy consuming*" (H3). "*Significant efforts and investments in time and expertise were needed*" (H6). However, the "*novelty of technologies used posed a reluctance in the staff to embrace new technologies, with little evidence that this technology will indeed be useful*" (H6).

In H4, the GM reports an irregular focus on the Information System build out. The lack of IT/IS knowledge among the practicing staff presented significant challenges. To encourage adoption and assimilation, training and awareness sessions were necessary to highlight the importance of the implementation for the wellbeing of the patients.

Resistance to adoption was at different levels in each hospital; at some, the senior staff was non cooperative at the start (H2 & H5) and for others healthcare personnel's attitude towards new technology was not always positive (H1, H6, H4). Used to their traditional pen and ink methods, they were slow to assimilate the benefits of ICT usage for patient care (H6). The cooperation among project implementation teams and between IT teams and the hospital's staff at all levels was a key issue in order to reduce this resistance (H2).

# 5 CONCLUSIONS

Lebanese public hospitals are short of capable to face the needs of an ageing population. This work has achieved two objectives to help answer the research question:

First, the use of RBV theory as a backdrop for this study has underlined major challenges facing public hospitals in their ability to use their resources to care for an aging population. Informants have reported resource constraints and limitations of multiple dimensions, namely (1) high occupancy ratios and a low capacity in terms of number of beds, (2) short-staffed on nurses, (3) lack of adequate infrastructure due to a reliance on self-financing, and (4) little government support. Hospitals in the study communicated occupancy ratios between 80% and 90%. Most of their patients are above 65 years of age (reaching 65%). Major ailments treated are heart failures, chronic disease (obesity, diabetes, blood sugar, nervous system etc.) that require close supervision and long residency periods. With this burden, all hospitals express a lack of adequate infrastructure and a relatively low capacity to handle patients (number of beds); they are short-staffed on nurses and face higher cost with little financing support from the government. In spite of this self-funded effort, most of these hospitals recognize the value of an integrated information system in lowering their costs and increasing their capabilities to deliver quality elderly care.

Second, this paper extends the body of knowledge of healthcare ICT in the country of Lebanon to provide an account of ICT use in healthcare for the ageing in public hospitals in settings where uses of ICT could contribute to the effectiveness of Lebanese public hospitals to provide quality care for an aging population.

## 5.1 ICT as a Dynamic Capability in Lebanese Public Hospitals

The informants of this study reported use of ICT to care for the aging to have four significant benefits: (1) Provide better access to data for an enhanced decision making for treatment and medication; (2) potentially reduce the time of treatments; (3) improve the practice of preventive care, disease management and promote wellbeing; and (4) improve the resource allocation for better quality care. However, such benefits were recognized to impose a significant investment on hospitals. Investments that await governmental or donor funding support must be supplemented by commitment from senior

management, continual development plans, and a collaborative approach between all hospital staff to raise the awareness on the benefits of ICT usage for elderly patient care.

## 5.2 Research Limitation

This research presents a methodological limit that may open up new avenues of future research. One of the limitations of the study that worth mentioning is that it relies on information provided by general managers, thus potentially limiting the credibility of information. The opportunity to expand the interviews to IT managers of the hospitals would provide deeper insights. Further, we are aware that the results presented in this research depend strongly on the context of the country. Their generalization thus requires a certain reserve. Further research must be done to advance the results of this study possibly through action research potentially exploiting principals of organizational development. This approach could deepen the understanding on how hospitals are transforming their healthcare practices to improve their capacity for solving problems and managing the challenges of care for the aging.

# REFERENCES


Abdulrahim, S. Ajrouch. K., Antonucci, T., 2015 Aging in Lebanon: Challenges and Opportunities. The Gerontologist doi:10.1093/geront/gnu095.

Adler-Milstein J, Salzberg C, Franz C, Orav EJ, Newhouse JP, Bates DW , 2013 Ann. Intern. Med.159:97–104.].

Ajrouch, K. J., Yount, K., Sibai, A. M., Roman, P., 2013. A gendered perspective on well-being in later life: Algeria, Lebanon, and Palestine. *In S. McDaniel, and Z. Zimmer, Eds., Global Ageing in the 21st Century, pp.49–77*. Surrey, UK: Ashgate Publishing.

Ammar, W., 2009. Health beyond Politics. WHO and Ministry of Public Health in Lebanon.

Anwar, F., Shamim, A., Khan, S., 2011. Barriers in adoption of health information technology in developing societies. Int J Adv Comput Sci Appl, 2, 8.

Aoun, M., and Hasnan, N., 2015. Health-care technology management: developing the innovation skills through implementing soft TQM among Lebanese hospitals. Total Quality Management and Business Excellence, ahead-of-print, 1-11.

Bing-Jonsson, P. C., Foss, C., Bjørk, I. T., 2015. The competence gap in community care: Imbalance between expected and actual nursing staff competence. Nordic Journal of Nursing Research, 0107408315601814.

Blumenthal D. and Glaser J., 2007. Information Technology Comes to Medicine", NEJM Vol. 356, 24th, June 2007.

Bolle, S., van Weert, J. C., Daams, J. G., Loos, E. F., De Haes, H. C., Smets, E. M., 2015. Online Health Information Tool Effectiveness for Older Patients: A Systematic Review of the Literature. Journal of Health Communication, 20, 9, 1067-1083.

Boulos, C., Salameh, P., Barberger-Gateau, P., 2014. Factors associated with poor nutritional status among community dwelling Lebanese elderly subjects living in rural areas: results of the AMEL study. The Journal of Nutrition, Health and Aging, 18, 487–494.

Bryson J, Ackermann F, Eden C. Putting The Resource Based View of Strategy and Distinctive Competencies To Work in Public Organizations. Public Adm Rev. 2007; 67:702–17.

Buntin M. B., Burke, M. F., Hoaglin M. C. Blumenthal D., 2011. The Benefits Of Health Information Technology: A Review Of The Recent Literature Shows Predominantly Positive Results. Health Affairs, 30, no.3, 2011:464-471.

Burton CR, Rycroft-Malone J., 2014 Resource based view of the firm as a theoretical lens on the organisational consequences of quality improvement. Int J Health Policy Manag. 2014; 3:113–5.

Chaaya, M., Sibai, A. M., Fayad, R., El-Roueiheb, Z., 2007. Religiosity and depression in older people: Evidence from underprivileged refugee and non-refugee communities in Lebanon. Aging and Mental Health, 11, 37–44.

Cimino, J. J., Patel, V. L., & Kushniruk, A. W. (2002). The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. Int J Med Inform, 68(1), 113-127.

Cohen, J. F., Coleman, E., Abrahams, L., 2015. Use and Impacts of E-health Within Community Health Facilities in Developing Countries: A Systematic Literature Review. AISeL. ECIS 2015 Completed Research Papers ECIS 2015 Proceedings.

Eisenhardt K.M., 1989, Building theories from case study research, AOM review, Vol 14, N 4, pp. 532-550.

Eisenhardt, K. M. and Martin, J. A., 2000, Dynamic capabilities: what are they? Strat. Mgmt. J., 21: 1105–1121.

ESA, 2012 Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat. World Population Prospects: The 2010, 2012 Revision.

Esmaeilzadeh, P., Sambasivan, M., Kumar, N., Nezakati, H., 2015. Adoption of clinical decision support systems in a developing country: Antecedents and outcomes of physician's threat to perceived professional autonomy. Int J Med Inform.

Fagerberg J., D.C. Mowery, R.R. Nelson, 2005, *"Systems of innovation: Perspectives and challenges"* Oxford handbook of innovation, Oxford University Press.

Hasanain, R. A., Vallmuur, K., Clark, M., 2015. Electronic Medical Record Systems in Saudi Arabia: Knowledge and Preferences of Healthcare Professionals. Journal of Health Informatics in Developing Countries, 9, 1.

Haux, R., Health information systems, past, present, future. Int J Med Inform, 2006, 75, 268—281.

Hennington, A. and Janz, B. D., 2007 "Information Systems and Healthcare XVI: Physician Adoption of Electronic Medical Records: Applying the UTAUT Model in a Healthcare Context," Comm. of the AIS: Vol. 19, 5.

Hillestad, R. et al, 2005 Health Information Technology: Can HIT Lower Costs and Improve Quality? Santa Monica, CA: RAND Corporation, 2005.

HIMSS EMR Adoption Model - HIMSS Analytics.

Khajouei, R., Wierenga, P. C., Hasman, A., & Jaspers, M. W. (2011). Clinicians satisfaction with CPOE ease of use and effect on clinicians' workflow, efficiency and medication safety. Int J Med Inform, 80(5), 297-309.

Khosravi, P., and Ghapanchi, A. H., 2015. Investigating the effectiveness of technologies applied to assist seniors: A systematic literature review. Int J Med Inform.

Kronfol, N., 2006, "Rebuilding of the Lebanese health care system: health sector reforms", Eastern Mediterranean Healh Journal, vol 12, pp 459-73.

Lluch, M., 2011. Healthcare professionals' organisational barriers to health information technologies. A literature review. Int J Med Inform, 80, 12, 849-862.

Lucas, H., 2008. Information and communications technology for future health systems in developing countries. Social Science and Medicine, 66, 10, 2122-2132.

Maamari, B., & Chaanine, J., 2013. Job satisfaction of the modern information-system-using nurse in Lebanon. Journal of Tech. Management in China, 8, 2, 120-136.

Makoul G, Curry RH, Tang PC. The use of electronic medical records: communication patterns in outpatient encounters. J Am Med Inform Assoc. 2001; 8, 6:610-5.

Marful, W. A., and Winter, A., 2015. When Information Technology meets Healthcare in West Africa: A Literature Review. Journal of Health Informatics in Africa, 2, 2.

McCarthy, G. (2009). Deliver tangible ROI. Three healthcare organizations see reduced costs, enhanced efficiency and increased compliance with CPOE systems. Health Manag Technol, 30(6), 26, 28-29.

Merino, M., Marqués, M., Egurbide, M., Romo, M., Rodríguez, I., García, M., de Manuel, E., 2015. Deploying Integrated Care Models for Frail Elderly Patients. International Journal of Integrated Care, 15, 5.

Mugdha R Oak, 2007. A review on barriers to implementing health informatics in developing countries, Journal of Health Informatics in developing countries, vol 1, No. 1, 2007.

Naranjo-Gil, D., 2009. The influence of environmental and organizational factors on innovation adoptions: Consequences for performance in public sector organizations. Technovation, 29, 12, 810-818.

Nicolian, N., Welch, C., Read, M., Roberts, M., Portsmouth, U. K., 2015. Critical organizational challenges in delivering business value from IT: in search of hybrid IT value models. Electronic Journal Information Systems Evaluation Volume, 18, 2.

Panir, J.H., 2011 Role of ICTs in the Health Sector in Developing Countries: A Critical Review of Literature. February 15, 2011. Journal of Health Informatics in Developing Countries.

Quivy R., Van Campenhoudt L., 1995, Manuel de rechercheen sciences sociales, Dunod.

Qureshi, N. A., Kundi, G. M., Qureshi, Q. A., Akhtar, R., Hussain, L., 2015. An investigation Into the Adoption and Use Issues of E-Health in Public Sector Hospitals of Developing Countries. Mediterranean Journal of Medical Sciences, 2, 1.

Rosenberg Hansen J, Ferlie E., 2014 Applying strategic management theories in public sector organizations: Developing a Typology. Public Management Review 2014. Sep17.

Saxena, P., 2008. Ageing and age-structural transition in the Arab countries: Regional variations, socioeconomic consequences and social security. Genus, 64, 37–74.

Shahin, A., Moudani, W., Chakik, F., Khalil, M., 2014, April. Data mining in healthcare information systems: Case studies in Northern Lebanon. In e-Technologies and Networks for Development, ICeND, 2014 Third International Conference on , pp. 151-155 . IEEE.

Sibai AM, Rizk A, Kronfol NM. , 2014. Ageing in the Arab region: Trends, implications and policy options. UNFPA, ESCWA, and CSA 2014.

Sibai, A. M., Yount, K. M., Tohme, R., Yamout, R., Kronfol, 2011: The older persons: from veneration to Vulnerability? In *S. Jabbour et al., eds Public health in the Arab Word, Cambridge,* Cambridge University Press.

Silbermann, M., Fink, R. M., Min, S. J., Mancuso, M. P., Brant, J., Hajjar, R., Strode, D., 2015. Evaluating palliative care needs in Middle Eastern countries. Journal of palliative medicine, 18, 1, 18-25.

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Shekelle, P. G. (2006). Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Annals of internal medicine, 144(10), 742-752.

Smith, D. H., Perrin, N., Feldstein, A., Yang, X., Kuang, D., Simon, S. R., Soumerai, S. B., 2006. The impact of prescribing safety alerts for elderly persons in an electronic medical record: an interrupted time series evaluation. Archives of Internal Medicine, 166, 10, 1098-1104.

Teece D, Pisano G, Shuen A., 1997 Dynamic Capabilities and Strategic Management. Strategic Management Journal. 1997; 18:509–33.

Waked, M., Khayat, G., Salameh, P., 2011. Chronic obstructive pulmonary disease prevalence in Lebanon: a cross-sectional descriptive study. Clinical Epidemiology, 3, 1315–323.

Wernerfelt, B., 1984, A resource-based view of the firm. Strategic Managements Journal, 5: 171–180.

Wildevuur, S. E., and Simonse, L. W., 2015. Information and Communication Technology–Enabled Person-Centered Care for the "Big Five" Chronic Conditions: Scoping Review. Journal of medical Internet research, 17(3).

# Sink or Swim: Connected Health Software
## *Grasping the Innovation Opportunities by Mitigating Risk*

Ita Richardson[1], Ahmed AbuBaker[1], Patrick O'Connor[2], Jerry O'Mahony[2] and Pádraig O'Leary[3]
[1]*Lero the Irish Software Research Centre, University of Limerick, Limerick, Ireland*
[2]*ADA-Security Ltd., Rockchapel, Co Cork, Ireland*
[3]*Emergent Research Ltd., Hillcrest, Dromhale, Killarney, Co Kerry, Ireland*
*ita.richardson@lero.ie, patrick@ada-security.com, jerry@cctv-ireland.net, padraigoleary@gmail.com*

Keywords: Connected Health, Innovation, Risks, Mitigation, Malnutrition, Software System.

Abstract: Connected Health innovation can be an opportunity for companies to develop and grow, if they take opportunity to develop solutions for healthcare. In this paper, we discuss a case study where a very small company in Ireland developed a connected health solution, but in doing this, discovered a number of risks which they faced. Working with a research from the University of Limerick (author 1), they developed mitigation strategies to avoid these risks, and subsequently developed an updated version of their initial connected health solution. This software, Global-MN, has been implemented by a charity in India, Varanasi Children's Hospital. We present information about both the initial and updated product, illustrating how overcoming the risks has resulted in the company redesigning their product for a global market. Data entered via this software is now providing Varanasi Children's Hospital with information and analysis, which, in turn, is allowing them to provide a better service and improve the nourishment of children in India.

## 1 INTRODUCTION

The term "Connected Health" describes a new form of healthcare service that depends on technology innovation to deliver healthcare. This is defined by Richardson (2015) as:

*Connected Health is where patient-centred care results from process-driven health care delivery undertaken by healthcare professionals, patients and/or carers who are supported by the use of technology (software and/or hardware).*

The development and implementation of connected health solutions cannot be undertaken by technologists alone – it must be carried out hand-in-hand with patients and healthcare professionals (HCP). For any connected health solution to work efficiently and effectively, the processes through which they work must also be developed. Therefore, while connected health products include software, e-health (electronic health) and m-health (mobile health), connected health must be recognised as a much wider concept.

Furthermore, for companies to move into the Connected Health market, they need to develop their products with healthcare in mind. Many innovative ideas have been proposed. With the global older population growing dramatically, and the costs of healthcare rising, connected health is an innovative marketing opportunity. To support this growth, the first author is on a Science Foundation Ireland (SFI) Industry Fellowship, where she has worked for the past 2 years (part-time) to carry out research within ADA-Security. She has developed insights into how connected health innovations should be structured so that both the healthcare consumer and the company can benefit. In this research (Richardson et al., 2016a), we are also interested in understanding how companies make the transition from one product to the next. Through the case study presented in this paper, we have been enabled to investigate how a very small company can shift extensively to a different product line through mitigating risk. This paper discusses these mitigations which ensure that the company could complete product development. Our research question is:

*What mitigations should a very small enterprise implement to ensure success in the growing Connected Health market place?*

The remainder of this paper discusses how the innovation process works for very small enterprises, presents a case study, outlines the company products, presents the research method used and



Richardson, I., AbuBaker, A., O'Connor, P., O'Mahony, J. and O'Leary, P.
Sink or Swim: Connected Health Software - Grasping the Innovation Opportunities by Mitigating Risk.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 213-221
ISBN: 978-989-758-213-4

discusses the risks faced and mitigations implemented.

## 2 BACKGROUND

In Ireland, small-to-medium sized enterprises (SMEs) account for 97% of all enterprises and the most important contributors to the economy (Forfás, 2007). The need for SMEs to become more innovative has probably never been greater given the new economies, new technologies, and hyper competition with which they are challenged (Drejer, 2002). Rogers (2004) regards innovation "*as a key ingredient in business success*".

Foster (1986) advocated that the diffusion of innovation over time follows an S-Curve (Figure 1). Under the S-Curve paradigm, firms go through various stages in the diffusion of innovation (Silveria, 2001), namely start-up, scale, maturation (compete) and decline (transition). Small to medium sized enterprises (SMEs) experience different market challenges at each stage ranging from survival and market validation, transitioning to challenges in increasing market share, and entering and expanding into the new markets.

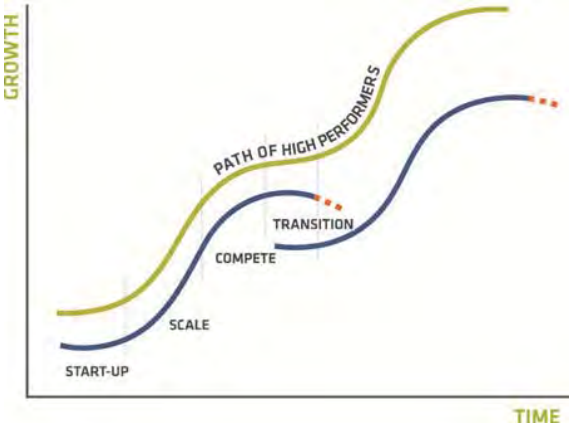



Figure 1: Double S-Curve Model (Squarespace, 2016).

SMEs need to recognise how to operate in markets with constraints, e.g. regulatory challenges, so as to retain their competitive advantage as they climb the S-Curve. Coupled with the diffusion of innovation is market adaptation, with particular attention being given to the stage and the focus of the market at each stage of the S-Curve. Credence must be given to the market forces at play given the very strong influence they may assert on the success of innovations in new targeted markets (Bolton and Thompson, 2013).

## 3 RESEARCH PROJECT

In her role as an SFI Fellow, the first author of this paper has spent 2 years part-time as a participant observer within ADA-Security, a very small enterprise in rural Ireland. Working closely with the managing directors (authors 3 and 4), she studied their connected health software development and innovation processes. She carried out interviews, attended meetings, discussed strategy, analysed the existing product, Local-Health and was involved in the development of their new product Global-MN.

## 4 CASE STUDY

ADA-Security is a company, whose focus has been on installing security systems within homes and companies nationally. It is a very small enterprise located in rural Ireland, over 50 miles from the nearest city. The local area has intermittent access to broadband, which is not uncommon in rural Ireland [Irish Farmers' Association, 2015].

The directors noted an opening in the market for the installation of home care systems, such as panic buttons and home monitoring. Innovatively, company has developed a number of services through which people can feel more secure and cared for in their homes. For example, they have developed activity monitors which can detect inactivity – if an older person has not risen by a certain time during the morning, and alarm can be activated. Their service also includes a friendly-call system which would have someone ring the older person during the day, reducing their feeling of social isolation. They now provide a combination of security, social and healthcare requirements.

Maintaining the security business and expanding into home care systems have ensured that the company has continued to be successful for over 2 decades. Examining the S-Curve in Figure 1, it is obvious that the company have shifted from one 'S' to another – from security to home care systems.

However, as innovators, and to ensure sustainability and continued growth of the business, the directors were soon looking at other business opportunities. They worked with the local community, investigated requirements from older persons and established that there is a need for further healthcare support within their homes, which could not be provided through hardware products – software was needed for this purpose. In rural areas, there is a growing trend towards home monitoring,

with many ratified medical devices available for use, by the older person in the home.

Home monitoring, then, brings with it another need – that of transmitting biometric data such as blood pressure measurement and blood sugar levels to the patient's health professional. General Practice has the potential to change – at times, there may not be a need for the patient, in this case the older person, to travel to the practice to have these measurements taken on a sporadic basis. Alternatively, they can take their own measures and transmit these to medical care. Consequently, ADA-Security directors, in conjunction with another company, developed Local-Health, a prototype for initial testing, with a view to supplying the connected health market.

## 4.1 Local-Health: Biometric Measurement System

Local-Health (described in Richardson et al., 2016b) allows individuals to text their biometrics from their mobile phone to the local General Practice. The information is coded for reading by the bespoke software where the text is received.

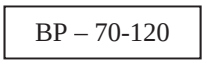



Figure 2: Blood pressure reading text message.

As an example, in Figure 2, the text message shows a Blood Pressure (BP) reading of 70/120, which was taken by the patient using a home blood pressure monitor. The receiving system recognises the mobile number from whom the text came, and the data is collected within the General Practice.

The data is compared against an expected reading for the particular patient. If the actual reading is abnormal, an alert is sent to that user, asking them to follow the suggested preventative advice, for example to seek medical advice. An alert can also be sent to individual users if they have not sent their results at the time expected. The General Practitioner can monitor each patient's readings on a regular basis, receive alerts if someone's readings go out of control, and conduct up a follow-up if they deem it to be needed.

Through modifying their healthcare process, the implementation of this simple connected health solution has introduced technology for use within the General Practice,. The patient no longer has to attend the practice on a regular basis, for example, weekly, and yet the Doctor and Nurse can effectively monitor the patient. The consequence of this is that the patient can be monitored without having to leave their own home, while the queues and load for the medics in this rural practice has been alleviated. The General Practitioner can view their patient's data as a line graph e.g. blood pressure over a period of time. Therefore, a trend is illustrated and support making medical interventions in a convenient fashion.

## 4.2 Local-Health: Potential Business Risks

As time progressed, the company directors recognised a number of potential business risks which arose with Local-Health system:

- Aimed towards individuals and small healthcare practices;
- Market for Local-Health is national;
- ADA-Security expertise is in hardware;
- System developed on a known platform;
- Regulation required for the system.

**Aimed towards individuals and small healthcare practices:** Once patients have mobile-phone text available they can submit readings, and they are not required to buy Local-Health software. Software cost, therefore, is totally borne by the healthcare practice. The expectation is that this will become a high-volume, low-cost product. Due to high software development costs, the business model is not very cost-efficient.

**Market for Local-Health is national:** Due to the nature of its hardware products, which includes physical installation of security and social systems, the company has dealt mainly locally and nationally. The market needs to become international. However, going global with a low-cost product can be an expensive undertaking.

**Company expertise is in hardware:** Those working in the company have a background and experience in hardware installation and hardware attributes. The nature of the new product requires software engineering expertise. This has not been readily available, and development to-date has been sub-contracted. However, without in-house expertise, other business requirements take priority.

**System developed on a known platform:** Local-Health was to be used by older persons within the local community and it was considered best to develop the product as a mobile-phone text-based system integrating with healthcare practice software. While this works effectively, there is no income due to selling product to individual patients.

**Regulation required for the system:** Within their hardware business in the security industry, ADA regularly implement regulations. Under European Council (2007) directive, a medical device means

(amongst other things) "software… intended by the manufacturer to be used for human beings for the purpose of diagnosis", and "software … is a medical device". As Local-Health is dealing with the transmission, collection and diagnosis involving patient data and HCPs, it is a medical device. Regulations need to be integrated into the software which is time-consuming and expensive, thus pushing up the cost of production.

## 5 MAKING THE TRANSITION

Reflecting on the S-Curve (Figure 1), the *Transition* required for ADA-Security to move from their main businesses of security and home care systems to Local-Health is substantial, and the company needs to find a way to overcome the problems identified. They had put many resources into the development of Local-Health to both fulfil a social need and be profitable. This was an opportunity for them to move across to yet another 'S' curve which is vital for the survival of any small company - it is the innovative companies who survive. The leap from security systems to homecare systems was achievable for the company. The leap from their current offerings to the Local-Health software system was difficult.

So how could the company get over this chasm? How could they mitigate the risks they had identified? Considering their options, the directors with the first author of this paper made strategic decisions to change the direction of the product, allowing the company to develop a new innovation, while also presenting opportunities to consider other markets with which they could be involved. The following modifications were made to mitigate the difficulties identified above:

- Seek global market opportunities
- Develop solution for an organisation
- Acquire software engineering skills
- Investigate non-regulated possibilities
- Investigate mobile solutions

**Seek global market opportunities:** Global opportunities present much larger markets than local, and companies have the ability to expand beyond their local area. For ADA-Security, the installation of hardware systems provided the natural consideration of national solutions. However, moving to a software-based product provides the potential to exploit global expanding markets. In particular, the company recognised that developing countries can offer large populations, therefore much greater sales potential.

According to World Food Programme (WFP) "there are around 800 million people in the world who are malnourished" (WFP, 2016). This includes 200 million children under the age of five suffering from undernutrition (USAID, 2016). A quarter of the world's malnourished children live in India (WFP, 2016), and there are 473,000 malnourished children in Kenya (UNICEF, 2016). Charities and Non-Government Organisations (NGOs) provide health check-ups to diagnose malnourished children and nutrition programmes to resolve the problem.

The development of a system, Global-MN, to support nourishment programs globally has the potential to provide a large market to ADA-Security's software solution.

**Develop solution for an organisation:** Local-Health was developed for individual use, but, selling to an organisation would be more profitable. Therefore, charities in developing countries were an opportunity that should be explored further, and the development of Global-MN was undertaken. This is an innovative mobile health software application that can store, track and monitor details of malnourished children.

NGOs hire Community Health Workers and health clinics to reach children, test and diagnose them, enrol them into nutrition programmes, and schedule further visits to monitor their progress. The test process normally includes taking the child's height, weight, middle-upper arm circumference (MUAC) and age. To ensure user input to the product, at development stage, ADA-Security teamed with an Indian based charity, Varanasi Children's Hospital, whose work with these children has been hampered by the following problems:

- The manual process currently used to document children's readings is time consuming, thus decreasing productivity of Community Health Workers;
- Retrieval of paper files when a child returns to the nutrition clinic is cumbersome and difficult;
- Paper-based process does not support efficient monitoring of children's progress;
- Community Health Workers are unable to identify previously registered children

As a result, Varanasi were not able to reach and help as many children as they would like. Additionally, there are difficulties in developing reasonable statistics to show how effective the work of the charity is. Varanasi were unable track how well individual children are progressing, nor could they see the success they are having in the field. Having

matched this situation with Local-Health, a decision was made within ADA-Security that Global-MN, would initially support the nutrition program within Varanasi Children's Hospital.

The identified end users are Community Health Workers, as Global-MN software will be downloaded into their phones, allowing them to register and perform screenings on malnourished children. The second end user group are charity and NGO management who ensure that their nourishment programme is effective. Also, to obtain investment from donors, they must demonstrate the effectiveness of their program. Therefore, Global-MN will follow a Business to Business (B2B) commerce model rather than Business to Consumer (B2C) commerce model.

In the first instance, ADA-Security has developed the product for Varanasi Children's Hospital, with plans to expand in the future to other charities internationally.

**Acquire software engineering skills:** This software-based innovation has potential to open doors to a global market. While, through their business and initial software development, they were in a position to identify the innovation, ADA-Security directors have recognised that they need to acquire software engineering skills and make this product a priority. The skills will ensure that the product they develop is marketable, secure and profitable. Therefore, to bring software engineering to the development process, ADA-Security have partnered with Emergent Research Ltd. (author 5), a high performance start-up software company headquartered locally.

**Investigate non-regulated possibilities:** While Global-MN must be secure and private, it is not a product that needs to be regulated. It is used to input, track and analyse food-related information – for example, information regarding children who are malnourished and amounts of food. Therefore, it is not a Medical Device and there are no regulations that need to be considered during its development.

**Investigate mobile solutions:** Local-Health was developed as a text-based system for the reasons stated earlier. However, within the global market for organisations within which Global-MN is being targeted, a text-based system would have many limitations, and a mobile internet-based product is the better option. We have undertaken research which demonstrates that there is internet coverage throughout both India and Kenya, countries where malnourishment programs with children are being undertaken, demonstrating sales' potential. Additionally, the company needed to understand the cheapest and most convenient method by which the data could be transmitted. In rural Ireland, given that there is often limited and sporadic internet access, the solution for Local-Health was to text data via mobile phone. In rural India, internet access is relatively stable, and cheaper than using a mobile phone. Therefore, Global-MH used the internet as its platform. Entering data in this manner means that once a child is registered with Varanasi Children's Hospital, their information is immediately available to the central office. The effectiveness of the program can be monitored on a regular basis. Children who are not progressing as expected can have their food source changed early on in the program, and interventions by the central office can happen quickly. Therefore, infant mortality rates are affected positively in the long-term.

## 5.1 Global-MN: A Software Solution

Global-MN system is a web-based system with data stored on a secure system in Ireland. This research project was carried out as one action research cycle, allowing the researchers to understand the difference that the implementation of the software made in the field and how the processes could be changed to make a difference. The charity director was trained in system use by author 3. He then travelled to India where he trained the Community Health Workers. They enter data to the system using software running on a smart phone. This data can be aggregated and analysed for Varanasi Children's Hospital allowing management to make informed decisions regarding their malnourishment program..

# 6 Global-MN: MANAGING MALNOURISHMENT

Using Global-MN, Community Health Workers add children's identification data and a photograph to the system in the rural villages and collect measures to determine malnourishment: MUAC, weight, height and age. Malnourished children are started on a special food program. They return to the clinic every two weeks where their signs are re-measured, further food provided, and updates added to the Global-MN system. Figure 3 shows relevant screen shots from the system. Once data is uploaded, management are provided with data analysis, allowing them to observe progress of children who are on the food program.

## 6.1 Data Analysis - Individual

Data illustrated is not real data to maintain confidentiality. However, it illustrates how the system can be used for the benefit of tracking the food program for malnourished children.

In Global-MN, data is stored about individual children, including: child ID, which is created by the system when child is first entered, child name, packets given to child per week, start weight, MUAC and height, weight, MUAC and height at each clinic visit, healthcare clinic to which the child is registered, and status, which can be active or discharged. Thus, the system allows the charity to visualise trends for individual children.

As an example, Figure 4 shows the percentage weight gained by individual children based on their start weight. We have highlighted the data for one child whose start weight was 3kg – this child gained 50% of start weight while on the program. Two children weighing over 15kg at the start of the program gained approximately 5% of start weight. Most children gained between 10-25% of their start weight, which is what would be expected from the food program. Figure 5 shows the weight and MUAC gain for a single child during 6 visits to the healthcare clinic. As weight increases, so does the size of the MUAC. This demonstrates that, for this child, the malnutrition program is having a positive effect.

Once a child attends the healthcare clinic, a target weight gain is determined for him / her depending on starting weight, MUAC and height. For the child illustrated in Figure 6, it was expected the child would gain over 4kg in the course of the program. However, it can be seen here that actual weight gain is much lower. With access to on-line real-time data, they can make decisions about this child. It could be, for example, that the child has an underlying condition, in which case she / he can now be sent for medical attention. Prior to Global-MN, this information was not available, further diagnosis was missing, thus perpetuating their problem.

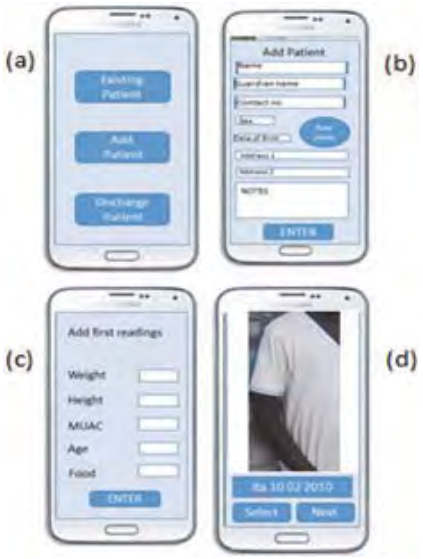



Figure 3: (a) Top level menu (b) Adding child's details (c) Adding information for a child, including number of food packets for that child (d) Information about existing patient including photograph.

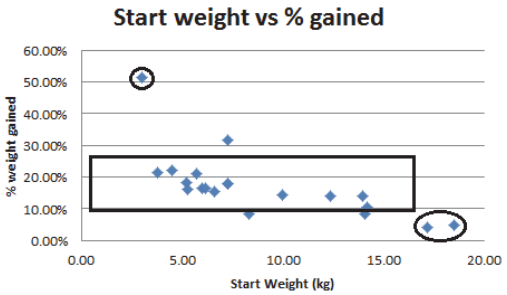



Figure 4: % weight gained based on start weight.

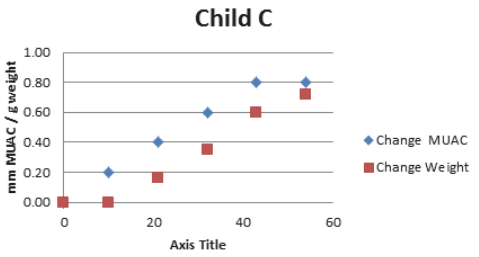



Figure 5: Weight (kg) / MUAC (cm) gained by child over 6 clinic visits.

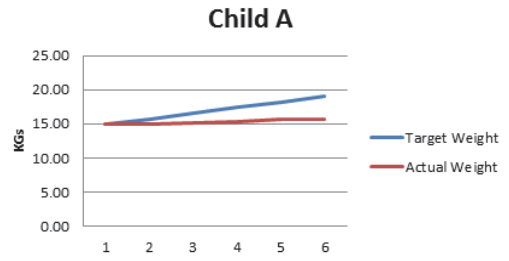



Figure 6: Actual weight gain mapped against Target weight for a child.

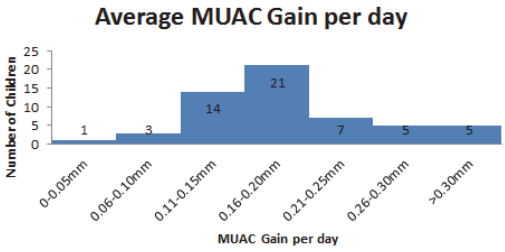



Figure 7: Average MUAC gain per day during program.

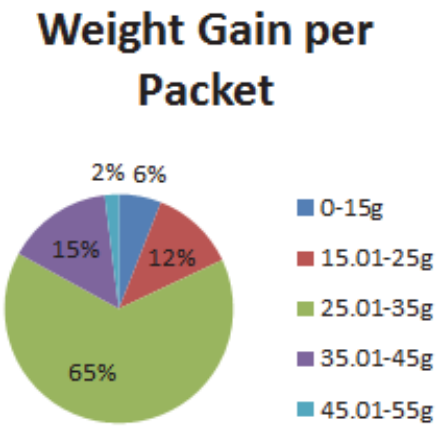



Figure 8: Weight gain per packet.

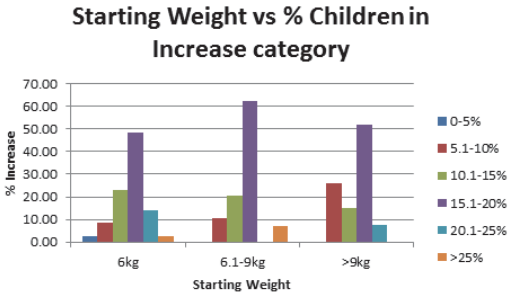



Figure 9: Weight increases mapped against starting weights.

## 6.2 Data Analysis - Collective

Data can also be aggregated to determine how well the food program is working. This information is used to ensure that children who can benefit are receiving food and that the investment in food is having the maximum required effect. In figure 7, the average MUAC gain per day illustrates that MUAC is increasing through the food program. Thirty-five children (out of 56), 62%, gain between 0.11 and 0.20mm per day. Figure 8 illustrates children's weight gain per food packet distributed. Sixty-five percent of children gained between 25-35g in weight per packet of food. However, six percent of children gained less than 15g per packet distributed. The charity can now further investigate as to why distributed food is not being effective.

Figure 9 shows data for discharged children categorised by starting weight. The expectation is that children's weight would increase by 15-20% of their starting weight. As can be seen in the graph, 48% of children who start at 6kg or less gain this amount of weight. Sixty-two percent of children who are in the 6.1-9kg category gain between 15-20%, while 51% of those whose starting weight is greater than 9kg gain similar amounts. Such aggregated data allows the charity to delve deeper into the success and difficulties within the program.

## 6.3 Case Study - Summary

To implement Global-MN within Varanasi Children's hospital, Community Care Workers had to be trained in system usage to ensure that data was being entered correctly. Initially, all records were also held on paper, requiring extra work for everyone involved. Trust in the system needed to be developed. At this point, we, the researchers, have been enabled to provide them with both individual and collective data, allowing them to make changes within the charity. For example, if a child is observed as not gaining the expected weight within the program they can be checked for other health issues. The analysis regarding weight per packet mapped against start weight is allowing the charity to consider whether they give different amounts of food to children who start with higher weights. They are collating data which can be used to support arguments for further funding for the charity.

Following the case study, the software is being updated to make it more user-friendly and to provide more on-line and visual reporting that currently exists within the system. Both Varanasi Children's Hospital

and ADA-Security are benefitting from the implementation and analysis of this case study.

## 7 CONCLUSION

In Ireland and globally, it is recognised that healthcare is a serious problem that needs fixing - "healthcare is the greatest immediate threat to the country" [USA]" (Chase, 2016). Many innovators see connected health in a variety of forms as a means to solving that problem. Innovators must decide the best way to mitigate risks that arise, ensuring that the return on investment is as expected. For the small to medium sized enterprise, and certainly for the very small enterprise, taking risks has a far-reaching effect.

Within ADA-Security, the directors recognised and mitigated the risks being faced by the company. Understanding that their business could grow significantly they sought global market opportunities, identifying that there was a need for data analysis and tracking in programs in developing countries. Their product needed to shift from supporting small to supporting large and from solutions for individuals to solutions for organisation. There are many NGOs working with malnourished children in developing countries, and, supported by a business plan, ADA-Security are now marketing internationally. They recognised that they needed to extend their skill set. Rather than take on new employees, they have partnered with Emergent Research to provide software skills. A further risk was entering the Medical Device software market which is heavily regulated. While being cognisant of regulation, they are developing software without having to obtain European Union or similar certification. In bringing in an established software partner, there are documented software development processes in place. From a technical perspective, the greatest change was that of the platform upon which the product would run - Global-MN is very different to Local-Health.

Throughout their 25 year history, ADA security have been an innovative company, expanding their product base, growing and providing employment in the local community. They recognised an opportunity for innovation, and, in exploiting this, have considered how to overcome the initial risks. In conclusion, Global-MN has given ADA-Security a further innovation opportunity. Considering the five risks and developing mitigation strategies has allowed them to leap the larger chasm between S-curves. As they move forward on their innovative journey, they can take the learnings from Global-MN development to ensure that they can continue to innovate and grow.

## ACKNOWLEDGEMENTS

This work was supported by Science Foundation Ireland (SFI) Industry Fellowship Grant Number 14/IF/2530. We would like to thank Pat McMahon, Varanasi Children's Hospital for invaluable discussion.

## REFERENCES

Chase, Dave, Why 98% of Digital Health Startups are Zombies and What They Can Do About it, http://www.forbes.com/sites/davechase/2016/05/18/why-98-of-digital-health-startups-are-zombies-and-what-they-can-do-about-it/#1f0ed5bf68f3, Forbes, 22th September, 2016.

Bolton, B., and Thompson, J., (2013) Entrepreneurs Talent, Temperament and Opportunity, Routledge 3rd Ed.

Drejer, A.(2002) 'Situations for innovation management: towards a contingency model', European Journal of Innovation Management, 5(1), 4-17.

European Union Council, 2007, EU directive 2007/47/EC.

Irish Farmers Association, Supporting Farms and Businesses in the Irish Countryside, 2015. https://www.ifa.ie/wp-content/uploads/2015/08/150805-IFA-Countyside-Policy-Document-Final.pdf

Forfás (2007) "Towards developing a policy for Entrepreneurship in Ireland", Dublin.

Foster, R.J., (1986) Innovation: The Attackers Advantage. New York: Summit Books.

Richardson I. (2015). Connected Health: People, Technology and Processes, Lero-TR-2015-03, Lero Technical Report Series. University of Limerick.

Richardson, Ita, O'Mahony, Jerry, Howarth, Paul, O'Connor, Patrick, Glenny, Liam (2016b), Connected Health: From Rural Ireland to Rural India, Software Quality in Healthcare Systems, co-located with IEEE International Conference on Software Engineering, Austin Texas, May 2016, pp 39-42.

Richardson, Ita, Costin, Yvonne, O'Connor, Patrick, O'Mahony, Jerry, Glenny, Liam (2016a) Connected Health Innovation within SMEs: Understanding the Marketplace, 4th International Conference on Innovation and Entrepreneurship, Toronto, Canada, 28-29th April, 2016, pp 63-71.

Rogers, M. (2004), 'Networks, firm size and innovation', Small Business Economics, 22, 141-153

Silveira, G.(2001), 'Innovation Diffusion: research agenda for developing economies', Technovation, 21, 767-773.

Squarespace (accessed September 2016) http://static1.squarespace.com/static/55d244b8e4b0f26e56cd2b32/t/55e51df8e4b09f6b9550a6d9/1441078778831

WFP (accessed September 2016). World Food Program: Hunger Statistics. https://www.wfp.org/hunger/stats.

UNICEF (accessed September 2016). UNICEF: Kenya. http://www.unicef.org/appeals/kenya.html.

USAID, (accessed September 2016). Home: GLOBAL NUTRITION. https://www.usaid.gov/what-we-do/agriculture-and-food-security/supporting-global-nutrition.

# Abstract Information Model for Geriatric Patient Treatment
## *Actors and Relations in Daily Geriatric Care*

Lars Rölker-Denker and Andreas Hein
*Department of Health Services Research, University of Oldenburg, Oldenburg, Germany*
{*lars.roelker-denker, andreas.hein*}*@uni-oldenburg.de*

Keywords: Information Model, Geriatric Care, Knowledge Processes, Organisational Learning.

Abstract: The authors propose an abstract information for geriatric care, the geriatric information model (GIM). They adopt an information model from cancer care and introduce characteristics for geriatric care (patient population, multidisciplinary and multi-professional approach, cross-sectoral approach). Actors (patients, physicians, therapists, organisations), information objects, and information relations are defined. The GIM is validated by mapping four typical knowledge processes (multi-professional geriatric team session, interdisciplinary clinical case conferences, tumor boards, transition management) onto the model. The GIM is stated as useful for understanding information flows and relations in geriatric care. All processes for validation can be mapped onto GIM. In future work the GIM should be tested with more knowledge process and could also be used for identifying gaps in the IT support of geriatric care. A study on high and low information quality in geriatric care is also proposed.

# 1 INTRODUCTION

Patient treatment is a heavily data, information and knowledge driven process with inter- and multidisciplinary cooperation (Chamberlain-Salaun et al., 2013, 74ff.). The amount of available data, information and knowledge is increasing due to ongoing technological developments and medical research. Medical information gathered in the domestic and mobile environment of the patient will tighten this process in the future.

These challenges also apply for geriatric patient treatment (Mangoni, 2014) (Rölker-Denker and Hein, 2015). Geriatric treatment is characterized by a target population with complex diseases and an increasing amount of patients, a multidisciplinary and multi-professional treatment approach and a cross-sectoral treatment (see section 3. For better understanding, managing and controlling of information flows under these constrains an abstract information model is needed.

In this work we adopt the approach of Snyder et al (Snyder et al., 2011) who introduce an information model for cancer care (see section 2). Afterwards, the principles of geriatric care in Germany are introduced (see section 3). The model is then modified to the needs of geriatric care based on literature review and results from observational studies and interviews with practitioners 4. The work is then validated with four typical knowledge processes being mapped to the model 5. The work then closes with a conclusion and outlook 6.

# 2 ABSTRACT INFORMATION MODEL

Snyder et al (Snyder et al., 2011) propose an abstract information model for cancer care. Information in cancer care is originated from clinician and patient side (actors) and there are different communication paths (relations). The actor-relations structure is depicted in figure 2.

## 2.1 Actors

Actors in the cancer care process resume have roles and functions. In detail these are:

- Patient: treated by a clinician;
- Other patients: patients with same or similar disease and/or treated by the same clinicians or hospitalized in the same health organisation;
- Patient's family and friends: people associated with the treated patient;



Rölker-Denker, L. and Hein, A.
Abstract Information Model for Geriatric Patient Treatment - Actors and Relations in Daily Geriatric Care.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 222-229
ISBN: 978-989-758-213-4

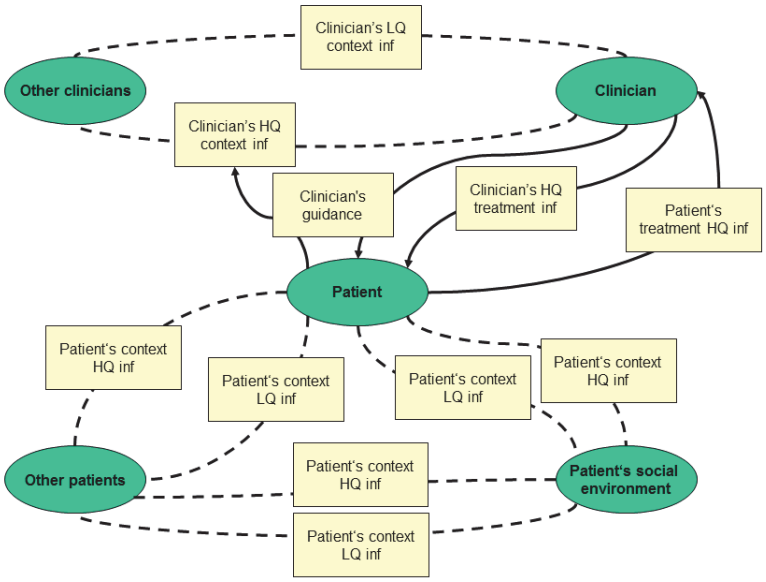



Figure 1: Abstract Information Model based on (Snyder et al., 2011).

- Clinician: treating a specific patient;
- Other clinicians: other clinicians from the same discipline (higher or lower rank), associated discipline or health organisation in contact with the clinician in charge.

## 2.2 Relations

Patients and clinicians exist in a universe of information. Snyder et al differentiate between high-quality (HQ) information and low-quality (LQ) information, with only a portion representing HQ information. HQ information relations are:

- Clinician's HQ treatment information: Combination of clinician's medical knowledge (gained from education and experience) and acquired medical information (laboratory, medical imaging, EEG, ECG) with the information gained from examining the patient (sensorial information);
- Patient's HQ treatment information: Information provided by the patient, e.g. drug intake, health-relevant behaviours (nutrition, smoking) or familial-genetic preload;
- Patient's HQ HQ information: information shared along the patient and its family and friends and along the patient and other patients;
- Clinician's HQ context information: Information shared along the care team;
- Clinician's guidance: Clinicians can direct their patients to appropriate information resources.

At the same time, it must be noted that much of the information available to both clinicians and patients is biased, incorrect, or otherwise not useful. LQ information is shared frequently among patients.

- Patient's LQ information: information shared along the patient and its family and friends, and along the patient and other patients;
- Clinician's LQ context information: Low quality information is shared even among clinicians.

# 3 CHARACTERISTICS FOR GERIATRIC PATIENT TREATMENT IN GERMANY

The following statements mainly focus on the specific situation in Germany which the specifics of the German health care system being separated into different sectors. Nevertheless the used information in geriatric care is comparable to other countries while crossing the sectoral boarders is the main challenge in Germany.

The information flows in geriatric treatment differ from the information flows in cancer care. There are four main reasons:

- Patient population;
- Multidisciplinary approach;
- Multi-professional approach;
- Cross-sectoral approach.

### 3.1 Patient Population

Geriatric patients often suffer from chronic conditions, multimorbidity, polypharmacy and cognitive deficits (Soriano et al., 2007, 15). They are often hospitalized in nursing or retirement homes and, due to cognitive impairments, not able to give proper information about their health status. This results in a strong demand on patients' information from clinicians' view. In addition the amount of geriatric patients is continuously rising with the demographic change in most industrial societies (Kolb and Weißbach, 2015). Therefor a structured information acquisition is essential for the future success of geriatric treatment.

### 3.2 Multidisciplinary Approach

Due to multimorbidity and chronic conditions, geriatric treatment follows a holistic and systemic approach including several different kinds of medical disciplines. The most frequent disciplines involved are internal medicine, family medicine, psychiatry and neurology followed by orthopaedics, surgery, trauma and abdominal surgery (Nau et al., 2016, 603ff.).

### 3.3 Multi-professional Approach

Geriatric treatment and geriatric care is a highly multi-professional process with several professions included (Tanaka, 2003, 69ff.). In Germany, geriatric is organised in different ways. In case of stationary care selected patients can be treated under supervision of the multi-professional geriatric team (MGT) in the so-called complex geriatric treatment (German: geriatrische frührehabilitative Komplexbehandlung) (Kolb et al., 2014) (Rölker-Denker and Hein, 2015, 314f.). The MGT consists of physicians, nurses, therapists (logopedics, physiotherapists, occupational therapists, psychologists) and social workers.

### 3.4 Cross-sectoral Approach

Geriatric patient are often treated over sectors borders and in other health care organisations (HCOs). In Germany, medical treatment is mainly separated into ambulatory care/out-patient care (general physicians, consulting/specialist physicians, ambulatory medical services provide by hospitals) and hospital care/in-patient care. Rehabilitation care, stationary care (nursing homes, retirement homes) and home care are other relevant sectors for patient treatment.

## 4 GERIATRIC INFORMATION MODEL

The key to the Geriatric Information Model (GIM) is depicted in figure 4, the GIM itself with actor-information relations in figure 4.2.

### 4.1 Actors

Within the GIM actors can be a single actor or a group, consisting of several single actors or other groups. Single actors describe a specific class of persons with similarities (e.g. patients, carers, clinicians) whereas a group subsume different actors. E.g. carers are one actor (because having the same characteristics) whereas therapists are group consisting of different kind of therapists.

- Patient (actor): geriatric patient treated by a clinician;
- Other patients (group): patients with same or similar disease and/or treated by the same clinicians or hospitalized in the same health organisation;
- Patient‘s Social Environment (group): people associated with the treated patient;
- MGT (group):The team consists of clinicians, nurses, therapists, and medical social workers;
- Clinician (actor): treating a specific patient and part of the MGT with specific geriatric education and training;
- Care (actor): nurses in charge for the patient, often with specific geriatric education and training;
- Therapists (group): logopedics, physiotherapists, occupational therapists, psychologists and also other therapists if needed. They perform their specialised assessments to monitor the treatment outcome;
- Social service (actor): medical social service workers are responsible for the social assessment, communication with other HCOs, with courts (in case of guardianship). They organise transition management to other HCOs (care home, ambulatory care);
- Other HCOs (group): these are other HCOs also responsible for the patient in the past and/or in the future, often with HQ information being important for the treatment. These HCOs can be from ambulatory care/out-patient care (general physicians, consulting/specialist physicians, ambulatory medical services provide by hospitals), hospital care/in-patient care (other hospitals), rehabilitation care, stationary care (nursing homes, retirement homes) and home care;

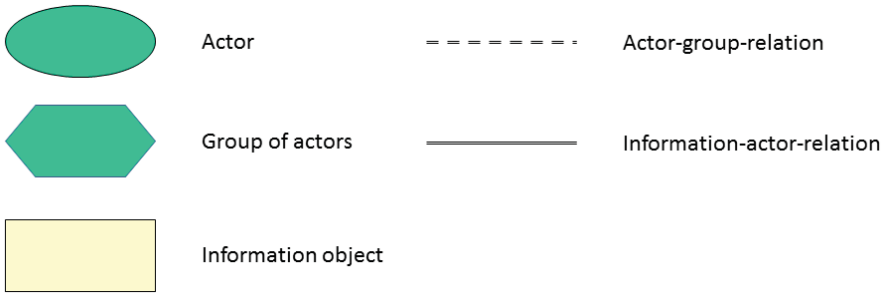



Figure 2: Key to Geriatric Information Model.

- Other clinical professions (group): these are all other clinical profession not in direct contact with the patient and not part of the MGT. This also includes other clinicians from the same discipline (higher or lower rank), associated discipline or health organisation in contact with the clinician in charge. They can be also from the geriatric discipline and/or internal medicine and share their information and knowledge in regular clinical conferences or they can be from other departments and disciplines and are often involved by consultation and/or patient transfer between the disciplines;
- Other knowledge actors (group): all other relevant knowledge actors outside the treating HCOs like medical societies, quality circles, medical specialist publishers, libraries, other hospitals from the same network (and not involved in the current treatment of the specific patient) etc. This group of actors could be also labelled as communities of practice (CoPs) (Wenger, 2000, 229ff.) (Li et al., 2009, 1ff.).

## 4.2 Information Objects

Information objects are shared between actors and groups (see section 4.3 below).

- Treatment Information: about current treatment, can contain diagnosis, treatment decisions, feedback from the patient about the progress, results of shared-decision, etc.;
- Context information: disease and behaviour related self-experiences (e.g. on procedures, medications), information about suitable contacts (specialised hospitals, physicians, disease-related support groups etc.);
- Patient‘s context information: health behaviour in the past, information on domestic and social environment;
- Clinical context information: laboratory findings, electroencephalography (EEG), electrocardiography (ECG), medical imaging, other information which is provided by specialised departments;
- Medical context information: medical background information, latest research results, clinical guidelines.

## 4.3 Actor-information Relations

The possible information relations are listed and explained below:

- Patient - MGT - Treatment Information: this is the main information relation in the geriatric treatment process. All necessary treatment from involved professions (clinician, care, therapists, social service) about the patient’s health status is resumed here;
- Patient - Other patients - Context information: this information relation contains all disease-related information, but also experience-related information like information from other patients being treated by the same HCOs or even the same clinician;
- Patient - Patient‘s Social Environment - Context information: this relation is comparable to the previous relation because persons from the patient’s social environment could be also suffering from a similar disease in the past or present;
- Other patients - Patient‘s Social Environment - Context information: in this relation other patients share their experience with the patient’s social environment. This could be information on how to act in critical disease-related questions;
- MGT - Patient‘s Social Environment - Patient‘s context information: through this relation information about the patient’s situation at home is shared. Treatment information could also be verified;

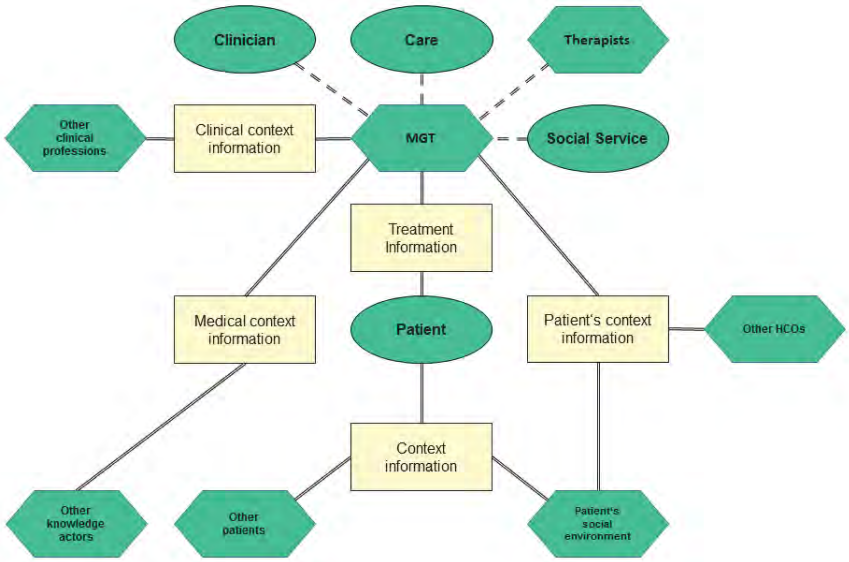



Figure 3: Geriatric Information Model.

- MGT - Other HCOs - Patient‘s context information: through this relation information about the patient’s previous treatments (other hospitals, general and specialist physicians), his domestic situation (in case of care or retirement home, or ambulatory care services) is shared;
- Patient‘s Social Environment - Other HCOs - Patient‘s context information: by this information relation the patient’s social environments shares patient’ context information with other HCOs like information on health behaviour in other contexts (previous disease, behaving in rehabilitation treatments, etc.);
- MGT - Other clinical professions - Clinical context information: this relation contains the information provided by consultations or morning, lunch or radiological conferences with other specialist clinicians but also with other professions like therapists not involved in the formal MGT;
- MGT - Other knowledge actors - Medical context information: MGT members communicate with other members of their COPs about their current treatment, they investigate in (online) libraries or journals.

# 5 VALIDATION OF GIM

To validate the GIM four typical knowledge processes are mapped to the model. The mapped knowledge processes are

- Multi-professional Geriatric Team Session (Rölker-Denker and Hein, 2015, 314f);
- Interdisciplinary Clinical Case Conferences (Rölker-Denker and Hein, 2015, 315);
- Tumor boards (Rölker-Denker et al., 2015b, 54);
- Transition management (Rölker-Denker et al., 2015a, ).

## 5.1 Multi-professional Geriatric Team Session

The MGT session is the regular meeting of the geriatric team 4.1. During this meeting all relevant information is discussed:

- Treatment information: Feedback from the patient on the health status is discussed as well as direct impressions from all persons in contact with the patient. Information passed towards the patient is also discussed as well as the further treatment process;
- Patient’s context information: this information is of very high relevance for the MGT session. This includes information about the domestic environment, e.g. how many stairs has the patient to climb at home, are there any assisting services or ambulatory care services, unhealthy behaviours and supply with medication and assisting devices;
- Clinical context information: This includes from other clinical professions like consultation results from other disciplines, blood values and medical

imaging. During the session information which will be forwarded to other clinical professions is also discussed, e.g. information for treating surgeons;

- Medical context information: This includes information stored in clinical guidelines, e.g. the guideline on urinary incontinence for geriatric patients (AWMF (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften) (engl: Association of the Scientific Medical Societies in Germany), 2016) but there also many other guidelines for age-related health issues and diseases (e.g. clinical nutrition, delirium, Parkinson disease, palliative care).

### 5.2 Interdisciplinary Clinical Case Conferences

Interdisciplinary clinical case conferences consist of members from different medical fields, the scope of these conferences is to discuss complex patient cases and to derive possible treatments (Feldman, 1999). The conferences are organised on a regular basis (Rölker-Denker and Hein, 2015, 315) (Rölker-Denker et al., 2015b, 54). During these conferences the following information is discussed:

- Treatment information: The MGT clinicians present their treatment information about the patient;
- Clinical context information: The other members of the clinical case conference provide their knowledge about the specific case and discuss with the inquiring clinicians possible treatment alternatives;
- Medical context information: other clinical professions provide and explain clinical guidelines the asking clinicians are not aware of.

### 5.3 Tumor Boards

Tumor boards are similar to clinical case conferences but focus on oncological diseases and overcome sectoral boarders by connecting clinical physicians with residential physicians and other oncological professions (Rölker-Denker et al., 2015b, 54). Geriatric oncological treatment is also multi- and interprofessional, includes the patients' social environment (Magnuson et al., 2016) and even allows patient participation (Ansmann et al., 2014, 865ff.). Mainly the same information is discussed as in the clinical case conference but in addition:

- Treatment information: in case of participation the patient can give information about the health status and also take part in the decision process on further treatment;
- Patient' context information: as residential physicians are also part of the clinical case conference (in terms of "other HCOs") they can provide more context information about the patient as clinical physicians could.

### 5.4 Transition Management

The goal of transition management is to ensure an optimal patient path through the different interfaces of cross sectoral care (Huber et al., 2016). Transition management does not only include communication between hospitals and downstream health care organisations (releasing a patient into rehabilitation or stationary/ambulatory care), it also includes the communication between hospitals and upstream health care organisations (moving patient from stationary care into hospitals) (Arve et al., 2009).

- Medical context information: up to now there is no national guideline on transition management by medical societies. But there are several local networks that develop such guidelines and make them publicly available;
- Patient's context information: this information is shared along all responsible HCOs and contains information about further medication, previous medication, recommendations on health-related behaviour (nutrition, physical activities, etc.).

## 6 CONCLUSION AND OUTLOOK

### 6.1 Conclusion

We developed an abstract geriatric information model (GIM) for the purpose of better understanding the typical actors of geriatric treatment and the information relations between them. The GIM was validated by mapping typical care settings which occur during the geriatric treatment. It was shown that all processes could be mapped into the GIM and all defined actors and information relations within the GIM are of relevance. Some knowledge processes are limited to a subset of actors (e.g. clinical case conferences do not imply the patient or the patient's social environment) whereas other knowledge processes include all actors and information relations (e.g. the MGT session).

The GIM is not intended to be used for developing sophisticated clinical information systems like other approaches, e.g. the HL7 Clinical Information Modeling Initiative (CIMI) (HL7 Clinical Information

Modeling Initiative, 2016). The purpose of CIMI is to develop interoperable healthcare systems on a technical basis. The focus is not on the communication between persons involved in the geriatric treatment. Nevertheless links to this work are mandatory in future work because geriatric treatment is cross-sectoral 3.4 and includes data and information from different IT systems.

### 6.2 Outlook

The GIM was only validated with four typical knowledge processes in geriatric treatment. Referring to previous studies of the authors (Rölker-Denker and Hein, 2015) (Rölker-Denker et al., 2015b) there are more knowledge processes to be mapped towards the GIM.

The approach of HQ and low quality information was not included in the GIM so far. There have been no dedicated studies on the information quality in daily geriatric treatment so far and, for thus, there are no validated results available. There are studies for general information quality, e.g. analyse the impact of internet health information (Laugesen et al., 2015) but there are no dedicated studies in the geriatric context.

The GIM can be also used for identifying gaps in the IT landscape (Snyder et al., 2011). Healthcare organisations can check all the actor-relation-couples and see if there are gaps.

## ACKNOWLEDGEMENTS

The authors would like to thank the Metropolregion Bremen-Oldenburg (reference number: 23-03-13) for partly supporting this work.

## REFERENCES

Ansmann, L., Kowalski, C., Pfaff, H., Wuerstlein, R., Wirtz, M. A., and Ernstmann, N. (2014). Patient participation in multidisciplinary tumor conferences. *Breast*, 23(6):865–869.

Arve, S., Ovaskainen, P., Randelin, I., Alin, J., and Rautava, P. (2009). The knowledge management on the elderly care. *International Journal of Integrated Care*, 9.

AWMF (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften) (engl: Association of the Scientific Medical Societies in Germany) (2016). Harninkontinenz bei geriatrischen patienten, diagnostik und therapie.

Chamberlain-Salaun, J., Mills, J., and Usher, K. (2013). Terminology used to describe health care teams: an integrative review of the literature. *Journal of Multidisciplinary Healthcare*, pages 65–74.

Feldman, E. (1999). The interdisciplinary case conference. *Acad Med*, 74:594.

HL7 Clinical Information Modeling Initiative (2016). Clinical information modeling initiative [online].

Huber, T. P., Shortell, S. M., and Rodriguez, H. P. (2016). Improving Care Transitions Management: Examining the Role of Accountable Care Organization Participation and Expanded Electronic Health Record Functionality. *Health Serv Res*.

Kolb, G., Breuninger, K., Gronemeyer, S., van den Heuvel, D., Lübke, N., Lüttje, D., Wittrich, A., and Wolff, J. (2014). 10 jahre geriatrische frührehabilitative komplexbehandlung im drg-system [german]. *Z Gerontol Geriat*, 47:6–12.

Kolb, G. F. and Weißbach, L. (2015). Demographic change : Changes in society and medicine and developmental trends in geriatrics. *Der Urologe*, 54(12):1701.

Laugesen, J., Hassanein, K., and Yuan, Y. (2015). The impact of internet health information on patient compliance: A research model and an empirical study. *J Med Internet Res*, 17(6):e143.

Li, L. C., Grimshaw, J. M., Nielsen, C., Judd, M., Coyte, P. C., and Graham, I. D. (2009). Use of communities of practice in business and health care sectors: a systematic review. *Implement Sci*, 4:27.

Magnuson, A., Wallace, J., Canin, B., Chow, S., Dale, W., Mohile, S. G., and Hamel, L. M. (2016). Shared Goal Setting in Team-Based Geriatric Oncology. *J Oncol Pract*.

Mangoni, A. A. (2014). Geriatric medicine in an aging society: Up for a challenge? *Frontiers in Medicine*, 1(10).

Nau, R., Djukic, M., and Wappler, M. (2016). Geriatrics – an interdisciplinary challenge. *Der Nervenarzt*, 87(6):603–608.

Rölker-Denker, L. and Hein, A. (2015). Knowledge process models in health care organisations - ideal-typical examples from the field. In *Proceedings of the International Conference on Health Informatics (BIOSTEC 2015)*, pages 312–317.

Rölker-Denker, L., Seeger, I., and Hein, A. (2015a). Überleitung aus sicht der krankenhäuser - ergebnisse aus semi-strukturierten leitfadeninterviews auf ebene der geschäftsführung in der region metropolregion bremen-oldenburg. In *14. Deutscher Kongress für Versorgungsforschung. Deutsches Netzwerk Versorgungsforschung e. V. 7. - 9. Oktober 2015, Berlin*.

Rölker-Denker, L., Seeger, I., and Hein, A. (2015b). Knowledge processes in german hospitals. first findings from the network for health services research metropolitan region bremen-oldenburg. In *eKNOW 2015, The Seventh International Conference on Information, Process, and Knowledge Management*, pages 53–57. IARIA.

Snyder, C. F., Wu, A. W., Miller, R. S., Jensen, R. E., Bantug, E. T., and Wolff, A. C. (2011). The role of informatics in promoting patient-centered care. *Cancer J*, 17(4):211–218.

Soriano, R. P., Fernandez, H. M., Cassel, C. K., and Leipzig, R. M. (2007). *Fundamentals of Geriatric Medicine : A Case-Based Approach*. New York : Springer Science+Business Media, LLC, New York.

Tanaka, M. (2003). Multidisciplinary team approach for elderly patients. *Geriatrics & Gerontology International, 2003, Vol.3(2), pp.69-72*, 3(2):69.

Wenger, E. (2000). Communities of Practice and Social Learning Systems. *Organization*, 7(2):225–246.

# Collaborative Reproducible Reporting
## *Git Submodules as a Data Security Solution*

Peter E. DeWitt[1] and Tellen D. Bennett[2]
[1]*Biostatistics and Bioinformatics, Colorado School of Public Health,*
*University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, U.S.A.*
[2]*Pediatric Critical Care, University of Colorado School of Medicine, Children's Hospital Colorado, Aurora, CO, U.S.A.*
{*peter.dewitt, tell.bennett*}*@ucdenver.edu*

Keywords: Data Security, Collaborative Authoring, Reproducible Reports, Workflow, Software.

Abstract: Sensitive data and collaborative projects pose challenges for reproducible computational research. We present a workflow based on literate programming and distributed version control to produce well-documented and dynamic documents collaboratively authored by a team composed of members with varying data access privileges. Data are stored on secure institutional network drives and incorporated into projects using a feature of the Git version control system: submodules. Code to analyze data and write text is managed on public collaborative development environments. This workflow supports collaborative authorship while simultaneously protecting sensitive data. The workflow is designed to be inexpensive and is implemented primarily with a variety of free and open-source software. Work products can be abstracts, manuscripts, posters, slide decks, grant applications, or other documents. This approach is adaptable to teams of varying size in other collaborative situations.

## 1 INTRODUCTION

Reproducible reporting, defined here as processing data and generating an abstract, manuscript, slide deck, or poster via a fully documented and automated process, is considerably more difficult when working with multiple authors and sensitive data, such as protected health information (PHI). Workflows for reproducible computational research using tools such as the Jupyter Notebook[1], the Galaxy project[2], or RStudio (Gandrud, 2015) are not consistently used in biomedical research (Peng et al., 2006; National Academies of Sciences, Engineering, and Medicine, Division on Engineering and Physical Sciences, Board on Mathematical Sciences and Their Applications, Committee on Applied and Theoretical Statistics, 2016). This may be due to concerns about slower production, the need for investigators to learn new tools, or barriers to collaboration between investigators with varying computational skills and development environments. Collaborative research involving sensitive data poses additional challenges.

One solution would be for a team to work in a single development environment hosted on a computational server with the necessary physical and electronic security standards for the level of sensitivity of the data. However, the financial investment required to build a full development environment behind an institutional firewall might be prohibitive for some research teams. Fortunately, a reproducible collaborative workflow that protects sensitive data is possible at much lower cost.

We minimize team hardware and software expenses in two ways. First, only those team members who require data access are provided with institutionally-owned laptops with licenses for whole-disk encryption and other proprietary software. Second, by using free and open-source software for version control, analysis, and manuscript authoring, we incur minimal financial expenses when new team members join or when we collaborate with external investigators.

Our solution to data protection and collaboration is to compartmentalize and distribute our project such that data resources, analysis scripts, and text are all linked together, version-controlled, and access-controlled via implicit and explicit read/write permissions. Raw data is stored on institutionally owned network drives and cloned on laptop computers which have been approved for storage of our data. Only team

[1]http://jupyter.org/
[2]http://galaxyproject.org/



Dewitt, P. and Bennett, T.
Collaborative Reproducible Reporting - Git Submodules as a Data Security Solution.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 230-235
ISBN: 978-989-758-213-4

members with institutional review board approval can access the data. Data analysis scripts and manuscript text files are available to all team members on public code hosting services. The linkage between the data and the code is made possible by a feature of the Git version control software: submodules. A 40-character hexadecimal sequence (SHA-1 hash) allows us to share the *version* of the data source publicly without compromising the data itself.

The objective of this manuscript is to present a workflow that we developed to 1) protect sensitive data from unauthorized access, 2) allow multiple authors, included those with and without data access rights, to contribute to a single set of files, and 3) minimize the financial commitment to hardware and software.

Our primary focus is on the use of the Git version control system and specifically, Git submodules. We will note other software tools and programs used in our workflow, but they can often be substituted for other similar software.

## 2 WORKFLOW OVERVIEW

Dynamic document authoring is a key component of the overall reproducible research paradigm. Variations on literate programming (Knuth, 1984) are ideal for this purpose. The R package `knitr` (Xie, 2015), an evolution of R's `sweave` (Leisch, 2002) package, provides a structured process for authoring a manuscript using a literate programming paradigm. `knitr` was highlighted several times at a recent workshop supported by the National Academies of Sciences, Engineering, and Medicine (National Academies of Sciences, Engineering, and Medicine, Division on Engineering and Physical Sciences, Board on Mathematical Sciences and Their Applications, Committee on Applied and Theoretical Statistics, 2016).

We typically perform data analysis with the statistical language R[3] and rely on either markdown or LaTeX for markup. The desired format of our deliverables dictates the markup language selection. Weaving R code with a markup language is well-described (Gandrud, 2015).

Our team manages collaborative projects using a distributed version control system, Git[4]. Git is free to use and is supported on all major operating systems. Distributed version control systems are becoming more common than centralized systems, although some distributed version control projects, including many of ours, have a centralized design (De Alwis and Sillito, 2009).

In the simplest centralized design, a Git server hosts the repository and each team member would *push* to, and *pull* from, that server. It is possible to have the individual team members' repositories directly linked, but we did not use this option because of network security concerns. Another option is to have a *bare* repository on a network drive act as the central code repository. We use that design for a minority of projects with unusually sensitive data. For most projects, our team takes advantage of the integrated issue tracker, web editing interface, and additional read/write permissions provided by a Git server.

Several public Git repository sites exist. We chose to use Atlassian's Bitbucket[5] to host our repositories. At the time this choice was made, Bitbucket allowed academic account holders unlimited private repositories and unlimited collaborators. Recently, Github.com has offered similar packages.

Code repositories solved the problems of dynamic document authoring and collaboration, but we also needed to track data set versions and limit data access to approved team members without preventing collaboration.

The solution was to use Git submodules. "Submodules allow you to keep a Git repository as a subdirectory of another Git repository. This lets you clone another repository into your project and keep your commits separate." (Chacon and Straub, 2014). Also, while the data files within the submodule exist in a subdirectory and are visible in the working directory, only the SHA-1 of the commit of the submodule is stored in the primary project repository. Thus, when the manuscript repository is pushed to bitbucket.org, the only reference to the data is a 40-digit hexadecimal number. The data never leaves the team members' machines, but the status of the data is shared and documented between team members.

## 3 INFRASTRUCTURE

Below we describe how we have used existing infrastructure, open source software, and free hosting services to generate reproducible reports while protecting sensitive data. We designed the workflow so that sensitive data is stored on a secure network hard drive or whole-disk encrypted personal machine. Data transfer between the network drive and a team member's machine only occurs on the institution's

[3] https://www.r-project.org/

[4] https://Git-scm.com/

[5] https://bitbucket.org

network. The following subsections describe the necessary hardware, repository design, and workflow for collaboration.

## 3.1 Hardware

Our institution maintains a Microsoft Windows network. We chose to work on Windows machines because they are available to all of our team members and because they support whole-disk encryption software that meets our institution's requirements for data security. Each team member with access to the data has a whole-disk encrypted laptop or desktop. This software costs approximately 100 US Dollars per machine, but allows each team member to have local copies of the data repositories relevant to their work. Like investigators at many academic institutions, we have access to secure network drives behind the university's firewall. We rely on the network drives for data repository hosting and backup.

## 3.2 Repository Design

Although Git is a distributed version control platform, we conceptually have central data and code repositories on a network drive or Git server, see Figure 1. Each collaborator has a local *clone* of the necessary data and code repositories on their machine that serve as distributed backups of the central data and code.

### 3.2.1 Data Repositories

Data are housed in .csv format within our local data repositories. For collaboration, team members with data access privileges *push* to and *pull* form *bare* Git repositories on our institution's secure network drives. Bare repositories do not contain a working directory: individual files are not visible when inspecting the contents of the directory and subdirectories. As such, inadvertently editing or over-writing the data files is very unlikely. We rely on the read/write access limits enforced by the institution's network to limit access to these bare repositories and entrust those with access to not manually edit files. The repositories theoretically could become corrupted. If that occurred, we would compare the distributed copies of the repositories between team members and re-initialize the repository from the most current local copy. This is an advantage of the distributed version control paradigm: every local copy is a backup for all others.

### 3.2.2 Code Repositories

In the simplest form of this workflow, a work product such as a manuscript has its own code repository. A basic repository design shown in Figure 2. An example code repository can be found at https://bitbucket.org/pedstbi/example_collaborative_report which has a data submodule available at https://bitbucket.org/pedstbi/example_collaborative_data_source. The analysis and manuscript authoring code is free of sensitive data. Therefore, the remote code repository can be maintained on a publicly available code development system such as GitHub or Bitbucket. We use private code repositories to maintain academic confidentiality prior to manuscript publication. Repositories on either GitHub or Bitbucket can be made public at any time, such as when a manuscript is submitted for publication.

A team member working on a project (manuscript, in this case) would have a local *clone* of the code repository on their machine. Their daily workflow would be to *fetch* and *merge* (*pull*-ing is shorthand for the *fetch* then *merge* process) any changes on the remote repository made by other team members, make changes to files using the text editor of their choice, *stage* and *commit* the changes using their local Git instance, then *push* those changes to the remote repository. Team members (clinical authors or collaborators at other institutions, for example) whose contributions are focused on writing the manuscript or who do not have a whole-disk encrypted machine might have a local copy of the code repository but not the data repository. Those team members can have the benefits of a version-controlled project without cloning the data submodule. One challenge introduced by this approach is that collaborators without local data repositories cannot compile manuscripts. Because the quantitative results in the manuscript are generated by embedded analytic code within the manuscript file, those results cannot be updated without a local data copy. Periodically, team members with both data and code access must compile the manuscript (which runs the embedded code) and *commit* the finished product to the central code repository for reference by collaborators who primarily write and edit manuscript text.

### 3.2.3 Limitations

The size of data submodules is the most important limitation of this repository design. Thus far, the largest data submodule in our system is approximately 10GB. Segmentation of the data repositories into, for example, a large raw data repository and a smaller analysis data repository for one project can improve efficiency.

Additional features of Git such as *branch*-ing strategies, forking, pull requests, *rebase*-ing, and others, provide additional levels of structure within the collaboration. However, such tools can be over-

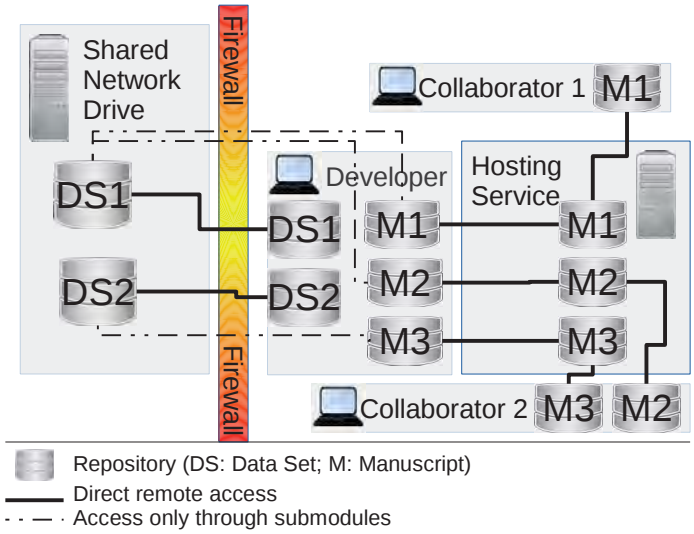



Figure 1: Collaboration Structure. Data is version-controlled in *bare* repositories on our institutional shared network drives. The datasets are tracked within projects as submodules. Each developer has access to the data on his or her whole-disk encrypted laptop or desktop. Non-sensitive code, manuscript text, references, etc. are hosted on bitbucket.org. Other authors are able to contribute by having access to the bitbucket.org code repositories. Note that the manuscript repositories, M1, M2, and M3, only have access to the data sets via git submodules. The copies of M1, M2, and M3 on the hosting service and on each collaborator machine have no access to the data sets. The hosting service and collaborator only see a 40-digit hexadecimal SHA1 to reference the version of the data repository.

whelming for a novice Git user. Increased training time, or limited participation, must be weighed against the benefit of Git feature use.

This workflow does include copies of sensitive datasets on the whole-disk encrypted local machines of selected team members. Our experience has been that data owners and institutional review boards are supportive of this approach. If a particular dataset was not permitted to be housed on a local machine with whole-disk encryption, then a computational server within the institution's firewall would likely be necessary.

Clinical members of our research team without computational backgrounds have been able to adopt most or all of this workflow with a modest time investment. However, like all complex tools, regular use is needed to maintain comfort. A more integrated environment that was friendly to the naïve user would increase the accessibility of a reproducible reporting workflow.

#### 3.2.4 Extensions/Other Options

Our team initially hosted code repositories on GitHub and moved to Bitbucket as the team grew and the number of projects increased. GitLab.com is another option that offers unlimited private repositories, unlimited collaborators, and up to 10GB disk space per repository (compared to Bitbucket's 1GB soft and 2GB hard limits). Placing a dedicated Git server behind our institutional fire wall would provide a solution for data management and access control and useful collaboration tools. Hardware and administrative support costs would need to be considered.

## 4 COSTS

This reproducible reporting workflow is powerful and also cost-effective. For many investigators, a Windows operating system and Windows Office software are supported by the institution. Whole-disk encryption software is inexpensive (100 US Dollars per team member). Other software needed to implement this workflow is free to use under the GNU General Public License (GPL)[6] or similar license. There are no hardware costs if investigators currently have individual computers capable of performing the planned analyses and access to a secure network drive. Many academic investigators already have this hardware in place.

The time and effort needed to learn the necessary tools to adopt this workflow are likely higher than the software and hardware costs. However, the return on investment can be high. Our experience in an academic research environment suggests that a team

[6] http://www.gnu.org/licenses/gpl-3.0.en.html

```
. <user-path>/project1/
|-- .git/                          # the Git repository
|-- analysis-scripts/              # data analysis scripts
|   |-- data-import.R
|   |-- primary-analysis.R
|   |-- secondary-analysis.R
|   `-  figures.R
|-- data/                          # A Git submodule
|-- products_donotedit/            # generated files
|   |-- cache/
|   |   |-- documentation-data-import-cache/
|   |   |-- documentation-analysis-cache/
|   |   `-- manuscript-cache/
|   |-- figures/
|   |-- tables/
|   |-- coverletter.docx
|   |-- coverletter.md
|   |-- documentation-data-import.html
|   |-- documentation-analysis.html
|   |-- manuscript.docx
|   |-- manuscript.md
|   `-- poster.pdf
|-- coverletter.Rmd                ## Files for authoring
|-- documentation-data-import.Rmd  ## coverletters,
|-- documentation-analysis.Rmd     ## documentation,
|-- manuscript.Rmd                 ## manuscripts, posters,
|-- poster.Rnw                     ## etc.
`-- README.md                      # project README
```

Figure 2: A generic repository layout for a manuscript written in `Rmarkdown`. Not shown in the graphic, but part of our overall design, are build scripts. A build script is a R script, .cmd or .sh file, or makefile. The format and location of the build script is project-specific. We decide which format to use based on the complexity of the build required, the development platforms (Windows, Mac, or Linux), the integrated development environments (RStudio or vim are used by our team), and ease of use.

adopting this workflow might see research production slow for up to six months, recover to initial levels within a year, and show potential increases after one year. Improvements in quality and reproducibility are difficult to quantify but are valuable.

# 5 ALTERNATIVE APPROACHES

Another solution to the simultaneous problems of multiple collaborators and sensitive data might be to run a local instance of Galaxy.[7] However, most Galaxy tools use Python.[8] Few clinical researchers have the training and experience to collaboratively develop analysis code in Python. Many more have been trained to use R. A capable computational server would solve the problems of multiple collaborators and data security. However, the purchase (5,000 US Dollars and up) and maintenance (varying, but potentially exceeding 1,000 US Dollars per year) costs for such a server are beyond the reach of most small research teams. Because many biomedical manuscripts are generated by small teams, we think it likely that the workflow we present here will be generalizable.

Existing cloud-based solutions such as RunMyCode.org[9] and the Open Science Framework[10] are reproducible and support multiple collaborators, but are not designed to protect sensitive data. Cloud-based computational server services, some of which now have robust data security features, are another option. Their utility will grow once institutional review boards and data owners (health care organizations, insurance companies, etc.) gain enough confidence in the data security measures used by those services that researchers are consistently permitted to analyze sensitive datasets in those environments.

We did not extensively test our Git-based solution against other possible solutions. This was primarily for two reasons. First, most available alternative approaches did not provide sufficient data security. Second, alternative approaches with sufficient data security required additional financial commitment beyond standard operating expenditures. We developed this workflow as part of an active academic research team and needed to maintain productivity in our content areas. The lack of formal method comparison is a limitation of this manuscript at this time. However, our team's ability to rapidly adopt this workflow and maintain productivity highlights the value and ease of use of this approach.

# 6 DISCUSSION

Collaborative and reproducible biomedical reporting can be inexpensive and have low barriers to entry even when working with sensitive data and a team with variable technical skills. Our goal is to introduce an overall workflow and one set of viable tools. Many data processing/analysis languages, markup languages, text editors, file formats, and file sharing systems can be used.

Peng (Peng et al., 2006; Peng, 2011) has suggested criteria for the reproducibility of epidemiologic and clinical computational research. The workflow we present here would meet the criteria for Methods (free, open-source software, public code repositories), Documentation (well-commented code in the repository), and Distribution (code repositories on

[7]https://galaxyproject.org/
[8]https://www.python.org/
[9]http://www.runmycode.org/
[10]https://osf.io/

public Git servers). However, due to the limitations regarding disclosure of data, our workflow would not meet Peng's Data Availability criterion. Summary statistics (Peng et al., 2006) could in some situations be posted publicly, but overall the balance between reproducibility and data privacy will need additional public discussion (National Academies of Sciences, Engineering, and Medicine, Division on Engineering and Physical Sciences, Board on Mathematical Sciences and Their Applications, Committee on Applied and Theoretical Statistics, 2016).

Rossini and Leisch described how "information and knowledge was divided asymmetrically between [collaborators]..." (Rossini and Leisch, 2003) and Donoho reported that one of the benefits of a reproducible computational workflow was improved teamwork (Donoho, 2010). Our experience would support both of those ideas, as team members with variable clinical, statistical, and technical backgrounds have all contributed to the development of this workflow and to the quality of the workflow's research products.

In conclusion, reproducible reporting is a key component of the reproducible research paradigm. This manuscript presents an inexpensive, practical, and easily adopted workflow for collaborative reproducible biomedical reporting when working with sensitive data.

## ACKNOWLEDGEMENTS

We thank Dr. Michael Kahn for his comments and suggestions on our manuscript. Dr. Bennett is supported by *Eunice Kennedy Shriver* National Institute for Child Health and Human Development Grant K23HD074620.

## REFERENCES

Chacon, S. and Straub, B. (2014). *Pro git*. Apress. Online at https://git-scm.com/book/en/v2.

De Alwis, B. and Sillito, J. (2009). Why are software projects moving from centralized to decentralized version control systems? In *Cooperative and Human Aspects on Software Engineering, 2009. CHASE'09. ICSE Workshop on*, pages 36–39. IEEE.

Donoho, D. L. (2010). An invitation to reproducible computational research. *Biostatistics*, 11(3):385–388.

Gandrud, C. (2015). *Reproducible Research with R and RStudio*. Chapman & Hall/CRC Press, second edition.

Knuth, D. E. (1984). Literate programming. *The Computer Journal*, 27(2):97–111.

Leisch, F. (2002). Sweave: Dynamic generation of statistical reports using literate data analysis. In Härdle, W. and Rönz, B., editors, *Compstat 2002 — Proceedings in Computational Statistics*, pages 575–580. Physica Verlag, Heidelberg. ISBN 3-7908-1517-9.

National Academies of Sciences, Engineering, and Medicine, Division on Engineering and Physical Sciences, Board on Mathematical Sciences and Their Applications, Committee on Applied and Theoretical Statistics (2016). *Statistical Challenges in Assessing and Fostering the Reproducibility of Scientific Results: Summary of a Workshop*. National Academies Press.

Peng, R. D. (2011). Reproducible research in computational science. *Science*, 334(6060):1226–1227.

Peng, R. D., Dominici, F., and Zeger, S. L. (2006). Reproducible epidemiologic research. *Am J Epidemiol*, 163(9):783–9.

Rossini, A. and Leisch, F. (2003). Literate statistical practice. Biostatistics Working Paper Series. Working Paper 194. accessed May 17th, 2016.

Xie, Y. (2015). *Dynamic Documents with R and knitr, Second Edition*. Chapman & Hall/CRC The R Series. CRC Press.

# Managing Provenance for Medical Datasets
## *An Example Case for Documenting the Workflow for Image Processing*

Ajinkya Prabhune[1], Rainer Stotzka[1], Michael Gertz[2], Lei Zheng[3] and Jürgen Hesser[3,4]
[1]*Institute for Data Processing and Electronics, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany*
[2]*Database Systems Research Group, Institute of Computer Science, Heidelberg University, Heidelberg, Germany*
[3]*Experimental Radiation Oncology, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany*
[4]*Experimental Radiation Oncology, IWR, Heidelberg University, Heidelberg, Germany*
{*ajinkya.prabhune, rainer.stotzka*}*@kit.edu, gertz@informatik.uni-heidelberg.de,*
{*lei.zheng, juergen.hesser*}*@medma.uni-heidelberg.de*

Keywords: Scientific Data Repository, ProvONE Provenance Standard, Open Annotation Data Model, DICOM Dataset, Angioscopy Workflow, Metadata Management.

Abstract: In this paper, we present a novel data repository architecture that is capable of handling the complex image processing workflows and its associated provenance for clinical image data. This novel system has unique and outstanding properties versus existing systems. Among the most relevant features are a flexible and intuitively usable data and metadata management that includes the use of a graph-based provenance management strategy based on a standard provenance model. Annotation is supported to allow for flexible text descriptors as being widespread found for clinical data when structured templates are not yet available. The architecture presented here is based on a modern database and management concepts and allows to overcome the limitations of current systems namely limited provenance support, lacking flexibility, and extensibility to novel requests. To demonstrate the practical applicability of our architecture, we consider a use case of automated image data processing workflow for identifying vascular lesions in the lower extremities, and describe the provenance graph generated for this workflow. Although presented for image data, the proposed concept applies to more general context of arbitrary clinical data and could serve as an additional service to existing clinical IT systems.

## 1 INTRODUCTION

Electronic health records (EHR) offer a digital documentation of the diagnostic and therapeutic history of a patient. Parts of these records are managed by Hospital information systems (HIS) and subsystems like radiology information systems (RIS). Over the initiative integrating the healthcare enterprise (IHE) defined workflows enable standardized records on one hand side and (as a final goal) a complete coverage of all procedures in a clinic. However, from the perspective of data provenance, these systems partially solve the data management problem. Data provenance is hereby of utmost relevance since it allows traceability with respect to validation and reproducibility of diagnostic and therapeutic procedures and decisions (Estrella et al., 2007). However, despite its relevance, this topic is scarcely discussed for clinical routine data reporting with some exceptions in bioinformatics (Davidson et al., 2007). Approaches that have been widespread used are neither complete, nor flexible and are therefore difficult to integrate in a clinical environment.

In the following, we propose a new architecture for medical data repository system with a dedicated focus on medical image data processing. Currently, the image data is mostly obtained in radiology departments, and the acquired data such as Magnetic Resonance Imaging (MRI) or Computerized Tomography (CT) data is typically stored in a Picture Acquisition and Communication System (PACS), backed by DICOM 3.0 format (Mildenberger et al., 2002). This format allows storing image data and the associated metadata such as patient id, type of acquisition, date, etc, that is embedded within the DICOM files. Thus, all relevant information required to repeat the acquisition is stored and accessible via reading and interpreting the DICOM header files.

Although being in widespread use, two major limitations of the PACS are, (a) the lack of a modern database (NoSQL) technology for storing, indexing and accessing metadata. (b) the brittleness of han-



Prabhune, A., Stotzka, R., Gertz, M., Zheng, L. and Hesser, J.
Managing Provenance for Medical Datasets - An Example Case for Documenting the Workflow for Image Processing.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 236-243
ISBN: 978-989-758-213-4

dling the workflow and provenance information for reproducing the results. For example, manipulations of image data via image processing routines (which will dominate in the future to automatically process standard cases), the processed result is stored as new DICOM data, even if only small modifications on the header information were performed.

From the perspective of repeatability, the image processing workflow would have to be stored in the DICOM data as well. This can be accomplished either by using private tags or by defining a new modality. The former step is an ad-hoc solution since there are no formal rules how this should be done and generally we cannot assume that all users are handling this issue with sufficient care. Hence, there is no guarantee that this image processing workflow can ever been repeated. The latter strategy requires a new DICOM modality standard, which is an immense overhead.

From this perspective, there is a demand for an easy-to-use technology that not only maps all metadata formats, workflows, and patient related data but also enables users to describe new workflows and still guarantee the provenance information.

This lack of flexibility of the current system and the brittleness of the DICOM standard led us to propose a different type of data repository architecture that is overcoming these limitations. Instead of introducing a disruptive solution in the existing PACS infrastructure, we present an auxiliary system that provides the following functionality;

- Metadata management for extracting, modelling, indexing and storing the metadata embedded from DICOM-file in a flexible and a scalable database
- Provenance tracking using standard provenance models such as ProvONE and PREMIS[1] for the image processing workflows.
- Data annotation (image annotations) for systematic capturing of vital details in the standard Open Annotation Data Model (Sanderson et al., 2013).
- Data preservation for allowing long-term access and reusability of the data.
- Data quality control and data curation for specific tools controlling rules of good practice and later diagnostic and treatment guidelines.

# 2 IMAGE PROCESSING DATA REPOSITORY

The goal of the Image Processing Data Repository (IPDR) is to provide the various auxiliary functionality for handling the lifecycle of DICOM dataset stored in the PACS server. The IPDR follows a modular architecture design, as shown in Figure 1. The modular design of IPDR allows us to extend the functionality to fulfil any new requirements put forth by the radiation therapy (RT) research community. The complete IPDR architecture is developed in Java, the various REST services are implemented in Jersey (Sandoval, 2009) and the front-end is developed in Vaadin framework. The IPDR is deployed on a web server (Tomcat 7). Following is the description of the core components of the IPDR architecture.

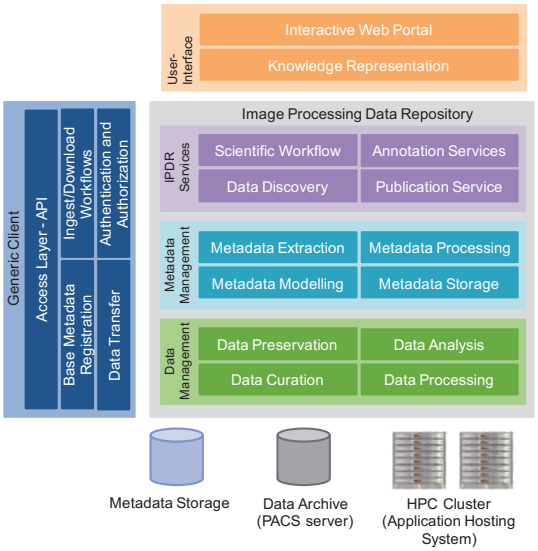



Figure 1: Image Processing Data Repository Architecture.

## 2.1 Generic Client

The Generic Client component provides a convenient solution for the RT medicine and research community to integrate their existing software/tools with the IPDR seamlessly. The Generic Client is an upgrade to the GCS-API (Prabhune et al., 2015) that was previously developed for handling the datasets of the nanoscopy research community. The Generic Client is extended to handle the DICOM files. The various modules of Generic Client are explained below:

- Access Layer-API: The access layer API exposes the various interfaces for connecting to the existing radiation therapy data processing tools. Currently, we have developed a command line interface (CLI) over the access layer API that allows the RT community to transfer the datasets.
- Base Metadata Registration: For long term archival, it is necessary to register of the DICOM files in IPDR. However, the DICOM files main-

[1] http://www.loc.gov/standards/premis/

tain its proprietary metadata schema that needs to be translated to the standard metadata schema supported by the IPDR. This module enables the automated registration of the DICOM datasets by extracting the administrative metadata concerning the DICOM Study, Series, and Image and mapping it to the Core Scientific Metadata Model (CSMD) (Matthews, 2015).

- Ingest/Download Workflows: Ingest/Download workflows module holds the predefined workflow that allow the ingestion of DICOM data from client systems to the IPDR or to download the DICOM datasets from the IPDR to the client system.
- Data Transfer: The data transfer module allows the transfer of DICOM datasets bi-directional from multiple endpoint, i.e. from client system or data acquisition systems to IPDR or vise versa. Currently, the HTTP WebDAV (Goland et al., 1999) protocol is supported for high-throughput transfer of the datasets. The various interfaces for integrating other transfer protocols such as, FTP and gridFTP (Allcock et al., 2005) are available. The data transfer module is designed to be fail-safe. For example, data transfers that are interrupted are automatically re-triggered for transfer. To optimize high-volume data transfers, the number of parallel WebDAV connections can be manipulated.
- Authentication and Authorization: The Generic Client follows a two-fold process. First, to enable the registration of DICOM datasets , for a registered member of the RT research community, an OAuth-secured RESTful connection is established. Each researcher is provided with a unique access token and access token secret. Second, for transferring the data, the WebDAV protocol authentication is required. Currently, WebDAV authentication for an entire research community is configured.

## 2.2 Image Processing Data Repository

The IPDR is a multi-module architecture that is separated into functionality specific module. The IPDR is deployed as a web application with access to the high-performance computing cluster, the large-scale data storage, and a dedicated metadata database. Following is the description of each module of the IPDR:
**IPDR Services:** The services module provides the various arbitrary high-end services for interacting with datasets stored in IPDR.

- Scientific Workflow (medical Image Data Processing workflow) submodule offers the integration of a workflow engine for automating the execution of the image processing workflows. Currently, the Apache ODE[2] BPEL workflow engine is integrated with IPDR and workflows defined in BPEL specification (OASIS, 2007) are supported. The Prov2ONE (Prabhune et al., 2016) algorithm handles the automatic creation of prospective ProvONE graph, which is stored in a graph data model of the metadata storage.
- Data discovery submodule provides the users to search the ingested DICOM datasets based on the metadata stored in the metadata storage. The metadata are indexed for enabling full-text search, for searching over provenance graphs various queries implemented as RESTful services are provided. Finally, to enable large scale metadata harvesting the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)[3] is implemented.
- Annotation service is an interactive submodule that allows enriching of the images through the annotation service. Valuable information provided in form of annotations by the researchers is modeled using the Open Annotation Data Model and maintained in the metadata storage.
- Publication service submodule allows users to share (provide open access) the experiment dataset with other research communities, which can be based on data exchange technologies like i2b2[4].

**Metadata:** The metadata module is responsible for handling all the metadata specific tasks in the IPDR. Metadata can either be embedded inside the DICOM files or for other file formats it can be ingested separately with the datasets. For enabling querying, sharing and reusing of metadata, DICOM metadata needs to be stored in a dedicated metadata storage. Furthermore, as DICOM metadata is subject to frequent changes throughout its lifecyle, it is necessary to have a flexible database model. Thus, we chose ArangoDB[5], a database which offers three different types of data models, namely, key-value, document and graph data model.

- Metadata Extraction: The various metadata extractors for extracting the metadata from DICOM, HDF5, TIFF and XML files are provided as independent micro-services by this module. For extending towards other file formats, this module provides a generic interface that can be implemented by any new metadata extractors.
- Metadata Modelling: The various metadata models (schemas) are registered through this com-

[2]http://ode.apache.org/
[3]https://www.openarchives.org/pmh/
[4]www.i2b2.org
[5]https://www.arangodb.com/documentation/

ponent. For example, the ProvONE provenance model, CSMD, Metadata Encoding and Transmission Standard (METS)[6] and PREMIS are some of the currently supported metadata schemas. Furthermore, application-specific metadata model of DICOM is also registered by this submodule.

- Metadata Processing: The metadata processing submodule provides the handling and assembling of the community specified METS profile. METS is metadata container format that comprises of various sections which allow encoding of administrative ⟨amdSec⟩, structural ⟨fileSec⟩, ⟨structMap⟩, ⟨structLink⟩ descriptive ⟨dmdSec⟩ and provenance ⟨digiprovMD⟩ metadata. The heterogeneous metadata comprising descriptive and administrative metadata from the ArangoDB document store and provenance metadata from graph store are assembled in the METS profile for enabling sharing of entire metadata for a dataset.
- Metadata Storage: The various database specific storage adapter for storing and querying the metadata are implemented in this submodule. The CRUD operations for document metadata representing the contextual information from the DICOM dataset, for graph metadata representing the provenance and workflows, and the six verbs of the OAI-PMH protocol are implemented in Arango Query Language (AQL).

**Data:** The Data module provides the integration with the low-level functionality that is responsible for storing the data in the cache storage and further in the tape storage for long term archival. This data storage represents the PACS server, which is enriched with the high-performance computing (HPC) cluster for enabling processing of the DICOM dataset, where these data processing services (algorithms) are deployed on the HPC. Following are functionalities that extend the underlying PACS server.

- Data Preservation: The Data Preservation submodule provides with the checksum of the dataset that is to be ingested in IPDR. This checksum is maintained in the PREMIS metadata schema for verifying the consistency of a file during data transfers through Generic Client.
- Data Analysis and Curation: The various community specific data analysis and curation algorithm are registered in these submodules. The data processing algorithms are deployed as individual, reusable processes that expose a unique REST endpoint, which is used when assembling a scientific workflow (necessary for composing the BPEL workflow).

[6]http://www.loc.gov/standards/mets/

- Data Processing: The integration with the HPC cluster and configuration of the execution environment is handled by this submodule. Moreover, when new intermediary datasets are generated as a result of an execution of data processing algorithm, the automated registration and ingest of these datasets is performed by this submodule.

### 2.3 User Interface

A web-based user interface developed in Vaadin is integrated with the IPDR. Feature such as free-text search over metadata and faceted search are available for discovering the datasets. Provenance graphs stored in ArangoDB are visually represented using the D3.js (Zhu, 2013) framework. The OAI-PMH metadata services are exposed through the user interface for harvesting the metadata. As the user interface is integrated with the workflow engine, the radiation therapy research community can perform remote execution of their image processing workflows.

## 3 DATA AND METADATA WORKFLOW

Datasets in interventional radiological clinics follow a systematic workflow, starting from the image acquisition in DICOM format to the final generation of a treatment record. During each step of the workflow, the DICOM images are subject to image manipulations and diagnostic information is extracted; these datasets are hence enriched with essential metadata describing the diagnostic details at each step. Moreover, additional and related datasets might be created as well. To manage these datasets, we described the complete flow of the DICOM datasets and metadata beginning from the Data Acquisition system (DAQ) and the treatment to making it accessible to clinicians for finding similar cases or for quality assurance reasons; and to use it for documentation. The complete workflow, described in our example case consists of eight steps as shown in Figure 2. (1) The workflows begins when the either raw DICOM dataset acquired from the CT scanner system is made available to the Generic Client. (2) The Generic Client is entirely automated, it performs the registration of the dataset to be ingested by extracting the base metadata from the DICOM metadata section, translates it into the CSMD standard and registers the dataset. (3) A successful registration of base metadata triggers the transfer of data to the PACS server. (4) The complete metadata from the DICOM dataset is extracted, modelled and

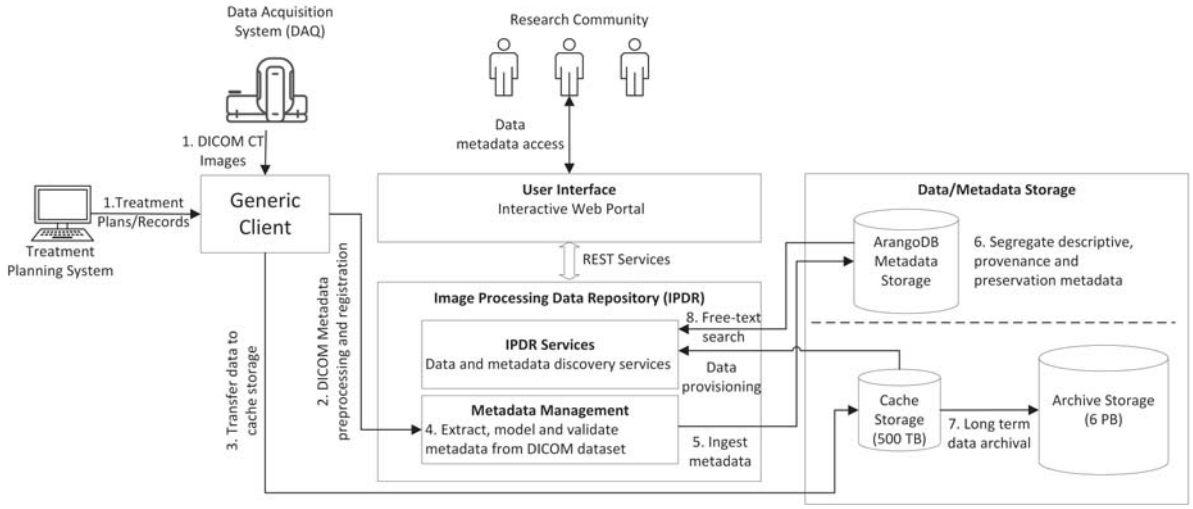



Figure 2: Flow of data and metadata from clinical data acquisition and treatment system to IPDR.

validated. (5) The metadata is ingested to the dedicated metadata storage database. (6) Metadata is segregated as descriptive or provenance metadata and stored either in document data model or graph data model of ArangoDB. (7) The DICOM dataset from the cache storage is transferred to archive storage for long-term preservation. The preservation metadata associated with the dataset is updated in the metadata storage. (8) The metadata is indexed for enabling free-text search and allowing discovery of the datasets from the IPDR. The data and metadata is accessible in the Vaadin based user interface which is connected to the IPDR through various REST services.

# 4 WORKFLOW AND PROVENANCE

A successful completion of the Schedule (Acquisition) Workflow (SWF) generates the various DICOM files that are stored in the PACS server. The Post-Processing Workflow (PAWF) is the logical extension of the SWF (Liu and Wang, 2010), which aims for deriving additional qualitative and quantitive data that is beneficial for improving the patient's treatment. Typically, in a clinical environment, there are two categories of post-processing workflows.

**Distributed application (agent) oriented workflows:** In the case of distributed agent oriented workflows, various information systems are controlled under the authority of different health care actors such as physicians, general practitioners and various hospital departments. Currently, distributed agent oriented post-processing workflow steps involves applications, such as Computer Aided Detection (CAD), Image processing, 3D reconstruction and surface rendering are available. However, these workflows are often distributed, and the various applications participating in the execution of the workflow are deployed on a stand-alone hosting system. Various endeavors address the handling of provenance and workflows in these distributed agent oriented workflows (Kifor et al., 2006) (Zhang et al., 2009).

**Custom data processing workflows:** In the case of custom data processing workflows the various data processing steps (algorithms) are typically defined and implemented for advanced data processing in the clinical environment. Currently, to coordinate and trace the execution of the workflow steps, the worklist embedded in the DICOM file is referred. The DICOM worklist (to-do lists) hold the list of tasks that are to be performed on DICOM datasets. However, DICOM does not have the capability to model and maintain the provenance traces associated with the data. Even though the worklist offers a convenient technique for maintaining the workflow steps, it obviously lacks an accepted workflow and provenance standard. For modelling comprehensive provenance information, i.e. both prospective provenance (workflow execution plan) and retrospective provenance (runtime events) (Zhao et al., 2006), the W3C ProvONE standard is adopted. Furthermore, to enable reuse of these data processing algorithms, they are registered as web-services in the Data Analysis and Curation sub-module of the IPDR. The intermediate data generated after each step in the workflow is ingested in the IPDR PACS server, thus preventing unnecessary repetition of the workflow.

Using an existing angioscopy image analysis workflow (Maksimov et al., 2009) (Brockmann et al., 2010), the various processing steps are deployed as

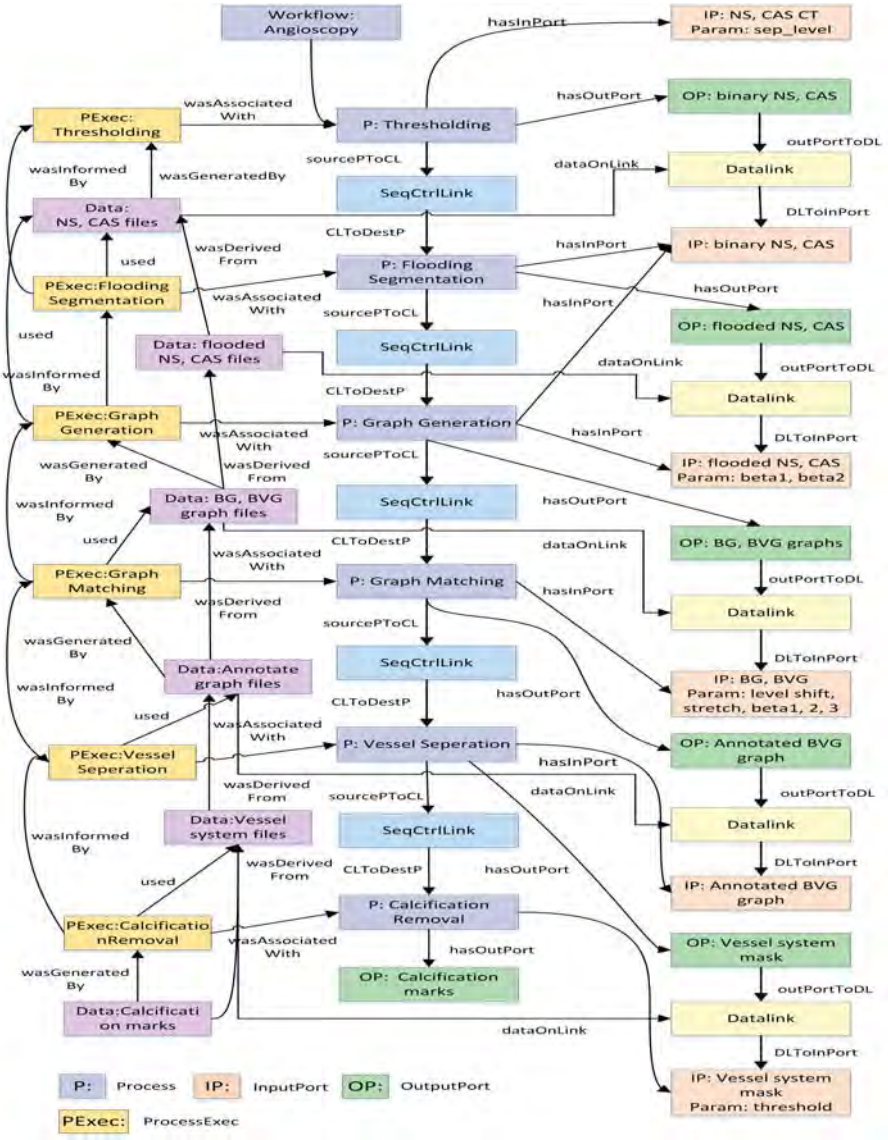



Figure 3: Angioscopy workflow provenance in ProvONE.

independent web services that are accessible through a REST URI. Based on these REST URIs, the complete workflow is modelled in BPEL specification and executed on the high-performance computing cluster. Before the execution of the workflow, the ProvONE prospective graph is generated by the Prov2ONE algorithm, and during the execution enriched with the runtime provenance information (retrospective provenance). The complete ProvONE provenance graph for the angioscopy workflow is shown in Figure 3.

During the execution of the image analysis workflow the intermediate result are ingested into the IPDR and the metadata is extracted, modelled and stored in the metadata storage, as shown in the Figure 2.

## 5 DISCUSSION

In the available systems, clinical process data is stored in diverse clinical information systems, for example in HIS including subsystems such as a RIS and PACS. The essence of these systems is not only the reporting but also to be able to retrospectively follow the course of the clinical decision and treatment. Whenever image data is manipulated such as via image processing or conclusions are driven via machine learning techniques, there is no adequate tool available yet that allows tracking these steps in an adequate way. The rationale for considering the IPDR is its high flexibility for different clinical processes and their adequate

documentation.

Furthermore, the benefit of being able to track previous diagnostic and treatment processes is to be able to assess the treatment outcome or the correctness of the diagnostic step. This allows both to increase the internal quality of clinical routines but also offers the chance to get a deeper insight when e.g. for cancer patients there is a relapse, which might offer different ways of further patient treatment.

However, to handle the frequent metadata changes in the DICOM dataset that are introduced during the clinical studies, we integrated a database system supporting multiple flexible data models. Additionally, the metadata is automatically indexed for enabling free-text search through the data discovery service. For capturing the dynamic information in the form of annotations, the W3C standard Open Annotation Data Model is implemented through the annotation service. The IPDR allows the entire modelling of the DICOM metadata in METS format and the retrospective provenance is modelled in PREMIS standard, thus allowing efficient sharing (harvesting) of the metadata using the OAI-PMH services. For handling the provenance of the image processing workflows, we integrated the ProvONE standard in IPDR, wherein both the workflow plan as well as the associated provenance are captured, thus, enabling reproducibility of scientific results.

We also presented the data and metadata workflow describing the integration of IPDR with the PACS server, see Figure 2. Our aim was to seamlessly integrate the entire IPDR into the existing PACS server and network infrastructure without any disruption in the existing execution environment. The DICOM dataset either from the data acquisition system (imaging modalities) or from the various workstations in the PACS network is processed through IPDR and forwarded to the PACS server. To handle the DICOM data the open-source DICOM DCM4CHE (Zeilinger et al., 2010) toolkit is integrated into the IPDR.

Regarding the regulations, documents concerning diagnostic and therapeutic procedures have to be filed for about 10-30 years, documents for quality control have to be archived for about 5 years depending on country and regulations. There is no formal requirement for a detailed documentation of medical processes so far but the availability of a flexible provenance software technique would foster regulation procedures if publicly available.

# 6 RELATED WORK

Currently, there are various commercial as well as custom implementations of PACS solutions available for clinical studies for handling the medical datasets.

Enterprise Imaging Repository (EIR) (Bian et al., 2009) is an alternative to the commercial PACS server and network solution that provides handling of DICOM files using the DCM4CHE toolkit. Various functionality like HL7(Dolin et al., 2006) interfaces, web access to DICOM objects and media creation services (CDW) are implemented by the EIR.

For handling the cardiology datasets (Marcheschi et al., 2009) have a PACS solution completely built using open-source technology. This solution is based on existing commodity hardware with a server farm comprising a storage pool of 24 TB for saving the cardiology datasets. A PACS based solution with automated workflow management techniques using the YAWL specification in radiology information system (Zhang et al., 2009) are implemented for allowing a workflow-aware and flexible integration of various components in building RIS.

The integration of a workflow system does offer novel approach for a flexible RIS design. However, the above mentioned PACS solutions fail to provide many of the critical aspects associated with the handling of the complete lifecycle of the medical datasets. The IPDR solution presented in this paper overcomes these limitations by providing functionalities such as, metadata handling based on standard metadata models, dedicated scalable and flexible metadata storage, annotation services based on standard Open Annotation Data Model, integration with high-performance computing cluster for performant execution of post-processing workflow tasks and capturing of provenance in ProvONE model.

# 7 CONCLUSION

In this paper, we presented the detailed architecture and the functionalities provided by the IPDR, which is a comprehensive repository system for handling the EHR datasets. The IPDR is based on the principle of modular architecture that enables easy extensibility by adding task-specific modules for handling any new requirements. The functionality provided by IPDR enables the RT researchers to: (a) automatically extract, store, and access metadata in standard metadata models, (b) allow reproducibility of scientific results by capturing provenance in ProvONE standard, (c) enrich the data quality by capturing annotations in the Open Annotation Data Model, (d) enable the repeata-

bility of complex image processing workflows using integrated workflow engine, (e) large-scale metadata harvesting through standard OAI-PMH protocol.

To demonstrate the handling of image processing workflow with capturing of its associated provenance, the angioscopy workflow modelled in BPEL was executed using a workflow engine integrated with the IPDR, and the associated provenance was automatically modelled in ProvONE (Figure 3).

However, the data exchange approach uses safe data transfer strategies but are not yet adapted to clinical needs. An integration of i2b2 and a data warehouse that collects all clinical data in combination with technologies such as PCORnet[7] offers the potential to standardised data exchange for studies and consultation and will thereby allow using our architecture to be integrated into a clinic-wide information system of the next generation. In particular, this strategy is considered to be integrated into data integration centres that are currently planned at the university medical centre Mannheim (UMM) and the MIRACUM[8] consortium that is currently funded by the German Ministry of Research and Education (BMBF). The technology is built upon developments of the BMBF-project LSDMA[9].

## REFERENCES


Allcock, W., Bresnahan, J., Kettimuthu, R., et al. (2005). The globus striped gridftp framework and server. In *Proceedings of the 2005 ACM/IEEE conference on Supercomputing*, page 54. IEEE Computer Society.

Bian, J., Topaloglu, U., and Lane, C. (2009). Eir: Enterprise imaging repository, an alternative imaging archiving and communication system. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2168–2171.

Brockmann, C., Jochum, S., Hesser, J., et al. (2010). Graph-matching-based computed tomography angiography in peripheral arterial occlusive disease. *Clinical Imaging*, 34(5):367 – 374.

Davidson, S. B., Boulakia, S. C., Eyal, A., Ludäscher, B., McPhillips, T. M., Bowers, S., Anand, M. K., and Freire, J. (2007). Provenance in scientific workflow systems. *IEEE Data Eng. Bull.*, 30(4):44–50.

Dolin, R. H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F. M., Biron, P. V., and Shabo (Shvo), A. (2006). Hl7 clinical document architecture, release 2. *Journal of the American Medical Informatics Association*, 13(1):30–39.

Estrella, F., Hauer, T., McClatchey, R., Odeh, M., Rogulin, D., and Solomonides, T. (2007). Experiences of engineering grid-based medical software. *International Journal of Medical Informatics*, 76(8):621 – 632.

Goland, Y., Whitehead, E., Faizi, A., Carter, S., and Jensen, D. (1999). Http extensions for distributed authoring–webdav. Technical report.

Kifor, T., Varga, L. Z., Vazquez-Salceda, J., Alvarez, S., Willmott, S., Miles, S., and Moreau, L. (2006). Provenance in agent-mediated healthcare systems. *IEEE Intelligent Systems*, 21(6):38–46.

Liu, Y. and Wang, J. (2010). *PACS and digital medicine: essential principles and modern practice*. CRC Press.

Maksimov, D., Hesser, J., Brockmann, C., Jochum, S., Dietz, T., et al. (2009). Graph-matching based cta. *IEEE Transactions on Medical Imaging*, 28(12):1940–1954.

Marcheschi, P., Ciregia, A., Mazzarisi, A., Augiero, G., and Gori, A. (2009). A new approach to affordable and reliable cardiology pacs architecture using open-source technology. In *2009 36th Annual Computers in Cardiology Conference (CinC)*, pages 537–540.

Matthews, B. (2015). Csmd: the core scientific metadata model. Online http://icatproject-contrib.github.io/CSMD/csmd-4.0.html.

Mildenberger, P., Eichelberg, M., and Martin, E. (2002). Introduction to the dicom standard. *European Radiology*, 12(4):920–927.

OASIS (2007). Standard, O.A.S.I.S: Web services business process execution language version 2.0. Online http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.

Prabhune, A., Stotzka, R., Jejkal, T., Hartmann, V., Bach, M., Schmitt, E., Hausmann, M., and Hesser, J. (2015). An optimized generic client service api for managing large datasets within a data repository. In *Big Data Computing Service and Applications (BigDataService), 2015 IEEE First International Conference on*, pages 44–51.

Prabhune, A., Zweig, A., Stotzka, R., Gertz, M., and Hesser, J. (2016). *Prov2ONE: An Algorithm for Automatically Constructing ProvONE Provenance Graphs*, pages 204–208. Springer Publishing.

Sanderson, R., Ciccarese, P., Van de Sompel, H., Bradshaw, S., Brickley, D., a Castro, L. J. G., et al. (2013). Open annotation data model. *W3C community draft*.

Sandoval, J. (2009). *Restful java web services: Master core rest concepts and create restful web services in java.* Packt Publishing Ltd.

Zeilinger, G., Montgomery, O., Evans, D., et al. (2010). The dcm4che project. Online Sourceforge project https://sourceforge.net/projects/dcm4che/.

Zhang, J., Lu, X., Nie, H., Huang, Z., and van der Aalst, W. M. P. (2009). Radiology information system: a workflow-based approach. *International Journal of Computer Assisted Radiology and Surgery*, 4(5):509–516.

Zhao, Y., Wilde, M., and Foster, I. (2006). *Applying the Virtual Data Provenance Model*, pages 148–161. Springer Berlin Heidelberg, Berlin, Heidelberg.

Zhu, N. Q. (2013). *Data visualization with D3. js cookbook*. Packt Publishing Ltd.


[7] www.pcornet.org

[8] www.miracum.de

[9] www.helmholtz-lsdma.de

# Internet of Things Controlled Home Objects for the Elderly

Research Dawadi, Zeeshan Asghar and Petri Pulli
*Department of Information Processing Science, University of Oulu, Oulu, Finland*
*research.dawadi@student.oulu.fi, {zeeshan.asghar, petri.pulli}@oulu.fi*

Keywords: Systematic Literature Review, Internet of Things, Visualization, Elderly Assistance, Caretaker, Remote Assistance, Assisted Living.

Abstract: The number of elderly people suffering from physical or cognitive difficulty is increasing continuously. Elderly people prefer to live in their familiar environment where they can easily perform different activities of their daily life which is also good for their mental and physical well-being. Internet of Things is a mechanism through which any objects can be monitored, controlled, and manipulated. In order to develop efficient application for the elderly living at home independently, the researcher should be aware of the home objects as well as of the living environment. This study uses systematic literature review to determine applications developed to assist elderly people inside their home. A total of 25 primary studies are identified. With the analysis of those studies, important and relevant objects in the daily life of the elderly are identified. Using the results from the review, a new scenario of home environment is visualized. The visualization is expected to provide caretakers with a better view of the living condition of the elderly and position and state of the home objects. This new home scenario is expected to offer a secure and easy living environment for the elderly, where Internet of Things can be used to control all the frequently used home objects by the elderly.

## 1 INTRODUCTION

Internet of Things (IOT) is a developing phenomenon in the field of technological advancement as well as in research domain (Atzori et al., 2010; Xu et al., 2013). The main principle of IOT is continuous monitor and control of everyday objects or things over the internet. The thing in IOT has no limitation and can be any objects that are used in day to day life (Bassi and Horn, 2008). The thing can be any living or non-living object; from electronic devices to foods, clothes and furniture we use in our daily life, from animals like cow, dog, cats, and rats, to plants and trees (Madakam et al., 2015). An object or thing has its own unique identity, which allows communication between not only humans, but also between objects (Madakam et al., 2015). With availability of continuous internet connection, all the connected objects can be monitored regularly through an interconnected network.

Over 7% of the world population are over 65 years of age and the number is expected to increase 20% by the year 2050 (Morris et al., 2013). Also, by 2035, the number of people affected by dementia is expected to double. Hence there is requirement of policies and resources to meet the need of the aging population as well as of those suffering from dementia. Elderly can live longer and safer if they stayed in their own home environment (Bassi and Horn, 2008). Their friends and family also feel secure to have them in their own house in comparison to hospitals and care homes (Arcelus et al., 2007). Hospitals and healthcare centres may not be able to provide services to all who require (Arcelus et al., 2007). Smart homes can provide automation of domestic tasks, easier communication, higher security, and are adaptive to modern human needs as well as social needs (Morris et al., 2013; Lê et al., 2012).

The aim of technological advancement for elderly is to provide them a sense of independence even if they are physically or cognitively incapable. Avoiding diseases and encouraging healthy living is also a prime concern (Tran, 2002). Everyday activities that are easy to accomplish can be problematic as people get older. Simple tasks like brushing teeth, turning the tap off, switching TV off or on, and taking medicines can get tougher with age (Morris et al., 2013). IOT is a step forward in helping elderly complete these activities without physical presence of a caretaker, which also helps to minimize their expenses (Arcelus et al., 2007). IOT will also help one caretaker to monitor and assist multiple elderly patients from a remote location. With IOT, caretakers have the freedom to



Dawadi, R., Asghar, Z. and Pulli, P.
Internet of Things Controlled Home Objects for the Elderly.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 244-251
ISBN: 978-989-758-213-4

manipulate the home objects themselves and control them.

Above mentioned concerns motivated us to conduct a study to find different IOT based applications targeted for the elderly people living alone. Moreover, the goal of the study is to determine different kinds of home objects dealt in those applications. Figure 1 shows the overall idea of the study. This study is beneficial for those researchers and developers who are using IOT for the care and well-being of the elderly. The paper is structured as follows: section 2 presents an overview of previous studies. Section 3 explains the research method carried out for the study, and section 4 presents the findings of the study, limitation of the study, and scope for future research. Section 5 concludes the paper highlighting the main contribution of our work.

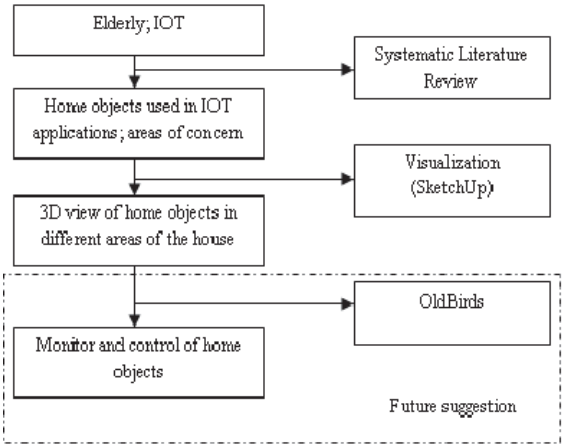



Figure 1: Overview of the study.

# 2 RELATED STUDIES

This section discusses about IOT and its all-round effectiveness. Furthermore, it accentuates some studies that highlights importance of IOT based applications for the elderly.

## 2.1 Internet of Things

IOT shows us a world where anything can be connected to any other thing. With embodiment of sensors and actuators in various physical objects, ranging from roadways to pacemakers, an object can both sense the environment and communicate. Information about the objects can be retrieved faster with IOT and people can make wise decisions quickly (Bassi and Horn, 2008). IOT enables inter-connectivity of different objects through a global infrastructure, internet (Madakam et al., 2015). IOT have been used successfully in various fields such as assisted living, e-health, domotics, enhanced learning, surveillance, environment monitoring, health monitoring, critical infrastructure monitoring, etc. (Atzori et al., 2010; Gubbi et al., 2013).

The main strength behind IOT is that using interconnected devices, it can have high impact on several aspects of our everyday life (Atzori et al., 2010). Considering the increasing number of research on IOT in recent years (Stankovic, 2014), it can be expected that IOT will play a greater role in our daily lives in upcoming years. IOT can help in digitalization of day to day activities, providing security in buildings, and decreasing the energy consumption of devices. IOT is not an independent system, but it can be regarded as a critical, integrated infrastructure by the use of which many applications and services can operate (Stankovic, 2014).

## 2.2 IOT and Elderly

As people grow older, their social circle and frequency of communication decreases on a regular basis. It can be caused by loss of family members or friends, diminishing eyesight or hearing ability, cognitive impairment, memory loss, etc. (Touhy and Jett, 2013). Moreover, their relatives and family members live far from them. It leads to a sense of isolation that may impact mental as well as physical health (Vardoulakis et al., 2012; Williams et al., 2014). The loss of physical as well as cognitive abilities affects their day to day activities and they require assistance in completing their daily tasks.

The introduction of IOT in the life of elderly people can help monitoring of chronic illness, on demand provision of fresh food, sending alarms and reminders, and enabling communication with family, friends or health care professionals (Dohr et al., 2010) as depicted in figure 2. Reminders of stove or iron left on, alerts at front door about visitors or intruders at home, etc. can be helpful for physically and cognitively disabled elderly people (Lê et al., 2012), but with IOT this service can be improved since required action can be taken by the caretaker remotely.

A caretaker can assist an elderly in completing tasks inside as well as outside their home environment for instance, guiding the elderly citizens to directions where they need to go (Firouzian et al., 2015). The concerns about privacy and security arise when we consider helping elderly in their home environment. Video monitoring is often regarded as an invasion to privacy by the elderly (Arcelus et al., 2007; Beunk,

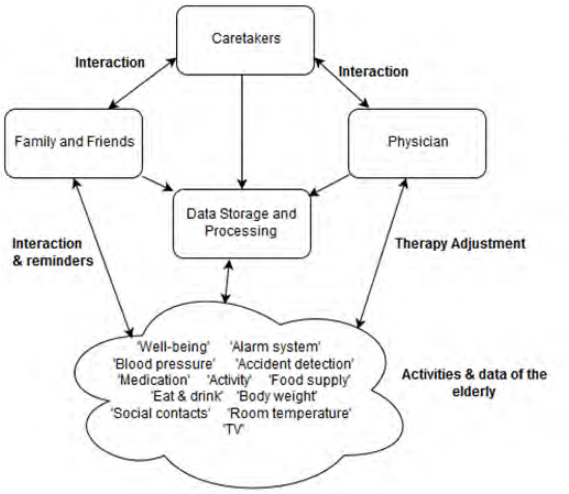



Figure 2: IOT and elderly, based on (Gubbi et al., 2013).

2015). To ease this concern, Old Birds, a web application prototype was implemented (Korvala and Raappana, 2015). It is run by a game engine that was originally built for remote care giving (Pulli et al., 2012). The elderly and their surrounding environment are presented in gaming avatars, which helps to minimize the concerns of privacy, as shown in figure 3.

Figure 3: Old Birds simulation environment, (Firouzian and Nissinen, 2013).

Currently, one of the main challenges for IOT system designers is to design a system that can help elderly in everyday activities as well as be supported by health care workers (Bassi and Horn, 2008). This rapid shift of the need of medical services for in-house care (Lee et al., 2013) provides the developers of IOT based applications with a new and productive area to deal with. Daily monitoring is enhanced by IOT thus increasing the quality of life and health of elderly people (Tran, 2002; Yang et al., 2014).

# 3 RESEARCH METHOD

This section explains the research method utilized in this study. Systematic Literature Review (SLR) was selected as a research method, which is a means of identifying, evaluating, and interpreting all research relevant to the particular area of interest or research question (Kitchenham, 2004). SLR can be used to verify or contradict any research hypothesis and can also lead to new research activities (Kitchenham, 2004; Biolchini et al., 2005).

The steps followed in the SLR were based on the guidelines for conducting a SLR (Kitchenham, 2004; Kitchenham and Charters, 2007). SLR, in general, is more time consuming than other traditional reviews and provides information about the effects of some phenomenon across a wide range of settings and empirical methods (Kitchenham, 2004). The first step in SLR is the formation of the research protocol. After a research area or a topic is finalized, it is necessary to create a review protocol that consists of research questions, inclusion and exclusion criteria, quality assessment criteria, keywords used to search literature, steps for reviewing the literature, and designated synthesis of the findings (Kitchenham, 2004; Biolchini et al., 2005).

The following two research questions were selected for the study:

- RQ1: What are the different areas inside the house where IOT can be applied for the elderly?
- RQ2: What kind of home objects necessary for the elderly are controlled by IOT based applications?

The inclusion and exclusion criteria used for the review are presented in table 1. Along with those, following quality assessment criteria were created to select the primary studies that were as relevant to the research questions as possible:

1. Is the application discussed in paper targeted to old people living alone?
2. How much is IOT discussed as part of everyday life for the user?
3. Was the application an experiment, controlled observation study, observation study without control groups or just theoretical?
4. Does the study clarify the problems faced during implementation?
5. Does the study provide details of the objects controlled by the applications?
6. Does the study is concentrated in daily activities inside the home or outside.The labels consists of sequential numbers?

Table 1: Inclusion and Exclusion criteria.

| Inclusion Criteria | Exclusion Criteria |
|---|---|
| • it covers the areas of both IOT and elderly citizens<br>• at least one home 'thing' is controlled remotely<br>• focus on elderly or elderly people with dementia<br>• it is a journal (peer reviewed), conference article, chapters from a book or a doctoral dissertation | • it is about IOT focused on all age group<br>• it is not in English<br>• it is a presentation<br>• it is a technical report<br>• it presents an incomplete project<br>• it is theoretical information only<br>• it covers IOT but not elderly assistance<br>• it covers elderly assistance but not IOT |

The use of inclusion and exclusion criteria along with the quality assessment criteria was completed following the steps shown in figure 4.

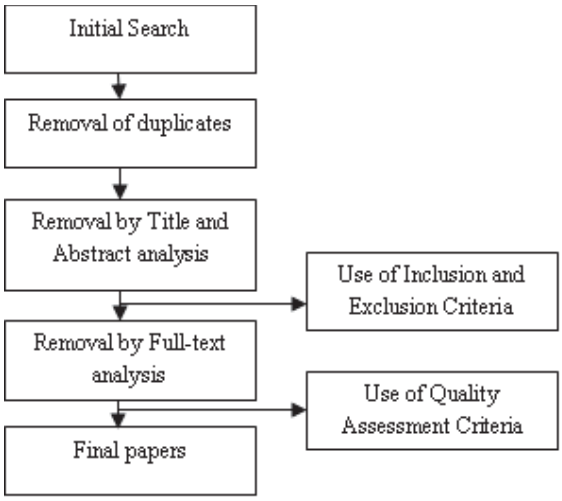



Figure 4: Steps to select primary studies.

# 4 RESULTS

This section discusses the findings of the SLR and relates the data to the research questions. Limitations of the study and scope for extending the research are also pointed out.

## 4.1 Findings

A total of 229 papers were identified in five different databases; Scopus, IEEE Xplore, Science Direct, Web of Science, and ProQuest. Following the steps presented in figure 4, a total of 25 papers were identified as the final selection of primary studies, which makes around 11% of the initial collection of papers. Among the 25 studies, 15 were conference proceedings paper, 8 were journal articles and 2 were monographs. Each of the studies explained applications that were developed to assist the elderly, make their life easier,improve their health, and take care of them in emergencies.

It was observed that most of the applications focused on the health aspect of the elderly.Body sensors were the most implemented sensors with applications in 11 of the 25 papers utilising it. Ambient sensors were second, with fall detection a common area of concern as well. Applications also dealt with providing proper medication to the elderly and maintaining a healthy lifestyle like guiding them in exercising. Table 2 lists the areas of the focus for the study and the number of paper that dealt with that area.

Table 2: Areas of Concern.

| Areas of Concern | Number of Papers |
|---|---|
| Kitchen | 7 |
| Health Care | 22 |
| Dining room | 1 |
| Living room | 6 |
| Bedroom | 4 |
| Bathroom | 3 |
| Shopping | 1 |
| Social connection | 2 |
| Security | 4 |
| Training and gatherings | 1 |
| Job Search | 1 |
| Exploration | 1 |
| Study room | 1 |

Table 3 lists all the home objects that were identified in the applications dealt in the primary studies. There were some studies that did not clearly indicate the objects controlled by the application, as they were denoted as smart objects or as devices in the house, and have been excluded.

## 4.2 Visualization

Most of the time, caretakers may not know the environment of the elderly well enough (Ikeda et al., 2011). With visualization of IOT objects, a caretaker

Table 3: Home objects used in IOT applications.

| Objects | Corresponding Paper |
|---|---|
| Kettle | (Brereton et al., 2015) |
| Mobile phone | (Dohr et al., 2010; Fortino et al., 2015; Gomes et al., 2015; Tang et al., 2015; Laranjo et al., 2013; Panicker and Kumar, 2015) |
| Blood Pressure meter | (Dohr et al., 2010) |
| Body sensors | (Fortino et al., 2015; Gomes et al., 2015; Guo and Bai, 2014; Dagale et al., 2015; Liang, 2016; Panicker and Kumar, 2015; Savola et al., 2015; Zgheib et al., 2015; Raad et al., 2015; Sung and Chang, 2014; Chen et al., 2015) |
| Ambient sensors | (Gomes et al., 2015; Luo et al., 2012; Boric-Lubecke et al., 2014; Wu et al., 2013; Savola et al., 2015; Zgheib et al., 2015; Rusu et al., 2015) |
| Gas | (Gomes et al., 2015; Lee, 2015; Cunha and Fuks, 2015) |
| Lights | (Gomes et al., 2015; Tang et al., 2015; Cunha and Fuks, 2015; Neßelrath et al., 2011; Zgheib et al., 2015; Rusu et al., 2015) |
| Exhaust (Smoke) | (Gomes et al., 2015; Lee, 2015) |
| TV | (Konstantinidis et al., 2015; Lee, 2015; Wu et al., 2013; Neßelrath et al., 2011) |
| Medicine bottle/drawer | (Laranjo et al., 2013; Cunha and Fuks, 2015; Neßelrath et al., 2011; Sohn et al., 2015) |
| Tablets | (Laranjo et al., 2013) |
| Camera | (Lee, 2015; Rusu et al., 2015; Sohn et al., 2015) |
| Food drawer | (Lee, 2015) |
| Chair/ Sofa | (Liang, 2016; Wu et al., 2013; Zgheib et al., 2015) |
| Bed | (Liang, 2016; López-de Ipiña et al., 2010; Wu et al., 2013; Zgheib et al., 2015) |
| Temperature sensor | (Cunha and Fuks, 2015; Rusu et al., 2015; Sung and Chang, 2014; Chen et al., 2015) |
| Humidity sensor | (Cunha and Fuks, 2015; Chen et al., 2015) |
| Infrared sensors | (Cunha and Fuks, 2015) |
| Noise sensors | (Cunha and Fuks, 2015) |
| Shoes | (Wu et al., 2013) |
| Door | (Wu et al., 2013; Savola et al., 2015; Rusu et al., 2015; Chen et al., 2015) |
| Oven/Microwave | (Wu et al., 2013; Neßelrath et al., 2011) |
| Electricity meter | (Tang et al., 2015; Wu et al., 2013) |
| Kitchen Surface | (Neßelrath et al., 2011) |
| Fridge | (Neßelrath et al., 2011) |
| Dishwasher | (Neßelrath et al., 2011) |
| AC | (Neßelrath et al., 2011) |
| Switches | (Zgheib et al., 2015) |
| Window | (Rusu et al., 2015) |
| PC | (Sohn et al., 2015) |
| Weight scale | (Sung and Chang, 2014) |

can know the status of the objects continuously and determine the changes in the state of the object due to any activity of the elderly. With such visualization, presence of camera is also redundant which will encourage old people to accept the technology as they regard technology, mainly video surveillance, as an intrusion to their privacy (Arcelus et al., 2007).

Visualization is a method to infer new knowledge from collected information to get a comprehensive view of the space of interest (Gershon and Eick, 1997). The home objects from table 3 were thus utilized to vision a scenario in the day of an elderly person. This visualization is intended to showcase an elderly person in different areas of the house and show different IOT enabled home objects that they have to deal with for their daily activities. The areas of the house where elderly require assistance from the caretakers were selected as: bedroom, living room, kitchen, dining room, and bathroom.

The visualization will help the caretaker to deter-


Figure 5: Visualization of different areas of the house.

mine the location of the elderly in the house and observe their actions. The caretaker can monitor the intake of medicine by analyzing weight of the medicine pill bottle, and provide notifications via smart phone or TV to take the medicine when necessary. Also, ambient sensors are enabled across all the rooms of the house to determine sudden emergencies like falling down. Wearable sensors will help to continuously monitor the state of health of the elderly and take action in case of sudden bodily changes. The home objects that are controlled by the above mentioned scenario are listed in table 4.

Table 4: Home objects in the visualized scenario.

| | | |
|---|---|---|
| Bed | Lamp | Mobile |
| Pill Bottle | Door/Windows | Chair |
| Stove | Dishwasher | Sink |
| Kettle | Fridge | TV |
| Food Cabinet | Table | Sofa |

### 4.3 Limitation and Future Scope

The main focus of the study is limited within the perimeters of the house of the elderly. The research can be extended by including the applications that are designed to assist elderly people outside their home. Since areas for use of IOT is quite large, future research can be done by including applications that have been designed not just for the elderly but for people of all age group. This will help to determine various other areas of concerns and home objects.

The data from SLR indicates that application developers consider health care as the major area of concern. For future applications, not just health care, but other aspects such as security can be integrated into the existing system, as shown in the visualization section. Also, there is a need of an IOT application that helps not only assisting when necessary but also constantly monitoring different parameters of their health. For e.g. a caretaker can monitor stress level or heart beat rate of an elderly while giving them instructions on how to cook a meal. The data from SLR and our visualization ideas both can be utilized to create a simulation of the home environment of the elderly.

The inclusion of scientific studies only in the review process leaves out many newspaper articles, electronic sources and other archives. It can be possible that all the IOT enabled applications designed for the elderly might not have been discussed in scientific literature. Future researchers are encouraged to include other sources as well, to broaden the range of application search.

## 5 CONCLUSION

The main contribution of this study is to provide the information about existing applications for the elderly in the context of their areas of concern and home objects controlled by them. The aim is also to visualize a new scenario of IOT enabled home environment for the elderly. Existing research have either focused on IOT, or some particular areas like health care or socializing. The visualization scenario integrates different areas of concern for the elderly and provide some data for the researchers as well as developers.

This study dealt with in-house environment only. Since elderly people spend most of their time inside their house, it is necessary to know about the objects they use regularly in order to create an autonomous environment for them. The data from SLR can be further utilized to develop new applications to improve quality of life and health of the elderly inside their home environment. These applications can help to assist elderly in their daily activities through technological assistance and also monitor them regularly.

The study also presents a 3D visualization of a safe and secure living environment across different areas of the house including the home objects. This can be utilized by application developers as a baseline to develop IOT based applications for the elderly. The visualized scenario is helpful for caretakers in monitoring and assisting the elderly.

# REFERENCES


Arcelus, A., Jones, M. H., Goubran, R., and Knoefel, F. (2007). Integration of smart home technologies in a health monitoring system for the elderly. In *Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st International Conference on*, volume 2, pages 820–825. IEEE.

Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15):2787–2805.

Bassi, A. and Horn, G. (2008). Internet of things in 2020: A roadmap for the future. *European Commission: Information Society and Media*.

Beunk, N. (2015). Visualize my data! : translating smart home sensor data into relevant feedback for elderly, informal caregivers and formal caregivers.

Biolchini, J., Mian, P. G., Natali, A. C. C., and Travassos, G. H. (2005). Systematic review in software engineering. *System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES*, 679(05):45.

Boric-Lubecke, O., Gao, X., Yavari, E., Baboli, M., Singh, A., and Lubecke, V. M. (2014). E-healthcare: Remote monitoring, privacy, and security. In *2014 IEEE MTT-S International Microwave Symposium (IMS2014)*, pages 1–3. IEEE.

Brereton, M., Soro, A., Vaisutis, K., and Roe, P. (2015). The messaging kettle: Prototyping connection over a distance between adult children and older parents. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 713–716. ACM.

Chen, Y.-H., Tsai, M.-J., Fu, L.-C., Chen, C.-H., Wu, C.-L., and Zeng, Y.-C. (2015). Monitoring elder's living activity using ambient and body sensor network in smart home. In *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*, pages 2962–2967. IEEE.

Cunha, M. and Fuks, H. (2015). Ambleds collaborative healthcare for aal systems. In *Computer Supported Cooperative Work in Design (CSCWD), 2015 IEEE 19th International Conference on*, pages 626–631. IEEE.

Dagale, H., Anand, S., Hegde, M., Purohit, N., Supreeth, M., Gill, G. S., Ramya, V., Shastry, A., Narasimman, S., Lohith, Y., et al. (2015). Cyphys+: A reliable and managed cyber-physical system for old-age home healthcare over a 6lowpan using wearable motes. In *Services Computing (SCC), 2015 IEEE International Conference on*, pages 309–316. IEEE.

Dohr, A., Modre-Osprian, R., Drobics, M., Hayn, D., and Schreier, G. (2010). The internet of things for ambient assisted living. *ITNG*, 10:804–809.

Firouzian, A., Asghar, Z., Tervonen, J., Pulli, P., and Yamamoto, G. (2015). Conceptual design and implementation of indicator-based smart glasses: A navigational device for remote assistance of senior citizens suffering from memory loss. In *2015 9th International Symposium on Medical Information and Communication Technology (ISMICT)*, pages 153–156. IEEE.

Firouzian, A. and Nissinen, H. (2013). A monitoring tool to support remote caretaking of senior citizens. In *14th International Conference on Intelligent Games and Simulation, GAME-ON 2013*.

Fortino, G., Giordano, A., Guerrieri, A., Spezzano, G., and Vinci, A. (2015). A data analytics schema for activity recognition in smart home environments. In *International Conference on Ubiquitous Computing and Ambient Intelligence*, pages 91–102. Springer.

Gershon, N. and Eick, S. G. (1997). Information visualization. *IEEE Computer Graphics and Applications*, 17(4):29–31.

Gomes, B., Muniz, L., e Silva, F. J. d. S., Ríos, L. E. T., and Endler, M. (2015). A comprehensive cloud-based iot software infrastructure for ambient assisted living. In *Cloud Technologies and Applications (CloudTech), 2015 International Conference on*, pages 1–8. IEEE.

Gubbi, J., Buyya, R., Marusic, S., and Palaniswami, M. (2013). Internet of things (iot): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645–1660.

Guo, Y. and Bai, G. (2014). An iot architecture for home-based elderly healthcare. In *International Conference on Management and Engineering (CME 2014)*, page 329. DEStech Publications, Inc.

Ikeda, S., Asghar, Z., Hyry, J., Pulli, P., Pitkanen, A., and Kato, H. (2011). Remote assistance using visual prompts for demented elderly in cooking. In *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, page 46. ACM.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26.

Kitchenham, B. and Charters, S. (2007). Procedures for performing systematic literature reviews in software engineering. *Keele University & Durham University, UK*.

Konstantinidis, E. I., Antoniou, P. E., Bamparopoulos, G., and Bamidis, P. D. (2015). A lightweight framework for transparent cross platform communication of controller data in ambient assisted living environments. *Information Sciences*, 300:124–139.

Korvala, T. and Raappana, H. (2015). Open visual guidance system for mobile senior citizen. Master's thesis, University of Oulu, Oulu, Finland, 2015.

Laranjo, I., Macedo, J., and Santos, A. (2013). Internet of things for medication control: E-health architecture and service implementation. *International Journal of Reliable and Quality E-Healthcare (IJRQEH)*, 2(3):1–15.

Lê, Q., Nguyen, H. B., and Barnett, T. (2012). Smart homes for older people: Positive aging in a digital world. *Future internet*, 4(2):607–617.

Lee, R. C. (2015). The new way of social connecting for the elderly through smart home applications. In *International Conference on HCI in Business*, pages 142–152. Springer.

Lee, S. I., Ghasemzadeh, H., Mortazavi, B., Lan, M., Alshurafa, N., Ong, M., and Sarrafzadeh, M. (2013). Re-

mote patient monitoring: what impact can data analytics have on cost? In *Proceedings of the 4th Conference on Wireless Health*, page 4. ACM.

Liang, P.-C. (2016). *Cost-effective design of real-time home healthcare telemonitoring based on mobile cloud computing.* PhD thesis, University of Surrey.

López-de Ipiña, D., Díaz-de Sarralde, I., and Zubía, J. G. (2010). An ambient assisted living platform integrating rfid data-on-tag care annotations and twitter. *J. UCS*, 16(12):1521–1538.

Luo, X., Liu, T., Liu, J., Guo, X., and Wang, G. (2012). Design and implementation of a distributed fall detection system based on wireless sensor networks. *EURASIP Journal on Wireless Communications and Networking*, 2012(1):1.

Madakam, S., Ramaswamy, R., and Tripathi, S. (2015). Internet of things (iot): A literature review. *Journal of Computer and Communications*, 3(05):164.

Morris, M. E., Adair, B., Miller, K., Ozanne, E., Hansen, R., Pearce, A. J., Santamaria, N., Viega, L., Long, M., and Said, C. M. (2013). Smart-home technologies to assist older people to live well at home. *Journal of aging science*, 1(1):1–9.

Neßelrath, R., Haupert, J., Frey, J., and Brandherm, B. (2011). Supporting persons with special needs in their daily life in a smart home. In *Intelligent Environments (IE), 2011 7th International Conference on*, pages 370–373. IEEE.

Panicker, N. V. and Kumar, S. (2015). Design of a telemonitoring system for detecting falls of the elderly. In *Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on*, pages 800–803. IEEE.

Pulli, P., Asghar, Z., Siitonen, M., Niskala, R., Leinonen, E., PitkPitkanen, A., Hyry, J., Lehtonen, J., Kramar, V., and Korhonen, M. (2012). Mobile augmented teleguidance-based safety navigation concept for senior citizens. In *2nd. International Conference on Applied and Theoretical Information Systems Research (2nd. ATISR2012)*, pages 1–9.

Raad, M. W., Sheltami, T., and Shakshuki, E. (2015). Ubiquitous tele-health system for elderly patients with alzheimer's. *Procedia Computer Science*, 52:685–689.

Rusu, L., Cramariuc, B., Benţa, D., and Mailat, M. (2015). Implementing bpmn 2.0 scenarios for aal@ home solution. *International Journal of Computers Communications & Control*, 10(2):230–237.

Savola, R. M., Savolainen, P., Evesti, A., Abie, H., and Sihvonen, M. (2015). Risk-driven security metrics development for an e-health iot application. In *Information Security for South Africa (ISSA), 2015*, pages 1–6. IEEE.

Sohn, S. Y., Bae, M., Lee, D.-k. R., and Kim, H. (2015). Alarm system for elder patients medication with iot-enabled pill bottle. In *Information and Communication Technology Convergence (ICTC), 2015 International Conference on*, pages 59–61. IEEE.

Stankovic, J. A. (2014). Research directions for the internet of things. *IEEE Internet of Things Journal*, 1(1):3–9.

Sung, W.-T. and Chang, K.-Y. (2014). Health parameter monitoring via a novel wireless system. *Applied Soft Computing*, 22:667–680.

Tang, L. Z. W., Ang, K. S., Amirul, M., Yusoff, M. B. M., Tng, C. K., Alyas, M. D. B. M., Lim, J. G., Kyaw, P. K., and Folianto, F. (2015). Augmented reality control home (arch) for disabled and elderlies. In *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2015 IEEE Tenth International Conference on*, pages 1–2. IEEE.

Touhy, T. A. and Jett, K. F. (2013). *Ebersole & Hess' Toward Healthy Aging: Human Needs and Nursing Response*. Elsevier Health Sciences.

Tran, B. Q. (2002). Home care technologies for promoting successful aging in elderly populations. In *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, volume 3, pages 1898–1899. IEEE.

Vardoulakis, L. P., Ring, L., Barry, B., Sidner, C. L., and Bickmore, T. (2012). Designing relational agents as long term social companions for older adults. In *International Conference on Intelligent Virtual Agents*, pages 289–302. Springer.

Williams, D., Ahamed, S. I., and Chu, W. (2014). Designing interpersonal communication software for the abilities of elderly users. In *Computer Software and Applications Conference Workshops (COMPSACW), 2014 IEEE 38th International*, pages 282–287. IEEE.

Wu, Q., Shen, Z., Leungy, C., Zhang, H., Cai, Y., Miao, C., et al. (2013). Internet of things based data driven storytelling for supporting social connections. In *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, pages 383–390. IEEE.

Xu, G., Ding, Y., Zhao, J., Hu, L., and Fu, X. (2013). Research on the internet of things (iot). *Sensors & Transducers*, 160(12):463.

Yang, L., Ge, Y., Li, W., Rao, W., and Shen, W. (2014). A home mobile healthcare system for wheelchair users. In *Computer Supported Cooperative Work in Design (CSCWD), Proceedings of the 2014 IEEE 18th International Conference on*, pages 609–614. IEEE.

Zgheib, R., Conchon, E., et al. (2015). A semantic web-of-things architecture for monitoring the risk of bedsores. In *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 318–323. IEEE.

# A Model Proposal for Augmented Reality Game Creation to Incentivize Physical Activity

José Antonio Cicció and Luis Quesada
*Escuela de Ciencias de la Computación e Informática, Universidad de Costa Rica, Montes de Oca, San José, 11501-2060, Costa Rica*
*{jose.ciccio, luis.quesada}@ecci.ucr.ac.cr*

Keywords: Gamification, Augmented Reality, Health, Fitness, Game Design, Mobile Applications.

Abstract: Obesity and a sedentary lifestyle are relevant issues in today's society. Even though different resources can be used to approach this problem, technology provides endless possibilities to fight against this problem. This article presents the results of a model to create augmented reality games where goals are achieved by doing physical activity (moving between different places). In order to evaluate the model, a prototype was built and presented to 50 participants. The results obtained indicated that an important percentage of the interviewees were attracted to the idea of playing a game to increase their physical activity.

## 1 INTRODUCTION

Nowadays more people are concerned about their physical fitness. In fact, activity trackers and some games to stand up and to move in a place are commonly used to keep people motivated; however, not all users feel the same about these long term commitment resources (Buchem, Merceron, Kreutel, Haesner, and Steinert, 2015).

Users, due to their human nature, tend to be motivated by challenges that come in a competitive, cooperative or solo form (Spillers and Asimakopoulos, 2014). For those that may require extra motivation, gamification is a possible way to incentivize them to do more exercise.

Related to the topic of using technology to keep people moving, the term of Augmented Reality (AR) is extremely useful. AR can be defined as "a technology that superimposes a computer-generated image on a user's view of the real world, thus providing a composite view" (Oxford Dictionary of English, 2010). This technique is becoming more and more popular among the users of new technologies like Google VR (Kipper and Rampolla, 2013).

The previous elements help developers to create appealing health video games that engage users while motivating them to do more exercise in a daily basis (González et al., 2016).

This fact led us to create a game model to encourage people to do exercise not only to be more active but also to continue with the physical work as time passes by.

By using AR, a GPS device, an accelerometer, a gyroscope and other elements, we propose a model where we identify the most important parts that a game should have in order to create a health-based video game. The latter will require users to move to specific points in a map by walking, running or climbing floors to get rewards and to advance in the video game story.

In sum, the main goal of this article is to create a model for the development of a generic video game that incentivizes users to be more active and to have a healthier lifestyle.

The following sections in this article include the related work in the area of game-based research, the purpose and the explanation of a generic model for AR health-based games, and the data related to the implementation of the model by creating a Role Playing Game (RPG), which may urge people to go to different places in order to level up and to get better gear to unlock game missions. Later, an evaluation of the proposed model will be conducted before discussing the final results.

## 2 RELATED WORK

Organizations, healthcare providers, and public initiatives promote healthy lifestyles through



Cicció, J. and Quesada, L.
A Model Proposal for Augmented Reality Game Creation to Incentivize Physical Activity.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 252-259
ISBN: 978-989-758-213-4

information technology. Alahäivälä and OinasKukkonen (2016) presented a systematic literature review that focused on "health gamification." The study discussed 32 studies conducted between 2011 and 2015. Nine of them were categorized as "increasing physical activity." Some of the studies in their literature review are described below.

Recent technological devices inspire new applications. By using Microsoft Kinect, Brauner et al. (2013) proposed a game to promote physical fitness that focuses on elderly people. An avatar is presented within a garden. The player, through the virtual avatar, can pick up carrots from the soil or fruits and vegetables from trees and the soil. An experiment was conducted with 70 users. Most of the participants (72%) stated that this game increased their motivation to exercise on a regular basis (Brauner et al., 2013).

Nintendo Wii Fit activity games were also evaluated. A group of beginners and non-beginners played with Wii Fit games in different settings. The former was positive about the games when combining fitness and fun. However, the latter was unsatisfied with the Wii Fit as a fitness tool (Reynolds, Sosik, and Cosley, 2013).

HealthyTogether is an Android application that uses an activity tracker (FitBit) that was developed in 2014. By following simple rules, the user can win badges and points (karma points).

The game is a step and floor counter. The application includes three settings: competition mode, collaborative mode, and hybrid mode. HealthyTogether was evaluated with 18 dyads. The study found significant step and floor increase in both cooperation and hybrid mode (up to 21.1% and 18.2%, respectively) but not in the completion mode (Y. Chen and Pu, 2014). Another study was carried out that considered solo mode in similar conditions. The researchers found an increase in the amount of exercise that was associated with gamification. Furthermore, they concluded that gamification affects software design (Giannakis, Chorianopoulos, and Jaccheri, 2013).

Spillers and Asimakopoulos (2014) discussed how extended gamification and social elements in the mobile and the fitness context can improve people's motivation. Indeed, a positive short term behavioural change was observed in their research project.

Zuckerman and Gal-Oz (2014) evaluated the effectiveness of virtual rewards. They argued that according to the self-determination theory (SDT) of human motivation, three innate psychological needs determine motivation: (1) competence, (2) autonomy, and (3) relatedness. Results showed that daily walking or running time while interacting with an application was significantly higher than walking time without any application interaction.

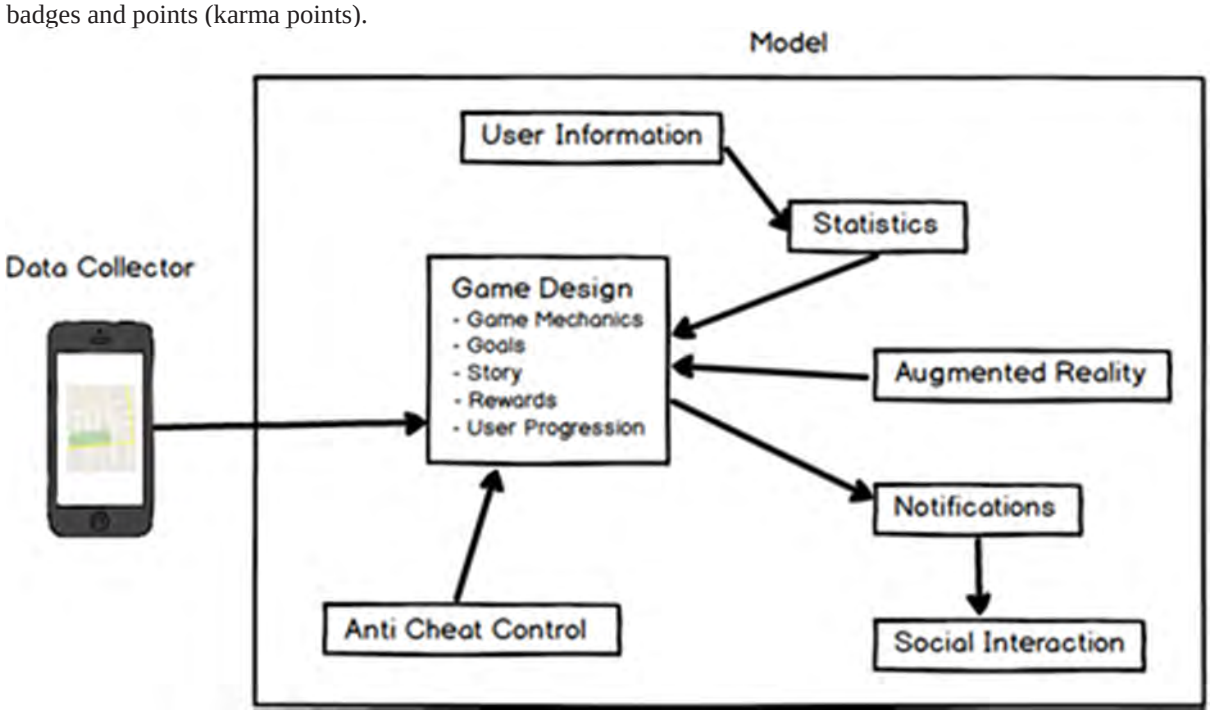



Figure 1: Video Game Model.

# 3 MODEL DESCRIPTION

This section describes the elements that all AR games that involve moving from one place to another should include. These elements are the result of years of developing different video games of diverse genres. Aside from the expertise acquired through the years, some of the key components explained in this article can be found in the most popular games available in the market (Novak, 2012).

The games that use this model should comply with the following requirements: a mobile device that has an active internet connection, a GPS, and a camera.

Figure 1 shows the proposed model with its components. These components include game design, statistics, user information, anti-cheat control, augmented reality, notifications, data collector, and social interaction. The following sections describe the previous concepts.

## 3.1 Game Design

Game design is an essential component that includes rules, gameplay, and storyline. All of them combined facilitate the interaction between players in case the game has health purposes (Novak, 2012).

### 3.1.1 Game Mechanics

There are several game mechanics that can be incorporated in order to incentivize people to do exercise. We consider that there are four main game mechanics—steps, walking, running, and floor climbing— that can be used in order let the player interact with the game. By using and mixing these four mechanics, different types of exercises can be covered (from casual training to more complex exercise routines).

Some actions can be followed in order to promote physical activity in the players (González et al., 2016). For instance, if it is assumed that users do not exercise regularly, they can start using the stairs instead of elevators to get to their destiny. Moreover, they can start walking short distances every day.

The game should be able to distinguish between walking and running. This can be achieved by using a combination of the GPS and the accelerometer of the mobile device. By using the correct type of movement and velocity, the game can provide users with different options in terms of the quests to be accomplished and the type of loot to be earned.

### 3.1.2 Goals

The game should have different goals that can be achieved as a result of the interaction between the user and the application (Novak, 2012). The game must have a list of customized daily goals for the player. The goals will be based on (1) how many steps have been made, (2) the distance, either by walking or running, that has been traveled, and (3) how many stairs or floors have been climbed.

The game should have an algorithm capable of creating customized goals based on the users' performance. The idea is to ask players for an extra effort every day they use the game. This idea can be achieved by presenting different places to visit in a single day or by including a far place to go.

When the goals depend on the users, the game provides a personalized experience and will get the highest possible retention ratio. If the user is falling behind, aspects such as distance and time should be easier for them to complete.

The game should spawn points on the map. Those points will represent the places the player should visit in order to make some kind of progress in the game. In addition to spawning points all over the map, the user can be encouraged to get to a specific place, within a time limit, by walking or running. Some goals can be based on the number of times a specific place is visited within a timeframe. Similarly, the player can achieve a goal by traveling the double or thrice of the standard goal.

### 3.1.3 Story

An essential part in today's games is their story. The story is the way people can get identified with the game and have an urge to keep playing. All games have a story which can be as simple or complex as the creators want (Dille and Zuur Platen, 2008). The idea is to have an easy-to-pick-up story to attract more people into the game. This should be aligned with the goals described in Section 3.1.2. In other words, the users have to be motivated during the game in order to keep moving through the different levels.

Another important aspect is to have a story that has a replay value. The story cannot be linear with a beginning and end because the main goal is to keep people interested in exercising from a mid to a long term. In order to do this, the story can have multiplayer elements that give an additional replay value to the players.

The two techniques that can be used with the players are to have them compete by themselves or to have them cooperate among each other to get a common goal (Novak, 2012). Aside from the

retention ratio, these techniques can help to attract more users into the game in less time because its existence can be easily spread (González et al., 2016).

### 3.1.4 Rewards

When a goal is achieved, the user should be rewarded by the effort made while playing the game. The reward can be a permanent upgrade—like gaining experience for the character, getting perk, or unlocking new weapons—or a temporary upgrade—like being healthier and causing more damage for a specific amount of time.

The rewards are an essential part of any game because users feel engaged with the story and may want to do more exercise to get more rewards (Novak, 2012). Additionally, they can have a positive impact on the replay value by making it easier for people to feel attracted to the content that the game offers.

### 3.1.5 User Progression

In order to have the highest percentage of retention ratio, the game should make the players feel that they are making progress. This can be done by unlocking the game content as the user goes through challenges or achieves a goal.

User progression creates the illusion that players are constantly completing different actions and that there is a constant purpose and motivation to finish the game (F. X. Chen, King, and Hekler, 2014). The story elements and the user progression can give a higher replay value to the game because the idea is for users to come and play again.

The game should be able to create new goals and objectives based on the characteristics of each user. By learning about the players' habits, the game can provide a personalized experience. For instance, the game can create goals or missions that require more effort when the user is more interested in the game.

## 3.2 Statistics

Every game that is intended to motivate people to be healthier by doing more exercise should collect and present statistical data (Brauner, Calero Valdez, Schroeder, and Ziefle, 2013). With this information, players can learn more about their performance in a specific timeframe.

The collected data can be presented in charts and in a granular way. The data can include what the user did in a particular day with an hourly breakout or with their performance within a month, a quarter, a semester, or any other long-term timeframe.

Developers should compare users' performance between different timeframes by using friendly-via graphics, which include the time users prefer to do exercise and the average workout time.

## 3.3 User Information

Health-related applications should create a customized experience for their users by gathering information about their physical activity performance (Alahäivälä and Oinas-Kukkonen, 2016). The game should ask the players for basic data that includes—but not limited to—age, sex, height, weight and how often he or she does exercise or intends to do it.

The weight data gathered from the user should be updated on a monthly basis in order to enable developers to create more accurate health plans for the players. Moreover, the progress made by the users can be tracked while they use the application.

The user information module should keep the raw data of how a player has been performing while using the app. Also, this module should be able to create an estimate of the calories the player has burned within a specific timeframe. Furthermore, if the users have a weight or distance target, this module can feed the game goals to help them achieve the health goal.

## 3.4 Augmented Reality

AR is intended to give users the bridge between reality and the game they are playing (Kipper and Rampolla, 2013). The idea is to collect objects in a real world location and to use them within the boundaries of the virtual game.

There are several ways AR can be incorporated into the game. For instance, when a player gets to a target destination, he or she will have the option to search a zone and find an object. Once the object is found by using the device camera, the users will get a reward in the game. By combining these features, the limits are only set by the developers' imagination.

## 3.5 Anti Cheat Control

All games need a mechanism that prevents players from cheating. This particular software piece should detect if the player is trying to achieve the goals by skipping the required exercises.

The game should have an algorithm mix that uses devices such as the GPS, the gyroscope, and the accelerometer to measure how fast the player is traveling. In fact, these mechanisms can detect if the movements made by the users are possible for human beings.

In this sense, the module should cap the maximum possible velocity of the players to detect if they are traveling by car instead of walking or running. By using the previous devices, the system detects more accurately if the user is doing exercise or not.

### 3.6 Notifications

Notifications remind users to stay active (Brauner et al., 2013). For instance, notifications let players know when they have to move either because they have been inactive for a long time or because they have to look for bounties or upgrades. Socially speaking, notifications can tell users when friends are asking for help, when they want to compete, or when a cooperative mission is needed.

Notifications should be customized in order to give users the option to enable or disable data. The players should decide what kind of information they want to receive after completing the tasks included in the notifications in order to avoid any unpleasant experiences. By providing this possibility, users will not feel overwhelmed or annoyed if they receive notifications of activities that they do not want to see.

### 3.7 Data Collector (Device)

One of the most relevant aspects to consider when creating a health-based video game is the device players need to use. The device has to be to capable of tracking the movement of the users, their velocity and the places they visit.

In order to fulfill the previous requirements, the device should include at least a GPS to have the geolocation of the user and to create target points where they should go within a distance range. In addition, an accelerometer and a gyroscope are required to avoid relying only on a GPS connection and to help reduce the possibility of cheating in the game.

Finally, the augmented reality functionality can be used if the data collector has a camera and an API that developers can access and use.

### 3.8 Social Interaction

Social networks aim at giving more visibility to applications or games to help them go viral; hence, a social module—planned in two directions— has to be included as part of the game.

The first part is an outflow where users can post achievements, rewards, perks, and missions in different social networks. In addition, players can share with other players their real life progress,


Figure 2: Objectives of the Application Mockup.

traveled distances, and time spent in a workout. The second part is an inflow where the users can ask for help in order to complete cooperative and competitive missions.

## 4 IMPLEMENTATION

The model of the game that we implemented was based on the model described in Section 3. The model can be used in different game genres such as action, simulation, and role plays.

Because a role-playing game (RPG) was set, a story of a brave warrior trying to reclaim his family honor and lands was developed. These lands were stolen by the evil wizard Magrathea. To reclaim them, the warrior needs to go through different dungeons. When going into the dungeons, the warrior has to meet some requirements like having a specific level and having a special gear in order to access different places in the game. In these situations, the user is required to do some kind of physical activity.

The game, with the help of AR, will spawn points of interest for the player to get different rewards—an armor, weapons or magical powers. The game also rewards persistency because if the players exercise every day, their virtual character becomes more powerful; otherwise, the virtual character becomes weaker and loses his or her lands.

The game includes competitive and cooperative elements. The players can conquer unclaimed lands and defend their lands from being conquered by other friends. The players can get better gear or upgrades for their gear by completing missions that may require some kind of help from friends.

We presented six mockups to illustrate the design of the game. Figures 2, 3, and 4 include three of them. Figure 2 shows the objectives screen. This screen includes the position of the players in the map and the nearest targets. Targets must have different levels of difficulty and should be selected prior working with near landmarks and sights (i.e. historic sites, fountains, monuments, statues, among others).

Figure 3 works with the augmented reality component. Once the player achieves the goal, a similar screen will be displayed. Figure 4 shows some sample statistics. The remaining mockups (not included in Figures 2, 3, and 4) are events related to the game.

## 5 EVALUATION

Fifty participants, aged from 18 to 42, were interviewed for this project. Sixteen of them were women, and 34 were men. Basic information—age, physical condition, and mobile telephone habits—was gathered. All the participants owned a smartphone and, after the required explanation, were familiarized with the key attributes of the game.


Figure 3: Rewards of the Application Mockup.

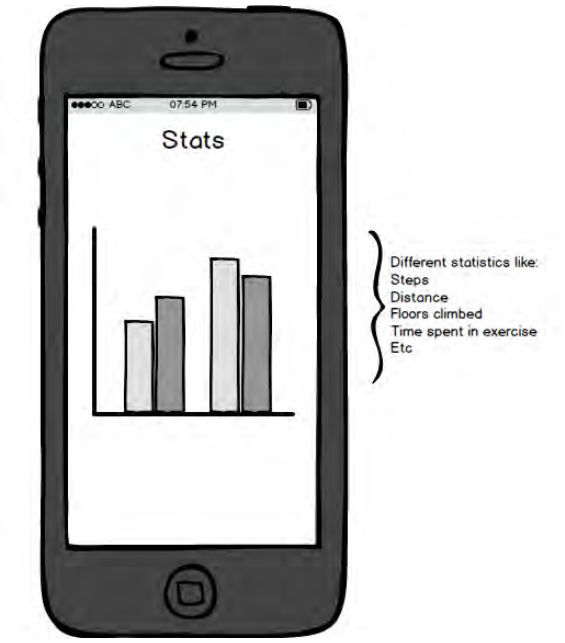



Figure 4: Statistics of the Application Mockup.

The game mechanics were also explained to each participant by using the mockups (see Figures 2 and 3). Then, the participants provided basic data that included the frequency in which they exercise, their mobile game interests, and the mobile fitness apps they use.

In terms of` how often they do physical activity, 52% (26 participants) do not exercise regularly. Approximately 85% (22 of 26 participants) reported some kind of interest in the game. Participants who exercise regularly (i.e. walk or run at least twice a week) were interested in the game. Figure 5 displays that 86% of the participants support the game as an instrument to promote physical activity by using AR components while eight participants indicated that they do not use mobile telephones to play.

When asked if they would play the game while exercising, seven participants (14%) provided a negative answer. The main reasons they provided to support their decision include one of the following: (1) “I prefer to practice different types of exercises (i.e. swimming, going to the gym, and riding a bicycle);” (2) “I feel insecure using the mobile phone outdoors in some places;” and (3) “I exercise but without any mobile application.”

Five participants (62.5%) indicated that they like a game that uses AR. These participants correspond to the ones who regularly use smartphones to play. Only three participants answered that they were not interested in this type of game.

From the total number of participants interested in the application, 100% mentioned that having access to statistics over time to see their monthly or quarterly progress is extremely useful. Approximately 92 % considered that they need daily statistics, and 85% of the participants need the application to tell them when to do exercise.

The participants were asked if they knew what AR is. For those who were unsure or did not know, a video with concrete examples was used to explain the concept to the participants. After seeing the mockups of the game and claiming that they were interested in it, 85% of the participants said that the AR component was essential and attractive in the game.

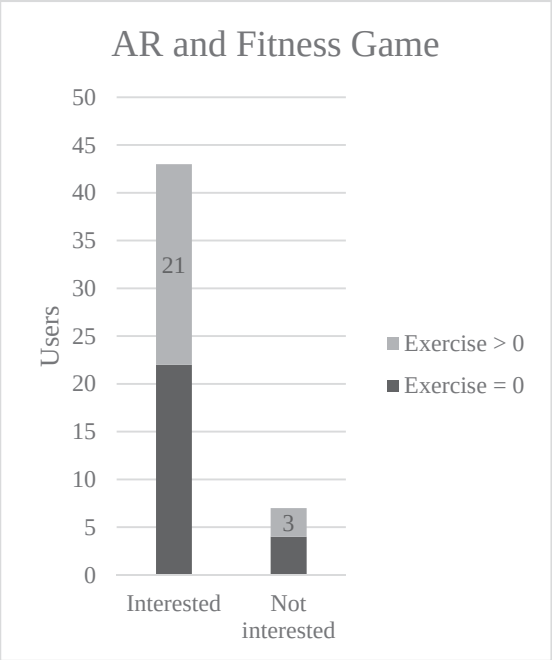



Figure 5: Participants' Interest in the Game.

## 6 DISCUSSION

The results revealed that people who do not exercise are more willing to change their routine if they find a strong reason to do so. In this project, 'this reason' is called the AR game. Similarly, people who already have an exercise routine also feel attracted to this kind of game application.

Based on the results obtained, the massive use of these technologies (AR and Smartphones) should be taken into account by many developers because they incentivize people do change their sedentary lifestyle. By changing users' habits, other problems such as diabetes and hypertension can be avoided or at least decreased. Due to the fact that being attracted to video games is not based on age or sex, new technologies have the possibility of reaching more people.

The findings also state that the proposed model provides the main components that a game should have in order to incentivize people to do exercise. This is possible because of the mechanics proposed in the game design section, which provide an overview of what a game should have in order to increase the number of engaged players. Users are attracted to a game when it has a high-replay value, an interesting story, and cooperative and competitive elements.

The participants also highlighted the importance of having player statistics which help to gather data of the users' real life progress in short and medium term. This part of the module is fundamental because if data is not quantified, users cannot determine if they are making progress in the game.

Some participants were really worried about the idea of playing in public places due to security problems. One possible solution is to consider places with low crime rates to be part of the game.

When developing a AR and Fitness Game, the lack of guidelines to design this kind of software increases the cost of the project and the effort that is needed to finish it. Therefore, having a model in mind will definitely help developers to design and to build AR games.

Finally, this model can be used in other types of exercises. For instance, roller skating, cycling or any other sport that involves moving from one place to another. For future projects, more participants can be included to test a possible prototype.

As shown in this article, new technologies can be used to reduce some of the most common problems related to a sedentary lifestyle.

## ACKNOWLEDGEMENTS

This work was partially supported by the Escuela de Ciencias de la Computación e Informática at Universidad de Costa Rica (ECCI-UCR).

## REFERENCES

Alahäivälä, T., Oinas-Kukkonen, H., 2016. Understanding persuasion contexts in health gamification: A systematic analysis of gamified health behavior change support systems literature. In *International Journal of Medical Informatics.*

Brauner, P., Calero Valdez, A., Schroeder, U., Ziefle, M., 2013. Increase physical fitness and create health

awareness through exergames and gamification: The role of individual factors, motivation and acceptance. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7946 LNCS, 349–362.*

Buchem, I., Merceron, A., Kreutel, J., Haesner, M., Steinert, A., 2015. Gamification designs in Wearable Enhanced Learning for healthy ageing. In *Interactive Mobile Communication Technologies and Learning (IMCL), 2015 International Conference on, (November), 9–15.*

Chen, F. X., King, A. C., Hekler, E. B., 2014. "Healthifying" Exergames: Improving Health Outcomes through Intentional Priming. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14, 1855–1864.*

Chen, Y., Pu, P., 2014. HealthyTogether: exploring social incentives for mobile fitness applications. In *Proceedings of the Second International Symposium of Chinese CHI on - Chinese CHI '14, 25–34.*

Dille, F., Zuur Platen, J., 2008. *The Ultimate Guide to Video Game Writing and Design*. Lone Eagle Publishing Company, 1st edition.

Giannakis, K., Chorianopoulos, K., Jaccheri, L., 2013. User requirements for gamifying sports software. In *2013 3rd International Workshop on Games and Software Engineering: Engineering Computer Games to Enable Positive, Progressive Change, GAS 2013 - Proceedings, 22–26.*

González, C. S., Gómez, N., Navarro, V., Cairós, M., Quirce, C., Toledo, P., Marrero-Gordillo, N., 2016. Learning healthy lifestyles through active videogames, motor games and the gamification of educational activities. In *Computers in Human Behavior, 55, 529–551.*

Kipper, G., Rampolla, J., 2013. *Augmented Reality: An Emerging Technologies Guide to AR*. Syngress, 1st edition.

Novak, J., 2012. *Game Development Essentials: An Introduction*. Cengage Learning, 3rd edition.

Oxford Dictionary of English, 2010. Oxford University Press.

Reynolds, L., Sosik, V. S., Cosley, D., 2013. When Wii Doesn't Fit: How Non-Beginners React to Wii Fit's Gamification. In *Gamification, 111–114.*

Spillers, F., Asimakopoulos, S., 2014. Does social user experience improve motivation for runners? a diary study comparing mobile health applications. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8520 LNCS(PART 4), 358–369.*

Zuckerman, O., Gal-Oz, A., 2014. Deconstructing gamification: evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. In *Personal and Ubiquitous Computing, (Idc), 1705–1719.*

# Medical Imaging: Exams Planning and Resource Assignment
## *Hybridization of a Metaheuristic and a List Algorithm*

Nathalie Klement[1], Nathalie Grangeon[2] and Michel Gourgand[2]
[1]*École Nationale Supérieure d'Arts et Métiers, LSIS CNRS UMR 7296, 8 boulevard Louis XIV, 59046 Lille Cedex, France*
[2]*LIMOS CNRS UMR 6158, Université Blaise Pascal, Complexe Universitaire des Cézeaux, 63178 Aubière Cedex, France*
*nathalie.klement@ensam.eu, {grangeon, gourgand}@isima.fr*

Keywords: Hybridization, Metaheuristic, List Algorithm, Medical Imaging.

Abstract: The presented work is about optimization of the hospital system. An existing solution is the pooling of resources within the same territory. This may involve different forms of cooperation between several hospitals. Problems of sizing, planning and scheduling may be considered. We define the problem of activities planning with resource assignment. To solve this problem, we propose a hybridization between a metaheuristic and a list algorithm. Single based metaheuristics are used. This proposition requires a new encoding inspired by permutation problems. This method is easy to apply: it combines already known methods. With the proposed hybridization, the constraints to be considered only need to be integrated into the list algorithm. For big instances, the solver used as a reference returns only lower and upper bounds. The results of our method are very promising. It is possible to adapt our method on more complex issues through integration into the list algorithm of the constraints. It would be particularly interesting to test these methods on real hospital authorities to assess their significance.

## 1 INTRODUCTION

Given the current economic situation, everything is done to move towards a better use of goods and services production systems. The hospital system also follows this trend as much or less resources are allocated to it but it should work more efficiently to meet a demand that is increasing. To do so, in 2015, the french government defined the HGT: Hospital Group of Territory, an evolution of the HCT previously presented (Gourgand et al., 2014a). It is a cooperation between public institutions, which are at different places, that implement a common strategy and jointly manage some functions and activities through delegations or skills transfer between them. Some decision support tools are needed to manage this new kind of organization.

The aim of our work is the development of a decision support tool to help to manage HGTs or any hospital cooperations. This tool should be used at different levels: strategic, tactical or operational, to deal with problems of sizing, planning, resources assignment or scheduling. It should be used to anticipate the creation of a new cooperation, to manage this organization day to day, or to react in case of hazard or crisis situation. In this paper, we take the problematic of medical imaging over a HGT as an example. Some material resources, such as X-ray, scanner, MRI, are located at different places belonging to the HGT. Human resources work there and have specific competences on these material resources. Some patients need to pass an exam on such a material resource. The planning horizon can be some days or weeks, divided in periods of half-days. Some incompatibilities and time constraints are defined. The tactical level will be discussed in this paper: the objective is to assign the exams to human and material resources during a period. The other levels can easily be solved by our method which uses some operational research purposes.

## 2 STATE OF THE ART

Since the last ten years, many methods of operational research have been used to solve hospital system problems. (Rais and Viana, 2011) referenced around two hundred and fifty articles about operational research for hospital system in general. (Car-



Klement, N., Grangeon, N. and Gourgand, M.
Medical Imaging: Exams Planning and Resource Assignment - Hybridization of a Metaheuristic and a List Algorithm.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 260-267
ISBN: 978-989-758-213-4

doen et al., 2010) referenced more than one hundred and twenty articles about planning and scheduling of operating theater. (Van den Bergh et al., 2013) made a literature review about scheduling human resources and referenced around three hundred articles.

But articles dealing with multi-place system in hospital network are scarce, or their case study are quite limited. Planning surgical vacations by specialty is dealt by (Santibáñez et al., 2007). Availability of operating theater, beds capacity, surgeons preferences and waiting list are considered. The proposed model allocates specialties to operating theater, and is applied to a Canadian study case, composed by eight hospitals, over four weeks. (Everett, 2002) developed a support aid tool to manage the waiting list of patients who need a surgical act. The system is made up by several hospitals, working in cooperation. The waiting list is common for all the hospitals. Each day, each patient is assigned to one place according to the availability of the hospitals. If no hospital is available on one day, patients are assigned the next day. Assignment of resources is not considered. (VanBerkel and Blake, 2007) developed a tool to reduce the waiting time and to plan beds capacity in a surgery service, over several places. It is a problem of allocation of the fixed resources, in an other Canadian example. This tool aims at studying a redistribution of postoperative beds between the places, using simulation tools. Problem of capacity in intensive care unit can result in the cancellation of programmed acts, an overload of medical team or a reject of urgent patients. Thus, urgent patients could be transferred to further places. A cooperative solution is studied by (Litvak et al., 2008), taking into example a case in the Netherlands. Some hospitals belonging to the same territory share some beds to urgent patients.

Researchs are done about the pooling of resources. (Trilling et al., 2006) dealt with the problem of scheduling human resources over different services in one hospital. The objective is to share resources within larger surgical suite, in order to reduce the costs. At another level, it can be seen as a sharing of resources from different places within larger organization such as HGT. In this paper, concerned resources are stretchers and nurses, who are common resources used for any hospital services and locations.

A lot of researches about hospital system are dedicated. Articles consider three problems: sizing, planning and scheduling. Most of the papers focuses on one particular problem. Their models and resolution methods are not easily reusable. Our proposed model and tool are generic, so they could be reused as often as possible.

# 3 ANALYSIS

To analyze our system, we split the system into three subsystems: the physical subsystem (physical entities used to perform all the activities, their geographical distribution and their interconnections), the logical subsystem (flows that the system should treat, all activities concerning the treatment of these flows and all entities in the system relating to them) and the decision subsystem (center of decision which contains all the decision rules).

## 3.1 Physical Subsystem

The HGT is composed by several places. There is a known distance between each place.

On each place, there are one or several material resources. A material resource belongs to a type (for instance X-ray, scanner or MRI). Each material resource has an opening schedule which defines the times when the material resource is available over each period. For example, a given material resource may only be available five hours on Monday because it needs a maintenance operation or because an external doctor reserved it. Overtime may be allowed but is limited in time.

Human resources compose a medical team. The composition of this team depends on the considered exam. This team should have a specific number of stretchers, specialist doctors, nurses, etc. Human resources belong to a given place but can work on other places belonging to the same HGT, allowing a pooling of human resources over the HGT. Moves are not allowed within the same period, but between two periods. A time is given to human resources to go from a place to another. A human resource can use one or several types of material resource according to its skills. A skillful human resource, who can work on several types of material resources, is potentially less efficient than a human resource who can only work on one type of material resource, or one particular material resource. This efficiency should be translated in the processing time of the concerned exams. Each human resource has a planning which defines its regular work time, taking into account break times and holidays. Time to move from one place to another is included in the work time of human resources. Overtime may be allowed but is limited in time.

## 3.2 Logical Subsystem

The logical subsystem defines the flow: the set of exams to plan and assign, and the relationship between these exams and the resources previously defined.

An exam should be done before a period at the latest, called a due date. Each exam has a known processing time which depends on the assigned human and material resources. Each exam starts at one period and ends at the same one. Each exam has a reference place, where it should be done, if possible.

An exam needs a given number of human resources and one material resource. All required resources must be compatible with each other. By definition, an exam is compatible with some material resources, so the assigned material resource must be compatible with the exam. This material resource belongs to a type, so the assigned human resources must have the needed skill to use this type of material resource. The place where the exam is done is deduced from the one where is located the assigned material resource.

## 3.3 Decision Subsystem

The objective is to develop a model which, from a set of exams, builds a planning associating the triplet $\{$exam, human resources, material resource$\}$ to a period. The study is made in a predictive approach, all the exams can be treated since the beginning of the planning horizon. The objective of this model is not to schedule exactly the exams but to assign one period to each exam. This planning must optimize some criteria and respect some constraints.

### 3.3.1 Criteria

Three categories of criteria can be defined: economical, about the comfort of the patient, and about the proper functioning of the HGT.

Concerning the economical aspects, criteria are about the costs. Occupation rates of each place, each material resource and each human resource help to ensure the proper use of these entities. To be the most economic, these occupation rates have to be maximized. However, it can be preferable to define a security margin, so the HGT can be reactive in case of hazard. All exams are planned during the considered planning horizon. The makespan is the period assigned to the last exam. It ensures that all exams are assigned as soon as possible. The smaller the makespan is, more time remains free at the end of the planning horizon to potentially treat the next exams. The number of moves of human resources in the HGT should be considered, as well as overtime of human and material resources. Overtime and moves have a cost for the HGT. It is better to minimize them. But it can be interesting to allow some overtime or some moves to increase the number of exams during the considered period.

About the comfort of the patient, the criterion is the number of exams done at their reference place. If some exams cannot be done at their reference place, the distances between the reference places and the effective ones may be minimized.

About the medical criterion, the number of exams done before their due dates has to be maximized. Thus, if a patient needs other exams, the next ones could be done on time. If an exam is planned after its due date, the tardiness may be minimized.

### 3.3.2 Constraints

**Constraints that Must Be Respected**

- Each exam must be assigned to human resources, one material resource and one period. The considered human resources must be assigned during the period at the place where the considered material resource is located.
- Compatibility between the material resource and the exam: the assignment must satisfy the given list of incompatibilities between exams and material resources.
- Compatibility between skill of the human resource and the type of the material resource: for each exam, the human resource must be able to work on the considered material resource.
- If a human resource can move during the planning horizon, its moves are constrained: a human resource can only work at one place during one period.

**Constraints that May Be Respected**

- Exams should be done before their due dates and at their reference place.
- Material resources and human resources may be used or work during their opening schedule. Otherwise, additional time is considered as overtime. Overtime is limited in time.

# 4 CONSIDERED PROBLEMS

The considered problem is defined as follow: exams planning and assignment of needed material and human resources. The previous analysis was about the complete problem. Some hypothesis are made to divide this complete problem into three problems.

## 4.1 Hypothesis

The following hypothesis are made:

- Only one human resource and one material resource are needed to perform exams. Human resources are compatible with one or several types of material resources.
- Processing time of the exams are given and fixed.
- Opening schedules of material resources are equal in every periods.
- To each exam, the release date is equal to the date of appointment decision. These dates are equal to zero: all exams are known at the beginning of the planning.
- Distances between places are taken into account in the time allocated to the human resources to move from one place to another. This time is assumed to be constant, all places are equidistant to the others.
- Overtime is not allowed.

## 4.2 Definition of the Considered Problems

The complete problem is divided into three problems of increasing difficulty:

- Problem 1 is the more basic: human resources are not considered. Only the material resources are considered.
- Human resources are considered in Problem 2. They can work on one or several types of material resources. They cannot move, they work all the time at the same place. The assignment of the human resources at the places is given.
- Human resources are mobile in Problem 3. They can work on several places, they can move from one place to another. The assignment of the human resources at the places has to be built by the model.

Two criteria are used in the following study:

- Sum of assigned periods to all exams, which ensure that all exams are planned as soon as possible.
- Number of exams assigned before their due date.

## 4.3 Analogy with the Bin Packing Problem

Our problem can be seen as a bin packing problem (Gourgand et al., 2014a). The bin packing problem considers $N$ items, with a given size, and some bins with the same capacity. The aim is to pack all the items in a minimum number of bins. The size of the packed items has to respect the capacity of the bins. Each item has to be assigned once and only once.

### 4.3.1 Without Human Resources

Considering Problem 1, the aim is to assign exams to a material resource during a period. The planning horizon is made by couples (resource, period). The objective is to assign exams to couples (resource, period). Exams have to be done as soon as possible: the aim is to minimize the number of couples, (= the number of bins). An example is given by Figure 1, where the assignment of exams to material resources $MR_1$, $MR_2$ and $MR_3$ is considered. Table 1 summarizes analogies between bin packing problem and Problem 1: exams planning with resources assignment.

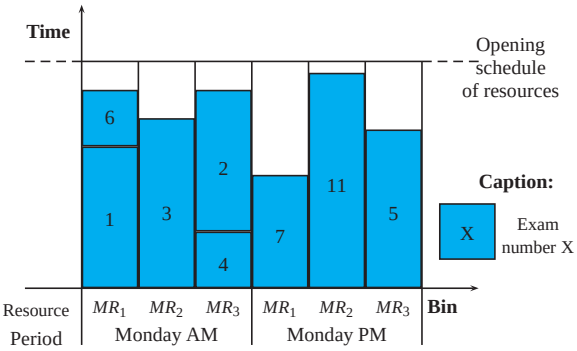



Figure 1: Representation of Problem 1.

Table 1: Analogies between both problems.

| | Bin packing problem | Problem of exams planning with resources assignment |
|---|---|---|
| **Data** | Item | Exam |
| | Bin | Couple (resource, period) |
| | Size of an item | Processing time of an exam |
| | Capacity of a bin | Opening schedule of resources |
| | - | Due date |
| | - | Reference place |
| **Problem** | Assign items to one bin | Assign exams to one couple (resource, period) |
| **Constraints** | Capacity of bins | Opening schedule of resources |
| | - | Constraint of compatibility |
| **Criteria** | Minimize the number of used bins | - |
| | - | Minimize the sum of assigned periods |

### 4.3.2 With Human Resources

Exams have to be assigned to material and human resources during one period. Analogy is made between this problem and interdependent bin packing problem.

Let's take $p_1$ and $p_2$ two bin packing problems, with a given number of bins. Groups of bins are defined in both problems. A group should be made by one or several bins. Number of bins can be different

for both problems, but number of groups is the same. Each item is assigned to one and only one bin in each problem. Interdependence between both problems is defined as follow: if an item is assigned to a bin from group $g$ in problem $p_1$, it must be assigned to a bin from the same group $g$ in problem $p_2$. The aim is to assign items in bins of both problem, by minimizing the number of used bins, satisfying capacity constraints and interdependence between both problems.

In our case, both problems $p_1$ and $p_2$ can be defined like this:

- $p_1$: assignment of each exam to a material resource during a period, respecting the opening schedule of the material resource during the period and the compatibility between the exam and the material resource.
- $p_2$: assignment of each exam to a human resource during a period, respecting the work time of the human resource during the period and the compatibility between the exam and the human resource.

Compatibility between exam and human resource is not directly defined but can be deduced: an exam is compatible with a human resource if and only if this exam is treated by a material resource from one type and this human resource can work on this type of material resource.

Thus, group $g$ is the couple (period, type). In both problems $p_1$ and $p_2$, exam has to be assigned during the same period to the same type of material resource. Figure 2 illustrates the interdependent bin packing problem in the HGT case. The lower portion of the figure is the assignment of exams to material resources and periods, in the same way as Figure 1. The upper portion is the assignment of exams to human resources and periods. In both portions, each exam has to be assigned to the same period and the same type (according to the type of the material resource and the competencies of the human resource to use this type).

# 5 RESOLUTION METHOD

The bin packing problem is NP-Hard (Garey and Johnson, 1979). Our problems are an extension of the bin packing problem, so our problems of exams planning with resources assignment are also NP-Hard. In the following, approximate methods are used to solve them.

Our proposed method is a hybridization of a metaheuristic and a list algorithm. Our tool is convenient because one part is generic: it can be used for any of the considered problems. Only the list algorithm needs to be specific to the considered problem.

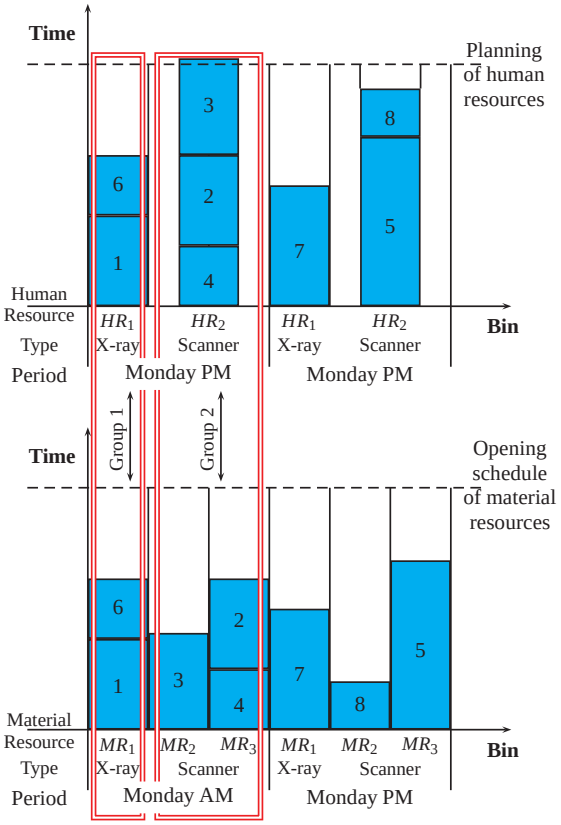



Figure 2: Interdependent bin packing problem.

## 5.1 Genericity

The proposed tool, illustrated by Figure 3, uses a hybridization of a metaheuristic and a heuristic, more precisely a list algorithm. A single solution based metaheuristic or a population based metaheuristic can be used. The encoding used by the metaheuristic is a list $Y$ of exams. The list algorithm $L$ considers the exams according to their order in list $Y$ to plan and assign them to the required resources, considering the problem constraints. This builds a solution $X$. The objective function $H$ evaluates the solution $X$. According to this evaluation, the solution is chosen or not by the metaheuristic. At the end of the running, the given solution by the hybridization is the best list $Y^*$ of exams: the one which optimizes the objective function by applying the list algorithm. This hybridization can be used to solve many problems: the specificity of a given problem is only considered in the list algorithm.

## 5.2 A List $Y$ of Exams

The general scheme of the encoding is given by Equation (1), with $\Omega$ the set of all the lists $Y$ and $S$ the set

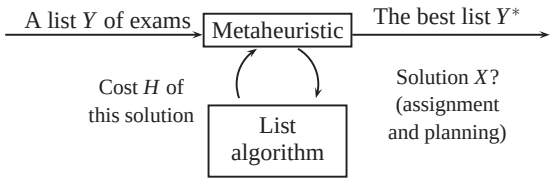



Figure 3: Hybridization metaheuristic - List algorithm.

of all the admissible solutions $X$ built by the list algorithm $L$.

$$Y \in \Omega \underset{Heuristic\ L}{\longrightarrow} L(Y) = X \in S \underset{Criterion\ H}{\longrightarrow} H(X) \quad (1)$$

The set $\Omega$ is the set of all permutations of exams. Cardinal of $\Omega$ is $N!$ with $N$ the number of exams. One solution $Y \in \Omega$ is a list of exams. More details about the encoding are given in (Gourgand et al., 2014b).

## 5.3 Metaheuristic

The metaheuristic performs in the set of solutions $\Omega$. An initial solution is randomly computed: a list of exams randomly sorted between one and the number of exams. Several metaheuristics have been used: some single solution based metaheuristics such as iterated local search or simulated annealing. A neighborhood system is used to visit the set of solutions, it allows to go from one solution to another one. Neighborhood system $V$ is a permutation of two exams in the list $Y$: the exam at position $i$ permutes with the one which is at position $j$. $V$ satisfies the accessibility and reversibility properties.

## 5.4 List Algorithm

A list algorithm is used to build the solution $X$ from the list $Y$: it assigns the exams to resources and to periods.

List scheduling algorithms are one-pass heuristics that are widely used to make schedules. A standard list scheduling algorithm constructs a schedule by assigning each job in listed order to the first machine that becomes idle (Zhu and Wilhelm, 2006). It is important to work with a list algorithm, because the metaheuristic browses the set of solutions. So the used algorithm needs to consider the order of the list to assign exams to resources and periods.

Our problem has be analyzed as a bin packing problem and some list algorithms have been proposed since the definition of this problem (Johnson, 1973). So our developed list algorithm is inspired by them. For instance, considering Problem 1, in which human resources are not considered, Algorithm 1 is an extension of the First Fit algorithm for the bin packing problem. Other list algorithms can be adapted from Algorithm 1 to solve the cases of Problems 2 and 3, considering human resources.

Algorithm 1: List Algorithm First Fit HGT.

**Data**: List of exams $(Y_i)_{i\in\{1,N\}}$; opening schedule of all resources during all periods; processing time of all exam
Occupied time := 0 for all resources and all periods
**forall the** $i$ **do**
  First resource, first period, *assigned* := $false$
  **while** (*assigned* = $false$) *AND current period* $\leq$ *max of periods* **do**
    **while** (*assigned* = $false$) *AND current resource* $\leq$ *max of resources* **do**
      **if** *exam* $Y_i$ *is compatible with current resource* **then**
        **if** *exam* $Y_i$ *fits in couple (resource, period)* **then**
          Assign exam $Y_i$ to couple (resource, period)
          Update occupied time of couple (resource, period)
          *assigned* := $true$
      Next resource
    Next period

## 5.5 Objective Function

Solutions are compared according to an objective function which characterizes the quality of the solution. In our case, the objective function represents the timing aspect of our problem. Exams have to be done as soon as possible, thus the makespan, the period assigned to the last exam, should be considered. Because many solutions may have the same makespan, we choose instead the sum of assigned periods to all exams, so the solutions can be dissociated. This criterion is written $H_S$. Another criterion is considered to ensure that most of the exams are assigned before their due date. This criterion, written $H_D$, is computed as the number of exams assigned after their due date. The weighed criteria method is used (Coello, 2000). The objective function is a weighed sum between both criteria, defined by Equation (2). $\omega_D$ is chosen equal to 5 because $H_S$ is always smaller than $10^5$ so both criteria are easily readable. This function has to be minimized.

$$H(X) = 10^{\omega_D} \times H_D(X) + H_S(X) \quad (2)$$

### 5.6 The Best List $Y^*$

Algorithm 2 describes the whole method with the example of stochastic descent as the used metaheuristic. Stochastic descent may be used in an iterated local search. The set $\Omega$ of the lists of exams is browsed thanks to the metaheuristic using neighborhood system $V$. Lists are compared thanks to the list algorithm $L$ and the objective function $H$. According to an acceptance criterion, some lists are selected. At the end, the metaheuristic gives the best found list $Y^*$.

Algorithm 2: Hybridization between stochastic descent and a list algorithm.

**Data**: Initial solution $Y$
$X := L(Y)$
**while** *necessary* **do**
  Choose uniformly and randomly $Y' \in V(Y)$
  $X' = L(Y')$
  **if** $H(X') \leq H(X)$ **then**
    $X := X'$
    $Y := Y'$
**Result**: $Y^* = Y$

## 6 EXPERIMENTS

The data are randomly generated but the characteristics and the size of the data represent real instances. The HGT is composed by 3 places. The planning horizon is made by 8 to 40 periods. As a remind, one period represents one half-day, thus the planning horizon is between 4 and 20 days. 4 to 8 resources are available. 50 to 500 exams need to be planned and assigned. Incompatibilities between exams and resources are randomly generated. Each processing time is between 5 and 100. Each material resource has an opening schedule equal to 300 minutes.

The results are detailed in Table 2. The host machine is powered by an Intel Xeon X5687 quad-core CPU running at 3.6 GHz. The computation has been stopped after thirty minutes. Each reported result is the value of the objective function for the best solution found in less than thirty minutes, but most of the time, the best solution is found in a few minutes. The results are presented as a couple of values $(H_D; H_S)$ with $H_D$ the number of exams assigned after their due dates, and $H_S$ the sum of assigned periods to all the exams. The results compare three methods:

- The resolution of the mathematical model with an exact method by using the solver CPLEX. If no optimal solution has been found in less than thirty minutes by the solver, no result is written.
- Our results from the method previously published in (Gourgand et al., 2014a), using two single solution based metaheuristics (iterated local search and simulated annealing) in a classical way: the best value found by all these methods is reported.
- Our results from our proposed method detailed in the current paper. The used metaheuristics are distinguished: iterated local search and simulated annealing, written ILS* and SA*.

The results are promising. Firstly, this problem has been solved by CPLEX thanks to our mathematical model previously proposed. The solver finds an optimal solution only for small size of problems (less than two hundred exams over four days). The solver does not find any solutions when the size of the problem increases. Then, it has been solved with two approximate methods: in a classical way, and with a hybridization. Both methods find an optimal solution for the small instances. For biggest instances, the hybridization between a metaheuristic and a list algorithm outperforms our previous method. Simulated annealing seems to work better than iterated local search.

## 7 CONCLUSION AND PERSPECTIVES

The current hospital context needs to find solutions to improve efficiency of hospital systems. Hospital cooperation has emerged, as Hospital Group of Territory. A pooling of resources may cause a better use of the different places in a same territory. But this cooperation needs some decision support tools to improve or optimize their running. In this paper, we defined the general problem of activities planning with resources assignment in a multi-place hospital context.

Because this problem is NP-Hard, we propose an approximate method to solve it: a hybridization between a metaheuristic and a list algorithm. The results are promising: our method finds good results in a few minutes. An improvement of the results is in process, using a population based metaheuristic: Particle Swarm Optimization. Using PSO, the results are very good for small instances: an optimal solution is found in a few seconds, but the method still needs some tuning for the biggest instances.

Thanks to the hybridization, our method can be easily reusable. Indeed, to solve other problems, only the list algorithm needs to be modified. The metaheuristic part will still be the same. Any kinds of planning, assignment or scheduling problem can be

Table 2: Results: (number of exams assigned after their due dates; sum of assigned periods to all the exams).

| Number of exams | CPLEX | (Gourgand et al., 2014a) | ILS* | SA* |
|---|---|---|---|---|
| 50 | (0;51) | (0;51) | (0;51) | (0;51) |
| 50 | (1;150) | (10;147) | (1;151) | (1;150) |
| 100 | (0;131) | (0;131) | (0;131) | (0;131) |
| 100 | (0;517) | (2;535) | (1;516) | (0;518) |
| 200 | (0;266) | (0;266) | (0;266) | (0;266) |
| 200 | - | (3;1197) | (0;1154) | (0;1135) |
| 300 | - | (0;548) | (0;537) | (0;534) |
| 400 | (0;830) | (0;890) | (0;841) | (0;835) |
| 500 | - | (0;1350) | (0;1241) | (0;1234) |
| 500 | - | (194;8218) | (19;6382) | (18;6659) |

solved thanks to this tool by changing the list algorithm: for instance, it has been used to solve an industrial problem (Silva et al., 2016). Problems with human resources can easily be solved by developing some new list algorithms dedicated to them. Then, a direct application to a hospital system could be envisaged. Other applications in the hospital field could be done, in other hospital services, with other resources, etc. We could extend our current work about medical imaging to medical surgeries. More constraints about medical team should be considered. The next problematic is to consider patients with several exams or surgeries, by taking into account precedence constraints between them.

## REFERENCES


Cardoen, B., Demeulemeester, E., and Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3):921–932.

Coello, C. A. (2000). An updated survey of ga-based multiobjective optimization techniques. *ACM Computing Surveys (CSUR)*, 32(2):109–143.

Everett, J. (2002). A decision support simulation model for the management of an elective surgery waiting system. *Health Care Management Science*, 5(2):89–95.

Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-completeness*. WH Freeman and Company, New York.

Gourgand, M., Grangeon, N., and Klement, N. (2014a). Activities planning and resource assignment on multiplace hospital system: Exact and approach methods adapted from the bin packing problem. In *7th International Conference on Health Informatics, Angers, France*, pages 117–124.

Gourgand, M., Grangeon, N., and Klement, N. (2014b). An analogy between bin packing problem and permutation problem: A new encoding scheme. In *Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World*, volume 438, pages 572–579. Springer Berlin Heidelberg.

Johnson, D. S. (1973). *Near-optimal bin packing algorithms*. PhD thesis, Massachusetts Institute of Technology.

Litvak, N., van Rijsbergen, M., Boucherie, R. J., and van Houdenhoven, M. (2008). Managing the overflow of intensive care patients. *European journal of operational research*, 185(3):998–1010.

Rais, A. and Viana, A. (2011). Operations research in healthcare: a survey. *International Transactions in Operational Research*, 18(1):1–31.

Santibáñez, P., Begen, M., and Atkins, D. (2007). Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a british columbia health authority. *Health care management science*, 10(3):269–282.

Silva, C., Klement, N., and Gibaru, O. (2016). A generic decision support tool for lot-sizing and scheduling problems with setup and due dates. In *International Joint Conference - CIO-ICIEOM-IIE-AIM (IJC 2016), San Sebastian, Spain*. ICIEOM.

Trilling, L., Guinet, A., and Le Magny, D. (2006). Nurse scheduling using integer linear programming and constraint programming. *IFAC Proceedings Volumes*, 39(3):671–676.

Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., and De Boeck, L. (2013). Personnel scheduling: A literature review. *European Journal of Operational Research*, 226(3):367–385.

VanBerkel, P. T. and Blake, J. T. (2007). A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. *Health care management Science*, 10(4):373–385.

Zhu, X. and Wilhelm, W. E. (2006). Scheduling and lot sizing with sequence-dependent setup: A literature review. *IIE transactions*, 38(11):987–1007.

# Uncovering Key Factors for a Hospital IT Change Strategy

Noel Carroll[1], Ita Richardson[1,2] and Marie Travers[2]
[1]*ARCH – Centre for Applied Research in Connected Health, University of Limerick, Limerick, Ireland*
[2]*Lero – the Irish Software Engineering Research Centre, University of Limerick, Limerick, Ireland*
*{noel.carroll, ita.richardson, marie.travers}@lero.ie*

Keywords Change Management, Healthcare, Hospital, Kotter's Model, Quality.

Abstract Changing an Information Technology (IT) system within any organisation is a difficult and complex process. However, within the hospital setting, additional complexities make such change more difficult. These complexities include the protection of patient safety and privacy, improving the quality of the patient experience, protecting information and supporting the clinician in their medical requirements. Our research indicates that uncovering the process of hospital IT change management is not documented – making it difficult to build on evidence-based research and instill a 'lessons learned' approach in publicly funded hospitals. We address this gap in this paper. Using qualitative research methods we present the results of observations carried out in healthcare settings as well as twelve structured interviews with hospital staff. We employ the Kotter Change Model as a lens to understand this change process. While benefiting from the structure that Kotter's model provides, we argue for the need to extend this model in an effort to capture the various influences of healthcare IT-enabled innovation which will, in turn, enable much needed change within hospitals. Building on our findings, we introduce a Healthcare IT Change Management Model (HIT-CMM).

## 1 INTRODUCTION

In recent years, much has been documented about the crisis which healthcare systems currently face due to growing demand and expectations from traditional healthcare models. Healthcare organizations now realize that innovation is increasingly required to sustain a quality healthcare service system (Cazzaniga and Fischer 2015). To be successful, innovations through the implementation and upgrading of Information Technology (IT) systems should align with practice and support the evolution of healthcare processes change.

Arguably, the healthcare system suffers from similar issues experienced by other sectors when implementing change through IT. For example, while healthcare service providers commit to improving a service and invest heavily in technological infrastructures to reach improved service levels, managing the change process of IT innovation is a complex task. Healthcare IT must protect patient safety and privacy, and in addition, there are clinical, technical and software regulations that need to be considered.

Thus, uncovering the process of IT change management draws on examining a wide range of perspectives to understand how change can be successfully managed. There are numerous models throughout the literature which guide the change process. Kotter's change model is one such change management model. The authors build on a recent study by Travers and Richardson (2015) which uses Kotter's change model (Kotter 2005) to examine change processes within a private sector medical device healthcare innovation context. Their study documented a single case study in a medical device company. They discovered that process improvement should be managed through the use of this model to ensure that change is implemented systematically throughout the whole organisation. In this paper, we use the same model as a basis to understand how IT change has been managed in public hospital departments. Our results contextualise the change process within the hospital domain and allows us to introduce a Healthcare IT Change Management Model (HIT-CMM).

The next section is divided in two, namely introducing IT systems in hospital settings and Kotter's model.



Carroll, N., Richardson, I. and Travers, M.
Uncovering Key Factors for a Hospital IT Change Strategy.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 268-275
ISBN: 978-989-758-213-4

# 2 LITERATURE REVIEW – IT CHANGE IN HOSPITALS

Change management requires a specific approach to transition an organisation to a desired future state (Benjamin and Levinson 1993). Within a hospital context, the various steps required to achieve a desired future state is of particular importance to ensure that patient safety is a priority and quality is not jeopardized (Cazzaniga and Fischer 2015). The objective of change management is typically to provide an approach to implementing change in a controlled manner while adhering to specific requirements such as functionality, budget and time through various deliverables or milestones. Change management is well documented throughout literature. For example, Lewin's Three Step Change Theory (Lewin 1947) and ADKAR Model (Hiatt 2006) are all applied to various dimensions of the change process.

## 2.1 Kotter's Change Model

Introducing change must be a formalised planned process (Forte 1997). Even though it is sometimes considered that having a process can be an overhead, change management techniques have shown that when change is planned it is more likely to be successful (Forte 1997). Therefore, most planning models assume that changes in organisations are planned changes (Hayes and Richardson 2008). The models stipulate that, for successful change, certain sequential steps need be executed. Kotter's change model is one such change management model (Kotter 2005).

We examine Kotter's change model (illustrated in Figure 1) within a publicly funded hospital setting. We refer to a publicly funded hospital as one where most of its funding comes from state funds. In our case study, state funding comes via the HSE. Using Kotter's eight steps, we conducted a case study to answer the following research question:

*How do clinical departments within a publicly funded hospital setting successfully implement an IT system?*

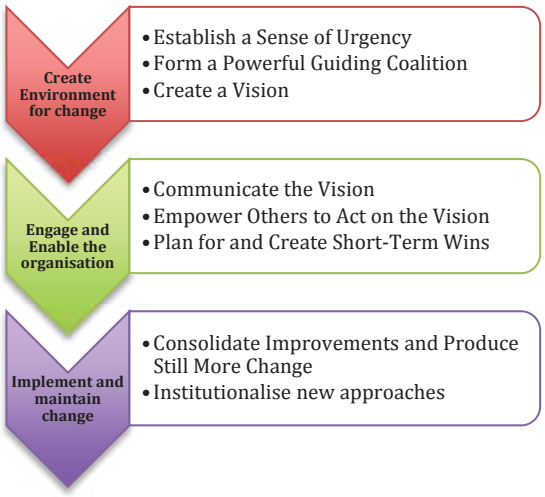



Figure 1: Kotter's Change Model (illustrated by authors).

# 3 METHODOLOGY

Qualitative research methods enjoy numerous approaches to capture raw and rich data. For example, adopting the case study method provided us with the structure to devise specific procedures to design a research strategy, collect data, analyse data, and present and report the results. We opted to undertake observational methods within a single case study considering the unique opportunity to capture an empirically rich account of specific phenomena (Yin 2013) within a healthcare context.

The authors carried out one-to-one interviews. The departments focused on were Radiology, Dermatology, Quality, Physiotherapy and IT. The interviews were held with twelve key staff members who were all involved in IT change to various degrees. Since the interviewees were healthcare experts within public hospitals, some were difficult to access. To overcome this, the authors employed a snowballing sampling strategy (Grbich 1999). This was used to identify other experts in this field within the sample population. This proved to be useful since each expert was able to recommend the next relevant expert. Through a structured interview technique, we were able to provide a more balanced insight to uncover the change process. The structured interviews supported our research methodology by ensuring consistency, i.e. each interviewee was presented with exactly the same questions in the same order. The questions had to be short since the health experts had limited time available to partake in the case study. The questions were as follows:

1. What are the current IT systems in place within your department?
2. Give examples of how new IT systems or processes were implemented? Specifically how was the change process managed? Give examples.
3. Kotter's (2005) is a change management model, which recommends 8 steps to follow to manage change. Kotter's Step 7 "Consolidate Improvements and Produce More Change" recommends that management or change advocates should be become more involved in the process thus ensuring continuation of changes. Kotter's Step 8 "Institutionalise New Approaches" recommends that for success change has to be implemented so that it is now part of the organisations culture. Is this true in your experience in regards to moving or changing to new IT systems/processes? Give examples.
4. Were there any unexpected problems or issues that affected such project changes? Give examples.
5. What is your opinion of the new IT system/process implemented?
6. What could or should have been done differently? Give examples.

The interviewees' answers were reliably aggregated and comparisons were made between the different interviewees. We identified a number of emerging themes using open coding to categorise the text – allowing us to build a story around specific events, facts, and interpretations.

The interviewees' work experience spanned from 4 to 30 years. Participant's interview data (Table 1) was analysed to understand the change process within the case study. We reviewed the data within the structure of Kotter's change model steps 1 to 8, which allowed us to understand how change had been made within the hospital setting. This facilitated our gaining a rich insight of the working environment.

Analysing the findings from the hospital study we identified key themes. We contextualized these findings and their implications on Kotter's change model. Our results indicate that some aspects of Kotter's change model is useful to successfully manage change but would need to be modified for a healthcare context. This case study facilitates analysis from a hospital perspective and the findings informed and enhanced a proposed model, which we call the HIT-CMM (see Table 3).

Table 1: Summary of Interviewee Profiles.

| Interviewee | Department | Yrs Exp. | Specialty |
|---|---|---|---|
| 1 | Quality | 23 | Nurse and Risk Manager with focus on use of IT systems |
| 2 | Radiology | 29 | Administration with focus on quality |
| 3 | Physiotherapy | 17 | General Administration |
| 4 | HR | 28 | Project Manager |
| 5 | Dermatology | 4 | Clinician |
| 6 | Radiology | 25 | Clinician/Project Manager |
| 7 | IT | 30 | Manager with focus on hardware and software deployment |
| 8 | Quality | 19 | Manager with focus on rick management |
| 9 | Laboratory | 28 | Manager with focus on deployment |
| 10 | Radiology | 28 | Clinician/Project Manager |
| 11 | IT | 20 | Project Manager |
| 12 | Quality | 10 | System user |

# 4 FINDINGS

Within the hospitals, we found that there were silos of IT innovation in which a clinician or manager championed IT change. Silos proved problematic when patients had to move between departments. The need for national or central rollout of projects was identified as a solution. National or central rollouts do take time so some departments would go ahead and implement new systems thus creating IT silos.

The interview findings identified various conduits of information on the real-world IT change management process, and enabled us to explain how change management may be viewed as a product of change leadership. Based on our analysis of the observations and the interviews, we identified a number of key themes, which we present as follows:

a) Requirement for Change
b) Attitudes towards new IT systems and processes
c) Lessons Learned

We provide a discussion to contextualize these findings and their implications on Kotter's change model.

## 4.1 Requirement for Change

The need for change was clearly highlighted from the interviews. For example, Interviewee 3 explained that "change is overdue as every evening each patient and the interventions delivered to them have to be input. This is very time consuming. Also a big change that is needed is with the problem of patients not having a unique identifier". As a solution to many of these issues, a number of projects were rolled out to improve services in Ireland and allow files to be viewed in more than one hospital. Interviewee 6 confirms that the project "was rolled out nationally with input locally". However, they caution that some form of "follow-up should have happened as staff are not using all the features of the system" (Interviewee 6). Targeted training and proper scoping of projects was identified as potential solutions by a number of interviewees. The findings indicate within various departments in the hospital, change is a forward planning process, which is well documented and audited through various stages. Change required a cultural commitment from the organisation as a whole to accommodate a new set of procedures, one of which is the use of auditing.

Stemming from a discussion on change, Interviewee 1 explained that change processes should be linked back to the concept of 'the Iron Triangle'. They explained that the Iron Triangle describes the relationship between cost, quality, and access within the hospital's department. The basic premise here is that a change (positive or negative) in one aspect of the triangle has a direct impact on the remaining two areas (Kissick 1994). Thus, while competing with each other, finding a balance and identifying what specific areas the department can trade-off becomes a key factor for change management teams. In addition, the reverse is also true – while improving one aspect of the Iron Triangle, change can also have a positive impact on the remaining two areas.

For the purpose of this research, we focus on the quality aspects associated with implementing change. The specific quality, safety and risk management software used has different sections for various quality documents on best practice. Interviewee 1 suggests that the documents should also link to audits to guide the change process. In addition, risk assessments are also conducted to provide a proactive management approach to assess issues, which may provide future challenges. All of these efforts support the hospitals quality improvement plan to identify what implementations are required and record incidences.

## 4.2 Attitudes towards New IT Systems and Processes

The interviewees reported mixed views with the introduction of new IT systems and processes. While some seemed relatively pleased with the new systems, others report disappointment with the overall change and the manner in which the change process occurred. Specifically, we revisit the Iron Triangle to highlight how Access can improve Quality, which is highlighted by Interviewee 4: "overall it is an improvement as images can be view from multiple locations".

Interviewee 4 explains that "involvement of staff is crucial for buy-in" which suggests that change management is a much wider collaborative effort within a department. Interviewee 5 highlights this and explains that the implementation of some new IT systems represents "silo thinking as lack of understanding of standards, networking, eco-system and health informatics". In addition, to accommodate a smooth change transition, training on a new system is vital. Interviewee 7 also shares similar concerns and highlights that "buy-in crucial to generate enthusiasm" about a change in service systems. In addition, they suggest "training should be relevant and timely" which may hamper user acceptance of IT-enabled innovation. Interviewee 10 also concurs "getting buy-in from stakeholders was crucial and management had to communicate well to do this. Without buy-in there is no engagement. Open meetings are useful".

We learn that with some projects there were "too long a time delay from training to using the system" (Interviewee 4) which can hamper the initial success of an IT change management programme. Interviewee 6 shares similar concerns regarding training and suggests, "more frequent staff sessions needed. Overall staff felt that training was not sufficient and more difficult for older people. Staged training sessions would have helped such as introduction, advanced, super user training". While some projects provide standard operating procedures (SOP) the inclusion of other software companies for supporting services may cause concerns for some users, for example, subcontracting support services (Interviewee 6).

Our findings also suggest that communication regarding the objective of implementing change is critical. For example, Interviewee 7 raises the question: "What are the objectives?" and goes on to

explain, “there is no point in implementing centrally and then letting people do what they want locally. You might as well have two systems”. Interviewee 10 states “communication was good with staff and team but could have been a lot better with the general public”. Interviewee 8 also highlights the importance of communication “very importantly for bringing in change that communication and team work essential”. This suggests that implementing change requires improved planning and communication strategies. This led us to consider whether change management requires a specific approach or whether it is a product of change leadership, which we examine further in the next section.

## 4.3 Lessons Learned

Interviewees were provided with the opportunity to explain what they might do differently if they were to undertake a similar change management task. Interviewee 1 explained that they would like to have more control of the chosen software vendors and suggest that not all users were happy with the software. Interviewee 2 raised more concerns with the overall change process. For example, interviewee 2 had concerns around the need to rebuild a service network, the need to fill out medical records (time-consuming) and the threat of personnel moving department or institution and in so doing, bring much needed competence out of the department. Therefore, more engagement of all parties and external expertise is a critical element of success in change management. Interviewee 5 explains that they could have “engaged with research centre…to get more visibility”. Building on this comment, the interviewee suggests that it should be a national competence approach to similar projects and explains, “we need a centre such as a medical software institute with wide stakeholder representation to oversee projects”.

Interviewee 2 highlights the usefulness of using a change model such a Kotter’s and indicates that the eights steps is “what should be done…but plans can change due to unexpected problems”. This suggests that there may be a need to offer greater flexibility or agility to change management models such as Kotter’s. Interviewee 5 also suggests that models such as Kotter offer a good basis to manage change. For example, interviewee 5 further explains, “for our system we were mobile and patient centric. We understood the people and their motivations. That is a platform for engagement and multi-disciplinary teams”. Adopting an improved structured approach was discussed by Interviewee 6 discusses this and suggest that a “well-structured maintenance service agreements especially out of hours service for example the previous system came from [Global Tech Company] and they had a person onsite to deal with issues”. Interviewee 7 suggests that the success in implementing change may be in the ability to understand user’s requirements and foster a relationship to ensure buy-in at the beginning of the project: “to implement change you have to talk to the end user and get buy-in. Start with what you want and work back. Successful projects always had buy-in”.

However, to facilitate an improved structured process, Interviewee 2 indicated the need to “encourage more trust” and avail of additional onsite support for the technology providers. One of the issues associated with the lack of support was the different time zones (i.e. Ireland and the USA) requiring out of office phone calls for long durations. The level of support provided was often unsatisfactory, for example, “they sometimes say the problem is our network when the network is working” (Interviewee 2). Interviewee 4 also shared these concerns and explained that if they underwent a similar project they would have “someone on hand instead of having to ring California with issues”. This would make a big difference.” The need for improved planning and greater stakeholder involvement was discussed. For example, Interviewee 9 discusses a failed project and suggests “it was not scoped well and users were not involved enough”. Interviewee 11 also acknowledges planning and suggests, “with any project there should be time given to planning the project timelines”. In addition, considering that one of the core objectives was to streamline healthcare processes, Interviewee 4 highlights their disappointment in that the project in question, “it is supposed to be paperless but it is not. Actually we are using more paper and ink now.” Interviewee 6 explains that going live presented some unexpected issues: “the initial go live took longer and also the bedding in period took longer than expected and more patient lists should have been cancelled beforehand. So what happened were lots of people waiting two weeks so it was not patient load effective at the beginning”. Interviewee 6 goes on to explain that planning and vision are often problematic: “we plan things and it takes so long that by the time it’s implemented the projects are too old.” Interviewee 6 also highlights some general issues associated with change management in the public sector such as:

- Not enough long-term strategic planning;
- Many projects are abandoned;
- Need be more proactive rather than reactive;
- Need to avail of informed expert opinion on change management.

Interviewee 7 shares similar concerns and suggests, "better long-term and short-term planning is needed". Thus, there is a clear indication that implementing change requires a structured approach, which communicates both the need and benefits of supporting change.

# 5 DISCUSSION

This study demonstrates that the need to manage such change is widely recognized. The interviewees confirm that management need to lead change. Reviewing Kotter's change model and eight phases of change, we learn that not all eight phases were necessary to successfully implement change in the hospital system. We highlight these as Strong Evidence, Some Evidence and No Evidence as detailed in Table 2. It also outlines the level of evidence of Kotter's change model using the eight phases, which we identified within our case study. Kotter's Step 4 'Communicate the Vision' stipulates that communication of the vision should come from senior management.

Therefore, staff were aware of relevant tasks to be completed in the project and of their roles within the project. This was not identified by any of the interviewees as a necessity, yet the hospital happened to successfully implement change and raises many questions as to how it could be improved and what key factors were in play from an organisational change perspective.

Table 2: Evidence of Kotter's 8 Phases.

| Kotter's Eight Phases | Evidence |
|---|---|
| 1. Establish a Sense of Urgency | |
| 2. Form a Powerful Guiding Coalition | |
| 3. Create a Vision | |
| 4. Communicate the Vision | |
| 5. Empower Others to Act on the Vision | |
| 6. Plan for and Create Short-Term Wins | |
| 7. Consolidate Improvements and Produce Still More Change | |
| 8. Institutionalise new approaches | |

The following steps were strongly identified by interviewees are being necessary during the implementation process:

- **Step 1: Urgency**. Hayes and Richardson (2008) state that, the need for such a change must be communicated to everyone in the organisation at the outset. This was confirmed by the interviewees, as there was an inherent imperative requirement for change to the current system in place.
- **Step 6: Plan**. Change should have clear goals and objectives and take place in small steps. The interviewees stated that there were clearly defined goals and that the objectives were all agreed on to be rolled out nationally.
- **Step 8: Institutionalise**. The interviewees remarked that the new approach is now part of normal way of working and is "bedded in well".

The following steps were identified by interviewees as being necessary during the implementation process but would require a greater presence throughout the change process:

- **Step 2: Coalition**. Kotter (2005) recommends progressively involving different members of the organisation in the change to form a project team. This was seen to be the case in one such project within the hospital, which was ultimately successful. Coalition was necessary as it involved numerous team members in various locations.

**Step 3: Vision**. Kotter (2005) recommends that a clear vision and plan for implementing change is required.

While **Step 5: Empower Others to Act on the Vision** was not obvious from our interviews, Kotter (2005) recommends that obstacles, such as organisational structure should be removed. The interviewees confirmed this as a requirement. For example, while the interviewees mentioned the various obstacles they would like to remove they were not empowered to instigate change to act on the vision.

Overall our findings suggest that there is a clear need to introduce a new model to support the implementation of change in a healthcare context. While Kotter's Steps 2, 3 and 7 were only partially implemented in successful projects the aims of these steps were achieved while carrying out other steps.

Table 3: HIT-CMM: Quality and Access.

| Health IT Change Factors | System Structure | | Process of Care | | Healthcare Outcome | |
|---|---|---|---|---|---|---|
| | Quality | Access | Quality | Access | Quality | Access |
| Establish a Sense of Urgency | Why does the quality of the current IT healthcare system prompt the need for change? | Will a sense of urgency of the structure create improved access to the IT healthcare system? | How will a sense of urgency improve the quality in the provision of care via the IT healthcare system? | How does the sense of urgency highlight the need for improved access to care via the IT healthcare system? | Can the sense of urgency lead to improved quality in healthcare outcomes using the IT healthcare system? | Will a sense of urgency in changing the IT healthcare system support the need for improved access to healthcare outcomes? |
| Form a Powerful Guiding Coalition | Can a coalition improve the quality by altering the structure of an IT healthcare system? | How does a coalition structure improve the accessibility of services the IT healthcare system? | How can a coalition improve the quality of care using the IT healthcare system? | How does a coalition improve access to the IT healthcare system improve the provision of care? | How does a coalition contribute towards the quality of healthcare outcomes though the IT system? | Can a coalition for an IT healthcare system lead to improve access to improved healthcare outcomes? |
| Create a Vision | How can a vision of an IT healthcare system address quality concerns within a service structure? | How does the IT healthcare system vision include grater access to the system structure? | How is the quality of the IT healthcare system captured in the vision? | Does the vision of the IT healthcare system propose greater access to healthcare service? | How does the vision for the IT healthcare system propose to improve the quality of healthcare outcomes? | Are there any specific factors described in the of IT healthcare system vision which proposes to provide greater access to improved healthcare outcomes? |
| Communicate the Vision | What processes are in place to communicate the quality factors of the IT healthcare system vision? | How does the vision inform stakeholders of for improved access to the IT healthcare system? | How is the quality of care suggested to improve through the IT healthcare system? | Is access to the IT healthcare system a key part of the vision statement? | How does the IT healthcare system vision propose to improve the quality of healthcare outcomes? | Does the IT healthcare system vision propose to improve access to improved healthcare outcomes? |
| Empower Others to Act on the Vision | What roles are assigned to stakeholders to act on the IT healthcare system vision to improve the service quality? | How are stakeholders empowered to have greater access to the IT healthcare system? | What processes are described in the IT healthcare system vision to empower stakeholders to improve the quality of care? | How does the vision empower stakeholders to have greater access to provide care using the IT healthcare system? | What roles are assigned to stakeholders to improve healthcare quality outcomes through the IT healthcare system vision? | How are stakeholders empowered to have greater access to the IT healthcare system in order to improve healthcare outcomes? |
| Plan for and Create Short-Term Wins | What short-term quality gains are defined for the change in the IT healthcare system structure? | What short-term access gains are planned for the change in the IT healthcare system structure? | What short-term quality gains are planned to be achieved using the IT healthcare system? | What short-term gains are planned for the provision of care via the IT healthcare system? | What short-term healthcare outcome quality gains are planned for using the IT healthcare system structure? | What short-term access gains can achieve improved healthcare outcomes using the IT healthcare system structure? |
| Consolidate Improvements and Produce Still More Change | What benefits to quality could be achieved through consolidated improvements in the IT healthcare system structure? | What improved levels of access could be achieved through consolidated improvements in the IT healthcare system structure? | What benefits to the provision of care could be achieved through consolidated improvements in the IT healthcare system? | How could improved access improve the provision of care using consolidated improvements in the IT healthcare system? | What healthcare quality outcomes could be achieved through consolidated improvements in the IT healthcare system? | What improved levels of access to healthcare outcomes could be achieved through consolidated improvements in the IT healthcare system? |
| Institutionalise new approaches | Are specific IT healthcare system quality metrics defined in the healthcare organisational strategy? | How is improved access to the IT healthcare system embedded into the organisational and IT strategy? | How does the organisation propose to improve the quality associated with the provision of care using the IT healthcare system? | How does the organisation propose to improve the access to care using the IT healthcare system? | Has the organisation established specific healthcare quality outcomes for the IT healthcare system? | How is it proposed that improved access to the IT healthcare system would improve the organisations overall healthcare outcomes? |

## 5.1 HIT-CMM

To develop a change model we identified an approach by O'Leary et al (2015) and Carroll et al (2016) which examines primary stakeholders to address their assessment needs from a multi-perspective viewpoint. We adopted a similar methodology to influence the development of HIT-CMM: Quality and Access (Table 3). Cost will be included in the next iteration of the model. The HIT-CMM acknowledges that change is multidimensional and occurs through a series of key management stages, combining Kotter's eight steps, which require assessment as per the Iron Triangle at various stages of the change management lifecycle. The questions presented throughout Table 3 are influenced case study data and constructed to support the hospital IT change strategy at various stages of the change process. We also found that some aspects of Kotter's change model is useful to successfully manage change but there are some shortcomings within a healthcare context. For example Kotter's step 5 Empower Others to Act on the Vision was seen as unnecessary within the medical device company while in the hospital it was not obvious from our interviews. Within each of the phases we assign the relevant Kotter steps to support change management along with steps identified in this case study such as Senior Management as supporters and staff buy-in. Communication of the vision was already identified as lacking in this case study, if the HIT-CMM were then used the assessment of this step should be in terms of cost, quality, accesses, structure, process and outcome.

# 6 FUTURE RESEARCH

It is planned to further develop the HIT-CMM and use it to guide change. This model would build on

the specific needs identified such as longer term strategic planning and more flexibility to manage unexpected issues. In particular, we will include Cost as the third element of the Iron Triangle.

The HIT-CMM will be incorporated into a more detailed strategy model, which also examines the process of innovation in healthcare. Specifically the HIT-CMM has already supported us to uncover key factors for a Healthcare Innovation Strategy and how we could begin to explore innovation opportunities. Given the small sample size a more complete picture will be facilitated by interviewing a larger number of participants.

# 7 CONCLUSIONS

This study demonstrated that the need to manage change is widely recognized. Different perspectives, methods and approaches (and the underlying theories that drive them) that are aligned cannot guarantee to deliver the required change in the time and on the scale necessary. Reviewing Kotter's change model and eight phases of change, we learn that not all eight phases are necessary to successfully implement change. Therefore a more tailored yet detailed framework was required. We present a mode suitable model to manage healthcare IT change through the introduction of our HIT-CMM.

# ACKNOWLEDGEMENTS

We would like to thank all the participating interviewees for their time and efforts.

This research is partially supported by Science Foundation Ireland (SFI) grant no 10/CE/I1855 to Lero (http://www.lero.ie), by Enterprise Ireland and the IDA through ARCH – Applied Research in Connected Health Technology Centre (www.arch.ie), BioInnovate and by Science Foundation Ireland (SFI) Industry Fellowship Grant Number 14/IF/2530.

# REFERENCES

Benjamin, R. I., Levinson, E., 1993. A framework for managing IT-enabled change. In *Sloan Management Review*, 34(4), 23-33.

Carroll, N., Travers, M. and Richardson, I., 2016. Evaluating Multiple Perspectives of a Connected Health Ecosystem. In *9th International Conference on Health Informatics (HEALTHINF)*, Rome, Italy, February.

Cazzaniga, S., Fischer, S., 2015. How ICH Uses Organizational Innovations to Meet Challenges in Healthcare Management: A Hospital Case Study. In Challenges and Opportunities in Health Care Management, Springer International Publishing, 355-361.

Forte, G., 1997. Managing Change for Rapid Development, In *IEEE Software* 14(6), 114–123.

Grbich, C., 1999. Qualitative Research in Health: An Introduction, Sage Publications. California.

Hayes, S., Richardson, I., 2008. Scrum Implementation using Kotter's Change Model, 9th International Conference on Agile Processes and eXtreme Programming in Software Engineering, Limerick, Ireland, Lecture Notes in Business Information Processing 2008, vol 9, Part 6, 10th-14th June, 161-171.

Hiatt, J. M., 2006. ADKAR: a model for change in business, government and our community, Prosci Learning Center. Loveland.

Kissick, W., 1994. Medicine's Dilemmas: Infinite Needs versus Finite Resources, Yale University Press. New Haven.

Kotter, J., 2005. Leading Change: Why Transformation Efforts Fail, Harvard Business School Press. Boston.

Lewin, K., 1947. Frontiers in Group Dynamics: Concept, Method and Reality in Social Science; In *Social Equilibria and Social Change. Human Relations*, June, 1(36).

O'Leary, P., Carroll, N., Clarke, P. and Richardson, I., 2015. Untangling the Complexity of Connected Health Evaluations. In *International Conference on Healthcare Informatics (ICHI)*, October, 272-281. IEEE.

Travers, M., Richardson, I., 2015. Medical Device Software Process Improvement – A Perspective from a Medical Device company, in *8th International Conference on Health Informatics, (Healthinf)*, Lisbon, Portugal, January.

Yin, R. K., 2013. Case study research: Design and methods, Sage publications.

# A Comparison of Statistical Linkage Keys with Bloom Filter-based Encryptions for Privacy-preserving Record Linkage using Real-world Mammography Data

Rainer Schnell[1], Anke Richter[2] and Christian Borgs[3]
[1]*City, University of London, Northampton Square, London EC1V 0HB, U.K.*
[2]*Institute for Cancer Epidemiology, Ratzeburger Allee 160, 23538 Lübeck, Germany*
[3]*University of Duisburg-Essen, German Record Linkage Center, Lotharstr. 65, 47057 Duisburg, Germany*
*rainer.schnell@city.ac.uk, anke.richter@uksh.de, christian.borgs@uni-due.de*

Keywords: Medical Record Linkage, Patient Identification Codes, Pseudonyms.

Abstract: New EU regulations on the need to encrypt personal identifiers for linking data will increase the importance of *Privacy-Preserving Record Linkage* (PPRL) techniques over the course of the next years. Currently, the use of *Anonymous Linkage Codes* (ALCs) is the standard procedure for PPRL of medical databases. Recently, Bloom filter-based encodings of pseudo-identifiers such as names have received increasing attention for PPRL tasks. In contrast to most previous research in PPRL, which is based on simulated data, we compare the performance of ALCs and Bloom filter-based linkage keys using real data from a large regional breast cancer screening program. This large regional mammography data base contains nearly 200.000 records. We compare precision and recall for linking the data set existing at point $t_0$ with new incident cases occuring after $t_0$ using different encoding and matching strategies for the personal identifiers. Enhancing ALCs with an additional identifier (place of birth) yields better recall than standard ALCs. Using the same information for Bloom filters with recommended parameter settings exceeds ALCs in recall, while preserving precision.

## 1 INTRODUCTION

Many medical studies link different databases containing information on the same patient (Jutte et al., 2010). If unique common identifiers are available, linking is trivial. However, in many situations in practice such unique identification numbers are not available. If privacy is not an issue, probabilistic record linkage based on pseudo-identifiers such as surname, first name, date of birth and address information can be used (Herzog et al., 2010). Under legal constraints demanding privacy for pseudo-identifiers, privacy-preserving record linkage (PPRL, for an overview see (Vatsalan et al., 2013)) is required.

In general, jurisdictions for linking patient data differ widely. Therefore, the technical details to comply with national legal requirements vary between countries. In the US, the HIPAA rules require the removal of nearly all information used for record linkage. The current legal situation in Europe has made pseudomysation of record linkage identifiers factually mandatory: Due to increasing privacy concerns of the population, the European Council, Parliament and Commission agreed on a new "General Data Protection Regulation" (Council of European Union, 2016), which will be part of the national jurisdictions in all 28 member states of the European Union by May 2018. The regulation clearly demands pseudonymisation techniques able to withstand re-identification attacks, but does not require absolute anonymization.Given this recent development, the demand for PPRL solutions will increase sharply.

Currently, due to the regional and organisational fragmentation of medical health care, the standard setting for medical record linkage is based on a one-time-exchange between otherwise computationally separated organizational units. This constraint restricts the number of potential PPRL solutions to a small subset of the many different PPRL approaches which have been suggested (for a review, see (Vatsalan et al., 2013)). Nearly all applied PPRL protocols use three types of actors: Two or more data holders, one linkage unit and a research group. In general, in such settings, all units interact only once. Most protocols assume that all partners act according to the protocol (but may keep track of all local computations). This assumption is called 'honest, but curious' or 'semi-honest' model (Goldreich, 2004).



Schnell, R., Richter, A. and Borgs, C.
A Comparison of Statistical Linkage Keys with Bloom Filter-based Encryptions for Privacy-preserving Record Linkage using Real-world Mammography Data.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 276-283
ISBN: 978-989-758-213-4

For such scenarios, only three approaches for linking medical data have been used repeatedly for real-word applications of large medical databases (Schnell, 2015): Using a third-party trustee, using encrypted identifiers[1] and using Bloom-filters.

If a third-party data trustee (Kelman et al., 2002) is used, unencrypted patient pseudo-identifiers are transferred to a trusted third party, which links the pseudo-identifiers and assigns a new identification number to the linked records. These newly constructed IDs are then used for linkage by a research group.

By far the most common approach in practical settings is the use of encrypted pseudo-identifiers. Here, the identifiers are concatenated into a single string which is then encrypted. The resulting encrypted string is called an anonymous linking code (ALC, (Herzog et al., 2007)).

Many of the more recent PPRL approaches (see (Vatsalan et al., 2013; Karapiperis et al., 2016) for reviews) have limited scalability, so they can not be used with large datasets. For example, although technically interesting, all homomorphic encryption methods are computationally expensive and do not scale well (Karakasidis et al., 2015). An exception are Bloom filter approaches. (Schnell et al., 2009) first suggested the use of Bloom filters for privacy-preserving record linkage. The approach is based on splitting each identifier into a set of substrings of length 2 (*bigrams*), which are mapped into a binary vector for each identifier with a linear combination of different cryptographic hash functions such as SHA-1 and MD-5. The similarity of these binary vectors (*Bloom filters*) approximates the similarity of the pseudo-identifiers, which makes Bloom filters attractive for error-tolerant PPRL.

Although using separate Bloom filters for each pseudo-identifier is the most common approach, the use of one *common* binary vector is harder to attack. The use of a single Bloom filter for all identifiers has been first proposed in (Schnell et al., 2011) and has been explored further by (Durham, 2012). The resulting *composite Bloom filter* is called a *Cryptographic Long-term Key* (CLK) in the original publication or 'record based Bloom filter' (RBF) by (Durham, 2012). CLKs have been used on real world data extensively (Randall et al., 2014; Schnell and Borgs, 2015; Schmidlin et al., 2015).

[1]Although data sets without direct personal identifiers, but containing indirect identifying information such as date of hospital admission and discharge are occasionally suggested (Karmel and Gibson, 2007) for record linkage, they rarely contain enough discriminating information for unique linkage pairs.

**Our Contribution.** No previous publication compared the performance of CLKs with the performance of the more traditional ALCs using real-world data. Therefore, we report on a new study assessing the performance of different variations of CLKs and ALC variants using real-world data from a large regional breast cancer screening program (Katalinic et al., 2007). Furthermore, for the first time, we compare the effect of including additional identifiers to linkage keys and Bloom filter encodings.

# 2 PREVIOUS WORK

Currently, only two different versions of encoding identifiers seem to be in practical use for PPRL: Anonymous Linkage Codes (ALCs) and Bloom filters. Both will be described shortly.

## 2.1 ALC Variants

ALCs are an encrypted single string formed by concatenating substrings or functions of different pseudo-identifiers. These pseudo-identifiers should be stable over time and free of errors. Most often, first name, surname, date of birth and sex are used for constructing ALCs. The resulting combination of identifiers is encrypted using cryptographic hash functions. The resulting hashed string is used as the linkage key. If two ALCs match exactly, the corresponding records are classified as representations of the same real-world entity. Due to the cryptographic hash function, it is nearly impossible to decrypt the identifiers directly.

The most simple and widely-used ALC is constructed in three steps (Herzog et al., 2007): all identifiers are preprocessed using a set of rules (for example, removal of non-alphabetical characters from names, removal of non-digits from dates, and capitalization of all characters). The resulting preprocessed identifiers are then concatenated to form one single string, which is finally encrypted with a cryptographic hash function. Examples of applications of *Basic ALCs* are described by (Kijsanayotin et al., 2007; Schülter et al., 2007; Johnson et al., 2010; Tessmer et al., 2011).

The design of the Basic ALC is not error-tolerant, since even the replacement of a single letter will result in an entirely different hash code. As spelling and typographical errors in patient identifiers are common, many true record pairs will not be classified as matches. Hence patients with variations in their respective identifiers might have different characteristics than patients with agreeing identifiers. Ignoring this problem can result in biased estimates (Ridder and Moffitt, 2007).

Different approaches to constructing ALCs allow for some errors in identifiers. The Swiss Federal Office for Statistics asked the Cryptological Unit of the Swiss Military to develop a privacy-preserving linkage method for medical patient data (Office fédéral de la statistique, 1997).To construct this ALC variation, the Soundex code of surname and first name are created after some preprocessing. The Soundex codes are concatenated with the date of birth and sex. The resulting string is encrypted using a cryptographic hash function (Office fédéral de la statistique, 1997). Applications and reviews of the Swiss ALC are discussed in (Borst et al., 2001; Holly et al., 2005; Eggli et al., 2006; El Kalam et al., 2011).

Another approach to construct more error-tolerant ALCs was invented by the Australian Institute of Health and Welfare (AIHW). Their solution uses substrings of first and last names instead of the full string. (Ryan et al., 1999) tested several variations and concluded that the second, third, and fifth character of the surname combined with the characters at the second and third position of the first name concatenated with sex and date of birth performed best. The resulting string forms the *Statistical Linkage Key* (SLK) which is often included in data published by the AIHW (Karmel et al., 2010). After applying a cryptographic hash function to the SLK, the *Encrypted SLK*, sometimes also denoted as 581-Key is the ALC variant that is widely used in Australian data linkage (Taylor et al., 2014). (Karmel et al., 2010) tested the effect of adding different versions of state and postcode to the 581-Keys. In general, 581-Keys don't seem to be considered as state-of-the-art any longer (Randall et al., 2016).

## 2.2 Simple Bloom Filters

Bloom filters have been used for calculating string similarities in privacy-preserving probabilistic record linkage (Schnell et al., 2009). A Bloom filter is an array of data proposed by Howard (Bloom, 1970) for checking the set membership of records efficiently (Broder and Mitzenmacher, 2003). It is represented by a bit array with a length of $l$ bits initially set to zero. For the mapping, $k$ independent hash functions $h \in \{h_1, \ldots, h_k\}$ are used.To store the set of entities $S = \{x_1, x_2, \ldots, x_n\}$ in the Bloom filter, each element $x_i \in S$ is hashed using the $k$ independent hash functions. The bit positions given by the hash functions are set to 1. If a bit was already set to 1, nothing is changed.

To store all elements of a set in Bloom filters, we apply the double hashing scheme proposed by (Kirsch and Mitzenmacher, 2006). They show that using two independent hash functions is sufficient to implement a Bloom filter with $k$ hash functions without an increase in the asymptotic false positive probability (Kirsch and Mitzenmacher, 2006). Therefore, the positional values of the $k$ hash functions are computed with the function

$$g_i(x) = (h_1(x) + i \cdot h_2(x)) \ mod \ l \qquad (1)$$

where $i \in \{0, \ldots, k-1\}$ and $l$ is the length of the bit array. We use two different keyed hash message authentication codes (HMACs), namely, HMAC-SHA1 ($h_1$) and HMAC-MD5 ($h_2$) (Krawczyk et al., 1997) to create the Bloom filters.

## 2.3 Composite Bloom Filters

For some applications, a single linkage key has to be used. If separate Bloom filters are used, for these applications, the set of Bloom filters has to be combined in a composite Bloom filter. Storing all of the identifiers used in a single Bloom filter was first proposed by (Schnell et al., 2011). This is called a *Cryptographic Long-term Key* (CLK), since they were intended for use in a longitudinal study of offenders.

For the construction of a CLK, each identifier is split into a set of $n$-grams. Each set is stored using $k$ hash functions using the same Bloom filter of the length $l$ for all $n$-gram sets of all identifiers used. This additive Bloom filter represents the CLK.

After preprocessing, first name and surname are split into bigrams, birth year into unigrams. In the second step, the first $n$-gram set (e.g. first name) is stored in the Bloom filter. Each bigram is hashed $k$ times. Bits having indices corresponding to the hash values are set to one. In the third step, the second $n$-gram set (e.g. surname) is mapped to the same Bloom filter. Finally, unigrams are mapped to the same bit array.

## 2.4 Cryptographic Attacks on ALCs

Frequency attacks on standard ALCs have not been reported in the literature so far. Discussions about the security of ALCs and 581-Keys up to now are hypothetical, not empirical (Randall et al., 2016).

However, since the same password is used for all records, within a combination of sex and date of birth, the most frequent name/surname combination will also yield the most frequent ALC. Therefore, given a large random sample, the most frequent name/surname combinations have a high risk of re-identification. Under the (unrealistic) assumption of uniformly distributed dates of birth, age and sexes, there are about $365 * 100 * 2 = 73.000$ combinations possible. This way, in a database of 10.000.000 records, about 137 records per combination are expected. If the frequency

distribution of names is skewed, aligning the most frequent name subsets could identify a large proportion of the records using this simple frequency alignment.

### 2.5 Cryptographic Attacks on Bloom Filters

Bloom filter-based PPRL has been attacked by two different techniques: by applying a Constrained Satisfaction Solver (CSS) on frequencies of entire Bloom filters (Kuzu et al., 2011; Kuzu et al., 2013) and by a interpreting the Bloom filter bit patterns as a substitution cipher (Niedermeyer et al., 2014).

The first attack is a variant of a simple rank swapping attack (Domingo-Ferrer and Muralidhar, 2016) which used the estimated length of the encrypted strings as additional information. (Kuzu et al., 2011) consider their attack on separate Bloom filters as successful, but not their attack on composite Bloom filters (Kuzu et al., 2013). It should be noted that this CSS attack is based on the entire data set of Bloom filters, therefore it is no decoding, but an alignment. This way of attack is impossible if many groups of similar cases generates a new bit pattern, for example by using salted encodings (Niedermeyer et al., 2014). In a salted encoding, a stable identifier such as date of birth, year of birth or place of birth is added to the password determining the hash functions.

The second attack attempted the actual revealing of all identifiers as clear text by a cryptanalysis of individual bit patterns within the Bloom filters (Niedermeyer et al., 2014). This attack is based on the limited number of bit patterns generated by the linear combination of two hash functions in the double-hashing scheme (Kirsch and Mitzenmacher, 2006) of the initial proposal. Exploiting this specific construction of the hash functions, (Niedermeyer et al., 2014) were successful with basic Bloom filters and (Kroll and Steinmetzer, 2015) with CLKs/composite Bloom filters. Therefore, replacing the double-hashing scheme by random hashing should prevent the success of this attack on Bloom filters (Niedermeyer et al., 2014). Random hashing is based on the idea of using bigrams as seeds for random number streams. This could be implemented by a linear-congruential pseudo-random number generator (LCG, (Stallings, 2014)), to generate a sequence $X$ with the length $k$ for each $n$-gram.Random hashing increases the number of possible bit patterns ($l = 1000,\ k = 15$) for a given $n$-gram from less than $10^6$ to more than $6.8 \cdot 10^{32}$. Therefore, the Niedermeyer-attack should fail for randomly hashed Bloom filters. This theoretical expectation has been empirically verified by (Schnell and Borgs, 2016).

In conclusion, for salted Bloom filter encodings using random hashing, no successful attack method is known. Of course, the number of records using the same salt should not exceed the minimum required for a frequency attack either on the whole pattern or the individual attributes mapped to the Bloom filter. Based on experiments reported by (Schnell and Borgs, 2016), this minimum number seems to be about 300 records. In most medical applications, this number is only exceeded in national databases. For this, an additional salt has to be used. Given this condition, we consider Bloom filter-based encodings as meeting the requirements of the EU Protection Regulation (Council of European Union, 2016) for a pseudonymisation method.

## 3 METHODS

Using real data from a German state-wide breast cancer screening program (Katalinic et al., 2007), we compared the CLK encryption with the Basic and Swiss ALCs and the encrypted SLK (*581-Key*).

The test data consists of mammography records of patients in a German state, covering about 3.4% of the total German population. File A consists of cases until the end of 2011 (with one record for each case) with $n = 138.131$ records, file B encompasses cases after 2011 (more than one record per case was possible) with $n = 73.004$ cases in 198.475 records.

The standard CLK is set up with a length of $l = 1000$. First name and Surname were padded with spaces before being split into bigrams (Robertson and Willett, 1998). The other identifiers were split into unigrams. Each set of $n$-grams is hashed using $k = 10$ HMACs (*Hash functions*) and a different cryptographic key. Since CLKs allow for matching strategies other than exact matching (Schnell et al., 2011), following (Schnell, 2014), Multibit Trees with various Tanimoto-thresholds were used. The statistical linkage keys were evaluated using exact matching.

The set of identifiers used consisted of *first name*, *surname*, *date of birth* and *sex*. According to recent studies, including more *stable* identifiers is desirable (Schnell and Borgs, 2015). Address information is very volatile, since places of residence may change during the course of a lifetime. Therefore, (Schnell and Borgs, 2015) suggested using places of birth as an additional identifier for Bloom filter-based PPRL. In the second experiment, we did this by adding *place of birth* to the set of identifiers for the CLKs and 581-Keys.

Since the real-world data sets used here contained only current places of residence, we simulated the place of birth according to German administrative population counts. We introduced artificial 10% address

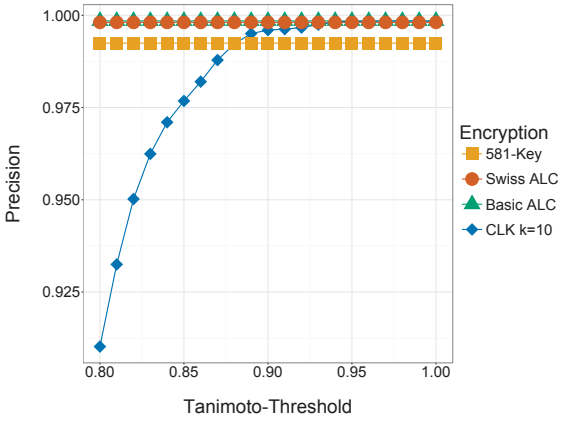



Figure 1: Precision of the CLK and encrypted statistical linkage key variants. Since the ALCs were matched exactly, their values are shown as constants, while several similarity thresholds were used for CLKs.

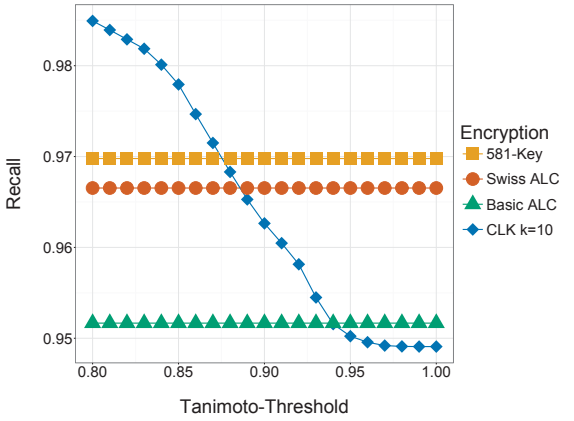



Figure 2: Recall of the CLK and encrypted statistical linkage key variants. Since the ALCs were matched exactly, their values are shown as constants, while several similarity thresholds were used for CLKs.

changes to the simulated data. As the two linked files refer to different years, this percentage should reflect a worst-case scenario for the amount of regional mobility in the population.

The current gold standard in use at the cancer screening program is considered as reflecting the true matching status. Based on this classification, the compared methods will yield true positive (TP), false positive (FP), true negative (TN) and false negative (FN) classifications of record pairs.

This way, we can compare the methods using precision (Precision $= \frac{TP}{TP+FP}$) and recall (Recall $= \frac{TP}{TP+FN}$) (Baeza-Yates and Ribeiro-Neto, 1999).

According to legal requirements, unencrypted identifiers were processed only at the office of the data holder. ALCs, 581-Key and CLKs were generated with Python 3, while R (R Core Team, 2016) was used for the matching and statistical computation.

## 4 RESULTS

Figures 1 and 2 show the results of the standard CLK ($k = 10$ hash functions) against the encrypted linkage keys in terms of precision and recall. Lowering the threshold improves the recall. Precision is stable until the threshold approaches 0.88. Above this threshold, precision drops considerably. Given this set of identifiers, CLK does not exceed the performance of the Swiss ALC and the 581-Key.

All in all, ALCs offer higher precision (less false positives) compared to the CLK. However, the CLK outperforms the ALCs in terms of recall as the similarity threshold is lowered below 0.88. At the recommended Tanimoto-threshold of 0.85 (Schnell, 2015),

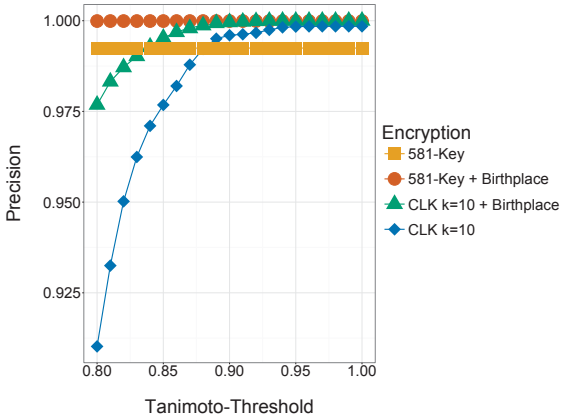



Figure 3: Precision of the 581-Key and CLK with and without inclusion of places of birth.

CLKs show more (0.7% – 2.8%) true positives than both standard ALC variants (see table 1), even outperforming the 581-Key. However, given this set of identifiers, the amount of false positives is considerably higher. Since CLKs should perform better if more (stable) identifiers are included. Therefore, for the second set of experiments, we included place of birth and hashed it into the original CLKs. We did the same with the 581-Key, concatenating place of birth to the 581-Key before hashing it again. Figures **??** and 4 show that the performance now exceeds the standard ALCs in terms of recall while showing improved precision values.

Table 1 lists the detailed classifications in terms of true (TP) and false positive (FP) record pairs, as well as missed record pairs (false negatives (FN)) along with recall and precision at a Tanimoto-threshold of 0.85 for all ALCs, the 581-Key and the CLKs. Details on the results for adding the simulated place of birth

Table 1: Classification results for all methods presented. CLK results are based on a Tanimoto-threshold of 0.85 using Multibit Trees.

| Variant | TP | FP | FN | Prec. | Rec. |
|---|---|---|---|---|---|
| Basic ALC | 51.587 | 79 | 2.620 | 0.998 | 0.952 |
| Swiss ALC | 52.454 | 101 | 1.816 | 0.998 | 0.967 |
| 581-Key | 52.633 | 400 | 1.640 | 0.992 | 0.970 |
| $CLK_{k10}$ | 53.012 | 1.260 | 1.196 | 0.977 | 0.978 |
| $581\text{-}Key_{+place\ of\ birth}$ | 51.945 | 5 | 2.328 | 0.999 | 0.957 |
| $CLK_{k10+place\ of\ birth}$ | 52.840 | 251 | 1.368 | 0.995 | 0.975 |

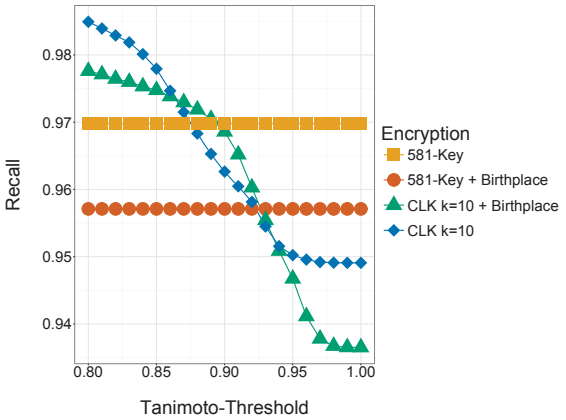



Figure 4: Recall of the 581-Key and CLK with and without inclusion of places of birth.

are shown as well. The CLKs consistently show more true positive classifications, while the ALCs and 581-Key perform better in terms of precision (fewer false positives).

It has to be noted that adding the place of birth to the set of identifiers improves the precision for both the 581-Key and the CLKs, while only decreasing recall marginally (likely due to the 10% errors simulated for the birth places). A CLK with birthplace information stored in it outperforms all standard ALC variants and the 581-Key without additional identifiers in both recall and precision.

Since the simulated birth places assumed a worst-case setting of 10% errors in the data, real-world applications using CLKs will benefit from including additional stable identifiers. These results show the potential of using Bloom filters for real-world privacy-preserving record linkage applications, especially if additional stable information is available.

## 5 DISCUSSION

In this paper, we showed a real-world application of the Cryptographic Long-term Key. Previously, ALCs were built by encrypting hashed or sampled identifiers. The CLK, representing an array of bits allows for similarity comparisons using Multibit Trees. The presented simulation results show better recall, but lower precision than best-performing ALCs. Since CLKs can be easily fine-tuned by selecting different thresholds, the impact of linkage errors on substantial results can be easily studied. Therefore, we consider the impact of increased false positives as not limiting the application of CLKs.

Precision and recall of CLKs will exceed ALCs and 581-Keys if more stable identifiers can be used. Recently, (Brown et al., 2016) showed that the optimal choice of identifiers and parameters is critical for the performance of Bloom filter-based PPRL. Their results vary, depending on the set of identifiers used. They also showed the need for *stable* identifiers, as errors and missing values (for example, in recent addresses) will reduce recall.

After fine-tuning parameters and identifier sets, PPRL linkage quality comparable to clear text linkage can be achieved with CLKs. Furthermore, using Multibit Trees as suggested by (Schnell, 2014), PPRL using CLKs can be done (without additional blocking) on standard hardware with two files containing 5 million records each in a little over 4 days (Brown et al., 2016). If additional blocks such as date of birth are used, linkage can be done in less than an hour (Schnell, 2015).

Bloom filters can be used to represent other data than strings: (Vatsalan and Christen, 2016) demonstrated the use of numerical and date information, (Farrow and Schnell, 2017) tested the inclusion of distance-preserving locational data. Both techniques will extend the number of possible applications for PPRL.

Currently, there is no known way of attacking CLKs and state-of-the-art variants of single Bloom filters (Schnell and Borgs, 2016). Therefore, they might be used to link files using personal identifiers according to the de-facto anonymity standard required by the new EU regulation on data protection.

## ACKNOWLEDGEMENTS

The study was supported by German Research Foundation (DFG) research grant SCHN 586/17-2 awarded to the first author. The first author would like to thank the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the programme 'Data Linkage and Anonymisation' (which was supported by the EPSRC grant Number EP/K032208/1) when work on this paper was undertaken.

## REFERENCES

Baeza-Yates, R. and Ribeiro-Neto, B. d. A. (1999). *Modern Information Retrieval*. Addison-Wesley, Harlow.

Bloom, B. H. (1970). Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13(7):422–426.

Borst, F., Allaert, F.-A., and Quantin, C. (2001). The Swiss solution for anonymous chaining patient files. In Patel, V., Rogers, R., and Haux, R., editors, *Proceedings of the 10th World Congress on Medical Informatics: 2–5 September 2001; London*, pages 1239–1241, Amsterdam. IOS Press.

Broder, A. and Mitzenmacher, M. (2003). Network applications of Bloom filters: a survey. *Internet Mathematics*, 1(4):485–509.

Brown, A., Borgs, C., Randall, S., and Schnell, R. (2016). High quality linkage using multibit trees for privacy-preserving blocking. International Population Data Linkage Conference (IPDLN2016): 24.08-26.08.2016; Swansea.

Council of European Union (2016). Council regulation (EU) no 679/2016.

Domingo-Ferrer, J. and Muralidhar, K. (2016). New directions in anonymization: Permutation paradigm, verifiability by subjects and intruders, transparency to users. *Information Sciences*, 337–338:11–24.

Durham, E. A. (2012). A framework for accurate, efficient private record linkage. Dissertation. Vanderbilt University.

Eggli, Y., Halfon, P., Chikhi, M., and Bandi, T. (2006). Ambulatory healthcare information system: A conceptual framework. *Health Policy*, 78:26–38.

El Kalam, A., Melchor, C., Berthold, S., Camenisch, J., Clauss, S., Deswarte, Y., Kohlweiss, M., Panchenko, A., Pimenidis, L., and Roy, M. (2011). Further privacy mechanisms. In Camenisch, J., Leenes, R., and Sommer, D., editors, *Digital Privacy*, pages 485–555. Springer, Berlin.

Farrow, J. and Schnell, R. (2017). Locational privacy preserving distance computations with intersecting sets of randomly labelled grid points. *Journal of the Royal Statistical Society, Series A, Under review*.

Goldreich, O. (2004). *Foundations of Cryptography. Volume 2, Basic Applications*. Cambridge University Press, Cambridge.

Herzog, T. N., Scheuren, F. J., and Winkler, W. E. (2007). *Data Quality and Record Linkage Techniques*. Springer, New York.

Herzog, T. N., Scheuren, F. J., and Winkler, W. E. (2010). Record linkage. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5):535–543.

Holly, A., Gardiol, L., Eggli, Y., Yalcin, T., and Ribeiro, T. (2005). Ein neues gesundheitsbasiertes Risikoausgleichssystem für die Schweiz. *G+G Wissenschaft*, 5(2):16–31.

Johnson, S. B., Whitney, G., McAuliffe, M., Wang, H., McCreedy, E., Rozenblit, L., and Evans, C. C. (2010). Using global unique identifiers to link autism collections. *Journal of the American Medical Informatics Association*, 17(6):689–695.

Jutte, D. P., Roos, L. L., and Brownell, M. D. (2010). Administrative record linkage as a tool for public health. *Annual Review of Public Health*, 31:91–108.

Karakasidis, A., Koloniari, G., and Verykios, V. S. (2015). Scalable blocking for privacy preserving record linkage. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 527–536, New York, NY, USA. ACM.

Karapiperis, D., Verykios, V. S., Katsiri, E., and Delis, A. (2016). A tutorial on blocking methods for privacy-preserving record linkage. In Karydis, I., Sioutas, S., Triantafillou, P., and Tsoumakos, D., editors, *Algorithmic Aspects of Cloud Computing: First International Workshop, ALGOCLOUD 2015, Patras, Greece, September 14-15, 2015. Revised Selected Papers*, pages 3–15. Springer International Publishing, Cham.

Karmel, R., Anderson, P., Gibson, D., Peut, A., Duckett, S., and Wells, Y. (2010). Empirical aspects of record linkage across multiple data sets using statistical linkage keys: the experience of the PIAC cohort study. *BMC Health Services Research*, 10(41).

Karmel, R. and Gibson, D. (2007). Event-based record linkage in health and aged care services data: a methodological innovation. *BMC Health Services Research*, 7:154.

Katalinic, A., Bartel, C., Raspe, H., and Schreer, I. (2007). Beyond mammography screening: quality assurance in breast cancer diagnosis (the quamadi project). *British Journal of Cancer*, 96(1):157–161.

Kelman, C. W., Bass, A. J., and Holman, C. D. J. (2002). Research use of linked health data: a best practice protocol. *Australian and New Zealand Journal of Public Health*, 26(3):251–255.

Kijsanayotin, B., Speedie, S. M., and Connelly, D. P. (2007). Linking patients' records across organizations while maintaining anonymity. *Proceedings of the 2007 American Medical Informatics Association Annual Symposium*, page 1008.

Kirsch, A. and Mitzenmacher, M. (2006). Less hashing same performance: building a better Bloom filter. In Azar, Y. and Erlebach, T., editors, *Algorithms-ESA 2006. Proceedings of the 14th Annual European Symposium: 11-13 September 2006; Zürich, Switzerland*, pages 456–467, Berlin. Springer.

Krawczyk, H., Bellare, M., and Canetti, R. (1997). HMAC: keyed-hashing for message authentication. Internet RFC 2104.

Kroll, M. and Steinmetzer, S. (2015). Who Is 1011011111...1110110010? Automated Cryptanalysis of Bloom Filter Encryptions of Databases with Several Personal Identifiers. In *Biomedical Engineering Systems and Technologies 2015*, pages 341–356. Springer.

Kuzu, M., Kantarcioglu, M., Durham, E., and Malin, B. (2011). A constraint satisfaction cryptanalysis of Bloom filters in private record linkage. In *The 11th Privacy Enhancing Technologies Symposium: 27–29 July 2011; Waterloo, Canada*.

Kuzu, M., Kantarcioglu, M., Durham, E. A., Toth, C., and Malin, B. (2013). A practical approach to achieve private medical record linkage in light of public resources. *Journal of the American Medical Informatics Association*, 20(2):285–292.

Niedermeyer, F., Steinmetzer, S., Kroll, M., and Schnell, R. (2014). Cryptanalysis of basic bloom filters used for privacy preserving record linkage. *Journal of Privacy and Confidentiality*, 6(2):59–69.

Office fédéral de la statistique (1997). La protection des données dans la statistique médicale. Technical report, Neuchatel.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Randall, S., Ferrante, A., Boyd, J., Brown, A., and Semmens, J. (2016). Limited privacy protection and poor sensitivity: Is it time to move on from the statistical linkage key-581? *Health Information Management Journal*, 45(2):71–79.

Randall, S. M., Ferrante, A. M., Boyd, J. H., Bauer, J. K., and Semmens, J. B. (2014). Privacy-preserving record linkage on large real world datasets. *Journal of Biomedical Informatics*, 50:205–212.

Ridder, G. and Moffitt, R. (2007). The econometrics of data combination. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6B, pages 5469–5547. Elsevier, Amsterdam.

Robertson, A. M. and Willett, P. (1998). Applications of n-grams in textual information systems. *Journal of Documentation*, 54(1):48–67.

Ryan, T., Holmes, B., and Gibson, D. (1999). A national minimum data set for home and community care. Canberra, AIHW.

Schmidlin, K., Clough-Gorr, K. M., Spoerri, A., and SNC study group (2015). Privacy preserving probabilistic record linkage (P3rl): a novel method for linking existing health-related data and maintaining participant confidentiality. *BMC medical research methodology*, 15:46.

Schnell, R. (2014). An efficient privacy-preserving record linkage technique for administrative data and censuses. *Journal of the International Association for Official Statistics*, 30(3):263–270.

Schnell, R. (2015). Privacy preserving record linkage. In Harron, K., Goldstein, H., and Dibben, C., editors, *Methodological Developments in Data Linkage*, pages 201–225. Wiley, Chichester.

Schnell, R., Bachteler, T., and Reiher, J. (2009). Privacy-preserving record linkage using Bloom filters. *BMC Medical Informatics and Decision Making*, 9(41).

Schnell, R., Bachteler, T., and Reiher, J. (2011). A novel error-tolerant anonymous linking code. Working Paper WP-GRLC-2011-02, German Record Linkage Center, Duisburg.

Schnell, R. and Borgs, C. (2015). Building a national perinatal database without the use of unique personal identifiers. In *2015 IEEE 15th International Conference on Data Mining Workshops (ICDM 2015)*, pages 232–239., Atlantic City, NJ, USA. IEEE Publishing.

Schnell, R. and Borgs, C. (2016). Randomized response and balanced bloom filters for privacy preserving record linkage. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDM 2016)*, Barcelona, Dec 12, 2016 - Dec 15, 2016. IEEE Publishing.

Schülter, E., Kaiser, R., Oette, M., Müller, C., Schmeisser, N., Selbig, J., Beerenwinkel, N., Lengauer, T., Däumer, M., and Hoffmann, D. (2007). Arevir: A database to support the analysis of resistance mutations of human immunodeficiency. *European Journal of Medical Research*, 12(Supplememt III):10–11.

Stallings, W. (2014). *Cryptography and Network Security: Principles and Practice*. Pearson, New Jersey, 6 edition.

Taylor, L. K., Irvine, K., Iannotti, R., Harchak, T., and Lim, K. (2014). Optimal strategy for linkage of datasets containing a statistical linkage key and datasets with full personal identifiers. *BMC Medical Informatics and Decision Making*, 14:85.

Tessmer, A., Welte, T., Schmidt-Ott, R., Eberle, S., Barten, G., Suttorp, N., and Schaberg, T. (2011). Influenza vaccination is associated with reduced severity of community-acquired pneumonia. *European Respiratory Journal*, 38(1):147–153.

Vatsalan, D. and Christen, P. (2016). Privacy-preserving matching of similar patients. *Journal of Biomedical Informatics*, 59:285–298.

Vatsalan, D., Christen, P., and Verykios, V. S. (2013). A taxonomy of privacy-preserving record linkage techniques. *Information Systems*, 38(6):946–969.

# A Modified All-and-One Classification Algorithm Combined with the Bag-of-Features Model to Address the Food Recognition Task

Kalliopi Dalakleidi, Myriam Sarantea and Konstantina Nikita
*School of Electrical and Computer Engineering, National Technical University of Athens, Iroon Polytechniou Str., Athens, Greece*
*kdalakleidi@biosim.ntua.gr, myriam.saradea@gmail.com, knikita@ece.ntua.gr*

Keywords: Diabetes, All-And-One, Bag-Of-Features, Food Recognition System.

Abstract: Dietary intake monitoring can play an important role in reducing the risk of diet related chronic diseases. Automatic systems that support patients to count the nutrient contents, like carbohydrates (CHO), of their meals, can provide valuable tools. In this study, a food recognition system is proposed, which consists of two modules performing feature extraction and classification of food images, respectively. The dataset used consists of 1200 food images split into six categories (bread, meat, potatoes, rice, pasta and vegetables). Speeded Up Robust Features (SURF) along with Color and Local Binary Pattern (LBP) features are extracted from the food images. The Bag-Of-Features (BOF) model is used in order to reduce the features space. A modified version of the All-And-One Support Vector Machine (SVM) is proposed to perform the task of classification and its performance is evaluated against several classifiers that follow the SVM or the K-Nearest Neighbours (KNN) approach. The proposed classification method has achieved the highest levels of accuracy (Acc = 94.2 %) in comparison with all the other classifiers.

## 1 INTRODUCTION

Diet related chronic diseases, such as obesity and diabetes mellitus, are expanding nowadays. Therefore, an urgent need for dietary intake monitoring arises that can reduce the risk of these diseases. Studies have shown that when patients with diabetes mellitus do significant errors in reporting their dietary intake, there is an increased risk of postprandial hypo- or hyperglycemia. Automatic systems, usually based on a mobile phone, can support patients that suffer from diet related chronic diseases with carbohydrates (CHO) counting. The user first takes a photograph of the upcoming meal with the camera of his mobile phone. Then, the image is processed so that the different types of food are divided from each other and segmented in different areas of the image. A series of features are extracted from each segmented area and are fed to a classifier, which decides what kind of food is represented by each segmented area. Then, the volume of each segmented area is calculated and the total CHO of the depicted meal are estimated.

Feature extraction can play a key role in dietary intake monitoring systems. Efficient feature descriptors could ensure stability and distinctiveness, where stability means that the extracted features are invariant to different photometric and geometric changes and distinctiveness means that the extracted features can be used to distinguish the specified object from other objects or the background. Features related to color and texture have been shown to ensure stability and distinctiveness. Moreover, a large variety of local feature descriptors has been proposed in the literature, like Gaussian derivatives (Florack et al., 1994), moment invariants (Mindru et al., 2004), complex features (Baumberg, 2000; Schaffalitzky and Zisserman, 2002), steerable filters (Freeman and Adelson, 1991), and phase-based local features (Carneiro and Jepson, 2003). A variant of Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), has the ability to capture spatial intensity patterns, while being robust to small deformations or localization errors and is shown to outperform the previous mentioned categories of features (Mikolajczyk and Schmid, 2003; Bay et al., 2008).

Classification results of food images in dietary intake monitoring systems can be improved when the dimension of the extracted feature vector is reduced. The use of the Bag-Of-Features (BOF) model (Peng et al., 2016), which is inspired by the Bag-Of-Words model for text classification (Cruz-Roa et al., 2011) has been reported to highly improve classification accuracy in food recognition tasks. The BOF model achieves dimensionality reduction by creating from



Dalakleidi, K., Sarantea, M. and Nikita, K.
A Modified All-and-One Classification Algorithm Combined with the Bag-of-Features Model to Address the Food Recognition Task.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 284-290
ISBN: 978-989-758-213-4

the extracted features visual words, and by describing the image content with the distribution of these visual words (Wang et al., 2016).

Another important aspect of the food recognition task is that it is usually a multiclass classification problem, as the used food datasets almost always contain more than two categories of food. There exist many classification approaches in order to address the multiclass recognition task, but the most prominent ones, like the One-Against-All (OAA), the One-Against-One (OAO) and the All-And-One (A&O) (Pedrajas and Boyer, 2006) descend from the binarization strategy, where the division of the initial multiclass problem to several binary class problems takes place (Galar et al., 2011).

Several attempts to implement automatic or semi-automatic systems for dietary intake monitoring have been reported in the literature. A food identification application called DietCam has been recently presented (Kong and Tan, 2012), which consists of three parts: image manager, food classifier and volume estimator. Images taken by the users are fed to the image manager, then SIFT features are extracted, clustered into visual words and fed to a simple Bayesian probabilistic classifier, which achieves high levels of accuracy (92%). The food volume estimator calculates the volume of each food item recognized by the food classifier and then the calorie content of the food is estimated. Another food recognition application has been recently proposed for the classification of fast-food images (Shroff et al., 2008). After segmentation of the fast-food image, color, size, texture, shape and context-based features are computed and fed to a feed-forward artificial neural network achieving a 90% accuracy. Moreover, a food identification system has been presented which consists of the following modules: image segmentation, feature extraction, food classification, and volume estimation (Zhu et al., 2010). Food description is based on a set of color and texture features, while classification is based on a Support Vector Machine (SVM) classifier, which has achieved high classification accuracy (95,8%). An automated Food Intake Evaluation System (AFIES) has been reported (Martin et al., 2009), which consists of reference card detection, food region segmentation, food classification and food amount estimation modules. The color RGB data are used as feature vectors for classification, which is performed using the Mahalanobis distance of pixels from food classes. The amount of calorie intake is estimated based on the assumption that food area is linearly proportional to the food volume. In another study, recognition of seven broad categories of food based on a representation for food items that calculates pairwise statistics between local features has been presented (Yang et al., 2010). These statistics are accumulated in a multi-dimensional histogram, which is then used as input to a SVM classifier. Food images are taken from the Pittsburgh Food Image Dataset (PFID) (Chen et al., 2009). This system has also achieved high levels of recognition accuracy (80%).

The use of the BOF model has been adopted in several food recognition systems recently, since food recognition does not presume any typical spatial arrangement of the food elements. Based on the BOF model, the Food Intake Visual and Voice Recognizer system which aims to measure the nutritional content of a user's meal (Puri et al., 2009) has been proposed. Given a set of three images of a user's plate of food, the system first asks the user to list food items through speech, then attempts to identify each food item on the plate, and finally reconstructs them in 3D to measure their respective volumes. Food images are collected by the developers of the system. Food classification is based on the combined use of color neighborhood and maximum response features in a texton histogram model, which resembles the BOF approach. Adaboost is used for feature selection and SVM for classification, which achieves recognition accuracy about 90%. Moreover, a food recognition system for the classification of Japanese food images has been introduced (Joutou and Yanai, 2009), based on the combined use of BOF of SIFT, Gabor filter responses and color histograms, which are then fed to a multiple kernel learning classifier, which has achieved acceptable levels of accuracy (61,34%). The BOF model has been used in another automatic food recognition system (Anthimopoulos et al., 2014). The system firstly computes dense local features using the SIFT on the HSV (Hue Saturation Value) color space, then builds a visual vocabulary of 10000 visual words by using the hierarchical k-means clustering and, finally, classifies the food images with a linear SVM classifier, which achieves high levels of accuracy (78%).

In the present study, a food recognition system is proposed which consists of two modules performing feature extraction and classification of food images, respectively (Figure 1). Motivated by the ability of SURF to capture spatial intensity patterns and the stability and distinctiveness provided by Color and Local Binary Pattern (LBP) features, the combination of SURF, Color and LBP features is examined in this study. Moreover, a novel modified version of the All-And-One (M-A&O) SVM classifier for multiclass classification problems is proposed and its performance is assessed against classification methods

based on SVM or the K-Nearest Neighbour approaches including the OAA SVM, the OAO SVM, the A&O SVM, the Weighted K-Nearest Neighbour (WKNN) classifier, the Dual Weighted K-Nearest Neighbour (DWKNN) classifier, and the K-Nearest Neighbour Equality (KNNE) classifier.

# 2 METHODS

## 2.1 Dataset

The Food Image Dataset (FID) used in this study consists of 1200 images, 260-by-190 pixels each, collected from the web. Each image belongs to one of six categories corresponding to bread, meat, potatoes, rice, pasta and vegetables (Figure 2). Each category is represented by 200 images. The food is photographed under different servings, view angles, and lighting conditions. The background of every image is edited so that it is completely black.

## 2.2 Feature Extraction

In the present study SURF, Color and LBP features have been combined to represent each food image in the proposed food recognition system.

SURF detects points of interest using an integer approximation of the determinant of Hessian blob detector, and, then computes the features based on the Haar wavelet response around each point of interest (Bay et al., 2008). Color features are calculated as the average value of color for every 4-by-4 pixel block of the image. LBP is a texture descriptor that provides a unified description, including both statistical and structural characteristics of a texture patch (Prabhakar and Praveen Kumar, 2012). The LBP feature vector is calculated by dividing the image into cells, and comparing the center pixel's value with the neighbours' pixel values of each cell. Then, a histogram of the numbers occurring over the cells is computed. A useful extension to the LBP is the uniform LBP, which reduces the length of the initial feature vector from 256 to 59 (Ojala et al., 2002).

The approach of BOF is used to decrease the input feature space, and deal with high visual diversity and absence of spatial arrangement encountered in food recognition. The BOF approach is influenced by the Bag-Of-Words representation for text classification (Cruz-Roa et al., 2011) and consists of the following two steps. Firstly, a set of small blocks are extracted from each image in the dataset, which are represented by feature vectors. Secondly, the visual dictionary of the image dataset is constructed and each image is represented by the frequency of the codewords of the visual dictionary. The visual dictionary is built with the use of the k-means clustering algorithm. The cluster centers of the feature points extracted in the first step of the BOF approach are defined as visual words. The visual dictionary is the combination of these visual words (Wang et al., 2016).

## 2.3 Classification

The classification task is performed using a modified version of the All-and-One SVM and its performance is assessed against several classification methods based on the SVM and K-Nearest Neighbours (KNN) approach, including the OAA SVM classifier, the OAO SVM, the A&O SVM, the WKNN classifier, the DWKNN classifier, and the KNNE classifier. All algorithms have been implemented with MATLAB 2015a, are trained with the 70% of the images of the FID, and tested with the rest 30% of the FID.

### 2.3.1 SVM-based Classifiers

***The OAA SVM Algorithm.***
The OAA SVM classifier (Galar et al., 2011) consists of K binary SVM classifiers, where K is the total number of classes. The i-th classifier is trained by labeling all the instances in the i-th class as positive and the rest as negative. Each test instance is classified to the class with the biggest score.

***The OAO SVM Algorithm.***
The OAO SVM classifier (Galar et al., 2011) consists of K(K-1)/2 binary SVM classifiers, where K is the number of classes. Each binary classifier learns to discriminate between a pair of classes. The outputs of these binary classifiers are combined so that the class with the highest score is assigned to the test instance.

***The A&O SVM Algorithm.***
The A&O SVM algorithm (Pedrajas and Boyer, 2006) combines the strengths of the OAO and OAA approaches. Taking into account that for a high proportion of miss-classifications of the OAA approach, the second best class is actually the correct class, and that the binary classifiers of OAO are highly accurate on their own, but may lead to incorrect results when combined, the A&O approach combines the results of K OAA classifiers and K(K-1)/2 OAO classifiers. The A&O approach first classifies a test instance using the K OAA classifiers and holds the two classes i,j with the biggest scores. Then, the binary classifier of the OAO approach is used to classify the instance among classes i,j.

***The M-A&O SVM Algorithm.***
The M-A&O SVM algorithm combines the strengths of the OAO and OAA approaches as the A&O SVM algorithm, but in a different way. The M-A&O SVM approach first classifies a test instance using the K OAA SVM classifiers and holds the scores. Then, the K(K-1)/2 SVM binary classifiers of the OAO approach are used to classify the instance. The test instance will be assigned to the class that will achieve the highest score from all (K + K(K-1)/2) classifiers.

### 2.3.2 KNN-based Classifiers

***The WKNN Algorithm.***
The WKNN algorithm is a modified version of the K-Nearest Neighbours (KNN) algorithm. According to the KNN algorithm, the k-nearest neighbours of the query instance are selected according to a distance criterion, such as the Euclidean distance. Then, the query instance is assigned to the class represented by the majority of its k-nearest neighbours in the training set. In the WKNN algorithm, the closer neighbours are weighed more heavily than the farther ones (Marinakis et al., 2009) and the distance-weighted function $w_i$ to the i-th nearest neighbor is defined as,

$$w_i = \frac{k+1-i}{\sum_{m=1}^{k} m}$$

where m is an integer in the interval (1,k) and k is the total number of the neighbours.

***The DWKNN Algorithm.***
In order to address the effect of the number of neighbours on the classification performance, a DWKNN algorithm has been proposed (Gou et al., 2011). The DWKNN algorithm gives different weights to the k nearest neighbours depending on distances between them and their ranking according to their distance from the query object (Dalakleidi et al., 2013). The distance-weighted function $w_i$ of the i-th nearest neighbor is calculated according to the following equation,

$$w_i = \begin{cases} \dfrac{d_k^{NN} - d_i^{NN}}{d_k^{NN} - d_1^{NN}} \times \dfrac{1}{i}, d_k^{NN} \neq d_1^{NN} \\ 1, \ d_k^{NN} = d_1^{NN} \end{cases}$$

where $d_i^{NN}$ is the distance of the i-th nearest neighbour from the query object, $d_1^{NN}$ is the distance of the nearest neighbour, and $d_k^{NN}$ is the distance of the k-furthest neighbour. Thus, the weight of the nearest neighbor is 1, and the weight of the furthest k-th neighbor is 0, whereas other weights are distributed between 0 and 1.

***The KNNE Algorithm.***
The KNNE algorithm (Sierra et al., 2011) is a variation of the KNN classifier for multiclass classification. It searches for the K-nearest neighbours in each class and assigns the query instance in the class whose K-nearest neighbours have the minimal mean distance to the test instance.

## 3 RESULTS

The FID is used for the evaluation of the proposed classification algorithm against the classification algorithms based on the SVM and KNN approach on the food recognition task. In order to improve the classification accuracy of the examined algorithms, several sizes of the vocabularies of the BOF model are tested. Table 1 shows the average accuracy of the OAO SVM classifier on the six food classes for different sizes of the vocabulary of the BOF model for SURF and Color features. The size of the vocabularies has been varied from 100 to 2000 words. As it can be observed from Table 1, the lowest accuracy (Acc = 85.0%) is achieved with the size of 300 for both the SURF and Color BOF vocabularies, whereas the highest accuracy (Acc = 93.9%) is achieved with the size of 1000 for both the SURF and Color BOF vocabularies. It is also important to note that among the three types of features, Color features contribute the most to the accuracy of the OAO SVM classifier.

Table 1: Average accuracy of the OAO SVM classifier on the six food classes of Food Image Dataset for varying size of the vocabulary of the BOF model for SURF and Color features.

| Features | | | Acc |
|---|---|---|---|
| **SURF** | **Color** | **LBP** | |
| 100 | 100 | 59 | 87.5 |
| 200 | 200 | 59 | 90.0 |
| 300 | 300 | 59 | 85.0 |
| 400 | 400 | 59 | 90.6 |
| 500 | 500 | 59 | 91.1 |
| 600 | 600 | 59 | 91.7 |
| 700 | 700 | 59 | 92.5 |
| 800 | 800 | 59 | 93.1 |
| 900 | 900 | 59 | 90.6 |
| 1000 | 1000 | 59 | **93.9** |
| 1100 | 1100 | 59 | 93.3 |
| 1500 | 1500 | 59 | 91.4 |
| 2000 | 2000 | 59 | 90.8 |

Table 2: The average accuracy (%) of the classifiers under comparison on the six food classes of the Food Image Dataset.

| Algorithm | Acc (%) |
|---|---|
| WKNN | 84.4 |
| DWKNN | 92.8 |
| KNNE | 93.9 |
| OAA SVM | 90.6 |
| OAO SVM | 93.9 |
| A&O SVM | 90.3 |
| M-A&O SVM | **94.2** |

Table 3: Confusion matrix of the M-A&O SVM for each food class (Bread, Meat, Pasta, Potatoes, Rice and Vegetables) of the Food Image Dataset.

| Confusion Matrix | | | | | | |
|---|---|---|---|---|---|---|
| **Acc (%)** | Br | M | Pa | Pot | R | Veg |
| Br | **93.3** | 0.0 | 0.0 | 0.0 | 6.7 | 0.0 |
| M | 1.7 | **95.0** | 0.0 | 3.3 | 0.0 | 0.0 |
| Pa | 0.0 | 0.0 | **93.3** | 6.7 | 0.0 | 0.0 |
| Pot | 0.0 | 1.7 | 5.0 | **93.3** | 0.0 | 0.0 |
| R | 0.0 | 0.0 | 6.7 | 3.3 | **90.0** | 0.0 |
| Veg | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | **100.0** |

In Table 2, the average accuracy of the classifiers under comparison on the six food classes is presented. The size of the vocabulary of the BOF method for SURF and Color features is 1000, thus a total number of 2059 features is used for the classification. Ten k-nearest neighbours are used for the WKNN, DWKNN and KNNE. As it can be observed in Table 2, the lowest average accuracy (Acc = 84.4%) is achieved by the WKNN classifier, whereas the highest average accuracy (Acc = 94.2%) is achieved by M-A&O SVM. The second best average accuracy is achieved by the OAO SVM and KNNE algorithms. The superiority of M-A&O SVM can be explained by the fact that it combines two very powerful strategies, the OAA SVM and the OAO SVM, for multiclass classification.

In Table 3, the classification accuracy of M-A&O SVM for each food class is shown in the form of the confusion matrix. It can be observed that the lower classification accuracy (Acc = 90.0%) is achieved for the class of rice. This is due to the fact that rice is mingled with several sauces which can be very different in color and texture. It is important to note that rice is misclassified to potatoes and pasta which are closer to it in terms of CHO than with meat or vegetables. The best classification accuracy is achieved for vegetables (Acc = 100.0%), this is due to the distinctive green color of vegetables.

## 4 CONCLUSIONS

Automatic food recognition systems can be used to estimate the content of a meal in CHO for patients with diet related chronic diseases, such as obesity and diabetes mellitus. In this study, an attempt to address the tasks of feature extraction and food image recognition was made. The use of the SURF, Color and LBP features in combination with the BOF model has proven to be particularly effective in terms of average classification accuracy. Several classification approaches for multiclass classification have been tested. The best classification accuracy (Acc = 94.2%) has been achieved by a modified version of the All-And-One SVM approach and is quite high as compared to reported values of classification accuracy for food images in the literature (60%-96%). The proposed system can be combined with an image segmentation module and a volume estimation module towards the development of an automatic food recognition system. Moreover, several other classifiers, like AdaBoost, Random Forests and Convolutional Neural Networks, can be used in the future for comparison purposes in the classification module.

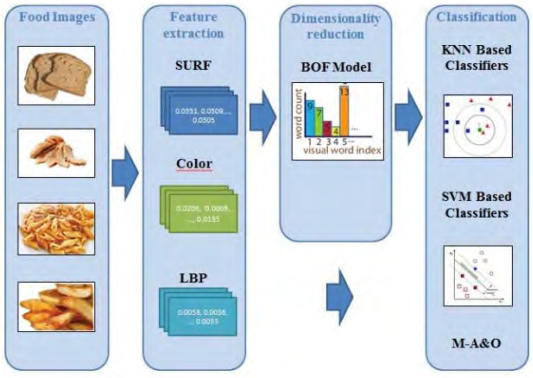



Figure 1: Block diagram of the proposed system.

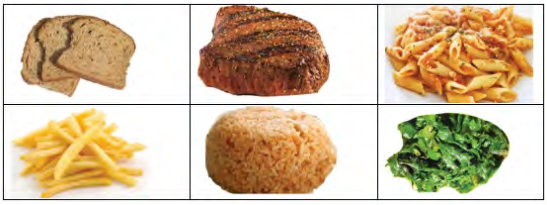


Figure 2: Example of images from each of the six categories of the Food Image Dataset.

## ACKNOWLEDGEMENTS

The work of Kalliopi Dalakleidi was supported by a scholarship for Ph.D. studies from the Hellenic State Scholarships Foundation "IKY fellowships of excellence for post-graduate studies in Greece-Siemens Program".

## REFERENCES

Anthimopoulos, M., Gianola, L., Scarnato, L., Diem, P. and Mougiakakou, S., 2014. A Food Recognition System for Diabetic Patients Based on an Optimized Bag-of-Features Model. *Journal of Biomedical and Health Informatics*, 18(4), pp. 1261-1271.

Baumberg, A., 2000. Reliable feature matching across widely separated views. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Hilton Head Island, South Carolina, USA: IEEE, pp. 774 – 781.

Bay, H., Ess, A., Tuytelaars, T. and Gool, L., 2008. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3), pp. 346-359.

Carneiro, G. and Jepson, A., 2003. Multi-scale phase-based local features. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Madison, Wisconsin, USA: IEEE, pp. 736 – 743.

Chen, M., Dhingra, K., Wu, W., Yang, L., Sukthankar, R. and Yang, J., 2009. PFID: Pittsburgh fast-food image dataset. In: *IEEE International Conference on Image Processing*. Cairo, Egypt: IEEE, pp. 289-292.

Cruz-Roa, A., Caicedo, J. and Gonzalez, F., 2011. Visual pattern mining in histology image collections using bag of features. *Artificial Intelligence in Medicine*, 52, pp. 91-106.

Dalakleidi, K., Zarkogianni, K., Karamanos, V., Thanopoulou, A. and Nikita, K., 2013. A Hybrid Genetic Algorithm for the Selection of the Critical Features for Risk Prediction of Cardiovascular Complications in Type 2 Diabetes Patients. In: *13th International Conference on BioInformatics and BioEngineering*. Chania, Greece: IEEE.

Florack, L., Haar Romeny, B., Koenderink, J. and Viergever, M., 1994. General intensity transformations and differential invariants. *Journal of Mathematical Imaging and Vision*, 4, pp. 171–187.

Freeman, W. and Adelson, E., 1991. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, pp. 891 – 906.

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H. and Herrera, F., 2011. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, 44, pp. 1761-1776.

Gou, J., Xiong, T. and Kuang, Y., 2011. A novel weighted voting for K-Nearest neighbour rule. *Computers*, 6(5), pp. 833-840.

Joutou, T. and Yanai, K., 2009. A food image recognition system with multiple kernel learning. In: *IEEE International Conference on Image Processing*. Cairo, Egypt: IEEE, pp. 285–288.

Kong, F. and Tan, J., 2012. DietCam: Automatic dietary assessment with mobile camera phones. *Pervasive and Mobile Computing*, 8(1), pp. 147-163.

Marinakis, Y., Ntounias, G. and Jantzen, J., 2009. Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification. *Computers in Biology and Medicine*, 39, pp. 69-78.

Martin, C., Kaya, S. and Gunturk, B., 2009. Quantification of food intake using food image analysis. In: *Engineering in Medicine and Biology Society Annual International Conference of the IEEE*. Minneapolis, Minnesota, USA: IEEE, pp. 6869–6872.

Mikolajczyk, K. and Schmid, C., 2003. A performance evaluation of local descriptors, In: *IEEE Conference on Computer Vision and Pattern Recognition*. Madison, Wisconsin, USA:IEEE, pp. 257 – 263.

Mindru, F., Tuytelaars, T., Van Gool, L. and Moons, T., 2004. Moment invariants for recognition under changing viewpoint and illumination. *Computer Vision and Image Understanding*, 94, pp. 3–27.

Ojala, T., Pietikäinen, M. and Mäenpää, T., 2002. Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), pp. 971-987.

Pedrajas, N. and Boyer, D., 2006. Improving multiclass pattern recognition by the combination of two strategies. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6), pp. 1001–1006.

Peng, X., Wang, L., Wang, X. and Qiao, Y., 2016. Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice. *Computer Vision and Image Understanding*, pp. 1-17.

Prabhakar, C. and Praveen Kumar, P., 2012. LBP-SURF Descriptor with Color Invariant and Texture Based Features for Underwater Images. In: *8th Indian Conference on Computer Vision, Graphics and Image Processing*. Mumbai, India.

Puri, M., Zhu, Z., Yu, Q., Divakaran, A. and Sawhney, H., 2009. Recognition and Volume Estimation of Food Intake using a Mobile Device. In: *Workshop on Applications of Computer Vision (WACV)*. Snowbird, Utah, USA: IEEE, pp. 1-8.

Schaffalitzky, F. and Zisserman, A., 2002. Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?". In: *7th European Conference on Computer Vision*. Copenhagen, Denmark: LNCS, pp. 414 – 431.

Shroff, G., Smailagic, A. and Siewiorek, D., 2008. Wearable context-aware food recognition for calorie monitoring. In: *12th IEEE International Symposium on Wearable Computers*. IEEE, pp. 119–120.

Sierra, B., Lazkano, E., Irigoien, I., Jauregi, E. and Mendialdua, I., 2011. K nearest neighbor equality: giving equal chance to all existing classes. *Information Sciences*, 181(23), pp. 5158–5168.

Wang, R., Ding, K., Yang, J. and Xue, L., 2016. A novel method for image classification based on bag of visual words. *Journal of Visual Communication and Image Representation*, 40, pp. 24-33.

Yang, S., Chen, M., Pomerleau, D. and Sukthankar, R., 2010. Food recognition using statistics of pairwise local features. In: *IEEE Conference on Computer Vision and Pattern Recognition*. San Francisco, California, USA:IEEE, pp. 2249–2256.

Zhu, F., Bosch, M., Woo, I., Kim, S., Boushey, C., Ebert, D. and Delp, E., 2010. The use of mobile devices in aiding dietary assessment and evaluation. *IEEE Journal of Selected Topics in Signal Processing*, 4(4), pp. 756–766.

# Static and Dynamic Approaches for Pain Intensity Estimation using Facial Expressions

Niloufar Zebarjadi and Iman Alikhani
*Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland*
{*nzebarja, ialikhan*}*@ee.oulu.fi*

Keywords: Regression, LBP-TOP, 3D-SIFT, LBP, DSIFT, Feature Extraction, Facial Expression Analysis.

Abstract: Self-report is the most conventional means of pain intensity assessment in clinical environments. But, it is not an accurate metric or not even possible to measure in many circumstances, e.g. intensive care units. Continuous and automatic pain level evaluation is an advantageous solution to overcome this issue. In this paper, we aim to map facial expressions to pain intensity levels. We extract well-known static (local binary pattern(LBP) and dense scale-invariant feature transform (DSIFT)) and dynamic (local binary patterns on three orthogonal planes (LBP-TOP) and three dimensional scale-invariant feature transform (3D-SIFT)) facial feature descriptors and employ the linear regression method to label a number between zero (no pain) to five (strong pain) to each testing sequence. We have evaluated our methods on the publicly available UNBC-McMaster shoulder pain expression archive database and achieved average mean square error (MSE) of 1.53 and Pearson correlation coefficient (PCC) of 0.79 using leave-one-subject-out cross validation. Acquired results prove the superiority of dynamic facial features compared to the static ones in pain intensity determination applications.

## 1 INTRODUCTION

Automatic recognition of a patient's pain level is a notable study and could have a large impact within health care centers and clinics. For instance, consistent monitoring of pain in severely ill or immature patients reduces the workload of medical staff and boosts the reliability of assessment. In addition, self-reporting of pain intensity is not an objective means of evaluation and is influenced by each patient's perception of pain (Khan et al., 2013).

A patient's facial expressions contain information about a subject's well-being (e.g. sickness, stress, fatigue), as well as pain intensity (Kaltwang et al., 2012), and have received increasing attention during last years. Four core facial actions representing lots of information about pain are brow lowering, eye closure, orbital tightening and upper lip levator contraction (Lucey et al., 2012).

Machine vision and facial expression analysis have been employed in recent years to 1) detect subjects suffering from pain (Ashraf et al., 2009; Lucey et al., 2011a; Lucey et al., 2011b; Khan et al., 2013; Roy et al., 2016; Neshov and Manolova, 2015) and 2) assess pain intensity level (Kaltwang et al., 2012; Rathee and Ganotra, 2015). One principal concern in facial expression assisted pain level estimation has been that whether a sample video should be analyzed frame-by-frame or sequence-based. Ashraf et al. proposed a pain detection technique in (Ashraf et al., 2009) based on active appearance model (AAM). A set of features are extracted from this model, including similarity normalized shape representation (S-PTS), similarity normalized appearance representation (S-APP) and canonical appearance representation (C-APP). They were mainly exploring to figure out whether the database should be labeled in a frame-level or in a sequence-level respect. In (Lucey et al., 2011a), S-APP, S-PTS and C-APP were utilized in order to build an automatic pain detection system using facial expressions. They studied the database proposed in (Lucey et al., 2011b), in a frame-by-frame level by analysis of action units (AUs) based on the facial action coding system (FACS) which properly detects movements of facial muscles. In (Lucey et al., 2012), the authors published their study on the same database using AAM/SVM pain detection system. The contribution of (Lucey et al., 2011b) was the 3D head pose motion data experimentation as a cue of pain. Later, Khan et al. in (Khan et al., 2013) suggested a new framework for pain detection on the same shoulder pain database. In that framework, following the face detection from each frame of input sequence, face was divided into two equal parts of



Zebarjadi, N. and Alikhani, I.
Static and Dynamic Approaches for Pain Intensity Estimation using Facial Expressions.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 291-296
ISBN: 978-989-758-213-4

upper and lower regions in order to assign equal significance to them. Then, Pyramid histogram of orientation (PHOG) and pyramid local binary pattern (PLBP) features were extracted from both regions and concatenated to reach a final descriptor. In (Khan et al., 2013), four different classifiers (SVM, decision tree, random forest and 2-nearest-neighbor) were employed to detect pain from facial expressions. Recently, several studies have attempted to enhance the performance of pain detection with different classifiers and descriptors (Neshov and Manolova, 2015; Roy et al., 2016).

There are also a few studies focusing on level-based pain intensity estimation which can propose more information to the medical staff (e.g. for prescribing appropriate drug dose). In (Kaltwang et al., 2012), they utilized facial landmarks, discrete cosine transform(DCT) and LBP method to extract features and relevance vector regression to determine pain intensity level. Recently, in (Rathee and Ganotra, 2015), a new method is proposed based on the modeling of facial feature deformations during pain using thin plate spline. They mapped the deformation parameters to higher discriminative space by the distance metric learning technique.

In this study, we aim to estimate the level of pain using four widely-used static and dynamic facial expression descriptors. To have a comprehensive comparison within two dimensional (2D) and three dimensional (3D) models, local binary pattern (LBP) and dense scale-invariant feature transform (DSIFT) are used as two frequently-used static features, as well as two corresponding dynamic features, including local binary patterns on three orthogonal planes (LBP-TOP) and three dimensional scale-invariant feature transform (3D-SIFT). Afterwards, support vector regression (SVR) is used to map the extracted features to the pain intensity level of subjects ranging from zero (no pain) to five (extreme pain) using leave-one-subject-out-cross validation.

# 2 UNBC-McMasterSHOULDER PAIN EXPRESSION ARCHIVE DATABASE

UNBC-McMaster shoulder pain expression archive database contains 200 video sequences of spontaneous facial expressions (48,398 frames) of 25 patients suffering from shoulder pain. In this database, participants performed a variety of motion tests, including abduction, flexion, internal and external rotation of arms (Lucey et al., 2011b).

Figure 1: Example frames of a sequence from the UNBC-McMaster shoulder pain archive database.

Figure 2: Example cropped frames of a sequence from the UNBC-McMaster shoulder pain archive database.

Besides, there are observed pain intensity (OPI) sequence-level rating from 0 (no pain) to 5 (extreme pain) provided in this database which is used as the reference value for the system. The distribution of the sequences over OPI is provided in Table 1.

Table 1: The inventory on observed pain intensity (OPI) measures at the sequence level.

| OPI | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Sequence Number | 92 | 25 | 26 | 34 | 16 | 7 |

# 3 METHODOLOGY

In this section, we mainly explain the static and dynamic feature descriptors that we have extracted from cropped faces, the regression machine and performance measurement metrics.

## 3.1 Static Features

### 3.1.1 LBP

LBP (Ojala et al., 2002) is a robust appearance feature descriptor. This descriptor was initially proposed for texture analysis (Ojala et al., 1996), while recently it has been utilized in the analysis of facial expressions as well (Ahonen et al., 2006). To acquire LBP histogram of an image, the examined frame is divided into several cells and LBP histograms are obtained for each cell. The histograms of all cells are concatenated as a feature vector for the entire frame (Ahonen et al., 2004). In each cell of the image there are two variables, **P** and **R** which stands for the number of neighboring points around each central pixel and the radius, respectively. To calculate the LBP of each pixel, the central pixel value is compared to the neighboring pixels and the greater neighboring values than the central one are assigned as ”1”, otherwise ”0”. This leads to an 8-digit binary number which is converted

to decimal (Ahonen et al., 2004). We consider **P** as eight neighboring pixels and **R** as two and three pixels through our analysis. Additionally, each sequence is divided into a different number of cells along x- and y-axis, ranging from six to ten and along time-axis, ranging from four to six parts.

#### 3.1.2 DSIFT

DSIFT is a robust and popular feature in image processing. SIFT describes local features in a frame by extracting discriminative key-points and computing a histogram of orientation for every single of them. SIFT key points are invariant to viewpoint changes that induce translation, rotation, and re-scaling of the image (Lowe, 2004). DSIFT extracts a densely sampled SIFT feature from image which can be adjusted by sampling step, sampling bounds, size and descriptor geometry. Key-points are sampled in the sense that the center of spatial bins is at integer coordinates within the image boundaries (Vedaldi and Fulkerson, 2010). The main advantage of DSIFT compared to SIFT is its computational efficiency. In order to employ DSIFT in a video sequence, we divide the video sequence into a few number of segments and calculate the DSIFT for each frame in each segment. In the following step, the feature values of all frames are averaged within each segment and then concatenated together. By this approach, the dimension of final feature vector is reduced significantly. So, in this descriptor also x-, y- and time axis grid-size should be tuned.

### 3.2 Dynamic Features

#### 3.2.1 LBP-TOP

LBP-TOP is basically local binary patterns on three XY, XT and YT orthogonal planes (Zhao and Pietikainen, 2007). It is a dynamic texture descriptor using LBP in order to extract spatio-temporal features. To obtain LBP-TOP histogram of a video, a sequence is divided into non-overlapping block volumes separately and the LBP-TOP histograms in each block volume are computed and then concatenated into a single histogram (Zhao and Pietikainen, 2007). The number of divisions in row and column of XY plane and in time as well as radius around each central pixel are considered as important parameters of this method.

#### 3.2.2 3D-SIFT

3D-SIFT (Scovanner et al., 2007) technique expands DSIFT descriptor from 2D to 3D by encoding the information in both space and time. In this method, a video sequence is divided into rectangular cubes and direction of gradient in each 3D sub-volume is indicated by two angular values $(\theta, \phi)$.

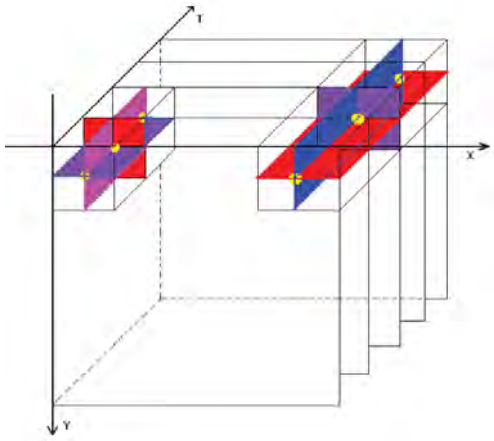



Figure 3: Computation of the LBP-TOP using non-overlapping block volumes (Zhao and Pietikainen, 2007).

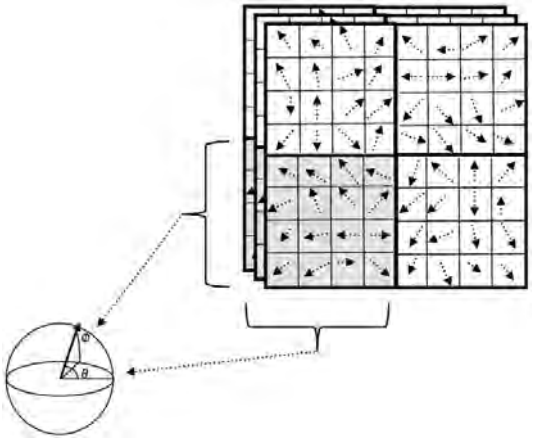



Figure 4: Computation of the 3D SIFT using two angular values $(\theta, \phi)$ (Krig, 2014).

Therefore, a single gradient magnitude and two orientation vectors provided in equations 1, 2 and 3 describe each point's characteristics.

$$m3D(x,y,t) = \sqrt{{L_x}^2 + {L_y}^2 + {L_t}^2}, \tag{1}$$

$$\theta(x,y,t) = \tan^{-1}\frac{L_x}{L_y}, \tag{2}$$

$$\phi(x,y,t) = \tan^{-1}\left(\frac{L_t}{\sqrt{{L_x}^2 + {L_y}^2}}\right), \tag{3}$$

### 3.3 Performance Measurement

The construction of feature vectors is followed by linear regression using SVR machine (Chang and Lin, 2011). The systems are trained using predefined OPI labels corresponding to each sequence pain level, ranging from zero (no pain) to five (extreme pain). We have considered leave-one-subject-out cross validation technique and thus, the system is iteratively trained using all except one subject's data and is tested on the excluded sample subject's data. The performance is then computed by mean squared error (MSE) and Pearson correlation coefficient(PCC), which are given in the following equations:

$$MSE(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(Y-X)^2, \quad (4)$$

$$PCC(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(\frac{X_i-\mu_X}{\sigma_X})(\frac{Y_i-\mu_Y}{\sigma_Y}), \quad (5)$$

where $X$ and $Y$ are the true OPI labels and estimated pain intensity level, respectively. $n$ is the number of sequences, $\mu$ and $\sigma$ correspond to the mean and standard deviation of their subscript vectors.

## 4 EXPERIMENTAL RESULTS

In this section, the results of proposed approaches are provided. Parameter adjustment should be conducted for all feature descriptors. Reasonably wide range of parameters are experimented to find efficient values for feature block sizes and SVR parameters. MSE and PCC of some tested parameters for all descriptors are depicted in Figures 5 and 6, respectively. The minimum MSE and maximum PCC are marked by triangles in each sub-figure.

Table 2 represents the best MSE and PCC results of all static and dynamic descriptors on UNBC-McMaster shoulder pain expression archive database. Best parameters of the features are provided as subscripts in this table. Parameters for both LBP and LBP-TOP are number of neighboring points around each central pixel (**P**), radius around each central pixel (**R**), number of divisions in row, in column and in time, respectively. In 2D and 3D-SIFT, parameters are the size of the extracted descriptor, number of bins along x axis, y axis and time divisions.

According to Figure 5 and the first four rows of Table 2, with respect to obtained MSE values, $LBP-TOP_{8,2,8,6,6}$ outperforms other models by 0.21 unit compared to the second best model. This result is in agreement with the acquired performance in the case of PCC measure for LBP-TOP model. However, optimal parameters of LBP-TOP model based on these two metrics are not the same.

Table 2: The best performance of all methods on UNBC-McMaster shoulder pain expression archive database. Subscripts are the parameters of each descriptor explained in section 4.

| **Feature descriptors** | MSE | PCC |
|---|---|---|
| $LBP_{8,2,8,7,5}$ | **1.81** | 0.76 |
| $DSIFT_{8,4,4,6}$ | **2.33** | 0.45 |
| $LBP-TOP_{8,2,8,6,6}$ | **1.53** | 0.74 |
| $3DSIFT_{8,3,3,10}$ | **1.74** | 0.61 |
| $LBP_{8,2,10,7,5}$ | 2.12 | **0.77** |
| $DSIFT_{8,4,4,10}$ | 2.40 | **0.48** |
| $LBP-TOP_{8,2,10,7,5}$ | 1.70 | **0.79** |
| $3D-SIFT_{8,4,4,10}$ | 1.80 | **0.64** |

From the least MSE point of view, dynamic features, including LBP-TOP and 3D-SIFT surpass the static feature descriptors, including LBP and DSIFT. Nevertheless, considering acquired PCC values, LBP family leads to superior outcome compared to the SIFT family.

Interestingly, with respect to either of the metrics, temporal feature descriptors in either of the feature families outperform the static feature descriptors of the same family. The reason is that, there is useful temporal information present in the sequences which boosts the performance of regression machine and this information might not be used by employing static feature descriptors. Although our obtained results are limited to UNBC-McMaster shoulder pain expression archive database, they are in agreement with (Zhao and Pietikainen, 2007; Scovanner et al., 2007) in this context.

Comparing 3D descriptor performances attained in our experiments, by either of the metrics, LBP-TOP gives superior results than 3D-SIFT. The same statement can be proposed for the corresponding 2D descriptors. This outcome shows the advantage of LBP family on the facial expression assisted pain intensity estimation applications and is correlated with the results obtained in many papers contributed in facial expression applications such as (Kaltwang et al., 2012).

## 5 CONCLUSION

Self-reported pain intensity level is not a reliable and always possible means of pain evaluation. Estimation of a patient's pain intensity using alternative solutions such as facial expression analysis is a functional and reliable indicator and this information can

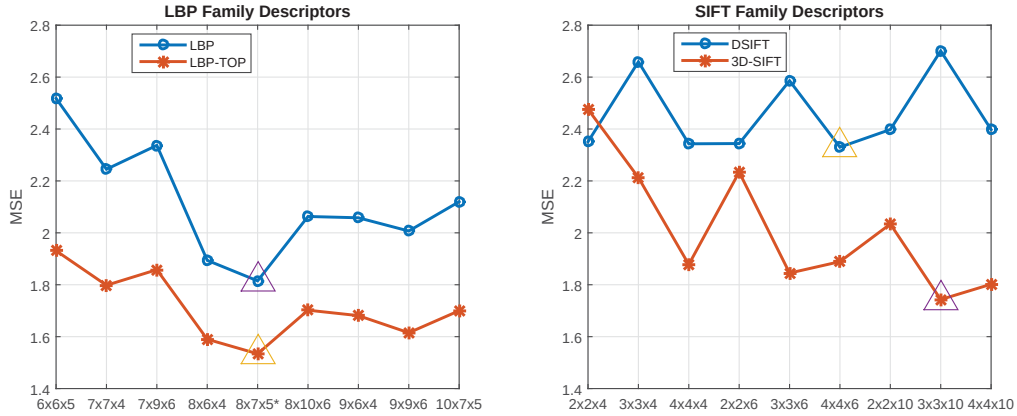



Figure 5: Acquired MSE over a different number of blocks. For each feature extraction technique(LBP, SIFT, LBP-TOP, 3D-SIFT), the minimum MSE is highlighted by a triangle. The x-axis tick labels are corresponding to row divisions (number of bins in x-axis) × column divisions (number of bins in y-axis)× time divisions regarding to LBP (DSIFT) and LBP-TOP (3D-SIFT). In the x-axis of left sub-figure, * corresponded to 8x6x6 for LBP-TOP and 8x7x5 for LBP.

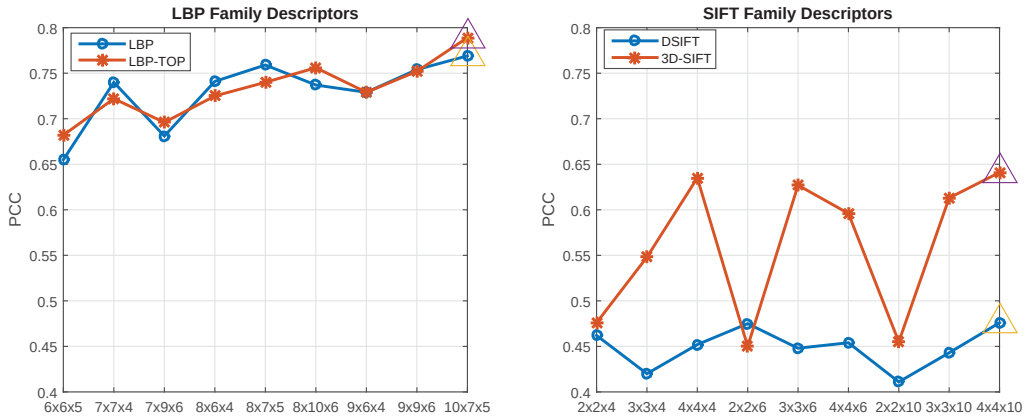



Figure 6: PCC over a different number of blocks. For each feature extraction technique(LBP, DSIFT, LBP-TOP, 3D-SIFT), the maximum PCC is accentuated by triangle. The x-axis labels are corresponding to row division (number of bins in x-axis) × column division (number of bins in y-axis)× time division with respect to LBP (DSIFT) and LBP-TOP (3D-SIFT).

be used for many clinical applications, e.g. drug dose management and monitoring. This solution is particularly advantageous for those patients who are not able to communicate reliably, including severely ill elder patients or immature patients. In this study, we employed four different feature sets, containing two static (LBP and DSIFT) and two dynamic descriptors (LBP-TOP and 3D-SIFT) in the application of pain intensity estimation from facial expressions. We have evaluated our models on the UNBC-McMaster shoulder pain expression archive database using SVR machine. Our experimental results underline the superior performance of dynamic models compared to the static ones. In addition, LBP family offers more descriptive information of facial expressions than SIFT family descriptors. LBP-TOP provides the most accurate results of regression by 1.53 and 0.79 as MSE and PCC values, respectively.

# REFERENCES

Ahonen, T., Hadid, A., and Pietikäinen, M. (2004). Face recognition with local binary patterns. In *Computer vision-eccv 2004*, pages 469–481. Springer.

Ahonen, T., Hadid, A., and Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041.

Ashraf, A. B., Lucey, S., Cohn, J. F., Chen, T., Ambadar, Z., Prkachin, K. M., and Solomon, P. E. (2009). The painful face–pain expression recognition using active appearance models. *Image and vision computing*, 27(12):1788–1796.

Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for

support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

Kaltwang, S., Rudovic, O., and Pantic, M. (2012). Continuous pain intensity estimation from facial expressions. In *Advances in Visual Computing*, pages 368–377. Springer.

Khan, R. A., Meyer, A., Konik, H., and Bouakaz, S. (2013). Pain detection through shape and appearance features. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, pages 1–6. IEEE.

Krig, S. (2014). *Computer Vision Metrics: Survey, Taxonomy, and Analysis*. Apress.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

Lucey, P., Cohn, J. F., Matthews, I., Lucey, S., Sridharan, S., Howlett, J., and Prkachin, K. M. (2011a). Automatically detecting pain in video through facial action units. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(3):664–674.

Lucey, P., Cohn, J. F., Prkachin, K. M., Solomon, P. E., Chew, S., and Matthews, I. (2012). Painful monitoring: Automatic pain monitoring using the unbc-mcmaster shoulder pain expression archive database. *Image and Vision Computing*, 30(3):197–205.

Lucey, P., Cohn, J. F., Prkachin, K. M., Solomon, P. E., and Matthews, I. (2011b). Painful data: The unbc-mcmaster shoulder pain expression archive database. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 57–64. IEEE.

Neshov, N. and Manolova, A. (2015). Pain detection from facial characteristics using supervised descent method. In *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2015 IEEE 8th International Conference on*, volume 1, pages 251–256. IEEE.

Ojala, T., Pietikäinen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59.

Ojala, T., Pietikäinen, M., and Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987.

Rathee, N. and Ganotra, D. (2015). A novel approach for pain intensity detection based on facial feature deformations. *Journal of Visual Communication and Image Representation*, 33:247–254.

Roy, S. D., Bhowmik, M. K., Saha, P., and Ghosh, A. K. (2016). An approach for automatic pain detection through facial expression. *Procedia Computer Science*, 84:99–106.

Scovanner, P., Ali, S., and Shah, M. (2007). A 3-dimensional sift descriptor and its application to action recognition. In *Proceedings of the 15th international conference on Multimedia*, pages 357–360. ACM.

Vedaldi, A. and Fulkerson, B. (2010). Vlfeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1469–1472. ACM.

Zhao, G. and Pietikainen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(6):915–928.

# WorkUp: A Mobile Application to Support Health Guidelines

Vinicius dos Santos[1], Henrique Yoshikazu Shishido[1], Francisco Pereira Junior[1] and Gabrielle Jacklin Eler[2]

[1]*Departament of Computing, Federal Technological University of Paraná, 1640 Alberto Carazzai Avenue, 86300000, Cornélio Procópio, Paraná, Brazil*

[2]*Collegiate of Nursing, Federal Institute of Paraná, 600 João XXIII, 86060370, Londrina, Paraná, Brazil*

*vinistos@gmail.com, {shishido, fpereira}@utfpr.edu.br, gabrielle.eler@ifpr.edu.br*

Keywords: Health Status, Indicators, Public Health, Monitoring.

Abstract: Objective: This paper presents a model of mobile application to assess patients and prescribe physical exercises offering interaction among health professionals and patients. Methods: The project is based on mobile platform and implemented using client-server architecture and cloud computing for data synchronization in different devices. Results: Health professionals and patients tested our application and answered questionnaire. The results indicate that the functionality and usability are satisfactory adhesion to our app design. Conclusion: Our approach may be a candidate model to government agencies to support in prevention of obesity and improve the health indicators of the patient to a healthier life.

## 1 INTRODUCTION

Communicable and non-communicable diseases are increasing in south of Brazil, becoming the main cause of death along with obesity risk factors (Capilheira et al., 2008). Prevention and early diagnosis of obesity are important for health promotion and the reduction of morbidity and mortality. Obesity has a direct impact in the individual's social acceptance due to the aesthetic concept widespread in contemporary society (Schmidt et al., 2011). A study conducted by the Brazilian Institute of Geography and Statistics showed overweight in 50.1% of men and 48% of women (da Saúde, 2014). Actually, there is evidence that the percentage of overweight in the global population has reached approximately 60% (Popkin, 2011).

Mobile health (mHealth) is the use of mobile computing and communication technologies in health care and public health (Free et al., 2010). Mobile applications are an option to support government agencies in order to monitor population health indicators, and manage the physical activities both of children and adults. In Brazil, government programs developed by the Ministry of Health, such as Food and Nutrition Surveillance, Family Health Strategy, and the School Health Program are part of the National Policy for Health Promotion. These programs have the physical activity as one of their priorities. Individuals diagnosed with chronic diseases and abnormal anthropometric data, should be referred to the Health Unit for treatment and monitoring (da Saúde, 2014).

Numerous systems have been developed in healthcare using mobile technologies. The applications enable collect data, usually by a questionnaire to assist public policies of disease control (Morrison et al., 2014). Recent studies have covered applications that promote the practice of physical activities that meet fundamental characteristics of mHealth and tracks physical activity and food comsuption behavior data (Al Ayubi et al., 2014). According to a recent mHealth apps review (Knight et al., 2015), there is no application that supports specifically public guidelines for aerobic physical activity (Tucker et al., 2011).

This study aimed to develop a mobile application model that supports public guidelines for aerobic physical activity. This paper is not focused neither in deploy nor test the app on site. It provides an opportunity to support government health programs. We show the development of application, called WorkUp. It enables the health professional such as physical educator linked to government health programs to evaluate patients and track medical histories with a variety of methods, such as: a) calculation of body fat index; b) determination of types of exercises and daily activities; c) class scheduling; d) interaction between both professional and patient; and, e) app data can be accessed with any mobile device connected on the Internet.



Santos, V., Shishido, H., Junior, F. and Eler, G.
WorkUp: A Mobile Application to Support Health Guidelines.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 297-304
ISBN: 978-989-758-213-4

The remaining part of the paper proceeds as follows: we discussed about related studies in second section. Section three presents the technologies, development process, software architecture, and tests and validation methods. In section four, the project implementation is presented. The fifth section is presented a qualitative evaluation about our design acceptance through users interview. Finally, we present our conclusion and limitations of our study in the section six.

## 2 RELATED WORK

Studies in several countries show success in combating obesity using physical activity together with combined interventions: changes to food/diet, increased physical activity, and behavioral strategies (Jebb et al., 2011; Turner et al., 2012). Face-to-face approaches are effective, but they are relatively expensive to implement, difficult to scale up, and do not suit those who work or live far from venues. The addition of mHealth apps in everyday life offers a practical and potentially cost-effective solution to the barriers of face-to-face approaches (Waterlander et al., 2014).

Researches demonstrated strong support for a mHealth weight management intervention; with 75% saying they would use a mobile app for weight management intervention (Gorton et al., 2011). Another study showed that the widespread use of information and communication technology tools offers an innovative and potentially beneficial avenue to increase the level of physical activity in Heart Failure (HF) patients (Franklin, 2015).

A survey performed in 2012, shows 50 applications available in Brazil within the health field, of which 20 belong only to the Android platform, 19 to the iOS platform, 6 on the Windows Phone platform, 1 supporting the three platforms (Diet and Health), and 4 common to Android and iOS platforms (SUS Procedures, SAESP, Measure One and Emagrecentro). They are mostly free (32 apps), but the Android platform leads to greater amount of paid applications. The content of more than half of these applications is aimed at professionals. The applications targeting the consumer public deal primarily with issues related to diet and physical conditioning (Bonome et al., 2012).

But still, comprehensive applications that enable a range of information about the patient and that allow the participation in the evolution process of his general state of health and weight are necessary. This is why we purposed and developed the WorkUp, based on users' needs becoming part of a National Policy for Health Promotion.

WorkUp, as other studies, has shown interest and adherence on the part of those involved. Some authors demonstrated that the success level of health services depends on the level of user acceptance and adoption (Sezgin and Yıldırım, 2014). Additional research is needed to explore several hypotheses. For example, if more engaging user interfaces, including easier navigation, simpler layouts, and refined aesthetics can contribute to the adherence of mHealth apps. Then, from this point of research design, the emerging studies should include more qualitative approaches and longitudinal studies in order to effectively understand the user's needs (Kaplan and Maxwell, 2005), which can strengthen the qualitative research conducted for the WorkUp.

## 3 METHODS

For the application's development, the system planning, design, and implementation were performed. Later, an evaluation of usability was applied in patients and health professionals.

### 3.1 Technologies and Tools

We adopted Android platform due to its highly used technology, its ability to support wide hardware compatibility, and to offer many components that provide a good user interaction. The codification is based on Java and eXtended Markup Language (XML) files describe in the user interface. The Android Development Toolkit (ADT) was adopted for application test through the use of its virtual device emulator.

The artefacts and requirements were modelled in Unified Modelling Language (UML) to design system diagrams using Astah Community Edition. The database was designed in BrModelo software that generates database script commands. SQLite was the system database employed for data persistence due to its native support for Android, good performance and low resources usage. In Android Development Tools (ADT) contains a manager for SQLite databases in order to handle data manually during development.

WorkUp data is stored on the cloud to get a wide geographic access and allow users access to their synchronized data, regardless of the device. Web service technology was chosen to synchronize the local data with the cloud, as it works as a neutral technology that can transmit data and be easily adapted to any other technology.

### 3.2 Data Synchronization

The application server was deployed on Amazon EC2 Services. The operating system was Linux Ubuntu 14.04, Java 8, and MySQL database to store all data through web services.

When our application is installed on a mobile device, it starts a service that periodically communicates with the data synchronization web service. Clients perform a query to check available updates in cloud database. The communication model is represented in Figure 1. If there is an update, the appropriate routine will be started automatically to synchronize the data.

The security of data communication can be treated in three ways. Firstly data is encrypted before it will be sent to server or retrieved from it. Before the data transmission to cloud, data is encrypted and its integrity is checked when arrives on the other side. In local SQLite database, there are libraries to encrypt data such as SQLCipher and SQLiteCrypt.

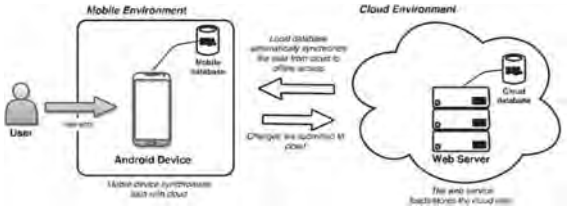



Figure 1: Data synchronization model of WorkUp.

## 4 PROJECT RESULTS

This section describes our app structure and use-flow. Our application has two user types: health professional (HP) and patient (P). Each one performs different activities on app. Both user type share functionalities such as account activities and view patient assessment, training set and class schedule. However, there are particular functions, which each user type can access. Health professional are able to create patient assessments, create or reuse training sets and manage class schedules. These functionalities are organized in modules, which are detailed in Table 1.

The basic use-flow consists in both user types create an account on app. After, health professional should add a relationship with his respective patients. The patients list of a health professional must be defined by the health unit. The patient must accept the request for relationship with the health professional. For each patient, the health team can schedule an assessment in order to verify patient health indicators. This assessment can be performed periodically in order to track the patients' health progress. Then, the physical educator can create a training set based on patient's physical needs and send it to him. The execution of exercises is registered by the patient on app, and it can be visualized by the physical educator. According to health program policy, there is an option to the HP schedule classes with patients individually or in groups. If the patient agrees with proposed schedule, he can accept it. Otherwise, the patient can message his health professional asking another available time.

### 4.1 Profile Management and User Communication

Local data persistence and guidelines to allow communication between the mobile app and web service to validate login (F1) were some of the features implemented. The Facebook API was integrated to provide access to the app using a Facebook account (F2). A notification and synchronization service (F16) had to be created in order to keep communication between the patient and the professional. We have also developed functions such as search user (F5), with the option of adding other users to the contact list. As stated earlier, a professional can establish connection with several patients. When clicking on a patient, the professional can manage his patients through an interface that allows access to assessments, training prescriptions, class schedule, and the removal of the patient (F6) from his list. The concern of the third sprint was the implementation of the interfaces for personal data handling through control panels, where the user has access to information and can update it as desired (F7).

### 4.2 Management of Assessments

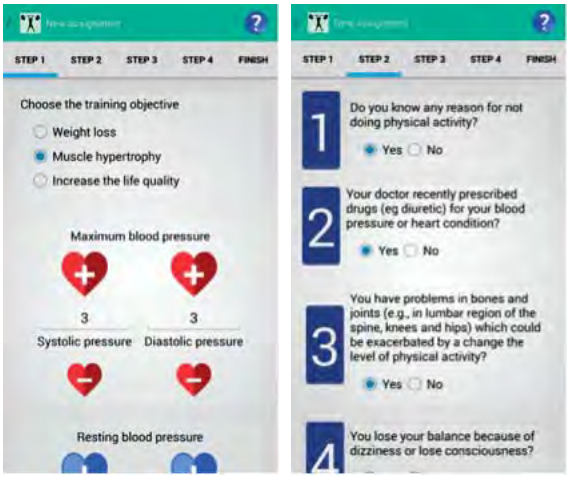
The second module deals with the management of assessments. It was concerned in keeping some characteristics in data collection, such as the interfaces understanding, separation of each type of evaluation, upload to web server, and both view and delete assessments.

The patient assessment tool (F8) was divided into four parts: a) verification of goals with the new training routine and verification of cardiorespiratory health through the maximum blood pressure during exercise and blood pressure at rest, as shown in Figure 2a; b) questionnaire of medical history through Pfeiffer questionnaire, which detect if there is any history or risk involving cardiorespiratory shortcomings, pain, recurrent chronic problems in implementing physical activity, loss consciousness, coronary problems or some other risk factors (Figure 2b); c)

Table 1: Functionalities of each use-case allowed for Health Professional (HP), Patient (P), and System (S).

| Use Case | ID | Functionality | HP | P | S |
|---|---|---|---|---|---|
| Manage Profile | (F1) | Standard login | X | X | - |
| | (F2) | Facebook login | X | X | - |
| | (F3) | Create a standard account | X | X | - |
| | (F4) | Request relationship to professional | X | - | - |
| | (F5) | Search a user | X | X | - |
| | (F6) | View assessments, training set, schedule of patient | X | X | - |
| | (F7) | Update account | X | X | - |
| Perform Assessment | (F8) | Create a new patient assessment | X | - | - |
| | (F9) | Send collected data to server | - | - | X |
| | (F10) | Notify patient if new assessment is available | - | - | X |
| Training Management | (F11) | Create a training set | X | - | - |
| | (F12) | Send completed exercise to server | - | - | X |
| | (F13) | Reuse a training set | X | - | - |
| Class Schedule | (F14) | Schedule a new class | X | - | - |
| | (F15) | Confirm a class schedule | - | X | - |
| Non-functional | (F16) | Notification and synchronization service | - | - | X |

the body fat assessment through a variety of skinfold methods; and, d) perimetry assessment, which enables log measures of body parts. Beyond the numerical evaluation functionality was integrated using the smartphone camera for photographic record of the evaluated body members.




(a) (b)

Figure 2: Evaluation management screens: (a) Objectives and blood pressure; (b) Questionnaire (QPAF); (c) Body fat calculation.

Existent apps adopt a unique skinfold method. However, each professional need a specific application, according to the physical characteristics of the patient. So, in this sense, WorkUp has a variety of skinfold methods to calculate body fat. We have implemented both Jackson Pollock 3 and 7-site, and Guedes methods. At the end of each assessment, the professional needs Internet access to send the collected data (F9) and automatically notify the patient (F10) that there is a new assessment available.

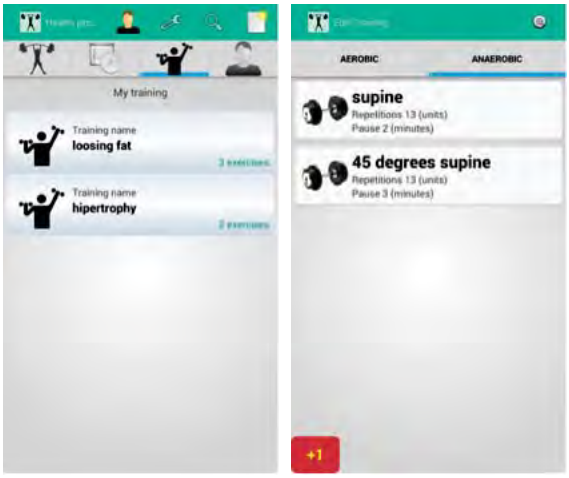
## 4.3 Training Management

The training has two types of exercises, which can be created by the own professional. Some rules were added to the application to improve system usability. After, it was marked by the inclusion of functionality to create a training list (F11) according to the purpose of the patient, inserting both aerobic and anaerobic exercises. Only the professional can mount the training set, leaving the patient with the exercises only. Following the conclusion of training, the patient's collected data are sent to a web database (F12) and will be available for professional access to track the patient's evolution.

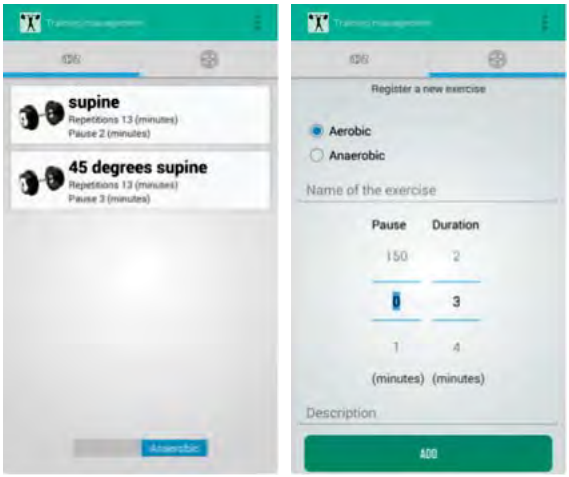
These functionalities were developed and are shown in Figure 3. The training sessions that have already been created can be reused by the professional (F13), as shown in Figure 3a. In order to add a new training, it is necessary to enter a name, add exercises by clicking on the +1 button (Figure 3b), and view the registered exercises classified between aerobic and anaerobic (Figure 3c), with name, duration, repetition times, and rest duration (Figure 3d).

## 4.4 Class Management

This module allows the application to control the class scheduling (F14) to improve communication between patient/professional. To use it, the patient needs to bound to the professional, so the patient can schedule




(a) (b)



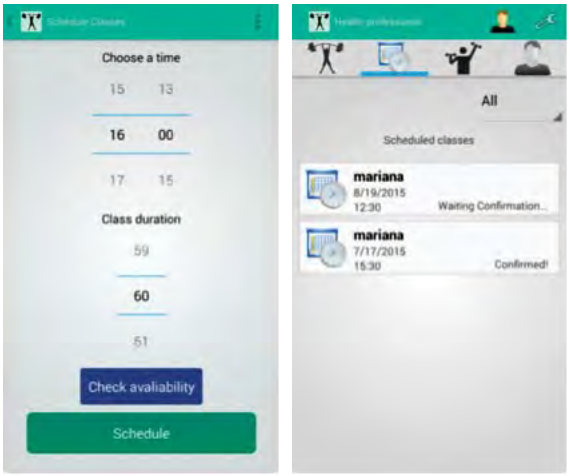

(c) (d)

Figure 3: Training management screens: (a) My trainings; (b) Edit training; (c) Registered exercises; (d) Characteristics of exercises.

a new class and the health professional can accept it or not. For a confirmation of a new class (F15), it is necessary that both agree and confirm the date and time. If there are any changes, both users will be notified by the system.

For the scheduling of a new class, the interface displayed in Figure 4a was built, requiring the patient/professional to select a date and time, with the ability to verify if the selected time is available. Users are notified to change the status of the class to confirmed, registering the schedule of classes. (Figure 4b).

In addition, we have considered good practices for interface design in order to assist the user of our application.




(a) (b)

Figure 4: Classes management screens: (a) Schedule new class; (b) User calendar.

## 5 EVALUATION

In order to validate our purpose, we adopted the qualitative research approach because it reveals a target audience's range of behaviour and perceptions. It uses in-depth opinions of small groups of people to support the construction of hypotheses through descriptive statements. The categories extracted from the answers of the participants were: a) effective registration method; b) ability to monitor more patients; c) accessibility of information; d) information distortion; and, e) interface and restrictions.

Sixteen participants, including health professionals and patients in both Cornélio Procópio and Sertaneja cities, located in south Brazil, tested our application.

Several interfaces were sketched and tested by four health professionals and twelve patients. The app validation was performed by 17 questions focused on usability of the system. Participants were divided into HP1 to HP4 (Health Professional) and P1 to P12 (Patients). Participants were asked to download the application in order to explore and test the app functionality. Next, the participants were requested to answer the questionnaire. All data, comments, suggestions and detected problems were analyzed and used as a source of information to evaluate the system. Responses were categorized and described19-21 as: a) effective registration method for physical educators; b) ability to monitor more patients; c) accessibility of information; d) information distortion; and, e) interface and restrictions.

- Effective registration method As main results,

participants emphasized that the control of physical activity is important for physical educators. Interviewed professionals highlighted the need to establish goals and analyze if the results obtained have achieved them. According to the answers, it is difficult to investigate the causes of inefficiency in the physical training because of the lack of effective registration methods to collect the patient's exercise routine. Table 2 presents the users reports referring to the registration method.

Table 2: User experience report about effective registration method from Health Professionals (HP).

| User experience report |
|---|
| HP 1: "It is difficult to monitor patients efficiently without a digital record. Most of the time there is only an initial assessment and a pre-prepared program for the patient to conduct over a period of time without a complete record of case histories and prescribed activities." |
| HP 2: "You need an environment where the teacher can record all information, such as lessons, exercises and schedules, in order to effectively monitor the patient. In addition, through the systematization of information, you can organize and visualize where the gaps are in training." |
| HP 3: "It is necessary to check how the patient's training was in relation to the time for without this analysis, all is vague." |
| HP 4: "I need to know if the patient is actually losing weight with the training, and if the tracing objective is being achieved." |

- Ability to monitor more patients It is common for a health professional to track multiple patients at the same time. This scenario is unfavorable for the quality of training provided to patients. Some interviewees pointed out that our app could help professionals to follow a larger number of patients with high quality. Table 3 shows the discourse of users that have tested the application, analyzing the resources to follow multiple patients.
- Accessibility of information Among the questions, the accessibility of the information required for patients' assessment was mentioned. The health professionals noted that some data are relatively complicated to extract, but are necessary for the proper development of activities, as described in Table 4.
- Information distortion The professionals point out that, in some cases, patients promote distortions in data collection, omitting important information about their health status (Table 5). There are cases

Table 3: User experience reports about ability to monitor more patients.

| User experience report |
|---|
| HP 2: "... I have on hand an easy way to record the data of my patients, which this training should follow. In addition, the schedule allows better distribution of times so I can organize myself." |
| HP 3: "It bothers me not keeping up with the patients, and not even knowing their workout for that day, because I cannot handle looking at the previous training and make the outline of the next." |

Table 4: User experience report about accessibility of information.

| User experience report |
|---|
| HP 1: "... I have on hand an easy way to record the data of my patients, which this training should follow. In addition, the schedule allows better distribution of times so I can organize myself." |
| HP 2: "It bothers me not keeping up with the patients, and not even knowing their workout for that day, because I cannot handle looking at the previous training and make the outline of the next." |
| P4: "I think the data is very specific and understanding it is complex." |
| P11: "At first sight, the need for data is difficult to understand, but once it is properly explained it is easy to understand its importance." |

where patients find unnecessary to fill out personal data in the application. The patient generally has no idea of the importance of evaluation for the prescription of activities.

Table 5: User experience report about distortion of information.

| User experience report |
|---|
| HP 1: "There are some points that can be filled incorrectly, which can work as a motivation to create new ways to reduce distortions." |
| HP 2: "The system as presented does not have methods that can prevent inconsistencies, for example, a patient who does not perform the exercises correctly." |
| P7: "We cannot guarantee that the data entered by the patients during the execution of exercises are really reliable." |
| P10: "This app would be of little use because I'm not used to perform this type of control." |

- Interface and restrictions When participants were

asked about the interface, system language, efficiency in performing the activities, arrangement of dialog boxes, error messages, and system iconography, there were no serious abnormality highlighted. But the health professionals pointed out that a more restrictive system would be interesting for more effectiveness in achieving results (Table 6).

Table 6: User experience report about interface and application restrictions.

| User experience report |
|---|
| HP 3: "The application is very pleasing to the eye and easy to understand." |
| HP 4: "It would be interesting if there was a way to make it easier for patients to access the training to be carried out." |
| P6: "The interface is pretty nice and does not show navigation problems." |
| P9: "I found no serious problems." |
| P12: "Some data is complicated to collect and understand, it needs an improved interface." |

# 6 CONCLUSION

Physical evaluation software has been developed to desktop computers. However, with popularization of mobile technologies, the adoption of mobile apps is undoubtedly a useful solution. For health professionals, the advantage of portability overcomes some barriers and allows the tracking of training, improving professional-patient communication, and allowing the follow-up of training guidelines with more effectiveness.

The WorkUp can be a model to support health agencies to promote a healthier life and keep record of the patient physical assessments. The application also provides users with the ability to keep tracking important health indicators that show the patient's evolution. As pointed out earlier, mobile technology is constantly evolving in the applications market, that is, every day there are new startups focusing on health and exercise routines. Therefore, WorkUp has to add up to this scenario, providing important information for health professionals. This information can assist professionals to achieve better results and improve the quality of life of their patients.

A limitation of this study relates to the performance of physical activities, as WorkUp currently restricts to patients the ability to perform physical activities proposed by a professional. However, these mechanisms are dependent on the use of the correct interface. The system is not integrated to sensors. Thus, for future work the integration of these technologies will be accomplished to provide better control of physical activities.

# ACKNOWLEDGEMENTS

The authors acknowledge the support of UTFPR.

# REFERENCES

Al Ayubi, S. U., Parmanto, B., Branch, R., and Ding, D. (2014). A persuasive and social mhealth application for physical activity: A usability and feasibility study. *JMIR mHealth and uHealth*, 2(2).

Bonome, K. d. S., Di Santo, C., Prado, C., Sousa, F., and Pisa, I. (2012). Disseminação do uso de aplicativos móveis na atenção a saúde. In *XIII Congresso Brasileiro de Informática em Saúde*.

Capilheira, M. F., Santos, I. S., Azevedo Jr, M. R., and Reichert, F. F. (2008). Risk factors for chronic noncommunicable diseases and the carmen initiative: a population-based study in the south of brazil. *Cadernos de Saude Publica*, 24(12):2767–2774.

da Saúde, M. (2014). *Strategies for the care of the person with chronic disease: Obesity*. Departamento de Atenção Básica.

Franklin, N. C. (2015). Technology to promote and increase physical activity in heart failure. *Heart failure clinics*, 11(1):173–182.

Free, C., Phillips, G., Felix, L., Galli, L., Patel, V., and Edwards, P. (2010). The effectiveness of m-health technologies for improving health and health services: a systematic review protocol. *BMC Research Notes*, 3(1):250.

Gorton, D., Dixon, R., Maddison, R., Mhurchu, C. N., and Jull, A. (2011). Consumer views on the potential use of mobile phones for the delivery of weight-loss interventions. *Journal of Human Nutrition and Dietetics*, 24(6):616–619.

Jebb, S. A., Ahern, A. L., Olson, A. D., Aston, L. M., Holzapfel, C., Stoll, J., Amann-Gassner, U., Simpson, A. E., Fuller, N. R., Pearson, S., et al. (2011). Primary care referral to a commercial provider for weight loss treatment versus standard care: a randomised controlled trial. *The Lancet*, 378(9801):1485–1492.

Kaplan, B. and Maxwell, J. A. (2005). Qualitative research methods for evaluating computer information systems. In *Evaluating the organizational impact of healthcare information systems*, pages 30–55. Springer.

Knight, E., Stuckey, M. I., Prapavessis, H., and Petrella, R. J. (2015). Public health guidelines for physical activity: is there an app for that? a review of android and apple app stores. *JMIR mHealth and uHealth*, 3(2).

Morrison, Z., Fernando, B., Kalra, D., Cresswell, K., and Sheikh, A. (2014). National evaluation of the benefits and risks of greater structuring and coding of the electronic health record: exploratory qualitative investigation. *Journal of the American Medical Informatics Association*, 21(3):492–500.

Popkin, B. M. (2011). Is the obesity epidemic a national security issue around the globe? *Current opinion in endocrinology, diabetes, and obesity*, 18(5):328.

Schmidt, M. I., Duncan, B. B., e Silva, G. A., Menezes, A. M., Monteiro, C. A., Barreto, S. M., Chor, D., and Menezes, P. R. (2011). Chronic non-communicable diseases in brazil: burden and current challenges. *The Lancet*, 377(9781):1949–1961.

Sezgin, E. and Yıldırım, S. Ö. (2014). A literature review on attitudes of health professionals towards health information systems: from e-health to m-health. *Procedia Technology*, 16:1317–1326.

Tucker, J. M., Welk, G. J., and Beyler, N. K. (2011). Physical activity in us adults: compliance with the physical activity guidelines for americans. *American journal of preventive medicine*, 40(4):454–461.

Turner, M., Burns, S. M., Knight, L., Ward, K., Garo, A., Morris, T., Hooper, E., and Conaway, M. (2012). Weight management practices among heart and vascular health care providers in an ambulatory setting. *Medsurg Nursing*, 21(4):222.

Waterlander, W., Whittaker, R., McRobbie, H., Dorey, E., Ball, K., Maddison, R., Smith, K. M., Crawford, D., Jiang, Y., Gu, Y., et al. (2014). Development of an evidence-based mhealth weight management program using a formative research process. *JMIR mHealth and uHealth*, 2(3):e18.

# Automatic Visual Detection of Incorrect Endoscope Adaptions in Chemical Disinfection Devices

Timo Brune[1], Björn Brune[2], Sascha Eschborn[2] and Klaus Brinker[1]
[1]*University of Applied Sciences Hamm Lippstadt, Marker Allee 76-78, Hamm, Germany*
[2]*Olympus Surgical Technologies Europe, Kuehnstraße 61, Hamburg, Germany*
*{timo.brune, klaus.brinker}@hshl.de, {bjoern.brune, sascha.eschborn}@olympus-oste.eu*

Keywords: Computer Vision, Feature Detection, Surf, Sift, Registration, Machine Learning, Supervised Learning, Endoscopes, Disinfection.

Abstract: This paper presents a complete analyzing system for detecting incorrect endoscope adaptions prior to the use of chemical disinfection devices to guarantee hygienic standards and to save resources. The adaptions are detected visually with the help of an image registration algorithm based on feature detection algorithms. On top of the processing pipeline, we implemented a k-nearest neighbor algorithm to predict the status of the adaption. The proposed approach shows good results in detecting the adaptions correctly.

## 1 INTRODUCTION

Endoscopic diagnostic is the main application for diseases of the gastrointestinal tract and has a huge clinical relevance. An important part of endoscopes is the quality of the preprocessing of the devices and the resulting hygiene, to minimize the contamination of the patient with microbes (Bader, 2002). In the past microbes have developed resistances against antibiotics. Consequent hygiene is therefore indispensable. Muscarella (Muscarella, 2014) showed that insufficient preprocessed endoscopes are responsible for the contamination with CRE-microbes. To guarantee an acceptable hygienic standard, we need to disinfect the endoscopes. To do this in a constant quality we apply supervising preprocessing of the endoscopes.

So called cleaning and disinfection devices for endoscopes (CDD-E) perform well in cleaning the endoscope's exterior and interior, where the procedure of adaption is rather complex. Medical employees often do not adapt the endoscopes correctly to the CDD-E because of this complexity. These adaption errors lead to a lack of hygiene. The CDD-E is able to detect these errors and can terminate the process of cleaning. An interruption always costs operational time of up to 20 minutes, water, cleaning chemicals and energy. We have implemented a system tailored to detect those adaption errors prior to the disinfection to save these resources and ensure the quality of the preprocessing.

We consider the margin between an endoscope adapter and its adaption counterpart in the chemical disinfection device in order to detect connection faults. We have transformed the underlying problem of determining the size of the gap between the respective parts into an image registration problem (Handels, 2009). Hence we want to try to align two reference images of the two sides of the adapter to an image of the disinfection device which contains the endoscope and the adaption counterparts. Please note that this new image is a 2d-projection of the underlying 3d-scene. In this paper we make use of a feature-based approach to image registration (Zitova, 2003). The first step in this processing pipeline is to detect feature points for each image independently. In a second step corresponding feature points on different images are matched.

We detect feature points with two different feature detection algorithms which detect, describe and match possible correspondences and compare their performance on our problem. More precisely we choose the algorithms scale-invariant feature transform (SIFT) (Lowe, 2004) and speeded up robust features (SURF) (Bay, 2008). We describe these feature detection algorithms in more detail in Chapter 2.2.4.

On top of the extracted features we use a simple k-nearest neighbor algorithm to classify correct / incorrect adaptions. More details are given in



Brune, T., Brune, B., Eschborn, S. and Brinker, K.
Automatic Visual Detection of Incorrect Endoscope Adaptions in Chemical Disinfection Devices.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 305-312
ISBN: 978-989-758-213-4

Chapter 2.2.5. Finally we show promising experimental results on the accuracy of our automatic error detection prototype in Chapter 3.

# 2 PROCEDURE

## 2.1 Hardware Setup

To accomplish the tasks of detecting errors in the adaption, a complete prototype was built. The system consists of a loading station wherein the detection occurs and the integrated software which is responsible for the processing pipeline of the detection. As a basis for our image processing are four images taken by four high-resolution cameras from four different perspectives in order to capture images without occlusions of relevant parts with a high probability. Figure 1 shows an example of an image which was taken by one of the cameras.

For optimal image processing it is important that the scene is illuminated using controlled light conditions (Bader, 2002). Because of that the detection system is sealed in a cube to prevent diffuse light entering the system. We enlighten the detection system with a planar LED-Panel.

Camera sensors take 50% of the information from the green interval of the spectrum. Green light transfers approximately 72% of the luminance and is there-fore most important for contrast and resolution (Fukanaga, 1975) Because of that we chose an LED-Panel with a color temperature of 6000 °K. According to Wiens Law, the maximum radiation is at 482.95 nm for 6000 °K which is within the green part of the spectrum.

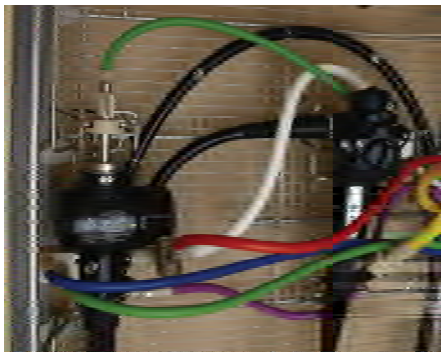


Figure 1: Example image of the scene, taken by one of the four cameras.

The cameras are controlled using a serial interface from the software to automate the system. With the software we are also able to manipulate the settings of the camera. So it is possible to adjust the aperture, the ISO-value and the white balance. The white balance is analogically set to 6000 °K. The other values were set empirically , so that there is no over-exposure and therefore no loss of information. We use Canon EOS 750D cameras with a resolution of 24 Megapixel. A high resolution is essential for an accurate detection.

Figure 2: Prototype.

We consider the task of detecting adaption errors for a variety of different endoscopes and adapters. In order to detect which endoscope is used, it is tagged with an RFID chip. We use an RFID-reader, also connected via a serial interface to differentiate between them through their integrated RFID-tags. On the basis of the detected endoscope the software determines which and how many adapters should be connected to the respective endoscope. The user gets visual information on how to attach these adapters to the detected endoscope. Figure 2 shows the general setup of our prototype. At the front there is a door which is not visible in the image.

## 2.2 Algorithmic Components

In order to classify whether an adapter is connected correctly or not we use a pipeline of algorithms which are explained in more detail in this chapter.

### 2.2.1 Image Registration

As explained in the introduction, we need to measure the size of the gap between the two sides of the adapter. We took several reference image pairs of each adapter as shown in Figure 3.

Figure 3: left: two parts of the same adapter, right: both adapters matched together show a complete view of the adapter.

If we combine those two parts of the image we get a complete view of the adapter. Our goal is to map and accordingly register these two-dimensional reference images independently on a two dimensional projection of the scene. For image registration, two images are needed. The first is called reference image denoted $B_R$, and the second is called template image denoted $B_T$. This leads to a mathematical optimization problem. We search for a linear mapping $t: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ that maps the object on $B_R$ the most exact to the object of $B_T$. Depending on the class of transformations we can differentiate between rigid, affine and perspective transformation. The positions of the endoscope and thus the adapters are completely unknown and only depend on the employee of the hospital. Therefore, we are not able to reach our goal with a simple rigid or affine transformation. We have to describe a three-dimensional transformation in space as a two-dimensional projection in the image layer (Schreer, 2005). A projection like this is defined as an endomorph description of a scene (Fischer, 2014).

A property of perspective projection is the loss of proportions. Objects that are further away from the center of projection appear smaller.

We can define this transformation as a $3 \times 3$ matrix

$$T_P = \begin{pmatrix} \cos\alpha\, C_x & -\sin\alpha\, S_v & T_x \\ sin\,\alpha\, S_h & \cos\alpha\, C_y & T_y \\ P_x & P_y & 1 \end{pmatrix} \quad (1)$$

which we apply to every position of the reference image.

It is now known which form the transformation matrix has to have to map a two-dimensional reference image into the three-dimensional space and map it again to a two-dimensional projection. We have to determine the nine degrees of freedom which uniquely define the matrix. To find the required parameters, we need two sets of corresponding points. $A$ is a set of points from the reference image and $A'$ is a set of corresponding points from the template image. Here $T_P$ is the transformation matrix of the perspective projection.

$$A \cdot T_P = A' \Leftrightarrow T_P = A^{-1}A' \quad (2)$$

To determine the transformation matrix we need to have correspondences of points in the reference and template image. In the following chapter we show how these correspondences are detected.

### 2.2.2 Feature Detection

Feature detection algorithms are methods from the field of computer vision. We use them to detect so called interest points and correspondences between points in two images. These images typically show the same object, but at a different time or from a different perspective. In the experiments we will analyze two well established algorithms to examine which one is more appropriate for this application field. The two algorithms are the scale-invariant feature transform (SIFT) (Lowe, 2004) and the speeded up robust features (SURF) (Bay, 2008). These two algorithms have the same general process, which is divided into three steps: 1. feature detection, 2. feature description and 3. feature matching.

The first step, feature detection, deals with the detection of so-called interest points. These are distinctive points in an image. They always depend on their neighborhood.

The feature description deals with the description of the detected interest points, enabling a comparison between the reference and the template images. The most significant feature is the surrounding of one point. Since the surroundings of the interest points are never exactly the same on the reference and the template image, a pixelwise comparison would not work robustly. Furthermore, the descriptors have to be invariant against geometric, perspective and illumination transformations and image noise. Both algorithms are based on computing one gradient of the complete neighborhood of an interest point, as well as in their sub regions.

The final step of the feature detection algorithms matches interest points in the reference and template image. The challenge is to find correct correspondences of points, which in fact show the same points of an object (Szeliski, 2011).

Every interest point is described through a multidimensional description vector. A similarity can be evaluated with the Euclidean distance between the two descriptors. The most accurate but slowest method is to compare every interest point of

the reference image with every interest point of the template image. Since accuracy is one of our main goals, we make use of this algorithm in the experimental section. Other possibilities are the randomized k-d tree (Silpa-Anan, 2008) and the priority k-means tree algorithm (Fukanaga, 1975). These algorithms are up to two times faster, approximate to 95% of correctness (Muja, 2014).

Despite the high accuracy of SIFT and SURF, there are always a few correspondence errors which have to be filtered. We filter these errors with the well-established random sample consensus algorithm (RANSAC) (Strutz, 2016). If these errors were not filtered they would have a bad influence on the computation of the transformation matrix.

### 2.2.3 Machine Learning

To classify if an adapter is adapted correctly we use a simple k-nearest neighbor approach. As features we use the Euclidean distance between the two projections and the number of detected correspondences. Figure 4 shows an example of registered features on one of the adapters. In this example the classes are linear separable.

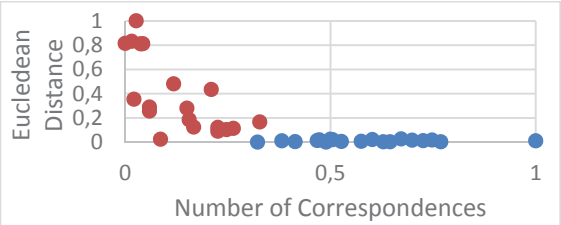



Figure 4: Normalized values of the Euclidean distances and the number of correspondences, red: incorrect, blue: correct.

### 2.2.4 Processing Pipeline

In this section we explain the complete processing pipeline. Figure 5 shows four template images made at runtime as explained previously.

We take five reference image pairs for every adapter. One part of the pair shows the part of the adapter at the endoscope, the other one the part of the tube. The process is the same for every adapter.

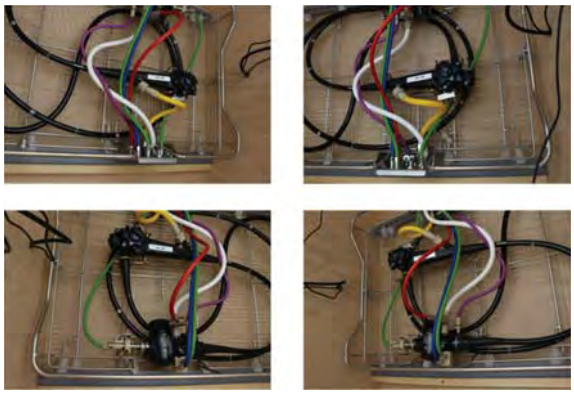


Figure 5: Template images from four different perspectives.

At first we intend to find the rough position of the endoscope. For this purpose we use the feature detection algorithm to find correspondences between a reference image of the part of the adapter at the endoscope.

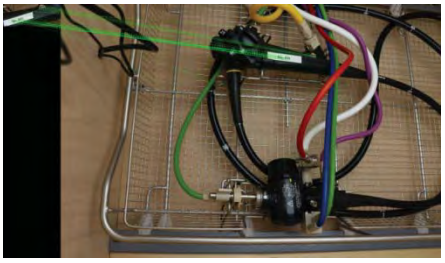


Figure 6: Correspondence pairs between the reference image and the template image to detect the rough position of the adapter.

These interest points and correspondences are easier to find because of the texts on the endoscope which are very distinctive. We compute the mean of all detected points in Figure 6 and crop that region of the image depending on the size and geometry of the adapter. This happens at a fourth of the resolution to save time. The following computations are made on the cropped images. This approach has two advantages. First: The complex computations are made on a much smaller image. Second: A smaller image minimizes the probability of correspondence errors.

When all subregions for all adapters were found, the accurate registration of the reference image pairs begins. For every adapter we have five reference image pairs. As one can see in Figure 7 the adapters not only have to be transformed in the space but may have to be rotated longitudinally as well. For the accurate registration we check all reference images and choose the one with the most detected

correspondences.

Afterwards we detect and describe the interest points of the chosen reference image of the adapter on the side of the endoscope and the associated reference image of the adapter on the side of the tube. Then we identify correspondences between the reference images and compute the transformation matrices as shown in equations (1) and (2). With the computed transformation matrices we map the reference images into the template image we identified previously. If the adaption is correct, the projected reference images should intersect on the inner edge as pictured in Figure 3. The same matrices for mapping the reference images are used to compute the center of the cutting edge. In Figure 8 one can see the two mapped reference images, bounded in green rectangles. The midpoints of the cutting edges are depicted by blue points. One can recognize only one point at the left image because of the optimal projections the two points overlap completely. In the right image one can see an incorrect adaption. The Euclidean distance between the points is one feature for the classification of the incorrect adaption.

Figure 7: A variety of reference images of the same adapter because of the longitudinal rotations, which are not detectable with the feature detection algorithms.

We use the two points to compute the Euclidean distance between the two projections. An explorative data analysis showed that the projection does not work perfectly at all times. So it is not reliable as single. Because of that we implemented a simple k-nearest-neighbor Algorithm to classify the adaptions. In addition to the Euclidean distance we use the number of correspondences as a second feature. This is depicted in Figure 4.

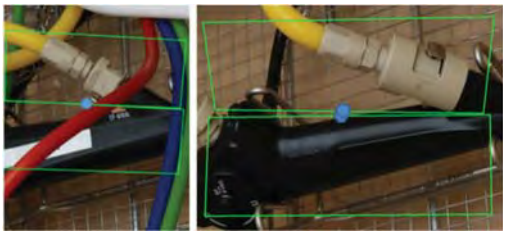


Figure 8: Mapped reference image pair on one template image. The blue point marks the middle of their cutting edges. Left: correct adaption, right: incorrect adaption.

# 3 EXPERIMENTAL RESULTS

As stated in the previous chapter, the detection of interest points and correspondences is essential for a correct projection of the reference images. The quality of the transformation matrix is significantly enhanced by detecting more correspondences. Vice versa a faultless projection is impossible if there are too many correspondence errors.. In this chapter we describe the results of the detection processes and of the classification.

We describe only the results of the reference images on the side of the adapter. Because of the high amount of letters on the endoscope itself, there are a lot of interest points and the projection always worked faultless.


Figure 9: Adapter set which we used for our experiments.

The quantitative results of the feature detection algorithms are our first criteria for the quality of the system. Figure 9 shows the adapter set we used for our experiments. In the following tables we show the statistic results of 40 processes for the first adapter set. Table 1 shows the results for correct adaptions generated with SURF.

The number of interest points and correspondences - absolute and per pixel – is of special importance here. In both categories one can see, that the values are in the same order of magnitude for most adapters. The adapters a, c, f and g have the lowest values. This is because of the little body and the few distinctive points on the adapters. So a correct correspondence is more difficult to compute. The adapters b, d and e have more distinctive points. So the correspondences are easier to find. In Table 2 we see the statistical values for incorrect adaptions, generated with SURF. The number of detected interest points is similar. This makes sense, because we use for both procedures images of the same size. One can see the difference at the inspection of the correspondences. For incorrectly adapted devices the algorithm detects much less correspondences. This is because the reference images were made with correct adaptions.

Table 1: Correct adaptions, generated with SURF as pre-processor.

| | Interest Points | | | | Correspondences | | | |
|---|---|---|---|---|---|---|---|---|
| Adapter | Mean # | Minimum # | Maximum # | Per Pixel | Mean # | Minimum # | Maximum # | Per Pixel |
| a | 824 ±90 | 764 | 995 | 3.4 E-3 | 139 ±25 | 105 | 189 | 5.8 E-4 |
| b | 367 ±24 | 331 | 382 | 2.1 E-3 | 97 ±12 | 81 | 115 | 5.0 E-4 |
| c | 349 ±8 | 339 | 357 | 1.0 E-3 | 52 ±13 | 23 | 65 | 1.6 E-4 |
| d | 572 ±12 | 561 | 586 | 1.,9 E-3 | 53 ±6 | 45 | 63 | 1.7 E-4 |
| e | 107 ±3 | 106 | 117 | 1.9 E-3 | 17 ±10 | 9 | 43 | 3.0 E-5 |
| f | 111 6 ±82 | 883 | 114 2 | 3.4 E-3 | 142 ±21 | 114 | 114 | 4.0 E-4 |
| g | 441 ±33 | 371 | 465 | 2.0 E-3 | 26 ±6 | 18 | 35 | 1.1 E-4 |

Table 2: Incorrect adaptions, generated with SURF as pre-processor.

| | Interest Points | | | | Correspondences | | | |
|---|---|---|---|---|---|---|---|---|
| Adapter | Mean # | Minimum # | Maximum # | Per Pixel | Mean # | Minimum # | Maximum # | Per Pixel |
| a | 794 ±33 | 780 | 888 | 3.3 E-3 | 38 ±16 | 24 | 73 | 1.6 E-4 |
| b | 356 ±26 | 323 | 382 | 2.0 E-3 | 15 ±5 | 10 | 25 | 8.0 E-5 |
| c | 342 ±6 | 339 | 354 | 1.0 E-3 | 7 ±2 | 5 | 10 | 2.0 E-5 |
| d | 566 ±29 | 540 | 602 | 1.8 E-3 | 19 ±9 | 12 | 40 | 6.0 E-5 |
| e | 159 ±58 | 106 | 227 | 2.8 E-3 | 35 ±38 | 4 | 86 | 6.0 E-5 |
| f | 873 ±310 | 142 | 114 2 | 3.0 E-3 | 60 ±27 | 27 | 91 | 1.8 E-4 |
| g | 424 ±49 | 324 | 465 | 1.9 E-3 | 9 ±4 | 7 | 19 | 4.0 E-5 |

If there is an incorrect adaption in the template image it is possible, that there are large perspective changes. If the perspective changes are too high, the SURF algorithm cannot detect them, so fewer correspondences are found as one can see in Figure 10. In six of the seven cases there are more than twice as many detected correspondences. The only exception here is adapter g because of its simple structure and few interest points. These experimental evaluations demonstrate that the number of correspondences is a meaningful feature for the machine learning algorithm.

Table 3: Correct adaptions, generated with SIFT as pre-processor.

| | Interest Points | | | | Correspondences | | | |
|---|---|---|---|---|---|---|---|---|
| Adapter | Mean # | Minimum # | Maximum # | Per Pixel | Mean # | Minimum # | Maximum # | Per Pixel |
| a | 402 ±5 | 391 | 407 | 1.7 E-3 | 80 ±13 | 54 | 91 | 3.3 E-4 |
| b | 135 ±8 | 121 | 140 | 8.0 E-4 | 46 ±5 | 38 | 51 | 2.6 E-4 |
| c | 157 ±13 | 140 | 170 | 5.0 E-4 | 14 ±5 | 7 | 26 | 1.0 E-5 |
| d | 321 ±20 | 278 | 336 | 1.0 E-3 | 23 ±7 | 15 | 33 | 7.0 E-5 |
| e | 52 ±8 | 41 | 58 | 9.0 E-4 | 9 ±3 | 5 | 13 | 1.5 E-4 |
| f | 697 ±185 | 206 | 776 | 2.0 E-3 | 208 ±10 | 191 | 220 | 6.3 E-4 |
| g | 238 ±40 | 199 | 339 | 1.0 E-3 | 17 ±5 | 10 | 28 | 7.0 E-5 |

In the following we outline the results of SIFT for the same adapter set. The statistical values of correct adaptions are shown in Table 3. Compared to the SURF algorithm it attracts attention that SIFT detects less interest points and correspondences than SURF. The exact quotient is shown in Figure 11.
One can easily see that the SURF algorithm detects more interest points and because of that more Correspondences.

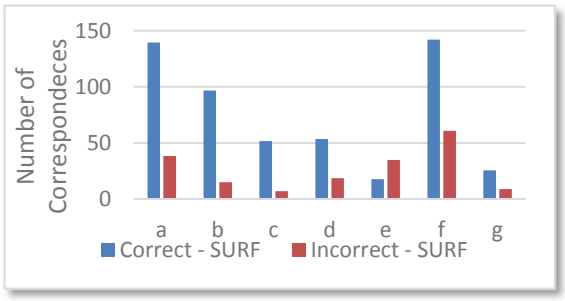



Figure 10: Comparison of the number of correspondences between correct and incorrect adaptions.

In Table 5 one can see the statistical values for incorrect adaptions, generated with SIFT. Analogue to the SURF algorithm it is obvious that fewer correspondences have been computed. This is because of the similar procedure. Striking are the detected minima of correspondences. If the algorithm detects less than three correspondences, a projection and following classification is impossible. In summary we conclude, that the SURF algorithm detects roughly twice as many correspondences than

the SIFT algorithm. If we have too few correspondences it is possible that the RANSAC algorithm cannot filter the correspondence errors. The result is an incorrect transformation matrix. Quantitatively the SURF algorithm has to be preferred.

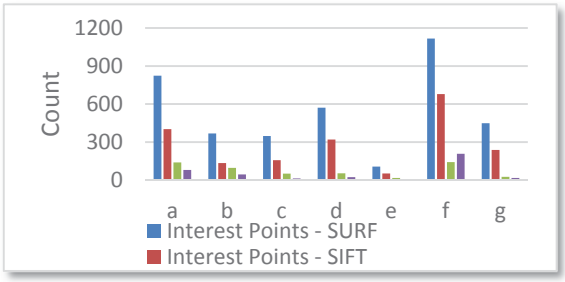



Figure 11: Comparison of the detected interest points and correspondences.

Now we will focus on the quality and precision of the classification. As outlined before we use a k-nearest neighbor Algorithm for the binary classification. The Euclidean distance of the projection pairs and the number of correspondences are the features of the algorithm. Experiments with validation data gave us an optimal value for $k = 5$.

A first estimation of the quality of the processes delivers the classification rate. In Figure 12 one can see the classification rate per adapter for both algorithms as pre-processing. All rates were determined on test data experiments.

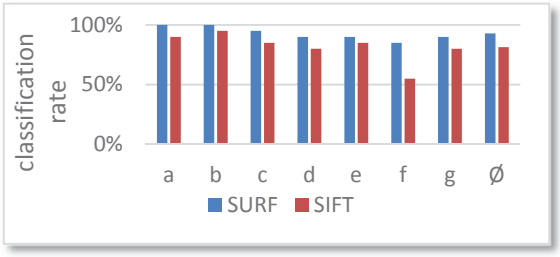



Figure 12: Comparison of the classification rates.

As predicted the process with SURF as pre-processing has a significantly higher classification rate of 92.86% compared to SIFT. The worst value occurs at adapter e. This is because of the geometry of the adapter. The Euclidean distance can be 0 for this adapter although it is adapted incorrectly. This adapter needs an additional physical rotation to be correctly adapted. This can only be determined through the number of detected correspondences.

Table 4: Confusion matrix of the classifications with SURF as pre-processing.

| SURF | CORRECT ADAPTIONS | INCORRECT ADAPTIONS | |
|---|---|---|---|
| PREDICTED CORRECT ADAPTIONS | TP = 64 | FP = 4 | P' = 68 |
| PREDICTED INCORRECT ADAPTIONS | FN = 6 | TN = 66 | N' = 72 |
| | P = 70 | N = 70 | $\sum = 140$ |

Fawcett explains another possibility to evaluate classifications (Fawcett, 2006). So called receiver operating characteristics (ROC) graphs can be used to evaluate and visualize the quality of a classification. We divide the results of our experiments in four groups: true-positive (TP), true-negative (TN), false-positive (FP) and false-negative (FN). These groups can be written in a so-called confusion matrix (Fawcett, 2006), as one can see in Table 4 and Table 6.

Table 5: Incorrect adaptions, generated with SIFT as pre-processor.

| | Interest Points | | | | Correspondences | | | |
|---|---|---|---|---|---|---|---|---|
| Adapter | Mean # | Minimum # | Maximum # | Per Pixel | Mean # | Minimum # | Maximum # | Per Pixel |
| a | 339 ±5 | 397 | 409 | 1.7 E-3 | 40 ±9 | 24 | 51 | 1.6 E-4 |
| b | 136 ±7 | 121 | 140 | 8.0 E-4 | 10 ±6 | 3 | 19 | 5.0 E-5 |
| c | 160 ±35 | 140 | 256 | 5.0 E-4 | 5 ±2 | 3 | 8 | 1.0 E-5 |
| d | 317 ±21 | 278 | 336 | 1.0 E-3 | 18 ±6 | 9 | 30 | 5.0 E-5 |
| e | 131 ±99 | 41 | 246 | 2.3 E-3 | 26 ±25 | 3 | 76 | 4.6 E-4 |
| f | 711 ±105 | 557 | 776 | 2.2 E-3 | 142 ±53 | 44 | 209 | 4.3 E-4 |
| g | 203 ±14 | 190 | 239 | 9.0 E-4 | 8 ±3 | 4 | 13 | 3.0 E-5 |

The contents of the matrices were generated in 40 experiments and are a summary of all adapters. $P$ is the number of correct, $N$ is the number of incorrect adaptions. $P'$ is the number of positive, $N'$ is the number of negative predictions. We can derive four statistical values from these matrices: the true-positive-rate (TPR), the false-positive-rate (FPR), positive-predictive-value (PPV) and the accuracy (ACC). We can now visualize the quality of the classification with a ROC-Graph as shown in Figure 13.

Table 6: Confusion matrix of the classifications with SIFT as pre-processing.

| SIFT | CORRECT ADAPTIONS | INCORRECT ADAPTIONS | |
|---|---|---|---|
| PREDICTED CORRECT ADAPTIONS | TP = 62 | FP = 18 | P' = 80 |
| PREDICTED INCORRECT ADAPTIONS | FN = 8 | TN = 52 | N' = 60 |
| | P = 70 | N = 70 | $\sum$ = 140 |

Ideally should the TPR be close to 1, the FPR close to 0. The more the point is in the North-West, the better is the classification. One can see in the visualization that the processes with SURF as Pre-Processing Algorithm is better than with SIFT. The system needed in average 71/42 seconds for the classification of the large/small adapter set.

Table 7: Statistical values from the confusion matrices.

| | SURF | SIFT |
|---|---|---|
| TPR | 0.914 | 0.886 |
| FPR | 0.057 | 0.257 |
| PPV | 0.941 | 0.775 |
| ACC | 0.928 | 0.814 |

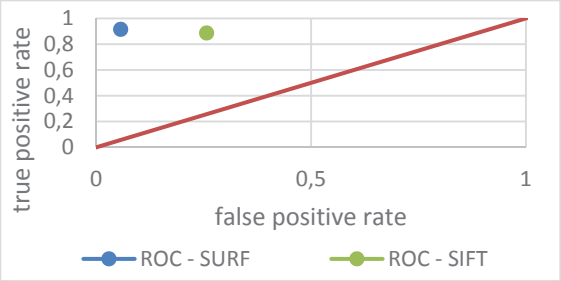



Figure 13: ROC Graph.

## 4 CONCLUSION

This paper presents an automatic visual system for detecting adaption errors in chemical disinfection devices for endoscopes. Our experimental evaluation shows promising results with respect to the classification accuracy. With the SURF algorithm as pre-processing tool, the prototype system yields a classification accuracy of 92.86% for determining the correctness of the adaptions in approximately one minute of processing. Future work will aim in enhancing the correctness of the prediction close to 100% and in installing the system directly into the endoscope thermal disinfector to save even more time and resources.

## REFERENCES

Bader, L. et al, 2002 HYGEA (Hygiene in Gastroenterology – Endoscope Reprocessing): Study on Quality of Reprocessing Flexible Endoscopes in Hospitals or Practice Setting. In *Journal for Gastroenterology 12 Pages 951-1000*.

Muscarella, L., 2014 Risk of transmission of carbapanem-resistant Enterobacteriaceae and related "superbugs" during gastrointestinal endoscopy. In *World Journal of Gastrointestinal Endoscopy 6 Pages 951-1000*.

Handels, H., 2009. Medizinische Bildverarbeitung, Vieweg und Teubner. Wiesbaden, 2nd edition.

Zitova, B., Flusser, J., 2003. Image registration methods: a survey. In *Image and Vision Computing Volume 21 Pages 977-1000*.

Lowe, D, 2004, Distinctive Image features from Scale-Invariant Keypoints. In *International Journal of Computer Vision 60 Pages 91-110*.

Bay, H. et al, 2008, Speeded-Up Robust Features (SURF). In *Computer Vision and Image Understanding Pages 346-359*.

Schreer, O., 2005. Stereoanalyse und Bildsynthese, Springer Verlag. Berlin Heidelberg, 1st edition.

Fischer, G., 2014. Lineare Algebra, Springer Verlag. Berlin Heidelberg, 18th edition.

Szeliski, R., 2011. Computer Vision, Springer Verlag. Berlin Heidelberg, 1st edition.

Silpa-Anan, C., Hartley, 2008, R., Optimised KD-Tree for fast image descriptor matching. In *Computer Vision and Pattern Recognition*

Fukanaga, K., Narendra, P., 1975, A branch and bound Algorithm for Computing k-Nearest Neighbors. In *IEEE Transactions on Computers C-24 Pages 750-753*.

Muja, M., Lowe, D., 2014, Scalable Nearest Neighbor Algorithms for High Dimensional Data. In *IEEE Transactions on Pattern Analysis and Machine Intelligence 36 Pages 2227-2240*.

Strutz, T., 2016. Data Fitting and Uncertainty, Springer Verlag. Berlin Heidelberg, 2nd edition.

Fawcett, T., 2006, An introduction to ROC Analysis. In *Pattern Recognition Letters 27 Pages 861-874*.

# MATCHuP: An mHealth Tool for Children and Young People Health Promotion

Beatriz López[1], Sreynoch Soung[2], Natalia Mordvanyuk[1], Albert Pla[3], Pablo Gay[1] and Abel López-Bermejo[4]

[1]*University of Girona, Girona, Spain*

[2]*Institute of Technology of Cambodia, Phnom Penh, Cambodia*

[3]*University of Oslo, Oslo, Norway*

[4]*Biomedical Research Institute of Girona, Girona, Spain*

{*beatriz.lopez, natalia.mordvanyuk, pablo.gay*}*@udg.edu, sreynoch.soung1@gmail.com, a.p.planas@medisin.uio.no, alopezbermejo@idibgi.org*

Keywords: Decision Support Systems, Telemedicine, Nursing Informatics, Wearable Health Informatics, eHealth Applications.

Abstract: The kids of European and occidental countries are threatened by obesity. They are potential persons to become chronic patients. mHealth technology can help them to change their nutrition and physical activity habits. This paper presents MATCHuP, a platform that involves several agents (kids, parents, healthcare providers) that collaborate and compete by games in a social network in order to create a enjoyable environment to promote a behavioural change towards a healthier life.

## 1 INTRODUCTION

Having care of our kids today is the best bid we can make to have a healthy society tomorrow [1]. However, three main diseases threaten kids health: obesity, diabetes and asthma. While parents are conscientious that diabetes and asthma are serious diseases, it seems that they are less aware about the harm of obesity. Obesity can be cured, avoiding reaching the adult age with related serious and chronic diseases with comorbidities.

Obesity is mainly caused by bad nutrition habits and the absence of exercise. The development of tools to support enhancing good nutrition and exercise habits would result in benefits for both, obesity and diabetes, known as metabolic diseases.

Nowadays, information and communication technology (ICT) is offering a means to support follow up of personal data, enabling the education on the right habits of obese children, but also offers clinicians a way to gather information about their patients, and other data coming from social workers. In that regard, Artificial Intelligence Techniques, including Machine Learning, have shown to be cornerstone to transform gathered data to knowledge, so as to support decision making towards a personalized health treatment (parents, clinicians) (Herrero et al., 2016; López et al., 2013). Games is another important technology that has been raised as a key issue for education, and can be a key issue for kids to be compliant with their treatment, especially if games are not designed as serious but popular games are paid via exercise or good diet instead of money.

This paper presents the MATCHuP platform centred on patient with metabolic diseases and their families to improve their education in the right habits towards a healthy future society. In so doing, MATCHuP aims to recommend actions connected to the patients community, using collaboration strategies to simplify input validation, and competition incentives so as to award the patient with gaming.

This paper is organized as follows. First, some related work is reviewed in Section 2. Next the description of the MATCHuP platform is presented in Section 3. The status of the current implementation is explained in Section 4. We end the paper on Section 5 with some conclusions and future work.

[1]The MOCHA project (Models of Child Health Appraised), http://www.childhealthservicemodels.eu/



López, B., Soung, S., Mordvanyuk, N., Pla, A., Gay, P. and López-Bermejo, A.
MATCHuP: An mHealth Tool for Children and Young People Health Promotion.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 313-318
ISBN: 978-989-758-213-4

## 2 RELATED WORK

The use of mHealth approaches to child obesity has been evaluated in (Tate et al., 2013). Among the advantages, the authors highlight four issues. First, that adolescent would prefer mobile-based and iterative technology for treatment and prevention. Second, that mHealth is a useful tool for monitoring adherence. Third, that the reachability enhancement of the population of a higher risk for obesity. And four, a similar enhancement could be observed regarding of the lowest educational level and income. However, several challenges were detected, as the sedentary behaviour of screen technologies and the detected decreased ability to focus attention of screen users. In that regard, wearables technology is arising as a new tool for measuring the real activity of kids, enabling the development of alternative mHealth platforms that tackle such challenges. MATCHuP uses them.

Of course there are myriads of mHealth solutions. In the recent report (Aitker, 2015) about 165,000 apps were identifyed. Among them, the authors confirmed 46,188 mHealth apps and they focus their stydy on the English Apps from which 26,864 where consumer/patient and only 8,965 apps were related to healthcare providers. Our app includes healthcare providers as well as consumers, and it is multilanguage, currently in English, Spanish and Catalan.

Regarding apps for kids, Table 1 shows a list of several apps. The elements that configure a mHealth success regarding nutritional habits, as for example, sugar ingesta, have been studied in (Sanders et al., 2009).The key issues to have success tools involves how the information is delivered. In that regards, several agents are identified in order to improve the literacy on health: caregivers, health systems, family health literacy skills, the educational system and the community system. In our work, we involve most of such actors: caregivers (endocrinologists), family (parents), and the community.

## 3 PLATFORM DESCRIPTION

The goal of MATCHuP is to improve nutrition and exercise habits of obese kids from 5 to 16 years. To that end, the platform gathers information about kids meals and physical activity, and according to the healthy quality of the data entered, kids awarded with some points, that are translated in skill scores regarding a virtual game. The game is not played in isolation but in teams. Therefore, kids should collaborate among her mates in order to have a competitive team that beats their adversaries. On the other hand, the validation of the information entered by the kids is performed in a collaborative way.

Therefore, several actors are involved in MATCHuP. First caregivers set up healthy targets to the kids according to their progresses. Second, the kids that self-monitor their progress toward the targets. Third, the parents that collaborates in the monitoring, by validating the inputs entered by kids. And finally, all the community of users (parents and kids) collaborate in different way inside a social network implemented for their community. Moreover, kids can set up teams in the community which compete in virtual games, and the skills of the avatars (virtual representation of the kid) depend on their healthy progress. An overview of the platform is shown in Figure 1. In the remaining of this section the different roles of the agents involved in the platform are described, including the social network in which they collaborate, and how the games are approached.

Table 1: Apps for healthy Children. Physical activity: v → virtual exercise; t → teach about exercise.

| App Name | Food | Nutrition | Physical Activity |
|---|---|---|---|
| Easy Eater 2 | x | x | |
| Eat and Move-O-Matic | x | x | v,t |
| Healthy Heores | x | x | |
| Perfect Picnic | x | x | |
| Smash Your Food | x | x | |
| Veggie Circus Farm | x | x | |
| Body Quest - Food of the Warrior. | x | x | |
| Grow It-Know It | x | x | |
| Catch the Carrot | x | x | |
| Snack Planet | x | x | v |
| Work It Off | x | x | t |
| Max's Plate | x | x | |
| Frutas y verduras para nios | x | | |
| Hora de Comer | x | x | v |
| Emma breakfast - KIDS | x | x | |
| EduKitchen-Kids Educational | x | x | |
| Veggie Bottoms Lite | x | x | |
| Sopa Hacedor | x | | |
| Awesome Eats | x | x | |
| Cocomong Season 2 | x | x | |

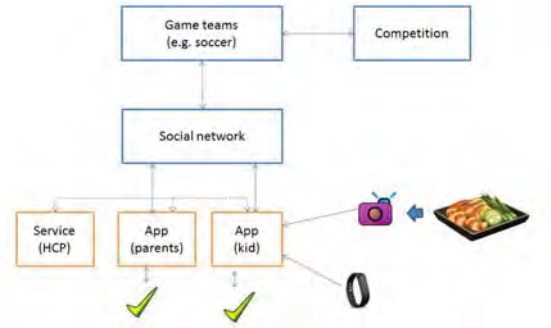



Figure 1: MATCHuP overview.

## 3.1 Kids App

The kids app includes several modules in order to respond to the following requirements:

- Nutrition: enable the entering of the different kid meals
- Exercise: register the activity of the kid
- Validation: check and receive information about the data entered as so far by other kids or by herself
- Assessment: acquaintance of the healthy behaviour progress

Every time the kid eats, she should register the nutrition information in the system. However, this information is not entered manually neither with a text, or by selecting photos in a library, as many other apps in the market, but by making a photo of the dish he is just eating. Next, the kid has three sliding buttons to label the photo, according to his knowledge, which is the amount of fruits and vegetables, carbohydrates, and meat that contains the served meal (see Figure 2). From the sliding button, a percentage on nutrition components is derived.

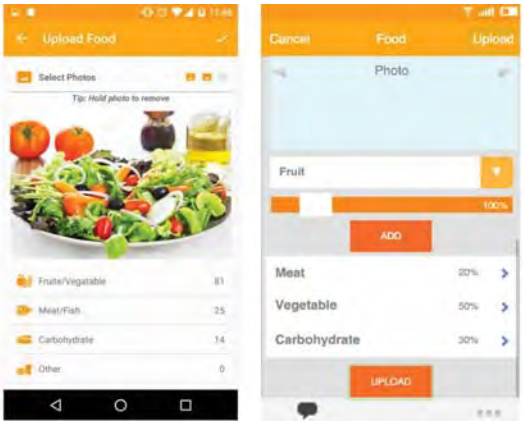



Figure 2: GUI for food labelling.

Of course, the information entered by the kid should be validated. Validation is performed in a collaborative way. That is, parents and other kids in the community validate the labels assigned to the photo (see Figure 3). Once a day, every kid receives a set of photos from other kids in the social network (see parents and community validation on Sections 3.2 and 3.4 correspondingly). The owner of the photos received for validation is unknown. They could come either from kids in the same team or from adversary teams. Therefore, the kid cannot manipulate the outcome to favouring her mates. In order to incentive kids in this validation process, some points are given to the kids that actively participate in this process that contribute to win the game match of the week (see Section 3.5).

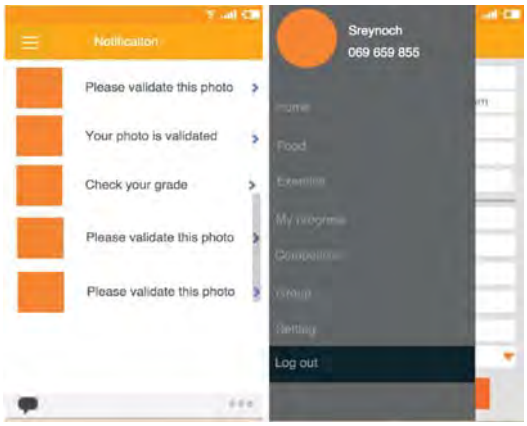



Figure 3: GUI for food validation.

Regarding physical activity, each kid defines a profile regarding her preferences about sports and the timetable they use to practice (supported by parents when under 12 years old). To validate the activity, wearables are offering a smart way of capturing it. To that end each type of activity, and its intensity is measured according to METs (the ratio of work metabolic rate to resting metabolic rate) (see (Ainsworth et al., 1993) and (Ainsworth et al., 2011) for further calculation descriptions). Some activities could come with non-scheduled hours (as for example, playing soccer in the school playground).

Kid assessment about her progress is provided by plots in which the differences between the current state and the targets is shown following a colour code (see Figure 4). The nutrition information is not taken from the kid' labelling, but from the outcome of the validation process. An aggregation method is used to combine the information of the kid (self information and the validation data from other users), giving a higher importance to the information coming from parents. The final nutrition fitness is provided in a

scale from A to E, being A the best value. A similar outcome is obtained for the physical activity, obtaining a second value defined in the same scale. Both fitness values, nutrition and physical activity, are finally aggregated, obtaining the kid current healthy state.

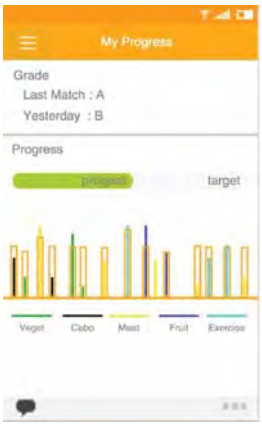



Figure 4: Kid assessment.

## 3.2 Parents App

Parents role is mainly focused on providing reliable validation information. In that regard, parents receive once a day a set of labelled photos that they need to revise. As in the kid case, they are not aware about the provenance of the photos (i.e. whenever they belong to their kid or do not).

## 3.3 Healthcare Professionals Web Service

Healthcare professionals are in charge of setting up the nutrition and exercise targets for the kids. They can also follow the kid progresses thanks to a visualisation screen that shows the distance between the target and the achieved results, in a colour code (see Figure 5).

Healthcare professionals access to the platform has been designed as a web service, instead of an mobile app because this facilitates the integration of the tool in the current Healthcare Information Systems of our region.

## 3.4 Social network

The community of users is managed by means of a social network, where collaborative and competitive events take place.

Regarding competition, users in the network are identified according to their sportive preferences (soccer, basket, dance, etc.), and her healthy status. This

Figure 5: GUI for healthcare providers.

data enables the configuration of sport teams and the corresponding game competitions (see Figure 6). In so doing, two conditions should be fulfilled:

- There should be enough teams in each sport to set up a game (sport competition matches)
- There should be a certain satisfaction degree among the kids preferences and the team assigned.

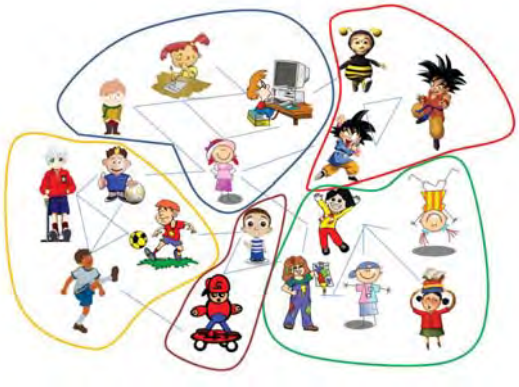


Figure 6: Physical activity teams.

On the other hand, collaboration arises in two directions:

- Help team mates to achieve their healthy targets (see Figure 8)
- Validate food photos from other users (see Figure 7 )

Regarding the validation of photos,

- 25% of the photos of a kid are validated by members of the same team
- 50% of the photos of a kid are validated by members of other teams of the same sport
- 25% of the photos of a kid are validated by members of other sports
- 50% of the photos of a kid are validated by parents, selected at random

Therefore, the number of validations obtained per photo is expected to be higher than 1. The worse case

scenario would be when no validation is achieved for a given photo. In that situation, the photo is scored neutral. Future work should include a monitoring module to facilitate the assessment and control such situations.

The validation feedback can help to understand the kid about the real contents of the meals, and learn about nutrition.

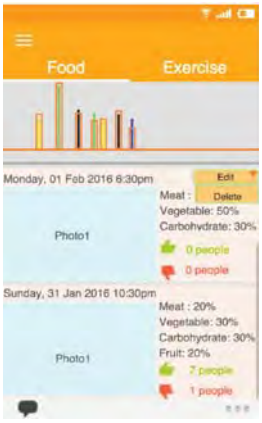



Figure 7: Food validated by the community.

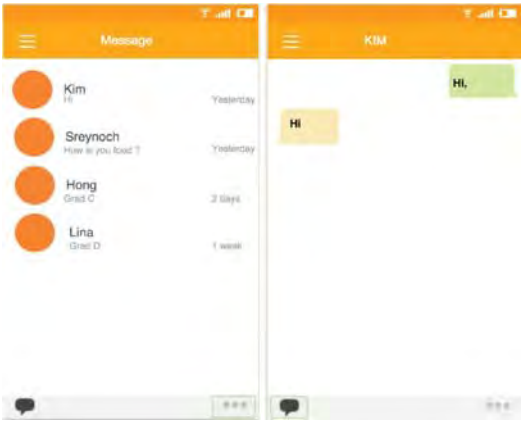



Figure 8: Chat functionalities for kid collaboration.

## 3.5 Games

In the social networks, there are $n$ games, according to the kids sportive preferences, although it depend on the number of users, too. For example, there could be a soccer game, or a skate game (see Figure 6). For each game, there are a given number $m$ of sport teams. Once a weak (e.g.Sunday),all teams play a match.

Ideally, the match should correspond to a market available game integrated in the platform. For example, there are soccer games which teams could be configured by the user. However, in the first approach of MATCHuP, the game is a simple rank of scores: the winner of the match is the team with the highest score (see Figure 9). The team score (defined in $\Im$)is obtained by the aggregation of the healthy state values of all of its members.


Figure 9: Competition outcome.

# 4 FIRST VERSION

The system has been developed using the Spring Java environment. The mobile application is deployed in Android (Figure 10). The first version deployed with all of the involved agents, but physical activity enter manually. Current languages are English, Spanish and Catalan.

The design goals of MATCHuP are simplicity and easy to use. Thus, photos favours usability and collaborative validation is simply. Of course, some image processing engines could be used to obtain nutrition components from photos, for example, but the state of the art of such engines are still under research for such purpose.

Next step will consider the inclusion of the exercise activities by means of a smartwatch or activity band. Other technological advances that are ready to use and that could be incorporated in a near future are Artificial Intelligence techniques to handle preferences to set up teams, as well as to aggregate information.

# 5 CONCLUSION

Obesity is a main issue for many people, specially children. It is becoming a big concern in European and occidental countries. Obese persons are not considered patients (as diabetic persons are), and therefore, they are very difficult to motivate for bringing them to healthier states. This paper presents the MATCHuP platform with the aim of helping obese


Figure 10: Home page for kids.

children and young persons (under 16) to reach a healthier state with the supervision of healthcare practitioners, families, and other users in a similar situation.

Along the paper, the system has been described, and the first version of MATCHuP presented. Next steps include the integration of smartwatches or similar wearable able to automatically detect the physical activity, as well as artificial intelligence tools to improve the aggregation methods and additional advices for further personalization and fast adaptation. Moreover, a the evaluation of the tool in for medical evidence is also required.

In that regard, the main challenge is to keep kids engaged in the platform. Clinicians argue that about 6 month of using the platform could be sufficient for obtaining some behaviour change. However, some studies have shown that having a kid engaged in a game more than 3 months is a great success. The long trial of the tool will provide inputs to that concern, and work for alternative artifaxts (Hevner et al., 2004) . A secondary, technological challenge, is the fact that sensitive data is stored in mobiles. The recent study (Blenner et al., 2016) highlights the necessity of consider privacy implications before using health apps.

## ACKNOWLEDGEMENTS

This project has received funding from the grant of the University of Girona 2016-2018 (MPCUdG2016), ans has been developed with the support of the research group SITES awarded with distinction by the Generalitat de Catalunya (SGR 2014-2016). Sreynoch Soung received financial support from the European Commission (Erasmus Mundus project Techno II, ref. 372228-1-2012-1-FR-ERA MUNDUS-EMA21).

## REFERENCES

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., Greer, J. L., Vezina, J., Whitt-Glover, M. C., and Leon, A. S. (2011). 2011 Compendium of Physical Activities: a second update of codes and MET values. *Medicine and science in sports and exercise*, 43(8):1575–81.

Ainsworth, B. E., Haskell, W. L., Leon, A. S., Jacobs, D. R., Montoye, H. J., Sallis, J. F., and Paffenbarger, R. S. (1993). Compendium of physical activities: classification of energy costs of human physical activities. *Medicine and science in sports and exercise*, 25(1):71–80.

Aitker, M. (2015). Patient Adoption of mHealth. Use, Evidence and Remainding Barriers to Mainstream Acceptance. Technical Report September.

Blenner, S. R., Köllmer, M., Rouse, A. J., Daneshvar, N., Williams, C., Andrews, L. B., J, W., and AM, B. (2016). Privacy Policies of Android Diabetes Apps and Sharing of Health Information. *JAMA*, 315(10):1051.

Herrero, P., Lopez, B., and Martin, C. (2016). Pepper: Patient empowerment through predictive personalised decision support. In *ECAI Workshop on Artificial Intelligence for Diabetes*, pages 8–9.

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1):75–105.

López, B., Coll, J., Gamero, F., BargallLó, E., and López-Bermejo, A. (2013). Intelligent systems for supporting premature babies healthcare with mobile devices. In *Mobilemed*, page 4 pages.

Sanders, L. M., Shaw, J. S., Guez, G., Baur, C., and Rudd, R. (2009). Health literacy and child health promotion: implications for research, clinical care, and public policy. *Pediatrics*, (124 Suplement 3):306–14.

Tate, E. B., Spruijt-Metz, D., O'Reilly, G., Jordan-Marsh, M., Gotsis, M., Pentz, M. A., and Dunton, G. F. (2013). mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. *Translational behavioral medicine*, 3(4):406–15.

# Statistical Analysis of Window Sizes and Sampling Rates in Human Activity Recognition

Anzah H. Niazi[1], Delaram Yazdansepas[2], Jennifer L. Gay[3], Frederick W. Maier[1], Lakshmish Ramaswamy[2], Khaled Rasheed[1,2] and Matthew P. Buman[4]

[1]*Institute for Artificial Intelligence, The University of Georgia, Athens, Georgia, U.S.A.*

[2]*Department of Computer Science, The University of Georgia, Athens, Georgia, U.S.A.*

[3]*College of Public Health, The University of Georgia, Athens, Georgia, U.S.A.*

[4]*School of Nutrition and Health Promotion, Arizona State University, Phoenix, Arizona, U.S.A.*

Keywords: Actigraph g3x+, Analysis of Variance, Body-worn Accelerometers, Data Mining, Human Activity Recognition, Random Forests, Sampling Rate, Weighted Least Squares, WEKA, Window Size.

Abstract: Accelerometers are the most common device for data collection in the field of Human Activity Recognition (HAR). This data is recorded at a particular sampling rate and then usually separated into time windows before classification takes place. Though the sampling rate and window size can have a significant impact on the accuracy of the trained classifier, there has been relatively little research on their role in activity recognition. This paper presents a statistical analysis on the effect the sampling rate and window sizes on HAR data classification. The raw data used in the analysis was collected from a hip-worn Actigraphy G3X+ at 100Hz from 77 subjects performing 23 different activities. It was then re-sampled and divided into windows of varying sizes and trained using a single data classifier. A weighted least squares linear regression model was developed and two-way factorial ANOVA was used to analyze the effects of sampling rate and window size for different activity types and demographic categories. Based upon this analysis, we find that 10-second windows recorded at 50Hz perform statistically better than other combinations of window size and sampling rate.

## 1 INTRODUCTION

The field of Human Activity Recognition (HAR) is dependent on a variety of instruments for data collection — heart rate monitors, GPS, light sensors, etc. — of which wearable triaxial accelerometers are the most commonly utilized (Lara and Labrador, 2013), (Preece et al., 2009). Accelerometers are commercially available in many formats, from modern smartphones and consumer-grade activity-monitoring products to high-grade research-oriented devices, the consequences of which are wide degrees of quality in data collection for HAR. When preparing for data collection in a HAR study, two aspects of the accelerometer to use should be strongly considered: the placement of the device and the sampling rate at which it gathers data.

The placement of the device depends greatly on the context of the study. Many studies focusing on ambulation activities (walking, running etc.) prefer hip-worn or wrist-worn devices (Lara and Labrador, 2013), both of which have advantages and disadvantages. Wrist-worn devices have trouble distinguishing lower-body activities (for instance, walking and stair climbing), while hip-worn devices can be problematic when recognizing upper-body activities (for instance, eating and brushing teeth). The impact of sampling rate is discussed in later sections.

Once data has been collected — typically at a fixed sampling rate — it is prepared for classification by extracting relevant features such as means and standard deviations and dividing the accelerometer readings into windows. Often, windows of fixed length are used.

Both the sampling rate and window size of data are crucial decisions in HAR which directly affect the accuracy of developed classifiers. Though a literature review revealed some relevant analyses (Section 2), there appears to be a relative dearth of work directly addressing sampling rate and window size in HAR. This study is an attempt to remedy what we perceive as a gap in the research. We have attempted to statistically identify the window size and sampling rate combination which best suits activity recognition across demographical and activity divisions.

The data used in this study was obtained from 77 demographically diverse subjects for 23 activities in studies performed at Arizona State University in



Niazi, A., Yazdansepas, D., Gay, J., Maier, F., Ramaswamy, L., Rasheed, K. and Buman, M.
Statistical Analysis of Window Sizes and Sampling Rates in Human Activity Recognition.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 319-325
ISBN: 978-989-758-213-4

Table 1: Description of Activities Performed.

| # | Activity | Duration or Distance | # of subjects |
|---|---|---|---|
| 1 | Treadmill at 27 młmin-1 (1mph) @ 0% grade | 3 min | 29 |
| 2 | Treadmill at 54 młmin-1 (2mph) @ 0% grade | 3 min | 21 |
| 3 | Treadmill at 80 młmin-1 (3mph) @ 0% grade | 3 min | 28 |
| 4 | Treadmill at 80 młmin-1 (3mph) @ 5% grade (as tolerated) | 3 min | 29 |
| 5 | Treadmill at 134 młmin-1 (5mph) @ 0% grade (as tolerated) | 3 min | 21 |
| 6 | Treadmill at 170 młmin-1 (6mph) @ 0% grade (as tolerated) | 3 min | 34 |
| 7 | Treadmill at 170 młmin-1 (6mph) @ 5% grade (as tolerated) | 3 min | 26 |
| 8 | Seated, folding/stacking laundry | 3 min | 74 |
| 9 | Standing/Fidgeting with hands while talking. | 3 min | 77 |
| 10 | 1 minute brushing teeth + 1 minute brushing hair | 2 min | 77 |
| 11 | Driving a car | - | 21 |
| 12 | Hard surface walking w/sneakers | 400m | 76 |
| 13 | Hard surface walking w/sneakers hand in front pocket | 100m | 33 |
| 14 | Hard surface walking w/sneakers while carry 8 lb. object | 100m | 30 |
| 15 | Hard surface walking w/sneakers holding cell phone | 100m | 24 |
| 16 | Hard surface walking w/sneakers holding filled coffee cup | 100m | 26 |
| 17 | Carpet w High heels or dress shoes | 100m | 70 |
| 18 | Grass barefoot | 134m | 20 |
| 19 | Uneven dirt w/sneakers | 107m | 23 |
| 20 | Up hill 5% grade w high heels or dress shoes | 58.5m x 2 times | 27 |
| 21 | Down hill 5% grade w high heels or dress shoes | 58.5m x 2 times | 26 |
| 22 | Walking up stairs (5 floors) | 5 floors x 2 times | 77 |
| 23 | Walking down stairs (5 floors) | 5 floors$\times$2 times | 77 |

2013 and 2014. Data was collected from a single hip-worn triaxial accelerometer, an ActiGraph GT3X+, at a sampling rate of 100Hz. By artificially downsampling the data and creating differently sized windows, we have obtained datasets at a cross section of 6 window sizes and 5 sampling rates. Multiple classifiers were tested out and random forests was selected as the standard classifier for this study. We used our standard classifier to train these datasets with 10-fold cross-validation and statistically observed the trends using repeated measures two-way ANOVA. We then further divided these datasets to observe how these effects change due to activity type or demographic features of the subject.

It should be noted that this study, by necessity, takes into account only certain aspects of HAR classification process. For example, we are utilizing data from a single hip-worn accelerometer, as opposed to other or multiple placements. Similarly, we use only time- and frequency-based features with a single classifier (Random Forests) to further standardize our tests. While feature sets and classifier selection certainly play a role in the outcomes of HAR classification research (Preece et al., 2009), to account for all of them would lead to an significant increase in complexity which could be better examined in future research.

Section 2 details the literature available in this domain. Section 3 describes the data collection and preprocessing done to the data to obtain our data sets. Section 4 gives the results of our classification and statistical analysis of these results. Finally, Section 5 states what we conclude from this work and how these conclusions can be implemented in HAR data classification.

## 2 RELATED WORK

While a considerable amount of research has been done in HAR using accelerometers, there has been a lack of consensus on the methodology of collecting and preprocessing data and thus this topic has largely remained unanalyzed (Preece et al., 2009). Lara and Labrador (2013) note that sampling rates in HAR studies vary from 10Hz to 100Hz while window sizes range from less than 1 second to 30 seconds. While there are some domain-related justifications for such decisions, there is a lack of standardization which likely impacts replicability.

Lau and David (2010) attempted a study similar

to ours, in the sense that multiple data sets of differing window sizes (0.5, 1, 2 and 4 seconds) and sampling rates (5, 10, 20 and 40 Hz) were generated from raw accelerometer data (gathered from a pocketed smart phone) and the effects studied. While they claim that these lower values are sufficient for good performance, their setup consisted of a single test subject performing 5 activities. Maurer et al. (2006), using 6 subjects, state that recognition accuracy does not significantly increase at sampling rates above 15-20Hz when their biaxial accelerometer is used in conjunction with 3 other sensors (light, temperature and microphone). Bieber et al. (2009) calculate that 32Hz should be the minimum sampling rate given human reaction time. Tapia et al (2007) varied window length from 0.5 to 17 seconds and tested the data sets with C4.5 decision tree classifiers, concluding that 4.2 seconds was the optimum window size for their needs. Banos et al (2014) created data sets with window sizes ranging from 0.25 to 7 seconds at interval jumps of 0.25. They found that 1-2 seconds is the best trade-off speed and accuracy for online training. Larger windows were only needed if the feature set was small.

Statistical analysis of classifier performance appears infrequently performed. Most studies, such as the ones cited above, simply state a performance measure (often accuracies and f-measures) but do not present any statistical evaluation. Demsar (2006) comments on the lack of statistical analysis of classifier performance and prefers non-parametric tests for comparing classifiers over parametric ones. The paper also notes that replicability is a problem for most experiments machine learning domain, hence experiments should be tested on as many data sets as possible.

# 3 DATA COLLECTION, PREPROCESSING AND METHODOLOGY

## 3.1 Collecting Data

The data used in the present study was collected in Phoenix, AZ from volunteers recruited through Arizona State University. Participants were fitted with an ActiGraph GT3X+ activity monitor positioned along the anterior axillary line of the non-dominant hip. The monitor was fixed using an elastic belt. The ActiGraph GT3X+ (ActiGraph) is a lightweight monitor (4.6cm x 3.3cm x 1.5 cm, 19g) that measures triaxial acceleration ranging from -6g to +6g. Devices were initialized to sample at a rate of 100hz. Accelerometer data was downloaded and extracted using Actilife 5.0 software (ActiGraph LLC, Pensacola, FL). The subjects performed a number of activities which can be observed in Table 1.

Data from 77 subjects (53 female and 24 male) was used to train the classifiers. The 77 subjects were taken from a larger group of 310 subjects who participated in the study. They were chosen for their relative diversity in both demographics and the activities they performed. Table 1 describes the activities performed while Table 2 provides demographic information on the subjects.

Table 2: Subject Demographics.

| | Mean | Standard Deviation | Range |
|---|---|---|---|
| Age (Years) | 33.2 | 9.7 | 18.2 - 63.2 |
| Height (cm) | 167.9 | 7.9 | 152.6 -188.9 |
| Weight (kg) | 72.1 | 12.1 | 48.3 - 105.5 |
| BMI | 25.6 | 3.9 | 17.7 - 35.4 |

## 3.2 Generating Datasets

As noted earlier, the raw data was collected at a sampling rate of 100Hz. From this, 30 data sets with varying window sizes (of 1, 2, 3, 5 and 10 seconds) with sampling rates (5, 10, 20, 25, 50 and 100Hz) were created. To create data sets for sampling rates $<$ 100Hz, we downsampled from the original data sets, e.g., 50Hz is generated by using every 2nd accelerometer record (100/50), 25Hz using every 4th record (100/25), etc. The number of records in a window then depends on the sampling rate as well as the window size. E.g., A 1-second window at 100 Hz contains 100 records (100x1), a 3-second window at 25Hz contains 75 records (3x25), and so on. As summarized in Table 3, the window size affects the number of records in the data set, a fact that will become significant during analysis.

It should also be noted that, in some situations, partial windows are formed; in these, not enough data exists to form a complete window. Such partial windows were discarded in order to provide the classifier a data set with a uniform format.

Table 3: Number of Records in the Datasets.

| Window Size (s) | No. of Records |
|---|---|
| 1 | 175284 |
| 2 | 88557 |
| 3 | 59666 |
| 5 | 36533 |
| 10 | 19186 |

### 3.3 Feature Extraction and Selection

246 features were extracted using the raw accelerometer data which were then reduced to a 32 feature data set with time- and frequency-based features. The 32 feature set was reduced through correlation-based feature selection, as well as from experts in the domain of human activity recognition. For more information on feature selection, see (Niazi et al., 2016).

- **Features in the Time Domain**: These features include the mean, standard deviation and 50th percentile of each axis (x, y and z) and their vector magnitude as well as the correlation values between the axes.
- **Features in the Frequency Domain**: These features include the dominant frequency and its magnitude for each axis (x, y and z) as well as their vector magnitude.

### 3.4 Methodology

Random forest classifiers perform very well with this data set (Niazi et al., 2016) and so this was chosen as our standard classifier. Each data set was divided and evaluated in 10 folds. Further divisions were carried out for certain activity groups (see Table 4) or demographic groups. The accuracy on the test fold was recorded. WEKA software packages (Hall et al., 2009) were used in conjunction with custom Java code for training and testing the data sets.

RStudio (RStudio Team, 2015) was used to evaluate results. A two-way factorial ANOVA was carried out with weighted least squares to calculate the expected average value (EV) for every combination. It was found that window size and sampling rate as well as their interaction were statistically significant. By determining the maximum expected accuracy (the maximum EV), we discovered the accuracy remained significant at the 95% confidence level. The next section details the analysis and results of our experiments.

Table 4: Division of activities in the clusters.

| Non-Ambulatory Activities | |
|---|---|
| 8,9,10,11 | |
| **Ambulatory Activities** | |
| Walking | 1,2,3,4,12,13,14,15, 16,17,18,19,20,21 |
| Running | 5,6,7 |
| Upstairs | 22 |
| Downstairs | 23 |

## 4 STATISTICAL ANALYSIS OF RESULTS

### 4.1 Weighting

From Table 3, it is clear that window size directly affects the number of records in the data set. Table 5 shows that the variance increases as window size increases, and so the weighting function should be inversely proportional to the variance. For the weighted least squres, we use 1/WindowSize as an approximation.[1] Although sampling rate can also be seen to have a small effect on the variance, it appears negligible. All experiments use this weighting function to normalize the distributions.

Table 5: Standard Deviations.

| | | Sampling Rate (Hz) | | | | | |
|---|---|---|---|---|---|---|---|
| | | 5 | 10 | 20 | 25 | 50 | 100 |
| Window Size (s) | 1 | 0.0035 | 0.0034 | 0.0032 | 0.0029 | 0.0027 | 0.0021 |
| | 2 | 0.0051 | 0.0031 | 0.0048 | 0.0032 | 0.0057 | 0.0032 |
| | 3 | 0.0049 | 0.0071 | 0.0076 | 0.0066 | 0.0040 | 0.0054 |
| | 5 | 0.0045 | 0.0057 | 0.0092 | 0.0108 | 0.0107 | 0.0071 |
| | 10 | 0.0091 | 0.0129 | 0.0074 | 0.0082 | 0.0098 | 0.0096 |

How the standard deviation varies according to window size and sampling rate for the data.

Subsection 4.2 describes in detail the statistical process followed by all the experiments.

### 4.2 All Activities and Demographics

Our first test evaluated all the data available, i.e., for 23 activities as performed by 77 subjects. The objective was to find the maximum average expected value ($EV$) and use this to determine if other values can be considered statistically significant. A two-way analysis of variance (ANOVA) on a Weighted Least Squares (WLS) linear regression model shows that both window size and sampling rate have a significant effect on accuracy with 99% confidence ($p<0.001$). The linear model is then used to obtain $EV$s for all window size/sampling rate combinations. These values are show in Table 6.

Table 6: All Activities/Demographics.

| | | Sampling Rate (Hz) | | | | | |
|---|---|---|---|---|---|---|---|
| | | 5 | 10 | 20 | 25 | 50 | 100 |
| Window Size (s) | 1 | 0.5858 | 0.6868 | 0.7893 | 0.8050 | 0.8251 | 0.8292 |
| | 2 | 0.6324 | 0.7355 | 0.8219 | 0.8334 | 0.8456 | 0.8435 |
| | 3 | 0.6544 | 0.7551 | 0.8269 | 0.8385 | 0.8488 | 0.8411 |
| | 5 | 0.6848 | 0.7752 | 0.8322 | 0.8379 | 0.8473 | 0.8282 |
| | 10 | 0.7316 | 0.8050 | 0.8474 | **0.8529** | **0.8583** | 0.8126 |

Values shown are the average expected value ($EV$) for accuracy on each dataset

[1]The weighting scheme was chosen after a consultation with the University of Georgia Statistics Consulting Center.

The $10s/50Hz$ data set has the highest expected value ($EV_{max}$) for accuracy (in **bold underline** in Table 6) in this experiment. Next we determine if other accuracy EVs are significantly different than the maximum $EV_{max}$. As the alternate hypothesis is that other combinations will have lower $EV$s, we use a 1-sided t-test with a 95% confidence interval.

$$\bar{X_{max}} - \bar{X_k} = t_{290,0.95} * \sqrt{MSE} * \sqrt{\frac{WS_{max}}{n_{max}} + \frac{WS_k}{n_k}} \quad (1)$$

Equation 1 is used to find the critical distance when the sample sizes are unequal but the variance is assumed equal. As each $EV$ represents 10 folds, we have 290 degrees of freedom. The value of $t_{290,0.95}$ is found as 1.651. The $MSE$ value is obtained from ANOVA. $WS$ represents window size of $EV_{max}$ while $WS_k$ and $n$ is the number of observations which in our case is always 10. Having found the critical distance, we can observe which $EV$ values fall inside the margin.

In this experiment, the $10s/25Hz$ value (in **bold** in Table 6) is less than the critical distance away from $EV_{max}$. Hence, it can be concluded that it is statistically as accurate as $EV_{max}$ with 95% confidence.

The procedure elaborated in this section is replicated for all of the following experiments.

## 4.3 Activity Groups

In Table 7, ambulatory activities were separated from non-ambulatory activities while in Table 8 they were classified as walking, running or stairclimbing activities. Both experiments represent a macro-classification and as such exhibit similar patterns to Table 6 — the $10s/50Hz$ has $EV_{max}$.

Tables 7-12 show the results of experiments on different activity group classifications. These groups were divided as shown in Table 4.

However, classifications at a micro-level, within these activity groups, exhibit different results. Classifying between ascending and descending stairs (Table 9) achieves $EV_{max}$ of 97% at $2s/50Hz$. However, statistically significant $EV$s for the experiment are spread across a wide range of window sizes and sampling rates. Interestingly data at lower sampling rates are also deemed significant for larger window sizes. Statistical values for non-ambulatory activities (Table 10) show similar patterns. For walking and running activities, the spread is smaller and concentrated towards higher sampling rates, though there is a lot of variation in window size. Running in particular prefers smaller windows. This is in agreement with the claim by Bieber, et al (2009) that the sampling rate should be more than 32Hz for ambulatory activities.

## 4.4 Demographics

For the next round of experiments, data was separated into demographic groups to observe any significant effects. The data sets were then used to classify all 23 activities.

Division by gender, female (53 subjects) and male (24 subjects) (Tables 13 and 14 respectively) display similar results. $EV_{max}$ is at $10s/50Hz$ for both experiments and there are very similar spreads in significant results. This indicates that there is an insignificant difference in HAR for genders and activity classification should be generalized for both cases.

Data was then divided into 4 age groups; $18-25$ (24 subjects), $26-32$ (24 subjects), $33-44$ (21 subjects) and $49-63$ (8 subjects). The results of these experiments are recorded in Tables 15-18, respectively. There is a visible trend of decreasing window size with increasing age. The spread of significant values gets larger as well.

Similar patterns are noted when the data is divided according to Body Mass Index (BMI) categories; Normal (40 subjects), Overweight (28 subjects) and Obese (9 subjects) (Tables 19-21). As BMI increases, the significance of the $EV_{max}$ decreases along with the window size. Subjects with lower BMIs fare better with larger windows than those with higher BMIs. This can suggest a correlation between age and BMI - elderly people are less likely to be active than young people and are thus more likely to have high BMIs. This hypothesis is supported in Figure 1 which shows that the proportion of normal weighted people decreases with age in the dataset.

## 4.5 Summary of Analysis

Viewing all experiments together suggests that $10s/50Hz$ is the optimal combination of window size and sampling rate, especially if the subjects of the study are young, able-bodied and physically active. Most high significant $EV$ are spread around high sampling rates and window sizes, although there is enough evidence to suggest there is not a very significant loss in accuracy if the sampling rate is decreased to $25Hz$ or window size is decreased to $2s$.

Table 7: Ambulatory vs. Non-Ambulatory Activites.

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.6408 | 0.7295 | 0.8228 | 0.8369 | 0.8559 | 0.8590 |
| 2 | 0.6812 | 0.7735 | 0.8521 | 0.8634 | 0.8754 | 0.8730 |
| 3 | 0.7016 | 0.7957 | 0.8605 | 0.8688 | 0.8791 | 0.8725 |
| 5 | 0.7319 | 0.8127 | 0.8656 | 0.8727 | 0.8796 | 0.8634 |
| 10 | 0.7792 | 0.8419 | 0.8805 | **0.8876** | **0.8913** | 0.8537 |

Table 8: Ambulatory Activity Groups.

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.8345 | 0.8720 | 0.9065 | 0.9106 | 0.9165 | 0.9170 |
| 2 | 0.8345 | 0.8872 | 0.9155 | 0.9177 | 0.9219 | 0.9195 |
| 3 | 0.8609 | 0.8951 | 0.9181 | 0.9211 | 0.9254 | 0.9200 |
| 5 | 0.8754 | 0.9045 | 0.9237 | 0.9267 | 0.9293 | 0.9180 |
| 10 | 0.9022 | 0.9264 | 0.9412 | **0.9411** | **0.9440** | 0.9169 |

Table 9: Stairs: Ascent vs. Descent.

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.9555 | 0.9640 | 0.9675 | 0.9682 | **0.9690** | **0.9694** |
| 2 | 0.9599 | 0.9652 | 0.9681 | **0.9686** | **0.9697** | **0.9690** |
| 3 | 0.9611 | 0.9651 | 0.9670 | 0.9675 | **0.9690** | 0.9673 |
| 5 | 0.9618 | 0.9655 | 0.9668 | 0.9672 | 0.9670 | 0.9647 |
| 10 | 0.9650 | **0.9676** | **0.9676** | **0.9690** | **0.9687** | 0.9624 |

Table 10: Non-Ambulatory Activites.

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.7854 | 0.8298 | 0.8609 | 0.8647 | 0.8711 | 0.8723 |
| 2 | 0.8086 | 0.8471 | 0.8726 | **0.8783** | **0.8795** | **0.8775** |
| 3 | 0.8161 | 0.8476 | 0.8734 | 0.8732 | 0.8780 | 0.8746 |
| 5 | 0.8246 | 0.8525 | 0.8682 | **0.8730** | **0.8726** | 0.8594 |
| 10 | 0.8406 | 0.8571 | 0.8713 | **0.8716** | **0.8716** | 0.8514 |

Table 11: Walking Activites.

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.5556 | 0.6656 | 0.7916 | 0.8105 | 0.8329 | 0.8385 |
| 2 | 0.5976 | 0.7162 | 0.8274 | 0.8407 | **0.8581** | **0.8574** |
| 3 | 0.6189 | 0.7415 | 0.8344 | **0.8460** | **0.8598** | 0.8543 |
| 5 | 0.6474 | 0.7594 | 0.8374 | 0.8408 | 0.8527 | 0.8353 |
| 10 | 0.6875 | 0.7746 | 0.8387 | 0.8491 | **0.8557** | 0.8159 |

Table 12: Running Activites.

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.7081 | 0.7795 | 0.8522 | 0.8688 | 0.9070 | **0.9140** |
| 2 | 0.7349 | 0.8191 | 0.8793 | 0.8961 | **0.9185** | **0.9210** |
| 3 | 0.7418 | 0.8321 | 0.8891 | 0.8968 | **0.9176** | **0.9177** |
| 5 | 0.7584 | 0.8266 | 0.8703 | 0.8863 | 0.8953 | 0.8972 |
| 10 | 0.7728 | 0.8333 | 0.8639 | 0.8714 | 0.8759 | 0.8553 |

Table 13: Gender: Female Subjects.

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.6037 | 0.7132 | 0.8128 | 0.8227 | 0.8405 | 0.8430 |
| 2 | 0.6509 | 0.7606 | 0.8388 | 0.8490 | **0.8599** | 0.8554 |
| 3 | 0.6762 | 0.7762 | 0.8433 | 0.8529 | **0.8598** | 0.8498 |
| 5 | 0.7052 | 0.7937 | 0.8441 | 0.8490 | 0.8539 | 0.8351 |
| 10 | 0.7521 | 0.8164 | **0.8586** | **0.8595** | **0.8667** | 0.8169 |

Table 14: Gender: Male Subjects.

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.6439 | 0.7248 | 0.8139 | 0.8265 | 0.8474 | 0.8508 |
| 2 | 0.6857 | 0.7633 | 0.8412 | 0.8506 | **0.8653** | 0.8624 |
| 3 | 0.7017 | 0.7815 | 0.8478 | 0.8569 | **0.8675** | 0.8597 |
| 5 | 0.7226 | 0.7984 | 0.8484 | 0.8547 | **0.8641** | 0.8408 |
| 10 | 0.7759 | 0.8183 | **0.8636** | **0.8678** | **0.8736** | 0.8253 |

Table 15: Age: 18-26 Years

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.6207 | 0.7174 | 0.8094 | 0.8236 | 0.8432 | 0.8457 |
| 2 | 0.6662 | 0.7620 | 0.8362 | 0.8488 | 0.8588 | 0.8553 |
| 3 | 0.6857 | 0.7824 | 0.8443 | 0.8559 | 0.8629 | 0.8551 |
| 5 | 0.7196 | 0.8024 | 0.8484 | 0.8542 | 0.8623 | 0.8424 |
| 10 | 0.7633 | 0.8292 | **0.8627** | **0.8717** | **0.8753** | 0.8250 |

Table 16: Age: 27-33 Years

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.6614 | 0.7513 | 0.8343 | 0.8428 | 0.8590 | 0.8618 |
| 2 | 0.7043 | 0.7891 | 0.8564 | 0.8676 | **0.8746** | **0.8731** |
| 3 | 0.7198 | 0.8051 | 0.8623 | 0.8678 | **0.8779** | 0.8677 |
| 5 | 0.7390 | 0.8117 | 0.8573 | 0.8643 | 0.8679 | 0.8488 |
| 10 | 0.7784 | 0.8292 | 0.8658 | **0.8695** | **0.8720** | 0.8250 |

Table 17: Age: 34-44 Years

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.6651 | 0.7660 | 0.8442 | 0.8547 | 0.8689 | 0.8722 |
| 2 | 0.7085 | 0.8038 | 0.8654 | 0.8730 | **0.8849** | **0.8805** |
| 3 | 0.7271 | 0.8193 | 0.8651 | 0.8730 | **0.8807** | 0.8696 |
| 5 | 0.7482 | 0.8226 | 0.8596 | 0.8624 | 0.8721 | 0.8533 |
| 10 | 0.7833 | 0.8424 | 0.8733 | **0.8792** | **0.8822** | 0.8375 |

Table 18: Age: 49-63 Years

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.7593 | 0.8382 | 0.8892 | 0.8981 | 0.9065 | 0.9063 |
| 2 | 0.7856 | 0.8581 | 0.9043 | **0.9084** | **0.9135** | **0.9146** |
| 3 | 0.8046 | 0.8689 | 0.9040 | **0.9067** | **0.9101** | 0.9030 |
| 5 | 0.8201 | 0.8730 | 0.9031 | 0.9017 | **0.9084** | 0.8855 |
| 10 | 0.8503 | 0.8986 | **0.9114** | **0.9114** | **0.9119** | 0.8725 |

Table 19: BMI: Normal

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.6031 | 0.7074 | 0.8056 | 0.8188 | 0.8363 | 0.8393 |
| 2 | 0.6531 | 0.7525 | 0.8320 | 0.8437 | 0.8531 | 0.8503 |
| 3 | 0.6776 | 0.7753 | 0.8395 | 0.8493 | 0.8553 | 0.8478 |
| 5 | 0.7138 | 0.7946 | 0.8446 | 0.8482 | 0.8549 | 0.8376 |
| 10 | 0.7617 | 0.8204 | **0.8614** | **0.8615** | **0.8678** | 0.8149 |

Table 20: BMI: Overweight

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.6419 | 0.7381 | 0.8256 | 0.8391 | 0.8564 | 0.8597 |
| 2 | 0.6831 | 0.7762 | 0.8520 | 0.8609 | **0.8714** | **0.8689** |
| 3 | 0.7002 | 0.7940 | 0.8523 | 0.8612 | **0.8701** | 0.8637 |
| 5 | 0.7225 | 0.8064 | 0.8549 | 0.8619 | **0.8696** | 0.8494 |
| 10 | 0.7612 | 0.8287 | 0.8607 | **0.8674** | **0.8732** | 0.8252 |

Table 21: BMI: Obese

| Window Size (s) \ Sampling Rate (Hz) | 5 | 10 | 20 | 25 | 50 | 100 |
|---|---|---|---|---|---|---|
| 1 | 0.7423 | 0.8279 | 0.8803 | 0.8900 | 0.8998 | 0.9015 |
| 2 | 0.7817 | 0.8532 | 0.9008 | 0.9039 | **0.9164** | **0.9115** |
| 3 | 0.7968 | 0.8648 | 0.9010 | **0.9098** | **0.9167** | **0.9098** |
| 5 | 0.8164 | 0.8663 | 0.8943 | 0.9001 | 0.9070 | 0.8878 |
| 10 | 0.8368 | 0.8774 | 0.8994 | **0.9125** | **0.9091** | 0.8648 |

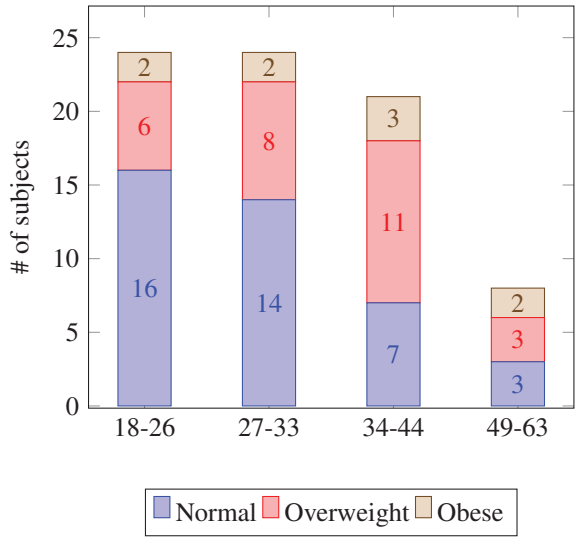



Figure 1: Distribution of BMI groups over age groups.

## 5 CONCLUSION

This study provides some basis for the selection of sampling rates and window sizes for human activity recognition. The analysis indicates that $10s/50Hz$ is statistically the best combination for data collected with a single hip-worn accelerometer. Most of the experiments carried out preferred larger windows and high sampling rates though some low intensity activities and demographics can perform better with smaller windows. Our analysis further suggests that window size can vary between 2-10 seconds and sampling rate 25-100Hz for different situations without a significant loss in performance. While our study has shown that larger windows are preferable, smaller windows can still provide significant results if power consumption is an issue. Additionally, lower values are preferable for studies involving the less dynamic activities or subjects who are more liable to be less active.

Future work in this field should be done to understand aspects of Human Activity Recognition better. This study was performed under some assumptions that can be scrutinized. The placement of the accelerometer could be shown to affect classifier performance for different activities — a combination of sensors can also be used. Other sensors, such as heart rate monitors or video image processors, provide new avenues. This study can also be replicated using different classifiers or learning methods with different feature sets. Extensive analysis on the statistical value of other machine learning and data mining methods could also help the field as a whole.

## REFERENCES

ActiGraph. Actisoft analysis software 3.2 user's manual. fort walton beach, fl: Mti health services.

Banos, O., Galvez, J.-M., Miguel Damas, H. P., and Rojas, I. (2014). Window size impact in human activity recognition. *Sensors*, pages 6474–6499.

Beiber, G., Voskamp, J., and Urban, B. (2009). Activity recognition for everyday life on mobile phones. *International Conference on Universal Access in Human-Computer Interaction*, pages 289–296.

Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18.

Lara, O. D. and Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communication Surveys and Tutorials*, 15:1192–1209.

Lau, S. and David, K. (2010). Movement recognition using the accelerometer in smartphones. *IEEE Future Network ands Mobile Summit 2010*, pages 1–9.

Maurer, U., Smailagic, A., Siewiorek, D. P., and Deisher, M. (2006). Activity recognition and monitoring using multiple sensors on different body position. *International Workshop on Wearable and Implantable Body Sensor Networks*.

Niazi, A., Yazdensepas, D., Gay, J., Maier, F., Rasheed, K., Ramaswamy, L., and Buman, M. (2016). A hierarchical meta-classifier for human activity recognition. *IEEE International Conference on Machine Learning and Applications*.

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., Howard, D., Meijer, K., and Crompton, R. (2009). Activity identification using body-mounted sensorsa review of classification techniques. *Physiological Measurement*, 30(4):R1.

RStudio Team (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.

Tapia, E., Intille, S., Haskell, W., Larson, K., Wright, J., King, A., and Friedman, R. (2007). Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. *IEEE 11th IEEE international symposium on wearable computers*, pages 37–40.

Yazdansepas, D., Niazi, A. H., Gay, J. L., Maier, F. W., Ramaswamy, L., Rasheed, K., and Buman, M. P. (2016). A multi-featured approach for wearable sensor-based human activity recognition. *IEEE International Conference on Healthcare Informatics (ICHI), Chicago, IL*.

# Laboratory Information Management System for NGS Genomics Labs

Jitao Yang[1,2], Shuai Zhu[2], Guangpeng Xu[2] and Chao Han[2]
[1]*School of Information Science, Beijing Language and Culture University, Bejing 100083, China*
[2]*Annoroad Gene Technology, Beijing 100176, China*
*jitao.yang@gmail.com*

Keywords: LIMS, NGS, Genome Sequencing, Genomics, Sample Tracking.

Abstract: The goal of genome sequencing is to unravel the ordered sequence of nucleic acids that form the DNA or RNA of a given sample. Genome sequencing lab requires the ability to select and track a large amount of samples through many experimental steps. Therefore, laboratory information management system (LIMS) is needed to provide a way of automating the laboratory experimental procedures and track the samples. LIMSs have been proposed and developed for many years, but still remain difficult for labs to implement successfully. In this paper, we demonstrate our genomic next generation sequencing (NGS) LIMS solution. We developed a web-based LIMS with flexible configuration and customization for NGS laboratories, and can help laboratories track samples and optimize experimental procedures and business workflows. We also describe our solution of integrating LIMS with the existing enterprise business information systems. Finally, we share our experience for the implementation of a successful LIMS.

## 1 INTRODUCTION

Genome next generation sequencing is now commonly adopted, and has a broad areas of applications, such as the Non-invasive Prenatal DNA Testing, ctDNA Testing for Non-invasive Tumor Personalized Therapy, Plant and Animal Molecular Breeding, Genetics and Evolution, Microorganism and Ecological Environment, etc.

To manage tens of thousands of samples that are subject to NGS analysis, it is inevitable to develop adequate laboratory information management system to track and manage the NGS workflows. However, it is extremely difficult to manage the NGS expriment and analysis procedures accurately and efficiently, given the issues of multiple-source samples enrolling, national and international logistical management and tracking, fragmented procedures for assessment and processing of samples, the intricacies of molecular experimental steps, and the complex and multiple pipelines of NGS processing.

LIMS provides many benefits for the users of laboratory, several of the main benefits identified are outlined below:

- brings accuracy and accessibility to the flow of samples and data in laboratory,
- universally accessible data via the web rather than digging through files,
- years of data can be kept and queried conveniently,
- business efficiency improvement,
- data quality control,
- efficient sample tracking and management,
- automated and in-depth customer reports,
- integration with laboratory instruments,
- experimental steps quality control,
- building automated analysis pipelines,
- status and results sharing with collaborators and clinicians,
- financial management,
- access control,
- track and analyze trends,
- error reduction.

In this paper, we demonstrate our genomic NGS LIMS solution. Our LIMS is a web-based system with flexible configurations and customizations for NGS laboratories and can support the integration of multiple NGS instruments, it can help laboratories track samples and optimize experimental procedures and business workflows. We also give solution of integrating LIMS with the existing enterprise business information systems. Finally, we describe the most



Yang, J., Zhu, S., Xu, G. and Han, C.
Laboratory Information Management System for NGS Genomics Labs.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 326-333
ISBN: 978-989-758-213-4

likely causes for a failed LIMS and share our experience for the implementation of a successful LIMS.

The rest of the paper is structured as follows: Section 3 introduces the design of LIMS, while Section 4 describes the implementation of LIMS, and Section 5 presents the integration of LIMS with existing information systems. Section 6 discusses the implementation of a successful LIMS. Section 7 concludes the paper and outlines future work.

## 2 RELATED WORKS

Some commercial and open source LIMSs (Bath et al., 2011; PerkinElmer, 2016; Grimes and Ji, 2014; Progeny, 2016; Illumina, 2016) are available but typically require extensive modification and extension to address the specific needs of NGS genomics labs. We therefore developed a web-based LIMS, which is robust and flexible for managing the samples and the NGS processes.

## 3 SYSTEM DESIGN

As described in Figure 1, the functionalities of our LIMS can be grouped into the following categories:

- Enrollment of Sample Information. Enrollment of sample information is not a simple form filling process but constituted by two or even three steps, and supports multiple business models. For instance, salesman who stationed at hospitals will input the basic sample information (e.g., sample code, sample type, photo of sample sheet) into system by mobile applications embedded in Wechat. Then, when the sample along with the sample sheet are transported to company, typists in company will finish the full sample information (in sample sheet) enrollment. Finally, the correctness of sample information enrolled in LIMS will be checked by another team in company.
- Sample Logistics Information Tracking. Samples come from different cities distributed in the country, and will be transported to our head quarter to be tested. Each sample logistics information will be tracked by system. When samples are packed and sent to our company, package logistics code will be scanned into system. Then, the system can get the newest logistics information from express companies through their open APIs. We can check the package logistics information status during the whole transportation process. If something unusual (e.g., delayed, destroyed) happened in logistics process, the recipients and senders can response quickly to take steps to minimize the losses.
- Sample Assessment and Processing. Depends on sample attribute, sample sheet, sample number, test type or more factors, the samples arrived at company can be divided to different processing directions including rejection, resampling, storage, or flow to the experiment steps. Sample recipients, experiment operators and genomic analysts can choose the processing directions and operate in the system.
- Experimental Management. Experimental management is the most important and complicated part of LIMS. Based on sample type, test type, previous processing result and operator's subjective judgment, the samples will experience multiple experimental steps such as plasma isolation, nucleic acid extraction, molecular library construction, molecular library quality control, and etc. Our system is also flexible to support the configuration of different experimental processes.
- Sequencing of Samples. After the experimental steps, the samples are ready to be sequenced by NGS instrument (e.g., Illumina HiSeq X Ten). To minimize the cost of using sequencing reagents, generally, one NGS instrument operation will be required to sequencing as many samples as possible, therefore, samples belonging to different test business line will be combined together in the pooling process. After sequencing, the raw .bcl data will be separated into deferent genomic analysis business processes. Abnormal result data will be auto labeled by particular rules configured.
- Genomic Analysis. Before genomic analysis, LIMS will prepare the raw .bcl data for the filter and quality control software to be processed. The generated FASTQ data will be further prepared for the genomic analysis pipelines operated in the cluster computing environment. After genomic analysis, LIMS will retrieve the analysis result set to system so as to wait for the genetic interpretation.
- Genetic Interpretation. Sequencing and analysis results are presented in a neat and orderly manner in LIMS for the genetic interpretation scientists. Additionally, LIMS can provide relevant knowledge base for genetic interpretation such as gene-disease associations from several public data sources (e.g. AutDB (Mindspec, 2016), DisGeNET (Pinero et al., 2015), OMIM (Hamosh et al., 2015)) and the literatures.

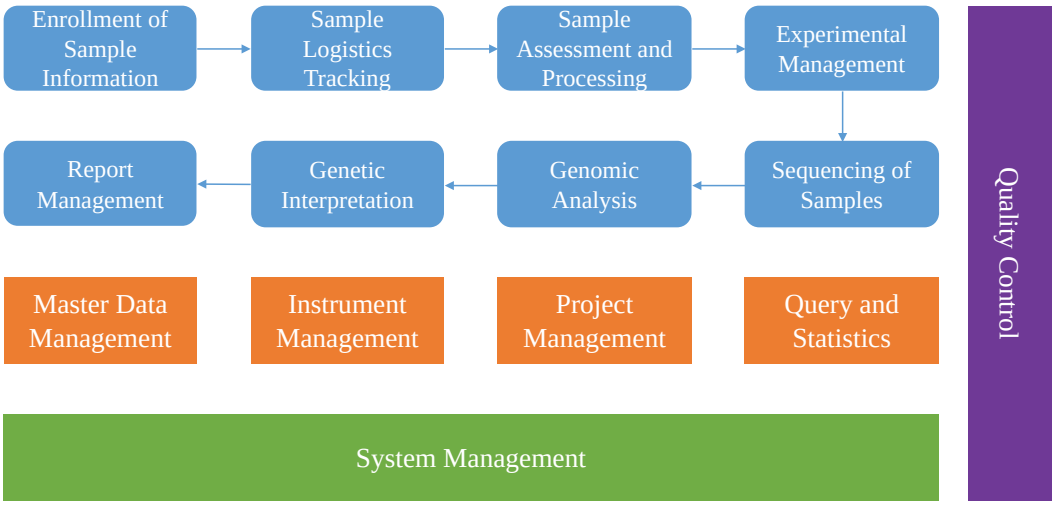



Figure 1: The Application Architecture of LIMS.

- Report Management. The genetic interpretation result will be classified by the report management module and transferred into different report generation pipelines of the reporting system.
- Quality Control Management. Quality control exists in the whole NGS work flows, including experiment steps quality control, reagents quality control, sequencing data quality control, genomic analysis quality control, genetic interpretation quality control, and etc. Different quality control processes can also be configured in system.
- Master Data Management. Master data management module includes tens of thousands of fundamental data for the operation of LIMS, such as the details data for experiment templates, barcode rules, agent information and so on. This module is usually maintained by department manager or system administrator, because master data include very important and sensitive data of the company.
- Instrument Management. Instrument management module not only records instrument property information like barcode, type, number, location, status, application methods and so on, but also associates instruments to a certain experiment step. An experiment step including its associated instruments could be exported as a detail guide which is very helpful for the operators.
- Project Management. Project management module provides the functions for project managers a convenient way to manage and monitor the projects, including project defining, project status monitoring, process intervening, and so on.
- Query and Statistics. Query and statistics module provides powerful functions of advanced data query and visualized multidimensional data statistic and analysis. Instead of querying through millions of data items to find meaningful results in experimental workflows, researchers can quickly identify the information of their interests through this module.
- System Management. System management module provides the features of user management, role management, privilege management, department management, business management, organization management. A user's role and his privileges could configured flexibly.

## 4 SYSTEM IMPLEMENTATION

Our LIMS is written in Java using the Spring (Pivotal, 2016), SpringMVC, Mybatis (Goodin et al., 2016) web application framework and implementation is platform-independent. Our web servers run Linux/CentOS and we use the MySQL relational database management system. The system architecture of LIMS is described in Figure 2.

The web interface is designed to handle a variety of functions in a modular format. The left column includes the function category of business process. The top navigation bar provides advanced management and analysis functions.

We have more than one hundred data tables in the system, Figure 3 shows part of our database schema for data tables.

HTTPS (Hypertext Transfer Protocol Secure) is supported and is implemented for the system. User's access to functionality is controlled via user roles

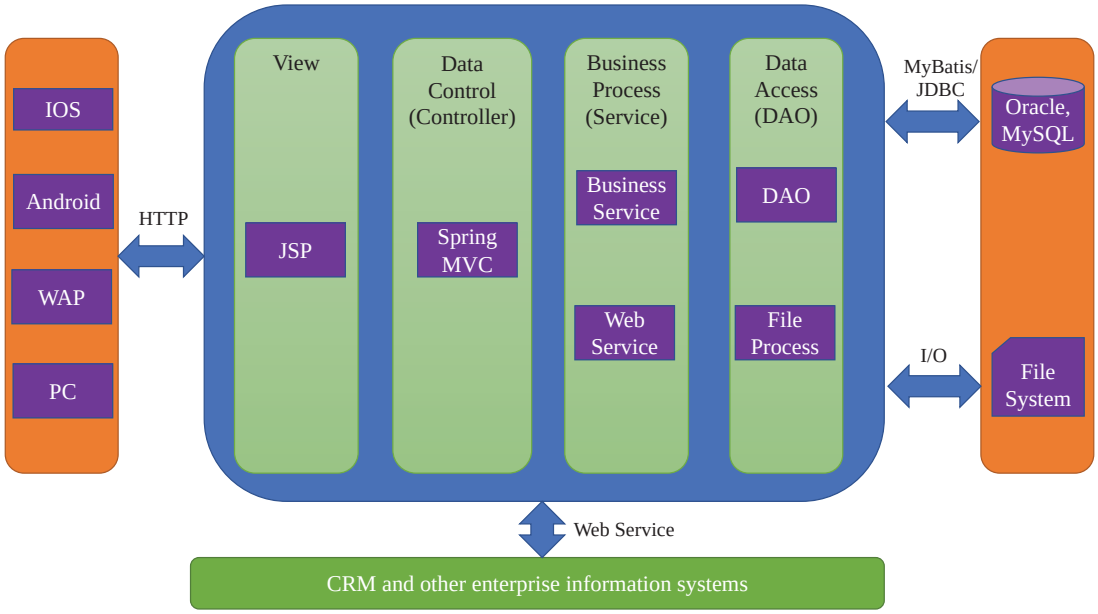



Figure 2: The System Architecture of LIMS.

which are defined and managed by the system management module.

Our LIMS currently supports the following Illumina NGS sequencing platforms: HiSeq X Ten, MiSeq, HiSeq, HiSeq 4000, HiSeq2500, NextSeq 500/550AR, and could be configured for other type of NGS instruments.

# 5 INTEGRATION WITH EXISTING SYSTEMS

Implementation of a LIMS requires a good degree of integration with the existing business information systems in enterprise. We integrate LIMS to our two existing business system platforms as demonstrated in Figure 4 and Figure 5.

## 5.1 Integrating with Company Systems

Figure 4 describes the business process in our company.

Customer Relationship Management (CRM) system serves primarily our clinical and health services, including Personal Genome Test, NIPT (Non-Invasive Prenatal DNA Testing), Cancer Gene Therapy, and etc. CRM supports sample information collection, hospital information management, agent management, product management, sales management, sample management, doctor management, financial management, customer reports management, and etc.

Project Management (PM) system serves primarily our technical services, including the research area of genomics, transcriptomics, epigenetics, and etc. PM includes the functions of sample information enrollment, contract management, project schedule management, project establishment, job order management, quality control, research achievement management, and etc.

Logistics system provides the management functions of express package, express company, alert time, receipt time, samples, logistics tracking, and etc.

All the information related to NGS experiments and analysis will be synchronized into LIMS, therefore, the sample information could be enrolled through:

- CRM web portal, or
- CRM mobile app, or
- PM web portal, or
- PM mobile app, or
- LIMS directly.

The sequencing result data will be stored in Network Attached Storage (NAS) system, and then filtered and classified to be analyzed by hundreds of different genomic analysis pipelines.

The genomic analysis result data will then flow to the reporting system, which can produce hundreds of different areas of professional reports automatically.

The reporting system will finally distribute the user reports to CRM and PM, then the user can receive

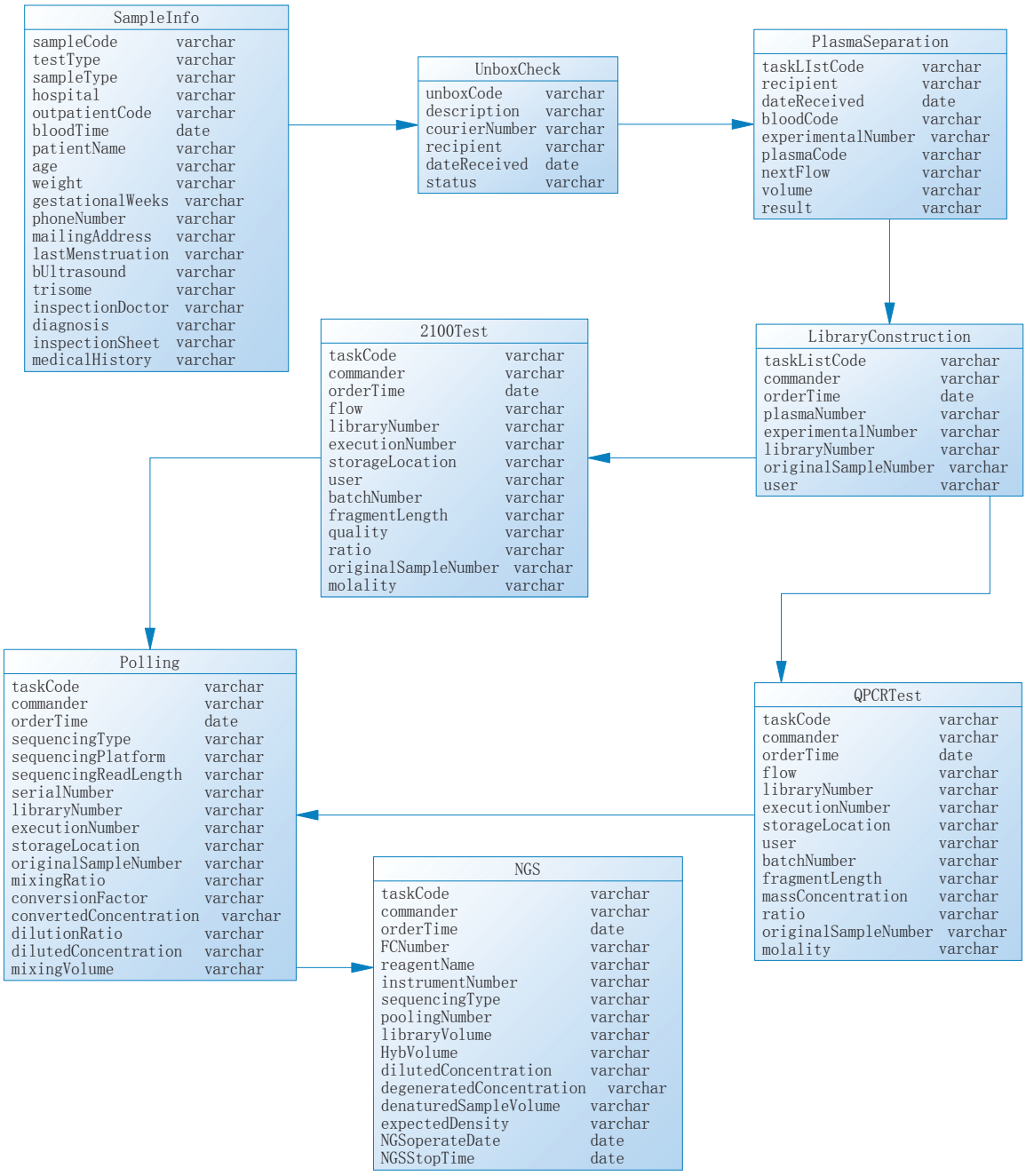



Figure 3: Part of Database Schema for LIMS.

their reports through our mobile app or web portal.

We have also constructed a hybrid cloud computing platform SolarGenomics (SolarGenomics, 2016), because our business can produce around 10 TB sequencing data everyday, both the local computing cluster and the local distributed storage can not provide computing and storage resources sufficiently and elastically.

## 5.2 Integrating with Hospital Systems

Figure 5 describes the business process in our collaborating hospitals.

We collaborate with more than 2000 hospitals. In some of the hospitals, we set up NGS laboratory and install our NGS instrument there so that the hospitals will have the abilities of NGS experiments, sequenc-

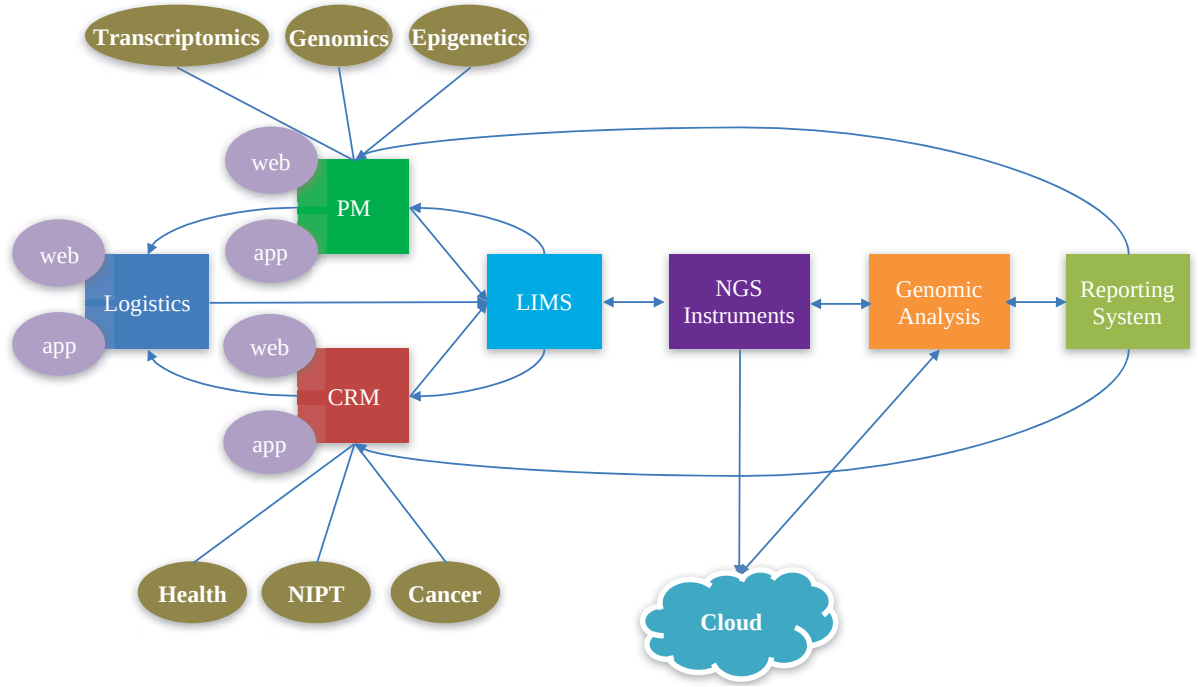



Figure 4: The Business Process System Platform in Enterprise.

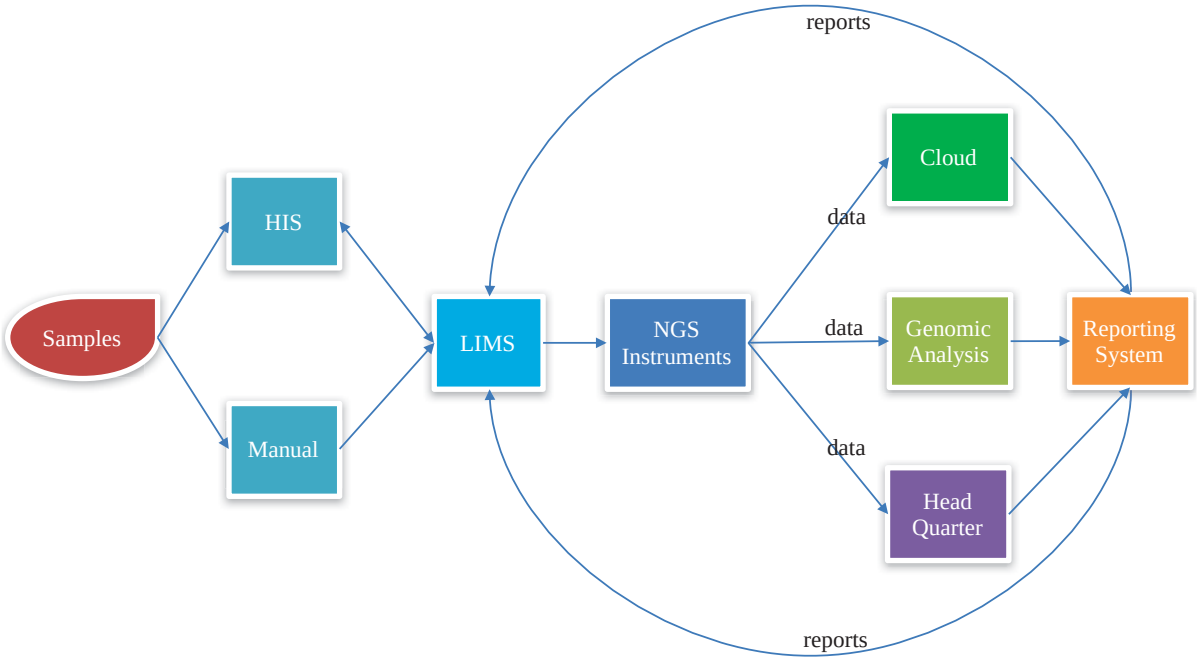



Figure 5: The Business Process System Platform in Hospital.

ing, genomic analysis, and genetic interpretation.

The samples in our collaborating hospitals could be enrolled to LIMS either manually or synchronized from the hospital information system (HIS). The samples will be processed by multiple experimental steps in the laboratory of hospital and then be sequenced by the NGS instrument placed in the hospital.

Further, the sequencing data will be:

- analyzed by our genomic analysis pipelines installed in the local hospital computing servers, or

- uploaded to our SolarGenomics cloud computing platform then analyzed by our genomic analysis pipelines on cloud, or
- transfered to our head quarter then analyzed by our bioinformatics scientists.

Finally, our reporting system will give professional reports to the clinical doctors or patients.

# 6 A SUCCESSFUL LIMS IMPLEMENTATION

A large number of LIMS implementations failed to meet the user's initial expectations, and this can be due to the lack of proficient user requirements specifications, the frequent requirements changes, and the technology-based shortcomings.

The absence of adequate requirements is the biggest reason why a LIMS may fail, as generally the software developers are seldom have sufficient knowledge of experiments and NGS, the success of a LIMS relies on the deep understanding of laboratory and genomic analysis business needs.

Frequent requirements changes is another reason causing the fail of LIMS, as the laboratory staffs are normally focused on a specific area of experiment, it is hard for them to identify clearly how a LIMS is going to fit into their laboratory situation and to propose a general function architecture of LIMS. Additionally, the lab staffs lack the IT knowledge and skills, they will think the modification of system functions as a simple task, therefore, the requirements will be modified frequently. This will postpone the deliver of project and even fail to publish LIMS.

Technology-based shortcomings should be emphasized, as the information technology department is often not the first class citizen in gene technology organizations.

Implementation of a LIMS project will also require a good degree of integration with the existing enterprise business information systems, data exchange mode and standard must be adequately addressed.

Mapping out clearly the requirements specifications, controlling the requirements changes effectively, addressing the problem of integrating LIMS into existing enterprise systems, relying on a strong IT team will stand a quite high chance of implementing LIMS successfully.

# 7 CONCLUSIONS

To meet the needs of managing tens of thousands of samples for genome sequencing, we developed a web-based laboratory information management system that is flexible to be configured to adapt to next generation sequencing technologies. A LIMS system is critical to the accurate and effective management of sample information, experimental data, genome sequencing data, and the reproducible analysis results. Our LIMS addresses all of these needs and seamless integrates with the existing systems in enterprise and in collaborating hospitals.

In our LIMS, all the data are stored into distributed authoritative repositories, samples are traceable from the enrollment, transportation, experiment, a sequencing run, quality control, genomic analysis, genetic interpretation, generation of report, to sample storage, and all the other processing steps in between. In conjunction with a sample identifier (QR code) encoding rules and advanced query and analysis capabilities, LIMS can quickly identify sample and significantly reduce errors in the whole steps of business process. Our LIMS provides a comprehensive and efficient management solution for NGS genomics labs.

LIMS is very complicated and difficult to implement, especially in the NGS research laboratory, we therefore share our experience of implementing a successful LIMS.

Regarding further work, it is promising in our schedule that we will provide a cloud-based LIMS solution in our SolarGenomics genome sequencing big data cloud platform, and open the service for more genomics labs.

# ACKNOWLEDGEMENTS

This work was partially supported by the Science Foundation of Beijing Language and Culture University (supported by "the Fundamental Research Funds for the Central Universities") (15YJ030001, 14YBB12, 16YJ030001).

# REFERENCES

Bath, T. G., Bozdag, S., Afzal, V., and Crowther, D. (2011). Limsportal and Bonsailims: development of a lab information management system for translational medicine. In *Source Code Biol Med*.

Goodin, B., Poitras, C., and Begin, C. (2016). Mybatis. In *http://blog.mybatis.org/*.

Grimes, S. M. and Ji, H. P. (2014). Mendelims: a web-based laboratory information management system for clinical genome sequencing. In *BMC Bioinformatics*.

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., and McKusick, V. A. (2015). Online mendelian inheritance in man (omim), a knowledgebase of human genes and genetic disorders. In *Nucleic Acids Research*.

Illumina (2016). Basespace Clarity LIMS. In *http://www.illumina.com/informatics/research/sequencing-data-analysis-management/genologics-lims.html*.

Mindspec (2016). Autdb: a Genetic Database for Autism Spectrum Disorders. In *http://www.mindspec.org/products/autdb/*.

PerkinElmer (2016). Genesifter Lab Edition. In *http://www.geospiza.com/Products/LabEdition.shtml*.

Pinero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., Sanz, F., and Furlong, L. I. (2015). Disgenet: a discovery platform for the dynamical exploration of human diseases and their genes. In *Database (Oxford)*.

Pivotal (2016). Spring framework. In *http://projects.spring.io/spring-framework/*.

Progeny (2016). Progeny LIMS. In *http://www.progenygenetics.com/lims/*.

SolarGenomics (2016). Solargenomics genome sequencing big data platform. In *www.solargenomics.com*.

# Acquisition of Confidential Patient Data Over Shared Mobile Device

Petr Vcelak[1,2], Martin Kryl[2], Ladislav Racak[2] and Jana Kleckova[1,2]
[1]*NTIS – New Technologies for the Information Society, University of West Bohemia, Univerzitni 8, Plzen, Czech Republic*
[2]*Department of Computer Science and Engineering, University of West Bohemia, Univerzitni 8, Plzen, Czech Republic*
{*vcelak, kryl, racakl, kleckova*}*@kiv.zcu.cz*

Keywords: Client-server, Data Acquisition, Mobile Application, Patient Data, Security, Shared Mobile Device, Smartphones, Tablets.

Abstract: Mobile devices have already been designed for many applications. Smartphones and tablet computers are modern, widespread and affordable solutions used for various purposes. Nowadays mobile devices are widely used in telemedicine. It is usually assumed, that the device is owned and used by a single person. We focus on security concerns and constraints from a different point of view – when the device is shared. In this paper, we are proposing a novel approach to prevent leakage of patient's confidential data when the device is used by multiple patients at the hospital's clinic or department. We present a prototype application and discuss its use case and designed workflow.

## 1 INTRODUCTION

Our goal is to enable gathering of patient's data via shared mobile device. Hospital would lend the configured device to patients and they would be able to fill out pre-defined medical questionnaires and personal information forms while waiting for their medical examination in situation, when medical personnel is busy examining another patient. Some types of data can be conveniently collected at this time. The doctor would then pass and verify data, possibly complementing deficiencies. It is not just about gathering general information like address, contact or health insurance situation. Forms could be used to gather more information about patient's perceived discomfort during the last period. Patient can also provide data for surveys, data for conducting studies or use mobile devices to measure additional data such as weight, blood pressure or pulse through wearable electronics. At this time, we do not expect the use of additional equipment like wearable electronics and sensors.

## 2 STATE OF THE ART

The main target in this paper are mobile phones such as smartphones and electronic tablets. Nowadays mobile phones are used in all healthcare areas including diagnostics, telemedicine, research, reference libraries and interventions. (Bastawrous and Armstrong, 2013) Currently, these devices are easily available, inexpensive, small, have enough computing power and provide sufficient space for the development of various types of new applications. An advantage is a user-friendly interface and availability of installed applications.

Different mobile applications are widely used in medicine with the aim to provide personalised approach or just for gathering health data. As Hayes et al. (2014) said, there were areas where patient-tailored risk prediction and treatment had been applied routinely in the clinic over mobile applications. Nevertheless, authors said, more work would be required to translate scientific advances into individualised treatment in other fields. (Hayes et al., 2014) There were publications regarding that, eg. a smartphone-centric platform for remote health monitoring of health failure (Bisio et al., 2015) or cloud-based smart health monitoring system for automatic cardiovascular and fall risk assessment in hypertensive patients (Melillo et al., 2015).

Android is the best selling operation system on tablets since 2013, and on smartphones it is dominant by any metric. (Manjoo, 2015) We can cite many articles and examples describing data acquisition via mobile devices on the Android platform, eg. a portable physical health monitoring system were proposed in (Tang et al., 2015) and continuous wireless monitoring of endogenous and exogenous bio-molecules on an android interface in (Stradolini et al., 2015).



Vcelak, P., Kryl, M., Racak, L. and Kleckova, J.
Acquisition of Confidential Patient Data Over Shared Mobile Device.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 334-339
ISBN: 978-989-758-213-4

On the other side, security and privacy issues are serious topics. Authors Baig et al. discussed mobile healthcare applications and its critical issues and challenges. As they said in (Baig et al., 2015), mobile phones were becoming important in monitoring and even in delivering of healthcare interventions. Results of testing mobile health applications on Android platform was discussed in (Knorr and Aspinall, 2015) where a number of serious vulnerabilities were discovered in the most popular applications. Dehling et al. (2015) did an overview of security and privacy infringements in mobile health applications on Android and iOS. They discovered that the majority of apps (95.63%, 17,193/17,979; of apps) had posed at least some potential damage through information security and privacy infringements. There were 11.67% (2,098/17,979) of apps that scored the highest assessments of potential damages. (Dehling et al., 2015) These results lead to belief that private or confidential information stored in a mobile device/application are at a risk.

# 3 SHARED DEVICE APPROACH

## 3.1 General Information

An obtaining data at own mobile device is an usual approach. Unfortunately, this may not be a safe way. Unlike the usual situation where everyone has their own mobile phone, we start from the opposite assumption. We consider the use of device that is not owned by the patient. It is only borrowed at a given moment. We refer to this as a **shared device** (SD). The patient has to fill new information or update existing data (delivered to/from the information system of the hospital or medical doctor) by the shared device.

An advantage of the SD approach is the ability to fully control its system and customise it. We can set up the environment and install all necessary tools, including our own custom applications. Through our custom application we can provide personalising. It can be determined in advance what information or questionnaire needs to be filled by particular patient. Different type and extent of data might be desired for different departments, patients, diagnosis or type of visit.

After the registration at the desk/office/nurse, patient can get pre-configured shared device, and can immediately begin checking and filling the form. Optionally, patient may be allowed to switch to other applications, eg. read news or play some simple games, while still waiting for examination.

Disadvantages of shared devices are especially the need to solve the issue of patient's privacy. Is it really a disadvantage though? Sensitive or protected data must not be available to the next patient/personnel using the device. We have to prevent data leaks when someone steals the shared device. Of course, data transfer have to be secured. What about other health mobile applications? Any kind of health application has to secure its data as a prevention to data leak, otherwise it leads to security and privacy infringements. Well, the data privacy have to be solved with any kind of application that works with patient health information. The difference for shared device approach is, that we are absolutely sure the application will be used by multiple users, therefore it has to be secured better.

## 3.2 Architecture

We have chosen client-server architecture where client will be responsible for a user interaction. The business logic will be on the server side only, eg. form definition and description, form source/input data and produced output. Client-server communication is based on REST API with public and private key-pairs encrypted messages. The HTTPS protocol is recommended.

### 3.2.1 Client

Client, as a mobile application, will provide direct interaction with users. User can be a staff member or a patient itself. We prefer to make client application as simple as possible. In the figure 1 you can see use case diagram. We expect no local configuration stored in a shared device with the exception of the URL server address. Security details are described in a separate section 3.4. Basically, the client application is prepared when URL address of the server is set. The shared device can operate in two modes:

**user** have to authenticate by username and password knowledge,

**delegated** no user authentication; server sends available forms and identifies retrieved data by device ID.

In user mode, there is no list of available users on SD for a user authentication. Instead an encrypted request (with username and password) is sent to server for authentication. After a user is authenticated, forms (only available to the user) are downloaded , and the user can choose either to be filled. At the same time the user can also choose to continue filling up an unfinished form, that has not been uploaded yet. Client

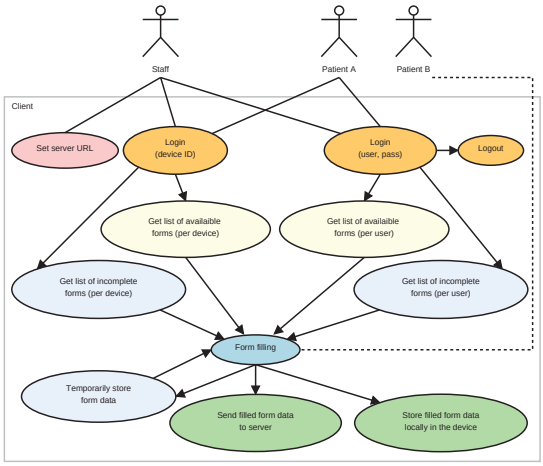



Figure 1: Use case of client.

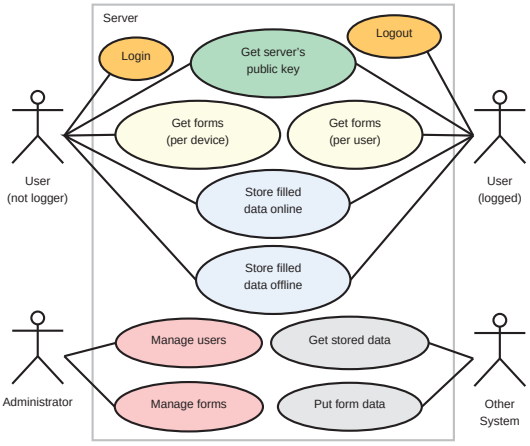



Figure 2: Use case of server.

requests list of available forms for the device directly when delegated mode is used.

List of available forms is user-dependent and device-dependent to allow for possible customisation. This is a quite simple solution for delivering different forms to the same patient on different clinics.

In the figure 1 you can see use case of delegated mode by *Patient B* actor. Shared device is prepared by Staff (actor) by using *Login (device ID)* method and the patient just fills out the form.

The *Temporarily store form data* use case can resolve any distraction (eg. the need to go to the toilet). Later, a patient can use *Get list of incomplete forms* use case (stored per device or per user).

Finally, user can finish completing form by two methods: (1) send data to the server or (2) store results data locally in the device. It is preferred to directly send data to the server. The second method is a backup solution, eg. when Wi-Fi signal is lost. Locally stored data is encrypted.

### 3.2.2 Server

REST API provided by server is the most important from the perspective of the client. There in the figure 2 you can see all of our actors. A client application represents both the top use cases, *User (logged)* and *User (not logged)* in the delegated mode. Both actors can get/download server's public key $S_{pub}$, list of available forms and upload/store filled form data. The form content can be uploaded online over Wi-Fi network or offline by connecting device by a cable.

Server must be properly secured on an operation system level. All retrieved data from the client applications will be stored there in an unencrypted form. For simplicity's sake, we do not describe details of server configuration, which may be implementation-dependent and include all activities related to the administration of users, forms and a description of their assignment to users or devices. In this context, *Administrator* and *Other System* actors use cases are out of scope of this paper.

## 3.3 Form Description

There are several ways to describe the form elements, their labels, groups, description and all the essentials including, eg. validation or enumeration values. There exist different forms description languages like XForms 1.1 (2009) and different libraries that helps building forms. In addition, there are number of differences across mobile platforms. We have made our own way of simplified definition and description of forms. The same form will need to be viewed and filled-in on different platforms. Our simplified definition and description of forms contains:

- form description – ID, name and description
- section – group of form elements has title,
- element – ID, label, data type, validation, default value, enumeration values,

All text labels may occur multiple times with different language attribute. Supported element data types are label, text, multi-line text, email, password, integer, number, currency, phone, boolean check or switch, date, time, selector box, selector list and URL. The form definition (description and content) have to be rendered by client application per mobile platform.

## 3.4 User Authentication Model Definition

The most important part of the shared device approach is communication protocol between client and

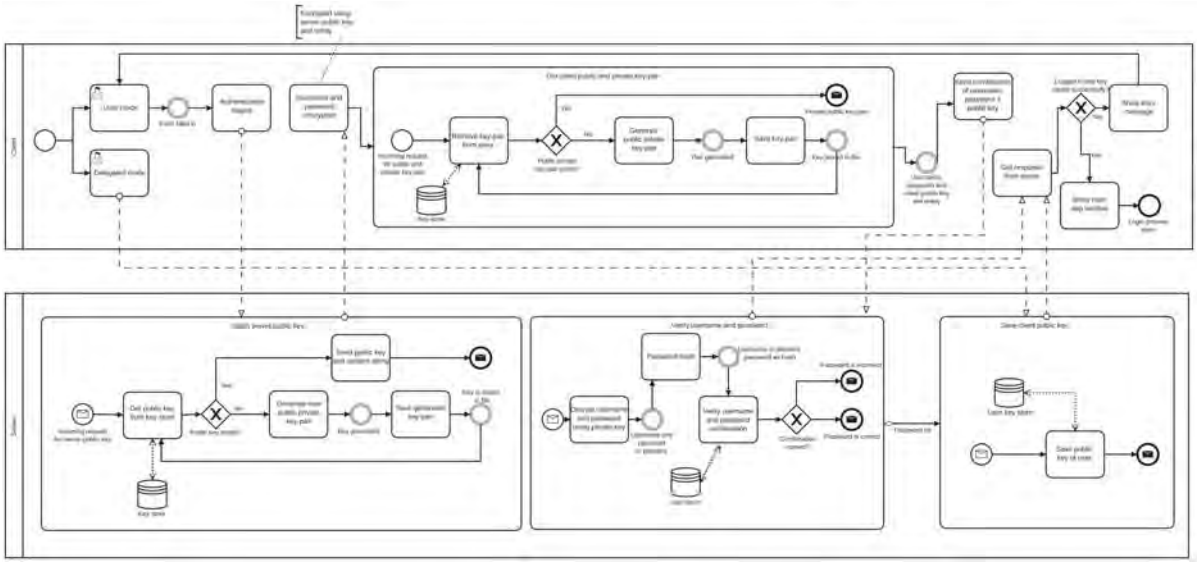


Figure 3: Diagram of the login process – BPMN v2.0 notation.

server. We need to ensure that data in the client application is protected against unauthorised access (read) or modification at the device or during transmission to the server. For this reason, we describe how user login process works on mobile devices. You can see the login model in the figure 3. Based on the model, encryption is enabled for stored data on client and communications between client and server.

As we mentioned in 3.2.1 and 3.2.2 sections, client application can login in two ways.

User does not need any username or password in the *delegated mode*, unique device ID is used instead. Private and public keys are required on the device. New key-pair ($C_{priv}$ and $C_{pub}$) is generated and stored in an internal memory of mobile device when no key-pair exist. Both, device ID $D$ and public key $C_{pub}$ are sent to the server. The client may see a list of currently-fill forms if the server successfully processes the request, so device identification and its public key are stored on a server. If communication with server fails, the client is informed of the error message and remains on the login screen.

A user login using username $U$ and password $P$ method need to have a public key of the server $S_{pub}$. Client application asks for server's public key $S_{pub}$. Server immediately sends an existing key or generates a new key pair by a configured algorithm. A cryptographic salt ($CS$) is sent together with the public key $S_{pub}$. Client stores both in an internal memory. Authentication then follows. Client application encrypts username $U$ and password $P$ by server's public key $S_{pub}$ and salt ($CS$) added. Server receives data and decrypts them by its own $S_{priv}$ key. Server checks that the received salt matches the salt sent to the device. Request is denied if the salt differs or the user does not exist on the server. A password hash stored on the server is compared with the one received from the client. The client's public key $C_{pub}$ is stored in case of the same password hashes. Client is notified that the user has logged in successfully and shows list of possible forms. Otherwise, client is informed of the error and remains on the login screen.

The salt ($CS$) is device-dependent. User's key-pair ($C_{priv}$, $C_{pub}$) depends on a username $U$ and device ID $D$. Each pair of user (as username $U$) and device ID $D$ has its own key-pair because of the same key reuse.

# 4 IMPLEMENTATION AND RESULTS

As the proof of concept we have created software corresponding to designed client-server architecture (3.2) and supporting the user authentication model (3.4). Server is a service-oriented Java web application with REST API and running over HTTPS protocol. The server needs Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files installed. We chose the Android platform, the world's most popular mobile operating system, for the client application. The implementation is based on Android version 7.0 with code name Nougat (API level 24). The five screenshots are shown on the figure 4.

The client application running on a shared device does not have any complicated setting, and provides only a few basic functions. The only configuration attribute is the server URL address. Mobile applica-

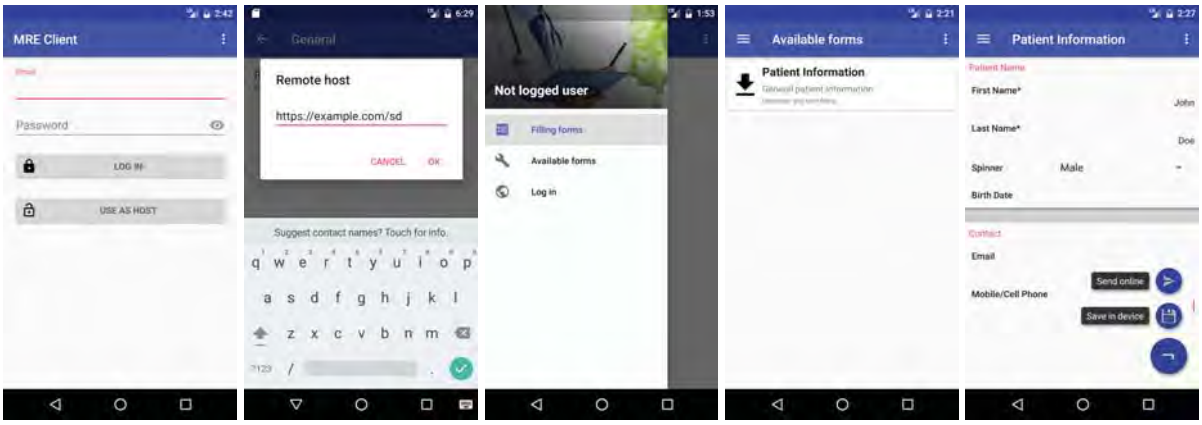



Figure 4: Mobile application (client) screenshots: (1) login screen, (2) configuration – set server URL address, (3) menu of actions in the delegated mode, user can fill-in per device forms or not finished ones, (4) list of available forms and (5) simple form/questionnaire generated and menu how to submit a result is shown.

tion has to download corresponding definitions of the forms that the user can fill in on the shared device. User cannot directly access data of any hospital information systems. Shared device is managed indirectly by the staff through the server. Staff member can assign a form type to the specific device or patient on the server. Client application downloads these form definitions and present an automatically generated forms to its user.

Retrieved, transmitted and permanently stored data on the shared device are encrypted. Only the patient himself can see his own data in a readable way. Form content is encrypted immediately, when form is validated and submitted. Online transmission is the preferred way of data transmission, but encrypted data can be stored locally when no network access is available. Data encryption is based on an asymmetric method using private and public key-pairs that are generated on the server or on the shared device.

## 5 DISCUSSION

We discuss the results and especially the security of the proposed shared device approach in this chapter. The most common and possible attacks include eavesdropping or modification of data and identity theft.

Client-server communication is designed to encrypt the whole communication, including form definitions, because it can contain personal information. The only unencrypted content is server's public key, client's public key and device ID. An attacker can get access to communication content on an unsecured network (eg. public Wi-Fi, HTTP protocol), but it is not possible to abuse this, because of the way asymmetric cryptography works (private key is needed for decryption). Eavesdropping is therefore not possible.

An another attack is by modifying request to obtain data access. It can happen on an unsecured network or when a shared device is infected by malware. Attacker can spoof the device ID via malware and request forms available for spoofed ID. Server sends forms belonging to someone else, however, they are not accessible, because the private key on attacker's device cannot de-crypt them. Attacker would need to possess private key of the device that is being imitated by spoofed device ID.

An entire request could be alternated by attacker, not only the device ID. Server process every request if its format is valid. Attack would be discovered, because server attempts to verify the electronic signature using the sender's public key. This verification would fail. Either electronic signature and request content or public and private key-pair will differ.

We also consider the case of identity theft. Attacker can monitor requests and reuse the message to repeat sending of that request with own fake public key. This situation is prevented by changing cryptographic salt, which is always used to encrypt user name and password. Accordingly, the resulting cipher names and passwords vary because of the different cryptographic salt.

Our, originally naive, solution of form definition/description proved to be very flexible. The form is dynamically generated on the mobile device (client), according to these definition. It is possible to dynamically create personalised form definition on the server which contain patient-oriented data (eg. different options per age groups or genders).

Mobile devices also have drawbacks. The client application needs to be online when loading the form definition. There are also issues with battery life, network availability, breakage or damage of borrowed device by patients. Network availability is important

when submitting form data to the server, but we can use offline data storage to temporarily alleviate connection problems.

At the testing phase, we identified an issue (occurred only once) with locally stored encrypted form content. There is a possibility of data loss, when user logs out of application and logs in again, because key-pair has to change and server does not store history of previously used keys. Our approach, originally devised to prevent attacks, may therefore lead to data loss on this occasion.

The mobile application has fully satisfactory and fluent response when connected via Wi-Fi. Only when using large forms (e.g. 1 000 items in select box) the response time worsened – increased time needed to download form definition and to render the form.

## 6 CONCLUSIONS

In this article we proposed an approach for secure health data acquisition using shared mobile devices. The data are confidential in general. The primary goal was gathering personal data and updating health status using form-oriented application. The security problems were discussed. We evaluated risk of data leak and designed data workflow for mobile devices that are shared across patients.

We designed a prototype and evaluated it as the real application on Android device. We identified that this workflow is properly secured. The discovered disadvantage is a possible data loss in special case when data were encrypted and key-pair changed before data was delivered to the server.

Obtaining data via electronic forms is easily customisable and extensible. There is a potential disadvantage of impersonal approach. Though, patients waiting for medical examination are usually feeling bored, and this interactive form might be therefore appreciated by them.

In the future we plan to expand the types of data that can be sent via secure forms and the presented approach. The logical extension is to support wearable electronics, sensors and other accessories connectable with a mobile device that will also acquire more data types in this way.

## ACKNOWLEDGEMENTS

This work was supported by the project PUNTIS (No. LO1506) of the Ministry of Education, Youth and Sports of the Czech Republic and by University of West Bohemia, project Data and Software Engineering for Advanced Applications (No. SGS-2016-018).

## REFERENCES

Baig, M. M., GholamHosseini, H., and Connolly, M. J. (2015). Mobile healthcare applications: system design review, critical issues and challenges. *Australasian Physical & Engineering Sciences in Medicine*, 38(1):23–38.

Bastawrous, A. and Armstrong, M. J. (2013). Mobile health use in low-and high-income countries: an overview of the peer-reviewed literature. *Journal of the royal society of medicine*, 106(4):130–142.

Bisio, I., Lavagetto, F., Marchese, M., and Sciarrone, A. (2015). A smartphone-centric platform for remote health monitoring of heart failure. *International Journal of Communication Systems*, 28(11):1753–1771.

Dehling, T., Gao, F., Schneider, S., and Sunyaev, A. (2015). Exploring the far side of mobile health: information security and privacy of mobile health apps on ios and android. *JMIR mHealth and uHealth*, 3(1):e8.

Hayes, D. F., Markus, H. S., Leslie, R. D., and Topol, E. J. (2014). Personalized medicine: risk prediction, targeted therapies and mobile health technology. *BMC medicine*, 12(1):37.

Knorr, K. and Aspinall, D. (2015). Security testing for android mhealth apps. In *Software Testing, Verification and Validation Workshops (ICSTW), 2015 IEEE Eighth International Conference on*, pages 1–8. IEEE.

Manjoo, F. (2015). A murky road ahead for android, despite market dominance. ISSN 0362-4331. (May 27, 2015) Retrieved Oct 17, 2016.

Melillo, P., Orrico, A., Scala, P., Crispino, F., and Pecchia, L. (2015). Cloud-based smart health monitoring system for automatic cardiovascular and fall risk assessment in hypertensive patients. *Journal of medical systems*, 39(10):1–7.

Stradolini, F., Riario, S., Boero, C., Baj-Rossi, C., Taurino, I., Surrel, G., De Micheli, G., and Carrara, S. (2015). Wireless monitoring of endogenous and exogenous biomolecules on an android interface. *IEEE Sensors Journal*, 16(9):3163–3170.

Tang, X., Hu, C., and Lin, W. (2015). Android bluetooth multi-source signal acquisition for multi-parameter health monitoring devices. In *Information and Automation, 2015 IEEE International Conference on*, pages 1790–1794. IEEE.

# A Business Model for Digital Healthcare Environments: An Organic Approach and a Use Case for Handling Cognitive Impairment

Andrea Pistorio[1], Paolo Locatelli[2], Federica Cirilli[2], Luca Gastaldi[1] and Simona Solvi[1]
[1]*Department of Management Engineering, Politecnico di Milano, Milan, Italy*
[2]*Fondazione Politecnico di Milano, Milan, Italy*
*{andrea.pistorio, luca.gastaldi, simona.solvi}@polimi.it, {paolo.locatelli, federica.cirilli}@fondazione.polimi.it*

Keywords: Healthcare Services, Business Models, Service Models, Digital Solutions, Cognitive Impairment.

Abstract: Ageing has significant impacts on the organization of healthcare systems and on social inclusion—especially for elderly people affected by Cognitive Impairment (CI). These people are significantly exposed to undeniable risks that can affect their health and wellbeing (falling, malnutrition, hygiene issues, etc.) – especially when living alone. This paper defines a Business Model (BM) allowing independent living for elderly people affected by CI. This BM include: (i) an up-to-date, modular, flexible and scalable organizational model describing the activities to be accomplished by regulators and service suppliers; and (ii) a digital platform based on innovative and easy-to-replicate information and communication technologies. The organic approach to the development of the BM is then focused in an Italian use case as a part of "DECI", a "Horizon 2020" project with four pilot projects in Israel, Italy, Spain and Sweden.

## 1 INTRODUCTION

The prevalence of Cognitive Impairment (CI) in elderly patients is one of the key issues in Western Countries since it involves loss of memory, cognitive slow down, aphasia, apraxia, sensorial and movement deficit, personality and mood disorders (Alexander et al. 2015; Rizzi et al. 2014). The elderly people with cognitive limits - especially those living in solitude - are significantly exposed to undeniable risks for their own safety: falling, malnutrition or unhealthy nutrition, hygiene issues due to lack of mobility, isolation and depression (Alzheimer's Association 2016).

The needs of these patients range from daily aid to dedicated medical assistance (Alzheimer's Association 2016; Miranda-Castillo et al. 2013). Though digital technologies can be one key lever to answer these needs, most of current solutions are immature for mass implementation.

In this context, the goal of this paper is to define a Business Model (BM) for supplying assistive services to elderly people affected by CI. This BM will include: (i) an up-to-date, modular, flexible and scalable organisational model describing the roles and the activities to be accomplished by policy makers and service suppliers; and (ii) a digital platform based on innovative and easy-to-replicate Information and Communication Technologies (ICT) streamlining and simplifying the flow of information and the communication among the various key stakeholders.

## 2 METHODS

This paper is based on DECI – Digital Environments for Cognitive Inclusion – a project funded by the European Commission under the Horizon 2020 program (grant agreement No 643588) that is aimed at improving a healthy lifestyle for elderly people affected by CI, passing through a system monitoring vital signs, treating and managing diseases (Locatelli et al. 2015).

DECI Consortium is led by Fondazione Politecnico di Milano (Italy) and involves partners from five different countries (Italy, Sweden, Spain, Israel and The Netherlands). The consortium merges academic and research competences, care and social-care providers, healthcare authorities, ICT industry and their broad network of stakeholders. Four pilots, involving 100 to 250 patients each, will allow assessing the feasibility, the effectiveness and the potential economic benefits of the proposed measures



Pistorio, A., Locatelli, P., Cirilli, F., Gastaldi, L. and Solvi, S.
A Business Model for Digital Healthcare Environments: An Organic Approach and a Use Case for Handling Cognitive Impairment.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 340-347
ISBN: 978-989-758-213-4

within specific local healthcare systems and real-life environments in Israel, Italy, Spain and Sweden.

The overall research process followed to develop the BM as well as the related processes encompassed three subsequent phases that are depicted in Figure 1 and described in the following paragraphs.

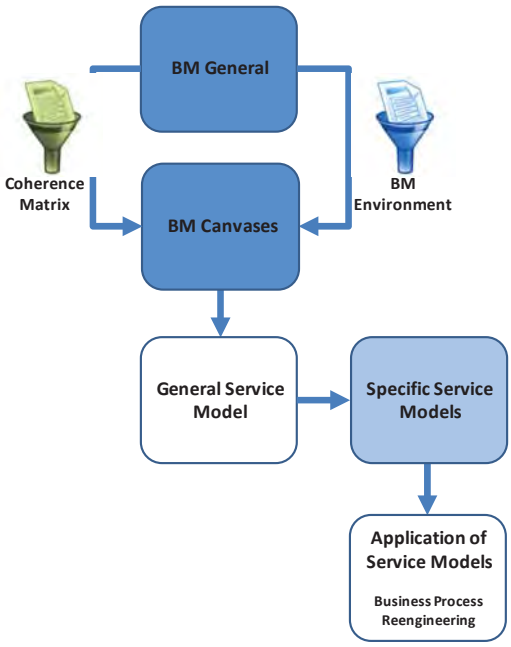



Figure 1: Research process.

## 2.1 First Phase

A literature review has been conducted to identify the most relevant articles dealing with the state of the art, good practices and trends related to the digital solutions for assisting elderly people with CI. The analysis of these articles allowed producing a general BM to focus the main variables that increase the likelihood of providing effective digital services.

Web of Science™ and Google Scholar™ have been used to research the articles. The literature has been analysed according to two complementary frameworks.

Initially, we relied on the STOF framework (Bouwman et al. 2008), which provides an overall picture of a BM from four interrelated perspectives—service, technology, organization and finance.

Next, we progressively deepened each insight through the Business Model Canvas (BMC) framework (Osterwalder and Pigneur, 2010), which allowed focusing on key design elements.

## 2.2 Second Phase

The second phase of the research was based on two steps that allowed to obtain, starting from the general BM developed through the first phase, the $BM_{Canvases}$ and, then, the Service Models (SMs) depicted in Figure 1. $BM_{Canvases}$ represent the Country-specific subsets of the elements included within the general BM. SMs contain configurations mainly focused on value propositions in all the considered Countries–with an emphasis on common aspects to be exploited.

The first step is made through two tools: (i) the Business Model Environment (BME) (Osterwalder & Pigneur 2010); (ii) a Coherence Matrix (CM). On the one hand, these tools have been used to reduce the multitude of elements proposed in the general BM. On the other hands, they allow prioritizing the required activities for the implementation of the pilot projects—distinguishing among relevant and irrelevant elements.

The BME organizes and describes the key contextual variables useful to synthetically characterize each Country involved in the research, according to the four macro-areas: (i) key trends; (ii) macro-economic forces; (iii) industry forces; (iv) market forces.

The CM has been developed by the research group and is characterized by the following five dimensions: (i) country; (ii) customer segment; (iii) functionalities; (iv) actions to address the needs of the patient and of the overall system; (v) relevance of the specific need. The CM describes the underlying interrelations among these dimensions and allows discerning between: (a) coherent and incoherent pairs functionality-action to address the need (i.e. need of the patient or of the overall system); (b) relevant and not relevant pairs functionality-action to address the need; (c) Country-specific and common among Country pairs functionality-action.

The second step allows switching from $BM_{Canvases}$ to SMs. It reduces the complexity of the vast number of elements of the $BM_{Canvases}$ focusing on the key building blocks of value proposition and customer segments. The path between $BM_{Canvases}$ and SMs requires to select the common value propositions among all considered Countries (general SM), and to consider also the peculiar aspects of the contexts (Country-specific SMs). All the functionalities, related to the pairs functionality-action have been clustered according to two levels: the main target of the solutions (i.e. patient, care provider, care pathway plan) and technological functionalities.

## 2.3 Third Phase

SMs are implemented within the third phase of the research project following a Business Process Reengineering (Hammer, 1990; Champy, 2002)

approach that includes the following phases: (1) design of the care process model, which firstly defines the macro-phases of a general care process for CI patients. (2) analysis of the AS-IS process models in each pilot site; (3) design of the TO-BE process models in each pilot site, underlining the difference between AS-IS and TO-BE process models. Considering also digital solutions (Locatelli et al. 2014).

# 3 RESULTS

This paragraph describes the results of the research. Firstly, the outcomes of the literature analysis and the evidences related to the STOF framework are grouped into the 9 building blocks of the BMC framework. Secondly, the results related to the application of the BME and CM are presented.

## 3.1 General Business Model

The general BM is characterized by the following 9 building blocks. For each of them we report the main results of the analysis of the literature accomplished.

1. Customer segments are convenient set of clients with common needs, behaviours or attributes. Following Petersen (2004), the main customer segments to be considered in CI are the following: (i) patients with dementia; (ii) patients with amnestic Mild CI (a-MCI); (iii) patients with non-amnestic Mild CI (na-MCI).

2. Value proposition is the reason why a customer chooses one product and/or service. In CI domain, a tool adopted to assess the various needs is the Camberwell Assessment of Need for the Elderly model (CANE) (Reynolds et al. 2000). The application of the model highlights that patients with dementia have more care needs than MCI patients. Furthermore, there are several solutions for the various actors involved into the care process.

3. Channels are the interfaces used to interact with the customers in order to deliver a value proposition. Literature shows that the main channels by which care services are delivered to elderly people with CI are the following: (i) caregiver; (ii) general physician; (iii) healthcare specialist; (iv) healthcare structure.

4. Customer relationships regards how relating with customer segments. Patients suffering from CI tend to communicate with their caregivers or social workers. These two support patients in adopting and using any digital solution for providing enhanced services. From the viewpoint of the actors involved in the care process, many companies maintain online most of their relationships with caregivers.

5. Revenue streams deal with the cash generated from every customer segment. Following Stroetmann et al. (2003), the main revenues for a digital solution in CI domain are the following: (i) service paid by insurance funds; (ii) service paid by governments; (iii) service paid by the patient or by his relatives/caregivers through out-of-pocket expenses. Three groups of people or organizations pay for the digital solution: (a) Business to Consumer (B2C): the service is sold directly to the patient or his caregiver. (b) Business to Public (B2P): the service is sold to public entities i.e. local authorities, NHSs and housing associations; (c) Business to Business (B2B): the service is sold to private companies and, in some countries, most of them are private medical insurance companies. Furthermore, there could be the following types of transactions between the provider of digital solutions and the healthcare organization: (1) healthcare organization purchases the entire system (one-time capital investment); (2) healthcare organization pays a license fee for each patient connected to the system.

6. Key resources are the essential assets necessary to create and offer the value proposition. From the perspective of the patient suffering from CI, human resources that will be crucial in the establishment of the BM are health professional and caregivers (or the social worker if the patient has not relatives that support her) (Robert et al. 2013). The human resource that is vital in the establishment of the BM is the specialist of digital services through which assisting people with CI (Kapadia et al. 2015). From the standpoint of physical assets required to provide the service, technologies play a key role both for the patient and for the overall system (Kerssens et al. 2015; Robert et al. 2013; Bharucha et al. 2009). Other key resources are patents, licenses and copyrights.

7. Key activities are actions that have to be performed in order to create and offer a value proposition. From the patient standpoint, the key activities are related to her involvement: (i) engagement of the patient; (ii) maintenance of the relationship with the patient. From the stand point of the actors involved in the care process, key activities are related to the creation and sustainment of the relationships among these key actors: (a) create the connection between healthcare specialist and caregiver/social worker; (b) maintenance of the relationship between healthcare specialist and caregivers/social workers. Following Ógáin and Mountain (2015), these activities can be supported through the contributions of governmental actors in terms of: (c) national/regional awareness campaigns; (d) financial incentives to healthcare organizations; (e) financial incentives to healthcare specialist for the

diagnosis of MCI and dementia; (f) financial incentives to healthcare specialists and healthcare organizations for the adoption of digital solutions to treat MCI and dementia. Following Robert et al. (2013), the key activities in the BM related to the technology adopted are the following: (1) production of the equipment and sensors; (2) selection of the most appropriate technology/sensor; (3) installation of technologies in patients' home; (4) calibration of the sensors.

8. Key partnerships refers to the network of suppliers useful to improve the BM. Literature on CI suggests that some key partnership for the BM could be: (a) government; (b) research center/university; (c) local regional community (Kapadia et al. 2015); (d) private organizations (König et al. 2015); (e) networks between the providers of digital solutions and healthcare organizations.

9. Cost structure deals with relevant costs characterizing a feasible BM. Literature suggests that the main costs for a digital solution in the CI domain are the following (Kapadia et al. 2015): (i) training costs; (ii) personnel costs; (iii) installation and maintenance costs; (iv) purchasing and manufacturing costs; (v) customer service costs.

## 3.2 Business Model Canvases

The next steps of the research allowed to move from the general BM to $BM_{Canvases}$ through the BME and CM. BME allows to describe the context in which BM will be implemented, considering the following 4 macro areas:

*1. Key trends*: (a) national socio-healthcare system overview; (b) general government expenditure on health; (c) private expenditure on health; (d) "Out-of-pocket" expenditure; (e) *per-capita* total expenditure on health (WHO 2015);

*2. Macro-economic forces*: (a) payment mechanisms; (b) policies regarding the sources of revenue and financial flow;

*3. Industry forces*: (a) balance between public and private healthcare; (b) centralization vs. decentralization (c) main actors of the NHS; (d) number of health and social care integration hospitals per 100.000 population (e) number of psychiatric beds per million population (WHO 2015);

*4. Market forces*: (a) percentage of population aged > 60 years compared to the overall population; (b) life expectancy at age 60 (WHO 2015); (c) estimated prevalence of dementia per 1.000 population (OECD 2015; Prince et al. 2013); (d) percentage of population living in urban areas (WHO 2015); (e) population ICT readiness; (f) tendency to informal care (Lupianez et al. 2013).

In addition to BME, the implementation of the CM returns different combinations of coherent, relevant and Country-specific technological functionalities and actions to address patients' needs. The tool allows highlighting the value propositions to focus the attention on while designing $BM_{Canvases}$. The synthesis of the findings related to BME and CM allows defining the $BM_{Canvases}$. The latter, though grounded on already established frameworks (e.g. Business Model Canvas, BME, etc.), take advantage of the flexibility of the tools adopted to build them, thus making them adaptable to the various possible contexts (i.e. specificities of the various countries) or other patients' clusters different from those adopted in this research.

## 3.3 Service Models

In order to design a general SM able to encase a value proposition common to all the four countries involved in DECI, the needs marked by clinicians as extremely relevant have been highlighted and clustered as follow:

*$N_1$. Diagnosis and assessment*: (i) overall medical condition; (ii) behavior and mood; (iii) assess the risk of malnutrition; (iv) Activities of Daily Living (ADL); (v) risk for falls.

*$N_2$. Patient psychological needs*: (i) cognitive stimulation; (ii) online cognitive training.

*$N_3$. Clinical team needs*: (i) coordination of care; (ii) clinical team information sharing; (iii) improve diagnosis method; (iv) advising on deciding course of action; (v) better access to and relevance of non-pharmacological therapies; (vi) standardized care pathway.

*$N_4$. Follow-up*: (i) monitoring overall condition; (ii) measurement of adherence and compliance of patients to treatment; (iii) assess timely changes evolving needs for social care support.

Starting from these clusters, it is possible to cluster also the technological functionalities that allow meeting these needs, and which are common to all Countries. Regarding the patient layer, the following 5 clusters have been identified (technological functionality in brackets):

*$TFP_1$. Patient's status (monitoring)*: (i) automatic remote-based measurement of patient's blood pressure; (ii) automatic remote-based measurement of patient's $O_2$ saturation; (iii) automatic provision and submitting of questionnaires (of various nature, including patient's health status and to support change detection) to care-involved subjects; (iv) gathering on non-structured information on patient's health status from informal caregivers or social caretakers; (v) evaluation and monitoring of cognitive

skills and monitoring of decay curves and other trends.

*$TFP_2$. Patient's status (alert)*: automatic provision of feedbacks and alerts on patients' progresses or deterioration.

*$TFP_3$. Patient's status (communication-cognitive stimulation)*: (i) cognitive games/exercises to stimulate patients to preserve cognitive/ executive functions; (ii) tele-consultation (tele-presence) functionalities allowing patients and professionals to communicate to each other visually.

*$TFP_4$. Patient's Activities (alert)*: (i) automatic provision of remote real-time feedback on patients' activities, including non-pre-scheduled activities; (ii) automatic provision of remote real-time patient-tailored motivational messages based on patients' activities, including non-pre-scheduled activities; (iii) automatic provision of remote feedback on patients' activities, building on long-time data analysis on patients' status.

*$TFP_5$. Patient's Activities (monitoring)*: activity monitoring through accelerometer for elderly monitoring (also outdoor, with batch data download once reconnected to base station): stand / sit / walk / steps + intensity of activity; GPS-based patient monitoring and structured health-based data gathering for outdoor step counting or activity monitoring (including detection of falls); registering of pre-scheduled activities performed by the patient (who is monitored real-time by sensors when performing the activity).

*$TFP_6$. Patient's status (storage and sharing information)*: (i) activity monitoring through accelerometer for elderly monitoring (also outdoor, with batch data download once reconnected to base station): stand / sit / walk / steps + intensity of activity; (ii) GPS-based patient monitoring and structured health-based data gathering for outdoor step counting or activity monitoring (including detection of falls); (iii) Registering of pre-scheduled activities performed by the patient (patient is monitored real-time by sensors when performing the activity).

Regarding the actors involved in the care process, the following 4 clusters have been identified:

*$TFS_1$. Care providers (communication)*: informal communication (messaging) among various actors (e.g.: family members and doctors);

*$TFS_2$. Care providers (teamwork)*: enablement of multidisciplinary teamwork across care providers, doctors and informal caregivers (or some of these);

*$TFS_3$. Care pathway/treatment plan (monitoring)*: coherence check between clinical guidelines/protocols and data gathered as part of care activities;

*$TFS_4$. Care pathway/treatment plan (sharing information)*: sharing of a treatment plan among caregivers, doctors and family members (or some of these).

Combining the clusters of common relevant needs with the clusters of common relevant functionalities, it is possible to point out the match between the two as number of notable crossings, in order to highlight packages of SM common to all four countries.

## 3.4 Italian Service Model

An example of the intersection between the two clusters and of the specificities of a Country is provided regarding the Italian context. Starting from the value proposition common to all the 4 countries, the general SM is enriched with further specificities related to the Italian Customer Segments.

Given Italian specific needs, further specific elements of the Italian SM are available: (i) immobility detection for elderly monitoring at home (indoor) for patient physical needs; (ii) fall detection for elderly monitoring at home (indoor) for patient physical needs; (iii) trend analyses performed on data gathered from various patients' monitoring activities for diagnosis and assessment & for caregivers needs; (iv) registering of pre-scheduled activities performed by the patient (patient is formally required to provide a yes/no answer) for patient environmental needs & for patient physical needs; (v) automatic reminder to patients for the performing of a scheduled activity for patient environmental needs & for patient physical needs; (vi) availability of personalized and adaptable remote-based training programs automatically tailored on individual patient's characteristics for patient physical needs; (vii) drug management for patient physical needs.

The functionalities clusters are enriched, for the Italian case, as follow:

*Patient's status (monitoring)*: (i) Immobility detection for elderly monitoring at home (indoor); (ii) Fall detection for elderly monitoring at home (indoor);

*Patient's activities (monitoring)*: (i) Trend analyses performed on data gathered from various patients' monitoring activities; (ii) Registering of pre-scheduled activities performed by the patient (patient is formally required to provide a yes/no answer);

*Patient's Activities (alert)*: (i) Provision of automatic reminder to patients for the performing of a scheduled activity; (ii) Availability of personalized and adaptable remote-based training programs

automatically tailored on individual patient's characteristics;

*Care pathway/Treatment plan (monitoring)*: drug management.

These packages complete the general SM, shaping the Italian specific one.

The other building blocks are grouped as follows: upstream building blocks (Key partners, Key activities and Key resources), downstream building blocks (Customer Relationships, Customer Segments and Channels), and cost and revenues (Cost Structure and Revenue Streams). The main impacts are:

*Upstream building blocks*: (i) Italian clinicians consider the diagnosis and assessment crucial (GPs should be involved for an early detection); (ii) given clinical team needs, it is important to provide the technological functionalities to the local health authorities to the municipalities and to the case service providers; (iii) the monitoring of the patient's status and activities meet all the 7 clusters of needs; (iv) to provide automatic reminders to patients for the performing of a scheduled activity is important to personalize the service; (v) from the point of view of the actors involved in the care process, is crucial to enable multidisciplinary teamwork across local health authorities, municipalities, case service providers, informal caregivers and voluntary associations.

*Downstream building blocks*: (1) from clinical standpoint, patient physical needs, caregiver needs and the ones regarding the patient environment are considered relevant; (2) the caregiver has a central role and must be engaged and informed about the evolution of symptoms and disease; (3) personal assistance based on human interaction should be preferred to meet the patient psychological needs.

*Cost and revenues*: (a) Revenue Streams should be a B2P and B2C mix: (a.i) B2P: some functionalities, for example the ones for the storage and sharing to external EMR or other care management tools, should be funded from general taxation; (a.ii) B2C: some services should be sold directly to the patient or his caregiver, for example the ones for monitoring the patient's status or for the cognitive stimulation; (b) Considering the central role of the caregivers, their training costs must be considered also to meet their needs.

## 3.5 Application of the Service Models

This paragraph is focused on the application of the SMs in the Italian pilot site, through three steps: (1) the design of a general Care Process Model for patients affected by CI; (2) the analysis of the AS-IS Process Models in each pilot site, and finally (3) the design of the TO-BE Process Models in each pilot site, starting from the Service Model resulted from literature analysis.

### 3.5.1 Design of the Care Process Model

The analysis highlighted four common phases to every care process for people with CI (Figure 2).

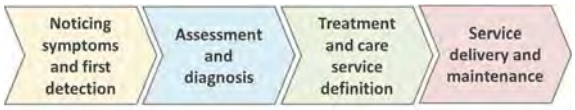



Figure 2: Care Process Model for patients affected by CI.

*A. Noticing Symptoms and First Detection:* this phase includes the access point of the patient with CI in the care process and it considers the first identification of the patient with suspected CI.

*B. Assessment and Diagnosis:* clinical activities aimed at the assessment and diagnosis of the CI in the patient identified in the previous phase. The phase includes: a first basic assessment of CI that can be owned by the GP, the socio-health care provider or external specialized physicians; and then, a comprehensive assessment that is usually owned the socio-health care provider, although some of the requested exams are provided by external physicians.

*C. Treatment and Care Service Definition:* analysis of the patient needs, both clinical (emerged from the clinical assessment delivered in the previous phase) and social needs (usually analysed with the patient and the family with a social assistant). When needs are defined, the care service is designed.

*D. Service Delivery and Maintenance:* delivery of the care service designed and continuous monitoring of the patient's status.

### 3.5.2 Analysis of the as-IS Process Models

We will describe in detail the Italian process model, in Palazzolo Institute of Fondazione Don Carlo Gnocchi Onlus in Milano, following the four macro-phases.

*A. Noticing Symptoms and First Detection:* the process starts though a first meeting between patients and physicians, like a check-up visit, or after an acute episode in a long term care facility. The visit can be owned by the GPs or specialized physicians (both within Palazzolo Institute as well as external). Once patients with certain kind of characteristics have been identified during a physician's visit, their path continues in the care process.

*B. Assessment and Diagnosis:* in case there are symptoms of CI, a basic assessment can be performed. Note that usually the GP points out the patient to specialists (at Palazzolo Institute or externally). This initial assessment consists of a basic examination of the patient with suspected CI, aimed at understanding more regarding patients' health conditions and social situation. In case the basic assessment strengthens the initial suspicion of CI, a comprehensive assessment follows. At Palazzolo Institute, for a comprehensive assessment, patients can be referred to a specialized geriatrics unit, which starts the assessment protocol aimed at reaching the diagnosis.

*C. Treatment and Care Service Definition:* in Italy, it is common that patients' relatives or caregivers take the responsibility for organizing the care pathway of the patient. The care pathway can be managed through the activation of home-based and facility-based services, whose activation is discussed together with the family.

*D. Service Delivery and Maintenance:* Palazzolo Institute delivers an integrated socio-healthcare service to patients affected by CI. During the service delivery the patient is continuously monitored by the professionals of the Institute.

### 3.5.3 Design of the to-BE Process Models

As regards the TO-BE process models, we will describe in detail the differences between the Italian AS-IS and the TO-BE process model. The fundamental differences from AS-IS and the TO-BE process models are linked to some new technologies, which ensure the sharing of information among the various actors enabling independent living.

*A. Noticing Symptoms and First Detection:* during the visit, physicians visualise the patient's medical records on the DECI platform connected with the databases of local hospitals and GP.

*B. Assessment and Diagnosis:* data collected by the specialized geriatrics unit are entered in a tablet and immediately shared (through a cloud service) with the other stakeholders of the care network.

*C. Treatment and Care Service Definition:* A first needs analysis can be performed by the patient/caregiver filling in on-line questionnaire on the DECI platform. Then a visit could confirm the results of the questionnaire.

*D. Service Delivery and Maintenance:* A wearable sensor could be used to detect abnormalities in the level of activity and send alert messages to family members or assistance operator. The GP and specialists can share the information collected through the DECI platform and plan together a revision of the medication and treatment plan, and schedule visits. The program can alternate sessions in physical presence of the physiotherapist with tele rehabilitation sessions. Some cognitive stimulation and rehabilitation exercises could be done also at home through the online platform. The platform can also be used by the case manager to communicate with the patient or the caregiver.

## 4 CONCLUSIONS

The approach adopted to design the BM is relevant because: (i) it adopts already established frameworks (e.g. Business Model Canvas); (ii) it highlights common traits and differences among the Countries within the DECI project; (iii) it allows distinguishing firstly between coherent and not coherent elements, then between high-priority and low-priority ones; (iv) it provides information about the impact of the various digital solutions; (v) it is applicable in other contexts (i.e. Countries) therefore it can overcome the boundaries of the DECI project thanks to its scalability; (vi) it supports decision-making processes also after the pilot phases because, once the first most relevant needs are addressed, it will be possible to proceed with the actions to address the needs with a lower relevance.

Furthermore, the approach was validated and refined also through a Scientific Advisory Board with the involvement of external stakeholders. Finally, the approach is aimed to support the definition and implementation of a comprehensive and multifaceted BM in a complex and continuously evolving context. The next steps of the DECI project are: (1) the implementation of the Pilot site (Italy, Sweden, Israel and Spain) adopting the proposed SMs and (2) the evaluation of key performance indicators considering different perspectives (e.g.: economics, social, etc.) in order to highlight the potential benefits of the designed approach.

## ACKNOWLEDGEMENTS

The authors would like to thanks the whole Consortium of the DECI project: Fondazione Politecnico di Milano, Consoft Sistemi SpA, Fondazione Don Carlo Gnocchi Onlus (Italy), Centre for Healthcare Improvement – Chalmers University of Technology, Västra Götalandsregionen (Sweden), Hospital Universitario de Getafe - Servicio de

Geriatría (Spain), Maccabi Healthcare Services (Israel), Roessingh Research and Development (The Netherlands).

## REFERENCES


Alexander, M., Perera, G., Ford, L, Arrighi, H. M., Foskett, N., Debove, C., Novak, G., and Gordon, M. F., 2015. Age-Stratified Prevalence of Mild Cognitive Impairment and Dementia in European Populations: A Systematic Review. *Journal of Alzheimer's Disease*, 48, pp. 355–359.

Alzheimer's association, 2016. 2016 *Alzheimer's disease facts and figures*. Alzheimer's & dementia, volume 12, issue 4, pp. 459–509.

Bharucha, A.J. et al., 2009. Intelligent Assistive Technology Applications to Dementia Care: Current Capabilities, Limitations, and Future Challenges. *The American Journal of Geriatric Psychiatry*, 17(2), pp.88–104.

Bouwman, H., de Vos, H., Haaker, T., 2008. *Mobile Service Innovation and Business Mod*els, Springer-Verlag Berlin Heidelberg.

Champy, J., 2002. *X-Engineering the Corporation: Reinventing Your Business in the Digital Age*, Warner Business Books. New York.

Hammer, M., 1990. Re-engineering work: don't automate, obliterate. *Harvard Business Review*, 68(4):104-112.

Kapadia, V. et al., 2015. Emerging ICT implementation issues in aged care. *International Journal of Medical Informatics*, pp.1–9.

Kerssens, C. et al., 2015. Personalized Technology to Support Older Adults With and Without Cognitive Impairment Living at Home. *American Journal of Alzheimer's Disease and Other Dementias*, 30(1), pp.85–97.

König, A. et al., 2015. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(1), pp.112–124.

Locatelli, P., Restifo, N., Cirilli, F., Moser, R., 2015. DECI: Digital Environment for Cognitive Inclusion - Improving Elderly Citizens' Quality of Life Through an ICT-enabled Patient-Centric Home Environment. *Proceedings of the International Conferences on e-Health 2015- Part of the Multi Conference on Computer Science and Information Systems 2015*, IADIS.

Lupianez, F., Maghiros, I. & Abadie, F., 2013. *Strategic Intelligence Monitor on Personal Health Systems, Phase 2 Citizens and ICT for Health in 14 European Countries: Results from an Online Panel* F. Abadie et al., eds.,

Miranda-Castillo, C., Woods, B. & Orrell, M., 2013. The needs of people with dementia living at home from user, caregiver and professional perspectives: a cross-sectional survey. *BMC Health Services Research*, 13, pp. 43.

OECD, 2015. *Addressing Dementia. In Health at a Glance 2015*. OECD Health Policy Studies. OECD.

Osterwalder, A., Pigneur, Y., 2010. *Business Model Generation* (pp. 1–288). Hoboken, New Jersey: John Wiley & Sons, Inc.

Ógáin, E.N., Mountain, K., 2015. *Remember me*. pp.1–33.

Petersen, R.C., 2004. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3), pp.183–194.

Prince, M. et al., 2013. The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer's & Dementia*, 9(1), pp.63–75.e2.

Reynolds, T. et al., 2000. Camberwell Assessment of Need for the Elderly (CANE) Development, validity and reliability. *British Journal of Psychiatry*, 176(5), pp.444–452.

Rizzi, L., Rosset, I., Roriz-Cruz, M., 2014. Global epidemiology of dementia: Alzheimer's and vascular types. *Biomed Res Int. 2014*; 2014:908915.

Robert, P.H. et al., 2013. Recommendations for ICT use in Alzheimer's disease assessment: Monaco CTAD Expert Meeting. *The Journal of Nutrition, Health Aging*, 17(8), pp.653–660.

Stroetmann, K.A., Stroetmann, V.N. & Westerteicher, C., 2003. Implementation of TeleCare services: benefit assessment andorganisational models. *Studies in Health Technology and Informatics*, 97, pp.131–141.

WHO, 2015. *World health statistics 2015*, WHO Press, Geneva.

# Voiceguard: Using mHealth Technology for Vocal Health Promotion

Christina Cesar Praça Brasil[1], José Eurico de Vasconcelos Filho[2], Daniele de Araújo Oliveira Carlos[1], Thiago de Oliveira Magalhães[2] and Raimunda Magalhães da Silva[1]
[1]*Pós-Graduação em Saúde Coletiva, Universidade de Fortaleza, Fortaleza-CE, Brazil*
[2]*Laboratório de Inovação Tecnológica - NATI, Universidade de Fortaleza, Fortaleza-CE, Brazil*
*danii_joly@yahoo.com.br, {thiago, euricovasconcelos, rmsilva, cpraca}@unifor.br*

Keywords: Voice, Mobile Technology, mHealth, Evaluation Studies, Content Analysis.

Abstract: This study aims to describe the development of an app for mobile devices to assist voice professionals in the management of vocal health. The research was held in two phases, from November 2014 to December 2015: 1) literature review and app stores search and 2) laboratory design, development and usability test. The multimedia feature was chosen for the app design and development, since it favours a motivating and dynamic environment. Teachers, when participating in the usability test, handled the tool for a few hours and issued their opinions. Data were analysed based on content analysis in the thematic mode. The results show the feasibility of the tool development to support and assist professionals in the care of their voice and open new perspectives to show that, in health promotion, technology can create new alternatives for health education and care, empowering the users.

## 1 INTRODUCTION

The use of m-Health technologies to promote vocal health is an issue that deserves attention in the current scenario, in view of the potential and vascularity of these technologies and the importance of voice for humans. The voice is a hallmark of the human being, is widely used in interpersonal, social and professional relationships. Characteristics such as sex, age, emotional state and personality are expressed through the vocal quality (Behlau et al., 2009). Moreover, the voice enriches the transmission of the word or message, both socially and professionally (Fabricio et al., 2010).

Among the professionals who use their voices as a working tool, there are the teachers, which intensively require this function and undergo several changes or even problems, due to lack of care and protective measures (Brazil, 2015). Vocal disorders in teachers represent a serious health problem, since it undermines the quality of life, work performance and compromises the quality of education (Fabricio et al., 2010).

Researches that investigated the vocal changes in teachers from different levels of education concluded that such prevalence is high, ranging from 21% to 80% (Roy et al., 2004; Strong et al., 2007, Brazil, 2015). Other studies that have captured information from teachers, showed a percentage of complaints related to voice that ranged from 54% to 79% (Gonçalves; Penteado; Silverio, 2005). These data reveal that there is a high incidence of voice problems among teachers, demonstrating that this profession can be demanding more attention from public policies to vocal health.

In this context, the technology brings new possibilities and resources that can make life easier for people and professionals to develop the health care, an example of this, are the mHealth technologies (Sarno et al., 2014).

In recent years, mobile health (mHealth), a branch of electronic health (eHealth), defined as "the use of computer technology and mobile communications in health care and public health" has been constantly expanding (Free et al., 2010). Thus, applications for mobile devices in health can cater to a heterogeneous audience (Free et al., 2010) and a wide variety of purposes (Riley et al., 2011).

The biggest advantages of using mobile applications in health consist in the fact that the devices are personal; that, nowadays, they have large processing capacity, several sensors (camera, microphone, accelerometer, gyroscope), internet connection; and are portable (Whittaker, 2012). Therefore, they can be used everywhere, even in everyday life and during hospitalization or rehabilitation. They can also meet the health care



Brasil, C., Filho, J., Carlos, D., Magalhães, T. and Silva, R.
Voiceguard: Using mHealth Technology for Vocal Health Promotion.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 348-353
ISBN: 978-989-758-213-4

providers during routine visits or emergency procedures. In addition, the technology has been widely used by health professionals in health promotion strategies.

The high frequency of voice disorders in teachers and the lack of public policies for the vocal health of these professionals (Brazil, 2015), show that the creation of technologies that empower the teacher to take care of the voice can contribute to health promotion, improving quality of life and occupational conditions of this population.

Given the above, it was thought in a technological feature that allows the teachers and professionals who intensively use the voice, to monitor vocal health (even in real time), creating an interface between the users and the environment. This way, the aim of this study was to develop a technology for mobile devices as a tool for promoting the teacher's vocal health.

## 2 METODOLOGY

A study of exploratory and experimental type was conducted, from November 2014 to December 2015 in the Application Center for Information Technology of University of Fortaleza - UNIFOR in the state of Ceará - Brazil.

The project, is still in development with two concluded phases – 1) literature review and app stores search and 2) laboratory design, development and usability test.

In the first phase, after listening to the difficulties faced by the voice professionals (future users of the app) due to the needs to take care of their voices, it was carried out a literature review on the subject and context proposed, allowing the research team a better understanding of concepts, methodologies and tools of the involved areas. The review included a survey of the existing applications dedicated to vocal health. The team investigated the existing vocal health technologies through a survey conducted in the app stores on the web, on the Android (Google Play) and iOS platforms, using the following terms: vocal health, voice management and voice (in Portuguese and English). As a result, we identified, among others, four applications directed to vocal health, which approached the purpose of this study; besides VoxMetria tool, that is a specific software for voice analysis and voice quality.

The second phase, held in laboratory, contemplated the development of the application, with the participation of a multidisciplinary team of speech pathologists, computer engineers and graphic designers.

For the development of the tool, it was adopted a methodology in Human Computer Interaction area, the participatory interaction design (Preece et al., 2013), which was chosen because it focuses on the user´s needs and in its continued participation during the process. The interaction design is divided into four activities (Figure 1): identify user´s needs and establish requirements (for the system), conceive the solution design and (when necessary) the redesign, build an interactive version (working prototype) and evaluate the version produced with users.

Once established the requirements and based on them, the activity of design began. At this moment, drawings of system interfaces were built, in order to achieve the requirements, as well as usability and aesthetics. These drawings were understood by all the members of the staff, allowing reviews with the participation of potential users of the proposal.

Finalized the revisions (design and redesign) of the artefact, it was started the construction phase of the interactive version. In this phase, the programming/coding of technological artefact was made. For this activity, the team used the best practices and tools of the areas of Human-Computer Interaction and Software Engineering. The platform chosen for the application was Android, since it dominates more than 80% of the Brazilian market for mobile devices.

The last activity of the laboratory process corresponded to application usability evaluation (Barbosa and Silva, 2010). This evaluation aims to determine whether the application is understandable, easy to use and generates a good user experience. This assessment, however, does not evaluate the content or if the artifact produced, reaches the proposed objectives, leaving this evaluation to the next stage of the methodology. In this activity, 6 female teachers, from Yolanda Queiroz Elementary School participated. They received smartphones with the app installed a script of activities to be performed in the app, the activities were associated with the main system functions: register in the system, make a vocal analysis and check the result, set the "water time" function and get tips on vocal health. In addition to the script, a form (answered individually by each participant) so that impressions and opinions on the user experience could be registered.

As result of the usability evaluation, positive evidence was obtained, since the teachers considered the application easy to handle, they used it in the proper way and had a good user experience, despite some minor changes that have been suggested.

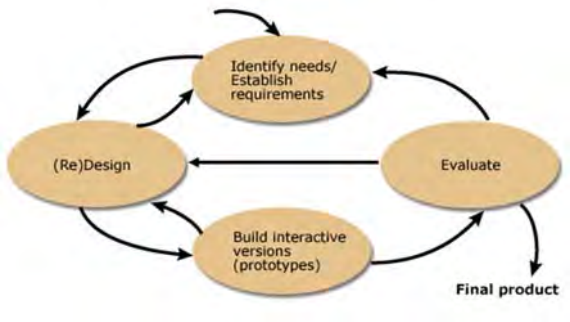



Figure 1: Interaction Design (Preece et al., 2013).

## 3 RESULTS

A first result of the project, from the literature review and the meetings with voice professionals (users) is the identification of the features/functions to be developed. A second result was the evaluation of the applications identified in the app stores based on the selected criteria. With this analysis, it was observed that the identified applications addressed the subject by focusing on specific points and did not offer features for managing, monitoring or a full self-care of vocal health, as can be seen in Table 1.

Based on the requirements and in the fact that there is no tool that would meet all these requirements, emerged VoiceGuard, a tool for mobile devices to the individualized support of the use of voice. The application is composed of 16 interfaces, each of them with a specific function. Table 2 shows the most relevant.

## 4 DISCUSSION

Throughout the research process, it was identified the lack of applications dedicated to vocal health. Thus, the VoiceGuard application brings new possibilities and helps professionals in the healthy use of this important tool that is the voice.

It is clear that technological advances have allowed the improvement of the media and bring to people the communicative interaction. Currently, the world has experienced the era of communication and health is one of the areas benefiting from the integration of these resources in the daily life of the population.

The VoiceGuard application fills a gap in the context of care to vocal health, giving users monitoring and self-management of vocal health through information, testing and alarm, setting an interface with the user. The application consists of a main menu covering six areas - voice analysis, sound level meter, time of water, results, tips and vocal heating - which unfolds in 16 interfaces, as outlined in Figure 1.

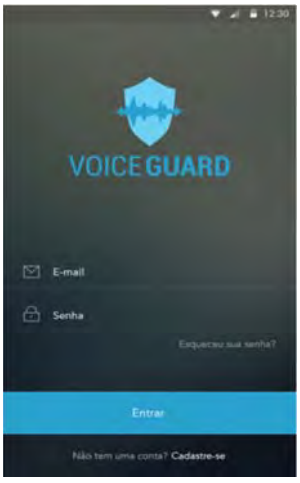
Each application area will be presented in detail in the following subsections:

Table 1: Evaluation summary of the applications studied based on the proposed requirements.

| FEATURES / SYSTEMS | Vocal WarmUp | Saúde Vocal | Vocal Ease | Warm me up for singers | VoxMetria |
|---|---|---|---|---|---|
| Vocal warm-up | X | - | X | X | X |
| Ambient sound measurement | - | - | - | - | X |
| Vocal health analysis | - | - | - | - | X |
| Care history | - | - | - | - | X |
| Vocal health tips | X | - | - | - | X |
| Water intake reminder | - | - | - | - | - |
| Level of vocal protection | - | - | - | - | - |

Legend: Symbol "X" mark the existence of the function in the application.

Table 2: Voice Guard interfaces.

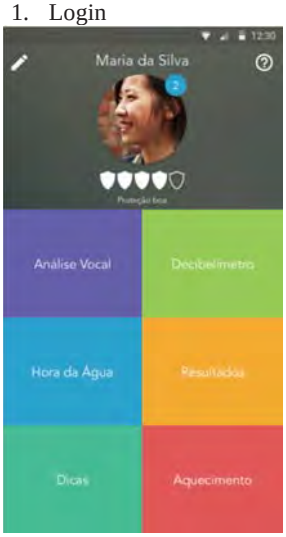



1. Login


2. Register

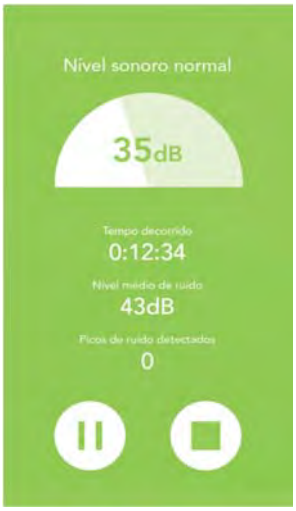

3. User classification


4. Vocal risk classification




5. System menu


6. Test of Maximum time of phonation




7. Measurement of sound/noise


8. Water time function

## 4.1 Register and Access Functions

There are seven interfaces that correspond to the functions for: the presentation of the application, login in the application (Table 2-1), user registration (Table 2-2), user classification (Table 2-3), data sharing, predisposition classification to vocal risk (Table 2-4) and main menu (Table 2-5).

Other system functionalities and associated interfaces are explained below.

## 4.2 Vocal Analysis

The first interface to support Vocal Analysis, presents an explanatory text on the operation of functionality. This explanation has its origin from users questions, recorded during testing, and shows to user the operation that occurs on the interface that executes the test used to check the maximum phonation time - TMF (Table 2-6) and that presents the result of the test with explanation.

Regarding the voice analysis, we used the test to measure the Maximum Phonation Time (TMF). TMF is widely used to verify the voice quality, using, among others, the fricative phonemes / s / and / z / (Behlau et al., 2001).

The test is performed as follows: the user, by issuing the phoneme /s/ in a sustained manner after a single exhalation, have the time of issue timed by the application. After finalize the issue of the first phoneme, other phoneme (/z/) should be required.

The use of phonemes /s/ and /z/ establish the link that checks the condition of glottal closure. Thus, the test based on parameters validated by experts

(Miglioranzi; Cielo; Smith, 2012), provides data on the dynamics of vocalization, being quite reliable in the evaluation of glottal efficiency. It is noteworthy that the normal range of s / z ratio is estimated at approximately 1 second (Gelfer; Pazera, 2006).

If the result of this test is less than 0.8, it indicates excessive coaptation of the vocal folds, which harms the vocal health of teachers, because it is perceived excessive effort to speak. Results from 0.8 to 1.2 is indicative of normality, suggesting that there is a normal operation of the vocal folds. Finally, the result of more than 1.2, indicates soprosity, i.e., there is an air leak between the vocal folds during phonation (Christmann et al., 2012).

The presented feature enables check and provide an overview of the glottal operation of each user, helping them in self-monitoring of vocal glottal efficiency. These results may alert about the signs of possible changes in the glottal closure, serving as a warning to seek expert help. The application also allows sharing the results in real time, with the speech therapist and/or physician, who are accompanying the teacher.

Another function is that, after the indication of the result of this test, the application presents the individual explanations of the identified condition, the interface results. With this, the user can access the tips interface where he/she will learn about the preservation of vocal health. It is clear that technology can influence and change the way of living and acting of people, even when these issues go through the care and health promotion context. Thus, it confirms the importance of incorporating advanced technologies in scenarios of assistance to vocal health as it implies a redefinition of self-care (Silva; Ferreira, 2009)

## 4.3 Decibelimeter

Another important function of VoiceGuard is the decibelimeter, which captures and shows the level of environmental noise. Moreover, it is possible to identify the noise spikes that exceed the acceptable intensity for the preservation of vocal health. For this purpose, the application presents three interfaces: explanations of the functions on the Main Menu, noise/sound measurement (Table 2-6) and the results of the noise measurement.

Studies show that noise is considered one of the most important risk factors for voice disorders in teachers. When the teacher is in the classroom and under the influence of noise, trying to be heard, he/she normally exceeds the voice intensity. Without realizing it, the teacher becomes vulnerable to the appearance of vocal disorders (Baring & Murgel, 2005; Brazil, 2015).

A study of Guidini et al. (2012) shows that, according to NBR 10,152, the acceptable noise level in the classroom must remain 40dB to 50dB (A). Thus, the application presents the following parameters to alert the user to the level of environmental noise: audible alarm - generates acoustic signals when the noise exceeds 50dB, conditioning students to reduce the parallel conversation or remain silent during the class; and visual alarm - reinforces the idea of the loud noise through the issuance of lights and aids in listeners behavior conditioning.

## 4.4 Water Time

The "water time" is another application function that helps professionals to condition themselves to drink water during vocal use. For this, the interfaces of "water time" (e.g. Table 2-8) are available.

The device enables the user to program the times for water intake over the work shifts. The application signals through visual and audible alarms the time for water intake, reminding the professional this as an important action for continued hydration and maintenance of vocal health. It is also possible that the user reports how many glasses of water he/she drank during the working day (Table 2-8).

Hydration is recommended in both the prevention and treatment of voice disorders. It is known that the benefits of hydration are many, such as: reducing the viscosity of mucus in the larynx and facilitating mucus-wave motion. Thus, a more hydrated mucosa provides greater flexibility to the vibration of the vocal folds, increasing vocal resistance and reducing the sensations related to vocal effort (Medaglia et al., 2008).

## 4.5 History Log

The interface of “History Log” provides access to the application user history and gives access to tips of voice heating and cool-down.

The interface that corresponds to the "results" is simple. It is possible to see the results of all tests, and enable visualization of comparative graphs. It also allows data sharing with the health professional who takes care of the app user. The VoiceGuard keeps in chronological sequence the results of the tests performed by the users and of other tests they want to register (reports or images).

Finally, the "heating and cool-down" is a tool that is still in the test phase for improvement. In this

interface, the user can perform simple exercises of voice heating and cool-down, before and after voice use, respectively.

## 5 CONCLUSION

The VoiceGuard application was designed in order to facilitate the improvement of the management of vocal health, aiming the adoption of habits and behaviors able to maintain healthy voice, which is consistent with the health promotion strategies.

As the mobile technology is present in everyday society, use it as a resource in health care can be a quick, easy and inexpensive strategy to achieve a considerable number of people. Therefore, we believe that the mobile application as a tool for vocal health promotion is a necessary resource and a new possibility for professional voice in the current context.

The usability testing gives us a positive and preliminary indication of acceptance and positive impact of the tool. All invited users were able to meet the proposed activities in an average time considered good, without doubts about the interaction with the tool and with some good suggestions (embraced by the staff) to change some terminologies and the location of information in the interfaces.

A third phase will be conducted after the development of the VoiceGuard, in which it will be validated with experts in voice and a larger number of elementary school teachers, in order to be available for free in the app stores.

## REFERENCES


Fabricio, Mariana Zerbetto; Kasama, Silvia Tieko; Martinez, Edson Zangiacomi. Qualidade de vida relacionada à voz de professores universitários. Rev. CEFAC, São Paulo , v. 12, n. 2, p. 280-287, Apr. 2010

Sarno F, Canella DS, Bandoni DH. *Mobile health* e excesso de peso: uma revisão sistemática. Rev Panam Salud Publica. 2014;35(5/6):424–31.

Christmann, M. K.; Scherer, T. M.; Cielo, C. A.; Hoffmann, C. A.. Tempo máximo de fonação de futuros profissionais da voz. Rev. CEFAC, São Paulo 2012.

Baring, J. G. & Murgel, E. Cuidado! Barulho demais faz mal à saúde. Revista Nova Escola, n. 179, p. 29, janeiro/fevereiro.2005.

Barbosa, S. D. J.; Silva, B. S. Interação Humano-Computador. Série SBC, Editora Campus-Elsevier, 2010.

Guidini, R. F. et al. Correlações entre ruído ambiental em sala de aula e voz do professor. Revista da Sociedade Brasileira de Fonoaudiologia, São Paulo, dez. 2012.

Behlau, M.; Madazio, G., Feijó, D.; & Pontes P. Avaliação de voz. (2001). In: Behlau M. organizador. *Voz: o livro do especialista*. vol 1. Rio de Janeiro: Revinter; 85-172.

Miglioranzi SL, Cielo CA, Siqueira MA. Capacidade Vital e Tempos Máximos de Fonação de /E/ áfonos e de /S/ em Mulheres Adultas Rev. CEFAC. 2012.

Medaglia, N.C. et al. Sintomas vocais relacionados à hidratação monitorada. IV Mostra Interna de Trabalhos de Iniciação Científica, 2008. Centro Universitário de Maringá, Paraná - Brasil

Behlau, M.; Pontes, P. Higiene Vocal: cuidando da voz. 4. ed. Rio de Janeiro: Revinter, 2009. 70p.

Brasil. Censo Escolar 2009. Brasília: Ministério da Educação, 2009.

Brasil. Conselho Nacional de Saúde. Diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. Resolução 466, de 12 de dezembro de 2012.

Brasil, C.C.P. A voz da professora não pode calar: sentidos, ações e interpretações no contexto da integralidade em saúde. 2015. 232 f. Tese – UECE, UFC e UNIFOR, Fortaleza, 2015.

Free, C. et al. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes, v.14, n.5, p.2-7, Out. 2010.

Gonçalves, C.G.O; Penteado, R.Z.; Silvério, K.C.A. Fonoaudiologia e saúde do trabalhador: a questão da saúde vocal do professor. Saúde em Revista, Piracicaba, v. 7, n.15,Jan. / Abr.2005.

Riley, et al. Health behavior models in the age of mobile interventions: are our theories up to the task? TranslBehav Med , v.1, n.1Mar. 2011.

Roy, N. et al. Prevalence of voice disorders in teachers and the general population. Journal of speech, language, and hearing research, Rockville, v. 47, n. 2, p. 281-293, Apr. 2004.

Preece, J. et al. Design de Interação: além da interação homem-computador. 3.Ed. Porto Alegre:Bookman, 2013.

Silva, R.C.; Ferreira, M. A. A tecnologia em saúde: uma perspectiva psicossociológica aplicada ao cuidado de Enfermagem. Esc. Anna Nery Rev. Enferm., Rio de Janeiro, v.13, n.1, p: 169-173, Jan./Mar., 2009.

Whittaker, R. Issues in mHealth: findings from key informant interviews. J Med Internet Res, Pittsburgh, v. 14, n 5, p.129, Oct. 2012.

# Identifying Key Components of Services in Healthcare in the Context of out-Patient in Norway

Eunji Lee
*SINTEF, Oslo, Norway*
*eunji.lee@sintef.no*

Keywords: Service Design, Healthcare, eHealth.

Abstract: This paper discusses components of service in healthcare. Four components of a service (service customer, service worker, service setting and service process) were introduced. Yet these components have not been explored in healthcare cases. We identified the key components through our case study with out-patient histories, involving electronic health record systems. Based on our analysis we propose a set of components to be considered for designing stakeholder-centred services in healthcare. The result of this study might be useful to the health informatics researchers to better understand the service interactions in today's healthcare in a more analytic and holistic way by taking the service engineering perspective, at the same time to the service engineering or design researchers to have a deeper insight of the services in healthcare and the components to be considered when designing the services.

## 1 INTRODUCTION

The service delivery process in healthcare is complex (Reichert, 2011); designing healthcare services can therefore be challenging. Healthcare services involve many actors, who work with different agendas, have highly specific knowledge, and who have tasks that are intertwined with other organisations. eHealth, a healthcare service that is supported by telecommunications and information technology (Mitchell, 2000), complicates the service delivery process even further. While eHealth technologies break down barriers of time and place, thus bringing people and resources together to provide healthcare services in more efficient ways (Hesse and Shneiderman, 2007), it also generates various interactions between many actors and systems which were absent in conventional health service situations.

Involving eHealth technologies in today's healthcare service is not uncommon. For instance, while a patient has a consultation with his/her general practitioner (GP), the GP looks up the information from the previous consultation(s) through an electronic health record (EHR) system. The use of such technology changes the healthcare practices and consequently can affect patient's life (Rodolfo et al., 2014). Therefore, there is a need to understand the complex service delivery process in healthcare in an analytic and holistic approach. Such approach might contribute to better assess the existing services in healthcare, which can be a starting point for designing improved services.

Gadrey (2002) introduced three components of a service: service provider, customer/client/user, and transformation of a reality. Fisk et al. (2013) presented and defined four components of service and the definitions are as below.

- Service customer: the recipient of the service
- Service worker: the contributor of the service delivery by interacting with service customer
- Service setting: the environment in which the service is delivered to the customer
- Service process: the sequence of activities essential to deliver the service

Yet these components have not been fully explored in today's complex healthcare settings. Our research question is "What are the key components in out-patient services?"

The rest of this paper is organised as follows: We first describe our research approach, context, methodology, and methods for data collection and analysis in Section 2. In Section 3, we introduce two out-patient histories. We then present the results from our analysis in Section 4 and discuss the results in Section 5. In Section 6, we discuss the limitations of this study. Finally, we conclude our study and suggest future research in Section 7.



Lee, E.
Identifying Key Components of Services in Healthcare in the Context of out-Patient in Norway.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 354-361
ISBN: 978-989-758-213-4

# 2 RESEARCH APPROACH

We applied a qualitative methodology to investigate our research question. We conducted a multiple case study using two out-patient histories in Norway from September-October 2013. Case study is defined as "scholarly inquiry that investigates a contemporary phenomenon within its real-life context (Yin, 1994)." Multiple case study is instrumental study which allows researchers to understand and analyse several cases across settings thus leading better theorising (Stake, 2005; Baxter and Jack, 2008).

Data was collected through conducting document analysis, observations and interviews at a surgical out-patient clinic in a hospital in Norway. Due to ethical consideration, a chief nurse explained two patients' histories by showing the electronic documents in an EHR system and other relevant paper documents; no direct access to the EHR system was given to the researcher. Semi-structured interviews with the chief nurse followed after the nurse's explanations. To obtain deeper insight in the histories, we conducted observations and unstructured interviews of a secretary working at the hospital's post/document centre, a medical doctor (specialist) and a health secretary working at the clinic. During the observations, the researcher took notes and some photos of the documents were taken. All interviews were audio-recorded. Email exchanges and telephone conversations supplemented the data after the interviews.

Document analysis is a systematic method for reviewing or evaluating documents, which is unobtrusive and nonreactive when obtaining empirical data (Bowen, 2009). Observation is a useful data gathering method in naturally occurring settings and it helps the researchers to understand the users' context, tasks, and goals (Rogers et al., 2011). Unstructured and semi structured interviews can be most suitable when the researchers want to have a deeper insight of a problem domain that is not familiar by giving the participants the chance to educate the researchers. (Lazar et al., 2010). Interviews and/or observation are often used to establish credibility and minimise bias of the data from document analysis, as a means of triangulation (Bowen, 2009). Triangulation is a process of using several sources of evidence to clarify meaning and verify the repeatability of an interpretation (Stake, 2005).

We analysed the collected data of two out-patient histories using qualitative content analysis (Graneheim and Lundman, 2004). Thematic analysis (Fereday and Muir-Cochrane, 2006) was used to fine-tune the analysis.

# 3 INSIGHT OF THE PATIENT HISTORIES

In this section, we introduce the patient histories and explain how we analysed our data. First, we briefly describe the two out-patient histories. Second, we present the process of our analysis.

## 3.1 The Out-patient Histories

The first patient history covered a period of ten and a half months. Different places were involved in this case, including a GP centre and two hospitals. Several stakeholders were involved: a patient, GP, secretary, radiologist, minimum two specialists, health secretaries, and nurses from the hospitals. Three different health information systems were used: a GP's EHR system, a radiology information system (RIS), and a hospital EHR system. These systems were used to store and share the patient related information. The GP's EHR system and the RIS could communicate with the hospital EHR system in a limited degree (e.g., sending and receiving electronic referrals or results of computed tomography (CT)).

The second patient history covered a period of two and a half months until the time of the interview and was still ongoing. Different places were involved in this case, including a GP centre and three hospitals. Even more stokeholds were involved: a patient, GP, radiologist, two pathologists, minimum three specialists, secretaries, health secretaries, and nurses from the different hospitals. Four different health information systems were used: a GP's EHR system, a RIS, and two different types of hospital EHR systems. The GP's EHR system and the RIS could communicate with a hospital EHR system in a limited degree, like in the first case. However, the other hospital EHR system could not communicate with the three other systems at all. Therefore, more interactions with physical evidence, such as a postal letter, were generated to cover the communication barrier (e.g., a specialist received a referral via postal letter).

Figure 1 shows the communications between the stakeholders in the first out-patient case and Figure 2 shows the communications between the health information systems in the first out-patient case.

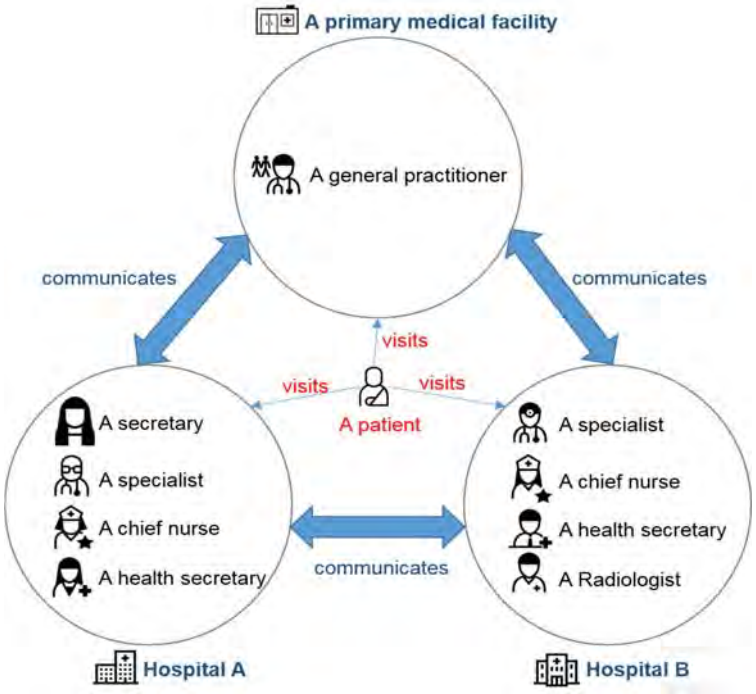



Figure 1: Communications between the stakeholders in the first out-patient case.

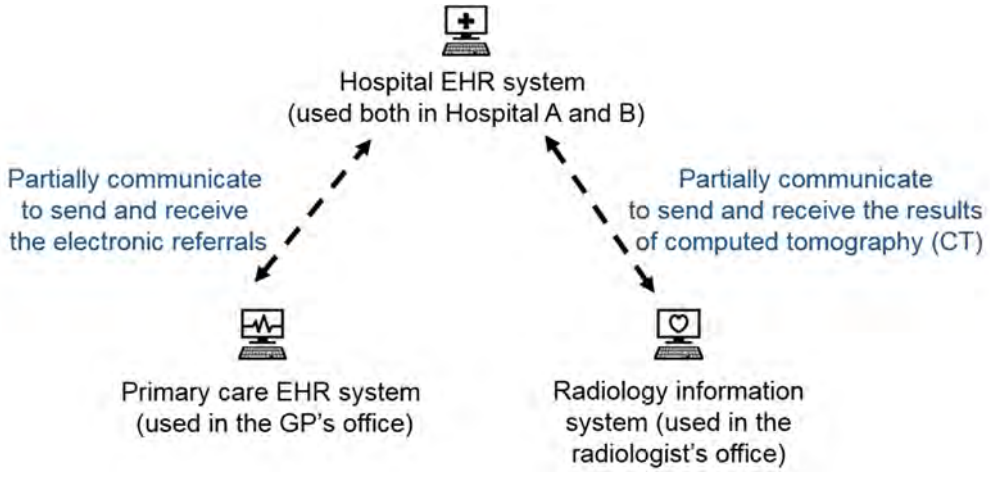



Figure 2: Communications between the health information systems in the first out-patient case.

## 3.2 Data Analysis Process

Based on the data collected in the researcher's notes, audio files, and photos taken, we constructed each patient's journey using excel spreadsheet. We identified key components of services in healthcare by improving the templates of the journeys in an iterative manner.

We constructed the first version of the journeys using a 'service blueprint (Stickdorn and Schneider, 2010)' method which includes the roles of the involved stakeholders, the places where the events happened, and the contexts of the events. We found that the stakeholder is either service customer or worker, and that the place is the service setting. We learnt the events can be recognised as small units constituting the entire service provision. Therefore, we call the context of the event as **sub-service provision context** and add it as a key component of services in healthcare.

We then constructed the second version of the journeys by improving the first version. While we

were doing this, we discovered that some events contain a sender, a receiver and an object. We also added the date for each event in the second version. We learnt that the date can be recognised as an indicator in the service process.

Finally, we could develop a systematic template that shows the patients' journeys (third version). We added the overall aim of the service and the identifier for each event. We distinguished the event involving a sender, receiver and object as a *touchpoint* that indicates an interaction between two stakeholders. We also found that each touchpoint contains a communication channel that is used in order to deliver the object to the receiver. We identified the other events as actions when there is no such interaction. We discovered that some touchpoints are electronic-based, occurring in or between the health information systems. For example, an electronic referral was sent from a GP's EHR system to a Hospital A's EHR system. We found that these health information systems can be seen as stakeholders that contribute to out-patient services. In addition, we identified various types of interaction in the patients' histories: human-to-human interaction (face-to-face or via telephone), human-to-physical evidence interaction, and human-to-computer interaction. We call the aim of the service as **service objective** and the type of interaction as **service interaction type** and add these as key components of services in healthcare.

People producing or maintaining an EHR system have influence on the interactions between a healthcare professional and the EHR system. We regard these people as **secondary service workers**. A patient can be affected by an interaction between a healthcare professional and an EHR system. In this context, we regard the patient as a **secondary service customer**.

# 4 COMPONENTS OF SERVICES IN HEALTHCARE

In this section, we first present the key components of services in healthcare, which we identified during our data analysis. We then present two examples (one for health service and one for eHealth service) of the services according to the key components we analysed from the patient histories.

## 4.1 Components of Health and eHealth Services in Out-patient Context

The out-patient histories include interactions situated in health service and in eHealth service. Here we define a health service as a conventional medical service not containing any interactions via electronic channels. We define an eHealth service as a service containing interactions via electronic channels. We identified the key components of health service and eHealth service separately.

The objective of the interactions situated in the health services was treatment. Thus, the service customers were the patients and the secondary service customers might be family members of the patients. The service workers were the healthcare professionals from different groups and organisations, like a GP and a nurse. The setting of the interactions situated in the health services were either a medical facility (e.g., a hospital) or a location where the patient has a touchpoint (e.g., a patient reads a postal letter at home or answers a phonecall at work). The processes of the health services were sequences of actions and touchpoints of the patients and the healthcare professionals. We found that the interaction type situated in the health services was either human-to-human interaction (e.g., a GP examines a patient.) or human-to-physical evidence interaction (e.g., a GP reads a postal letter from a hospital.). The health services involved sub-services (smaller units constituting the service) for the service objective (patient treatment). The sub-service provision context of the health services was either a service worker provides a service to a service customer (e.g., a surgeon operates on a patient to treat a disease.) or a service worker provides a service to another service worker (e.g., a health secretary in a hospital sends an out-patient note to a GP via postal letter.).

The objective of the interactions situated in the eHealth services was efficient communication among healthcare professionals. Therefore, the service customers were the healthcare professionals from different groups and organisations, while the patients became the secondary service customers. The service workers of the eHealth were the health information systems such as EHR and RIS, while the secondary service workers might be people producing or maintaining the health information systems. The setting of the interactions situated in eHealth service was the health information system software. The processes of the eHealth services were sequences of touchpoints via the health information systems. We found that the service interaction type

situated in the eHealth services was human-to-computer interaction (e.g., a specialist dictates an out-patient note through an EHR system). The sub-service provision context of the eHealth services was a service worker provides an e-service to a service customer (e.g., a GP's EHR system stores a referral, which can be seen electronically by a secretary in a hospital's post centre.) Table 1 shows the components we identified as the result of our data analysis.

Table 1: Components of health and eHealth services in out-patient context in a hospital in Norway.

| Service type / Component | Health service | eHealth service |
|---|---|---|
| **Service objective** | Treatment | Efficient communication |
| Service customer | Patient | Healthcare professional |
| **Secondary service customer** | Family member of a patient | Patient |
| Service worker | Healthcare professional | Health information system |
| **Secondary service worker** | None | People producing or maintaining the health information system |
| Service setting | A medical facility or a location where a patient has a touchpoint | Health information system software |
| Service process | Sequence of actions and touchpoints of a patient and health professionals | Sequence of touchpoints via health information systems |
| **Service interaction type** | Human to human or Human to physical evidence interaction | Human to computer interaction |
| **Sub-service provision context** | A service worker provides a service to a service customer or A service worker provides a service to a service worker | A service worker provides an e-service to a service customer |

## 4.2 Examples

In this section, we present two examples of the services according to the key components we identified. First, we show one example for health service and then we show one example for eHealth service.

The following example shows the components we identified using a part of a hypothetical episode, in which a patient visits a specialist in a hospital.

- Service process: A patient comes to a specialist's office room, the specialist talks with the patient about his/her condition, and then the specialist examines the patient using a stethoscope.
- Service customer: The patient
- Secondary service customer: A spouse of the patient who accompanies the patient
- Service worker: The specialist
- Secondary service worker: None
- Service setting: An office room for the specialist at an out-patient clinic in a hospital
- Service interaction type: Human to human (the specialist to the patient) interaction
- Sub-service context: A service worker (the specialist) provides a service (examination with stethoscope) to a service customer (the patient).
- Service objective: Treatment

The following example shows the components we identified using a part of hypothetical episode that a specialist writes an out-patient note.

- Service process: The specialist navigates to a dictation module in a desktop-based EHR system and dictates an out-patient note into the system.
- Service customer: The specialist
- Secondary service customer: The patient
- Service worker: The EHR system the specialist uses
- Secondary service worker: The people who produce and maintain the EHR system
- Service setting: A desktop-based EHR system software
- Service interaction type: Human to computer (the specialist to the EHR software) interaction
- Sub-service context: A service worker (the desktop-based EHR system software) provides an e-service (electronic dictation service) to a service customer (the specialist).
- Service objective: Efficient communication

# 5 DISCUSSION

In this section, we discuss the above-mentioned results. We especially focus on the additional components we identified in the out-patient services during the iterative process of our analysis.

## 5.1 Service Type: Service vs. e-Service

Characteristics of e-services are different from the ordinary services since e-services involve interactions via electronic channels. Väänänen-Vainio-Mattila et al. (2009) claim that the characteristics of the service experience (inseparability, variability, perishability, and intangibility) are recognised for ordinary services and do not apply directly to e-services. Therefore, the components affecting ordinary service and e-service experience might be different from each other. We identified components of health service and eHealth service separately. We found that these components are not contradictory each other, but rather complement each other. For example, the eHealth service in the sub-section 4.2 can be followed after the health service in the sub-section 4.2 is done. But, it is also possible that an eHealth service comes before or during a health service. For example, the specialist can check the patient's information via the desktop-based EHR system software (In other words, the desktop-based EHR system software provides an electronic patient information look-up service to the specialist.) before the patient comes into his/her office. We suggest that all of the components should be considered when designing services in healthcare, because today's healthcare involves both health service and eHealth service. Holmlid and Evenson (2008) also argued that identifying clear genres (in this paper, we call these *service type*) and the components offers efficiency in service design.

## 5.2 Service Objective

In our case study, the purpose of the health service was providing treatment to the patients. However, in the eHealth service perspective, the purpose becomes efficient communication among healthcare professionals. Concerning these service objectives, it might be beneficial to better orchestrate the actions and touchpoints in service experience when designing services in healthcare.

## 5.3 Secondary Service Customer and Worker

In service dominant logic (Chandler and Vargo, 2011), interactions hidden from customers are not considered in value co-creation (Wetter-Edman et al., 2014). However, those interactions can affect the customers' service experience. For example, a patient's experience can be affected by the interactions between his/her GP and an EHR system. Alsos and Svanæs (2011) introduced the concept of primary and secondary user in eHealth services context. A primary user indicates a person who uses an information system directly, and a secondary user points out a person who relies on the primary user to get information from the system and who is affected by the primary user's experiences with the system (Alsos and Svanæs, 2011). In the eHealth service context, the patient becomes a secondary service customer and people producing/maintaining health information systems become secondary service workers. On the other hand, in a health service context, the family members of a patient become secondary service customers. Holmlid (2007) argued that the customer's customer (secondary service customer) is as important as the customer in service design. We postulate that considering not only secondary service customer, but also secondary service worker when designing a service, might contribute to better understanding the whole service delivery.

## 5.4 Service Interaction Type

A service consists of different types of interactions. "The service perspectives become a challenge to interaction design, and technology usage becomes a challenge to service design (Holmlid, 2007)." Paying attention on those types and considering them in appropriate manners when evaluating and designing service might be helpful to create consistency in service provision.

## 5.5 Sub-Service Provision Context

In a broad and holistic perspective, a service can contain several sub-services. For instance, an air travel service consists of sub-services, such as check-in, providing meal on the plane etc. In healthcare, many actors are connected to each other to solve specific tasks and eventually pursuit the ultimate goal: maximising health of the population in the society (Coast, 2004). Considering such sub-service provision types, it would be helpful to better

coordinate various interactions between different actors and systems in services in healthcare. In our case, no ubiquitous computing or pervasive technology originated sub-service was found. However, it might appear more and more in future services as the technology advances. Since the interactions originated from ubiquitous computing or pervasive technology happen without the customer's direct control (Cellary, 2015), it can be more challenging for us to well integrate them in service delivery.

## 6 LIMITATIONS

There are different types of eHealth service depending on who communicates with whom. We conducted our case study with eHealth services where healthcare professionals communicate with each other. Thus, the key components in other types of eHealth service (e.g., telepsychiatry where a psychiatrist communicates with a patient) might be different from what we identified.

Our case study was conducted with desktop-based eHealth services. Conducting a case study with a mobile-based eHealth service might lead to the results that are not the same as what we found from our case study.

## 7 CONCLUSION

Our research reveals that out-patient care includes interactions situated in both health service and in eHealth service. We found that these two different types of service consist of different components. We expanded the Fisk et al. (2013)'s four components of service (service customer, service worker, service setting, and service process) for services in healthcare by adding five new components: service objective, service interaction type, sub-service provision context, secondary service workers, and secondary service customer. Considering these components when evaluating service experience might support an analytical way of understanding the complexity in service delivery process in healthcare. This understanding might contribute to designing more stakeholder-oriented services in healthcare.

There is a need for a holistic and stakeholder-centred approach in designing and evaluating eHealth services. "the effectiveness of emerging eHealth technologies in improving the processes or outcomes of healthcare is unproven (Pagliari, 2007)." We envision further research in the form of empirical studies that consider the key components of services in healthcare when evaluating or designing services in healthcare. Investigating how to present or document all the actions and touchpoints of a service delivery process in more holistic way might also be interesting. Our research is based on document analysis, observation, and interview because of the challenges in conducting ethnography study with patients due to ethical consideration. Thus, we are also interested in investigating how to collect richer data that can provide a deeper insight of services in healthcare.

## ACKNOWLEDGEMENTS

The research presented here has been conducted within the VISUAL project (project number 219606) funded by the Research Council of Norway. Thanks to Ragnhild Halvorsrud for helping me arranging data collection activities for the case study, to Maja Van Der Velden and Stefan Holmlid for guidance in writing this paper, to our industrial partner for the contribution. Above all, we thank the participants who provided data for the case study.

## REFERENCES

Alsos, O.A., Svanæs, D., 2011. Designing for the secondary user experience. In *INTERACT 2011, 13th International Conference on Human-Computer Interaction*. Springer-Verlag. pp. 84–91.

Baxter, P., Jack, S., 2008. Qualitative case study methodology: Study design and implementation for novice researchers. The qualitative report 13, 544–559.

Bowen, G.A., 2009. Document analysis as a qualitative research method. Qualitative research journal 9, 27–40.

Cellary, W., 2015. e-Service-dominant Logic. In *AHFE 2015, 6th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences*. Procedia Manufacturing. pp. 3629–3635.

Chandler, J.D., Vargo, S.L., 2011. Contextualization and value-in-context: How context frames exchange. Marketing Theory 11, 35–49.

Coast, J., 2004. Is economic evaluation in touch with society's health values? BMJ 329, 1233–1236.

Fereday, J., Muir-Cochrane, E., 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. International journal of qualitative methods 5, 80–92.

Fisk, R., Grove, S., John, J., 2013. *Services Marketing An Interactive Approach*, Cengage Learning. USA, 4th edition.

Gadrey, J., 2002. The misuse of productivity concepts in services: Lessons from a comparison between France and the United States. Edward Elgar Publisher.

Graneheim, U.H., Lundman, B., 2004. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Education Today 24, 105–112.

Hesse, B.W., Shneiderman, B., 2007. eHealth Research from the User's Perspective. American Journal of Preventive Medicine 32, 97–103.

Holmlid, S., 2007. Interaction design and service design: Expanding a comparison of design disciplines. Nordes (Nordic Design Research).

Holmlid, S., Evenson, S., 2008. Bringing Service Design to Service Sciences, Management and Engineering, in: Hefley, B., Murphy, W. (Eds.), Service Science, Management and Engineering Education for the 21st Century, Service Science: Research and Innovations in the Service Economy. Springer US, pp. 341–345.

Lazar, J., Feng, J.H., Hochheiser, H., 2010. *Research methods in human-computer interaction.* Wiley.

Mitchell, J., 2000. Increasing the cost-effectiveness of telemedicine by embracing e-health. J Telemed Telecare 6, 16–19.

Pagliari, C., 2007. Design and Evaluation in eHealth: Challenges and Implications for an Interdisciplinary Field. J Med Internet Res 9.

Reichert, M., 2011. What BPM Technology Can Do for Healthcare Process Support, in: Peleg, M., Lavrač, N., Combi, C. (Eds.), Artificial Intelligence in Medicine. Springer Berlin Heidelberg, pp. 2–13.

Rodolfo, I., Laranjo, L., Correia, N., Duarte, C., 2014. Design Strategy for a National Integrated Personal Health Record. In *NordiCHI '14, 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. ACM. pp. 411–420.

Rogers, Y., Sharp, H., Preece, J., 2011. *Interaction Design: Beyond Human Computer Interaction*, Wiley. 3rd edition.

Stake, R.E., 2005. Qualitative Case Studies, in: The SAGE Handbook of Qualitative Research. Sage Publications, pp. 443–466.

Stickdorn, M., Schneider, J., 2010. *This is service design thinking*. Wiley.

Väänänen-Vainio-Mattila, K., Väätäjä, H., Vainio, T., 2009. Opportunities and Challenges of Designing the Service User eXperience (SUX) in Web 2.0, in: Isomäki, H., Saariluoma, P. (Eds.), Future Interaction Design II. Springer London, pp. 117–139.

Wetter-Edman, K., Sangiorgi, D., Edvardsson, B., Holmlid, S., Grönroos, C., Mattelmäki, T., 2014. Design for Value Co-Creation: Exploring Synergies between Design for Service and Service Logic. Service Science 6, 106–121.

Yin, R., 1994. *Case study research: Design and methods*. Sage publishing. Beverly Hills. CA,

# Poke-R
## *Using Analytics to Reduce Patient Harm*

James P. McGlothlin[1], Evan Crawford[1], Jesse Wyatt[2], Carlos Samayoa[2], Yana Vaks[2], Brenda Bruneau[2], Merrick Lopez[2], Anthony Moretti[2], Michele Wilson[2] and James Pappas[2]
[1]*Fusion Consulting Inc, Irving, TX, U.S.A.*
[2]*Loma Linda University Health System, Loma Linda, CA, U.S.A.*
*{jmcglothlin, ecrawford}@fusionconsultinginc.com,*
*{jewyatt, csamayoa, yvaks, bbruneau, mlopez, amoretti, mwilson, jpappas}@llu.edu*

Keywords: Data Warehousing, Healthcare Analytics, Quality, Pediatric Intensive Care, Business Intelligence.

Abstract: Major events and surgeries are not the only sources of trauma during a hospital encounter. Many small, less invasive events such as shots, line placements, blood draws, and imaging studies happen throughout a patient's hospital stay. Many of these less traumatic events have the potential to negatively impact patient outcomes by increasing the risk of hospital-acquired infections through skin invasions and exposure to organisms, reducing the patient experience by causing pain and frustration, increasing cost and causing other complications. The goal of this project is to reduce such events when they are not clinically required. This is an analytics project so this goal is facilitated by making accurate and meaningful information available to the appropriate personnel. This includes timely information to clinicians so they can alter treatment, and retrospective trend analysis to enable and track performance improvement and identify opportunities for additional process improvement.

## 1 BACKGROUND

This project is based on the Prevent Pain and Organisms from sKin and catheter Entry (POKE) project initiated at Dixie Regional Medical Center. Dixie Regional implemented the POKE initiative within their Neonatal Intensive Care Unit (NICU) (Ridout, 2014). The results at Intermountain Health System show that the POKE project has resulted in reduced POKEs for NICU patients and significant financial savings estimated at $3.5 million over 5 years for a single hospital. Reduction in length of stay was also identified. Figure 1 and Figure 2 quantify cost reduction associated with the original Dixie Regional Medical Center project.

We are calling our project POKE-R because we are including Radiology events. Radiology images can cause serious complications later in life for pediatric patients due to the much higher sensitivity children have to radiation, and also reduce patient/parent experience (Medscape, 2014) (Slovis, 2002) (Brenner, 2002). We are leveraging this prior research and enhancing it. We are considering anything a POKE which invades the skin or opens a line or drain into the patient. This includes medication administrations, blood draws, placement of lines, drains and airways (LDAs), surgeries, and other invasive procedures.

The goal of our project is to reduce POKE-R events by providing detailed information to the clinicians. Often lab draws or procedures are not medically necessary and may cause more harm than good (Salisbury et al, 2011). Also, many times lab tests can be combined to use a single specimen collection. A patient sees many providers throughout a hospital stay and there may be redundant orders or orders which are no longer medically warranted.


Figure 1: Hospital savings experienced by Dixie Regional Medical Center POKE initiative.



McGlothlin, J., Crawford, E., Wyatt, J., Samayoa, C., Vaks, Y., Bruneau, B., Lopez, M., Moretti, A., Wilson, M. and Pappas, J.
Poke-R - Using Analytics to Reduce Patient.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 362-369
ISBN: 978-989-758-213-4

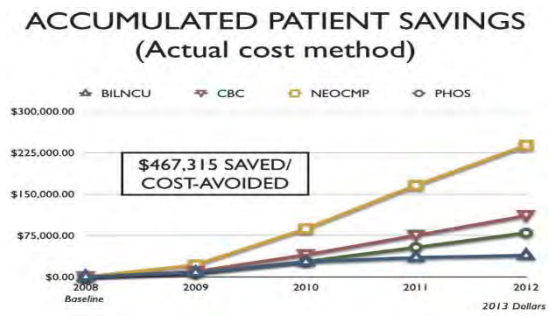



Figure 2: Patient savings experienced by Dixie Regional Medical Center POKE initiative.

By reducing POKEs we hope to achieve each of the following potential improvements:

- Reducing hospital-acquired infections

Every time a patient's skin is punctured, line is opened, or catheter is placed, there is an increased risk of a hospital-acquired infection. For example, there is significant evidence that risk of Central Line Associated Blood Stream Infections (CLABSI) is increased by repeated blood lab draws (Foster and Sabella, 2011) (Grissinger, 2011) (Sengupta et al, 2010).

- Improving the patient experience and the satisfaction of the patient and his/her family

The pain caused by invasive procedures, shots, and placement of lines, drains and airways has a clear effect on the patient experience and satisfactions.

- Reducing anemia, blood loss and blood transfusions

Each time blood is removed from a patient to perform a lab test, there is an increased risk of side effect or even the need for a blood transfusion. This is particularly prevalent with neo-natal patients and young pediatric patients and also with acute myocardial infarction (AMI) patients (Bateman et al, 2008) (Salisbury et al, 2011) (McVoy and Shandler, 2013)

- Reducing complications from radiology

Radiology imaging have been connected to complications for pediatric patients including cancer. Children have a greater sensitivity (10 times more than a middle aged adult) to radiation dose and computed radiography. Furthermore, the necessary movement required for a radiological exam can increase risks of further injury or disrupt lines, drains and airways (Foster and Sabella, 2011) (Brenner, 2002). Finally, the physical methodology of many imaging procedures, such as MRIs, can cause psychological trauma and reduce the quality of the patient experience.

- Reducing length of stay

Reducing POKEs has been shown to reduce the median length of stay for neonatal patients. Furthermore, POKEs increase the risk of infection, and hospital-acquired infections dramatically increase average length of stay for patients (Alharfi et al., 2014) (Foster and Sabella, 2011).

- Reducing cost

All procedures performed incur costs so the simple act of reducing the number of procedures directly reduces costs. Costs are also indirectly reduced through reduction in hospital-acquired infections and reduced length of stay.

The principle behind our approach is value-based medicine. The concept is to focus practice on patient and financial value of the medical interventions. The goal is to incorporate the highest level of evidence based interventions while ensuring adequate patient care and minimizing healthcare costs (Bae, 2015).

# 2 TECHNICAL IMPLEMENTATION

At Loma Linda, we have an enterprise data warehouse sourced from the clinical data in the Epic Electronic Medical Record (EMR) System. We implemented the POKE-R process by using the information in this data warehouse to build a new POKE-R fact table using standard industry dimensional modelling data warehouse practices.

The foundation for the data warehouse is provided by the EMR vendor, but we have extended it to include more detailed information useful for this project including

a) Lines, drains and airways

Needed to determine when a LDA was placed and when it was removed.

b) Procedure performance details

Needed to know if a procedure or image was actually performed, by who and when

c) Medication Administration Route

Needed to know how a medication was administered

Additionally, we had to add two extensions specifically for this project. We added an extension and modified EMR workflow specifically for LDAs to know how many attempts the LDA placement took. Furthermore, physician-performed LDAs such as central lines were documented in a different manner so we created a special extract to get the placement times and attempts. Finally, it was not enough to know when a specimen was taken. We needed to know which procedure orders shared blood draws and

which required separate blood draws. If 5 lab draws show the same collection time, it is important to know whether they were separately drawn, or all of the tests used the same blood collection.

With these extensions, all of the data needed to mine the POKE-R information was available in the data warehouse. However, before we could search for the POKE-R events, we had to configure which events were defined as POKEs. We did not want to hard-code this information and we did not want the information determined or maintained by IT personnel as it is clinical in nature. Therefore, we established an interface to configure POKE-R.

We needed to define every event which was a POKE-R event and whether it was painful. This needs to be configured using attributes of the data elements. The following attributes were identified by the clinician as identifying POKEs:

1. Medication Administration: Route and Administration Event
2. Lab Test: Specimen Type and Specimen Source
3. Procedure Order: Type and Code

Additionally, the presence of a line or drain prior to the event can impact whether the event is a POKE and whether it is painful. For example, blood tests and medication administrations are considered non-painful if they use an existing line. A urine sample is not a POKE at all unless there is a catheter used to obtain the specimen.

Finally, if the patient is under anesthesia at the time of the event, it is considered non-painful.

We created a simple secure interface for the Patient Safety and Reliability leadership to provide and administer this clinical information. This interface contains the data points listed above prepopulated from the actual clinical data warehouse. The user can then choose which values for each data point indicate a POKE and can combine data points.

We developed software code using Microsoft SQL Server Integration Services (SSIS) to read the POKE-R configuration file and then extract POKE-R events from the data warehouse into a new table within the data warehouse called PokeFact. This fact table contains the following information:

1. Encounter ID: The encounter the POKE happened during
2. POKE type: LDA Placement, LDA Placement Extra Attempts, Medication Administration, Image, Specimen Collection, Point of Care Test (POCT), Invasive Procedure, Surgery
3. Whether the POKE has already happened or is scheduled to happen in the future
4. When the POKE was ordered
5. The provider who ordered the POKE
6. When the POKE was scheduled to occur
7. When the POKE occurred
8. Whether the POKE is painful
9. Who performed the POKE

One thing that was very important was to determine the scheduled POKE-R events. Our goal was to show the clinician the upcoming POKE-R schedule so that treatment could be altered to reduce the POKEs. To do this we brought in every scheduled medication administration, procedure, surgery, image or lab test.

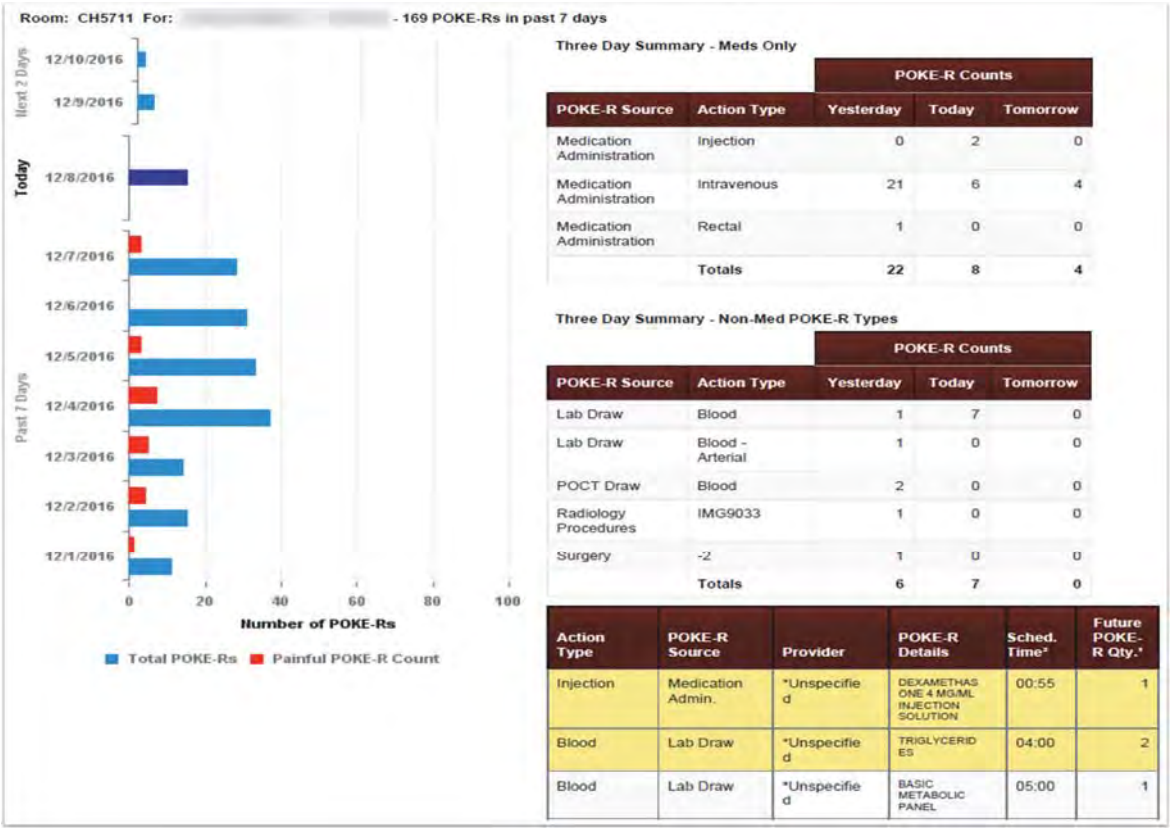
# 3 INFORMATION PRESENTATION

At this point, we had aggregated all of the information necessary to analyze POKE-R. The next step was to make this information useful to a clinician.

We developed three reports. The first report was a detailed report of patients currently in the hospital. This report lists for each patient the total number of pokes and painful pokes, the number of pokes and painful pokes in the last 7 days and the number of scheduled pokes for the next 3 days in graphical format. This poke counts are then shown grouped by the type of poke. Finally, every POKE performed in the last 7 days and every POKE scheduled for the next 3 days was individually listed with details. This reported was filtered by department so that an individual department could see each patient in the department. Scheduled POKEs are not always ever performed or cancelled. They can be left in pending status. So we dropped any scheduled POKE in the past which was never performed. The second report was the abridged version of the first report, showing only the number of POKEs over the past 7 days and what POKEs were scheduled for today. This made it more simpler for clinicians to digest the POKE-R information and make actionable decisions. Figure 3 shows an example of this report.

The third report was a trend of POKEs per Patient per Day over time so we could see if performance improvement was being achieved. This report was able to be filtered by location, unit, or attending provider. Figure 4 shows an example of this retrospective report.

Together, these reports enable process improvement and improved treatment. The provider and treatment team are supported by the detailed report, while the analysts in Patient Safety and Reliability have the aggregate and retrospective information to identify improvement opportunities.




Three Day Summary - Meds Only

| POKE-R Source | Action Type | POKE-R Counts Yesterday | Today | Tomorrow |
|---|---|---|---|---|
| Medication Administration | Injection | 0 | 2 | 0 |
| Medication Administration | Intravenous | 21 | 6 | 4 |
| Medication Administration | Rectal | 1 | 0 | 0 |
| | Totals | 22 | 8 | 4 |

Three Day Summary - Non-Med POKE-R Types

| POKE-R Source | Action Type | POKE-R Counts Yesterday | Today | Tomorrow |
|---|---|---|---|---|
| Lab Draw | Blood | 1 | 7 | 0 |
| Lab Draw | Blood - Arterial | 1 | 0 | 0 |
| POCT Draw | Blood | 2 | 0 | 0 |
| Radiology Procedures | IMG9033 | 1 | 0 | 0 |
| Surgery | -2 | 1 | 0 | 0 |
| | Totals | 6 | 7 | 0 |

| Action Type | POKE-R Source | Provider | POKE-R Details | Sched. Time* | Future POKE-R Qty.* |
|---|---|---|---|---|---|
| Injection | Medication Admin. | *Unspecified | DEXAMETHASONE 4 MG/ML INJECTION SOLUTION | 00:55 | 1 |
| Blood | Lab Draw | *Unspecified | TRIGLYCERIDES | 04:00 | 2 |
| Blood | Lab Draw | *Unspecified | BASIC METABOLIC PANEL | 05:00 | 1 |

Figure 3: Example of abridged Patient POKE-R report.

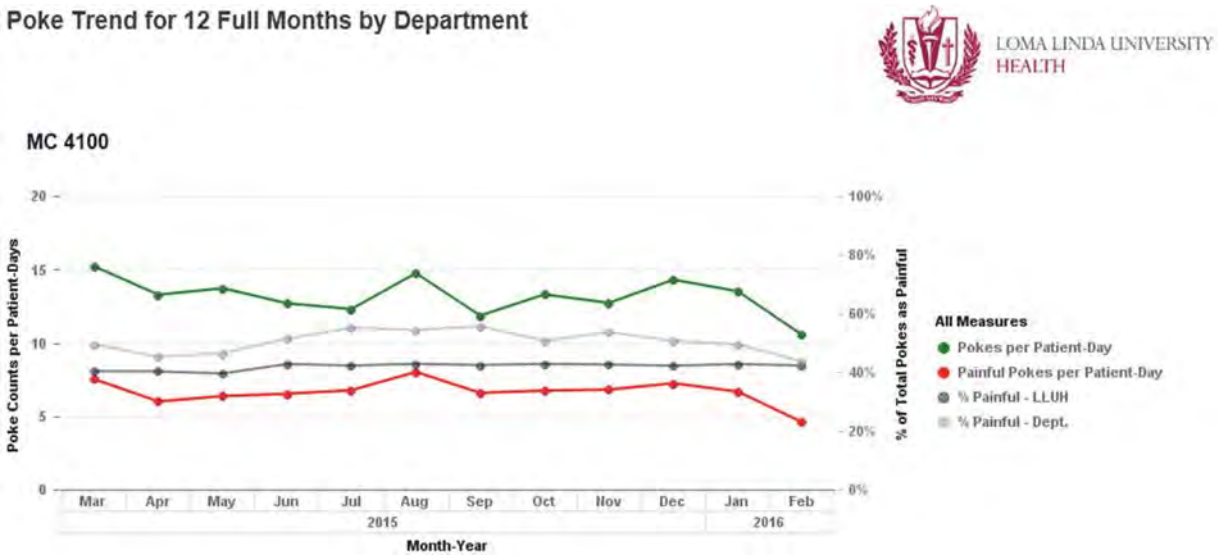



Figure 4: Example of POKE-R trend report.

# 4 PROCESS IMPLEMENTATION

To pilot the program, we chose a single department, the Pediatric Intensive Care Unit (PICU). We chose this department specifically because of the increased risk of complications in critical care pediatric patients including anemia and infection (Bateman et al, 2008) (Ridout, 2014) (Sengupta et al, 2010). The abridged report was scheduled to be automatically printed in the PICU at 5 am every morning so clinicians could (r to most POKEs being performed for the day. A resident fellow and a clinical nurse specialist were assigned specifically to manage the implementation of the program and received the daily detailed report. This allowed them to examine the most critical patients and suggest opportunities for POKE-R reduction.

At Loma Linda, the PICU uses structured interdisciplinary bedside rounds (SIBR). Under the SIBR methodology, all members of a patient's care team visit and communicate with the patient as a unit. Figure 5 shows the SIBR methodology. Because the SIBR methodology includes careful review of lab work, it provides a perfect opportunity to address potential POKEs. We have adjusted the SIBR methodology to include POKE-R.

The methodology includes:

1. Discuss and justify each care intervention.
2. Choose interventions that are:
    - Supported by evidence (consider pre-test probability)
    - Lead to change in treatment plan
    - Lowest cost
3. Considers cost in terms of financial burden and patient experience
    - Deliver best possible care, at the lowest cost to the healthcare system and the patient.
    - Reduction in patient harm, exposure, and pain are all considered.
4. Minimize ordering labs, instead perform a risk vs. benefit analysis for each test

Please see Figure 6 for an illustration of the SIBR POKE-R approach.

Additionally, three sets of patients were targeted as providing significant opportunity and actively managed using the daily report. These were patients with traumatic brain injury (TBI), patients with asthma and patients with external ventricular drain (EVD) placements. These patients are especially susceptible to infections and complications (Alharfi et al, 2008). Furthermore, asthma patients often experience an excessive number of lab tests in order to monitor the effects of medication on patient potassium levels (Schuh et al, 1989). Traumatic brain injury patients often experience sodium instability which requires monitoring (Atchinson et al, 1993). Therefore, these patients are likely to have a substantial number of POKE-R events and are particularly vulnerable to harm from these events.

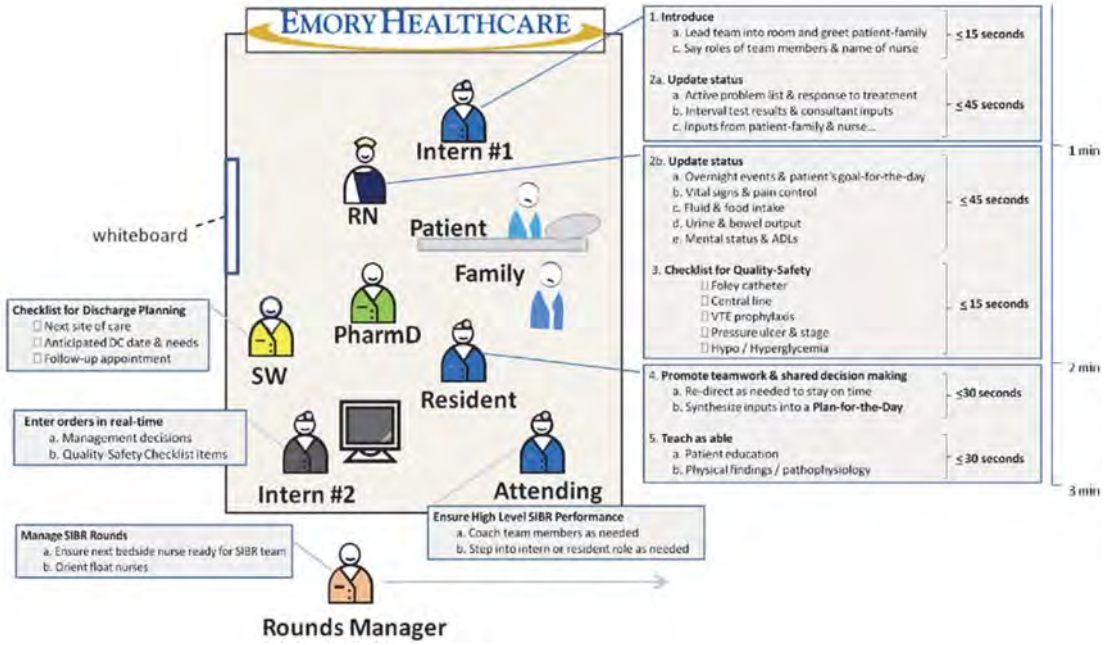



Figure 5: Structured interdisciplinary bedside rounds roles and process.


Figure 6: Order menu customized for POKE-R awareness.

## 5 RESULTS

Our project is in production in the PICU at Loma Linda University Children's Hospital. To analyze the success of the project, we compared patients prior to the introduction of both POKE-R and SIBR to patients after these programs were instituted. In all, we analysed 3,338 pediatric ICU patients.

We have seen a reduction in POKE-R events by 8.7%. Specifically, this was a decrease of 1.8 POKE-R events per patient per week. We compared this patient cohort with a historical control set using standard t-test methodology. This result was statistically significant with $p < 0.012$. Statistical significance for our purpose is defined as $p < 0.05$.

Furthermore, we saw significant reductions based on event type. Medication administrations were reduced by 1.2 administrations per patient per week. Specimen collections were reduced by 0.3 pokes. Radiology procedures and point of care tests were also reduced. We saw an increase in line pokes, but this was expected because inserting a line actually reduces the pokes for medication administrations and lab draws and therefore is not discouraged by the program. Finally, we saw a reduction in surgeries but we do not think this change was influenced by our POKE-R program.

To determine the statistical significance of the data we utilized t-tests and chi-square goodness of fit All of our results met the criteria for statistical significance except for painful poke counts and point-of-care testing. Table 1 shows the means, deltas and statistical significance from our analysis. The two rows marked in red are not considered statistically tests and one-way analysis of variance (ANOVA).

significant due to p>0.05. Figure 7 shows the the one-way analysis of variance (ANOVA) graph for all pokes, where stage 1 is prior to implementing our program and stage 2 is after program implantation. The graph has green diamonds on it which describe the 95% confidence intervals of the mean within the upper and lower peaks of the diamond. The line across the middle is the mean, and the width of the diamond is proportional to the sample size. Figure 8 shows the control chart for all pokes.

We also analyzed specifically the asthma patients and found a decrease in metabolic panels by 30% with p<0.0001. For patients with hypo/hypernatremia, the metabolic panel reduction was 16%. These results demonstrate drops in Basic Metabolic Panels performed during the post-implementation period. Figure 9 shows the ANOVA graph for asthma and cerebral salt wasting patients.

Table 1: Statistics for poke analysis.

| Statistical Variable | Mean Pokes /patient week | Δ Control Set | p (Prob > \|t\|) |
|---|---|---|---|
| All pokes | 14.372 | -1.76 | 0.003 |
| Painful | 2.638 | +0.10 | 0.472 |
| MAR | 10.57 | -1.19 | 0.007 |
| Specimen | 2.220 | -0.306 | .003 |
| POCT | 0.811 | -0.190 | .157 |
| BMP (Asthma and CSW) | 1.741 | -0.357 | .000 |

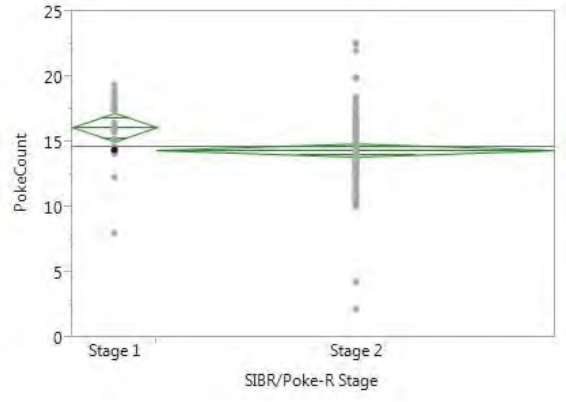



Figure 7: ANOVA graph for all pokes.

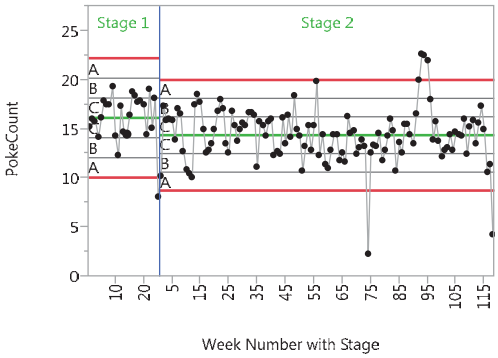



Figure 8: Control chart for all pokes.

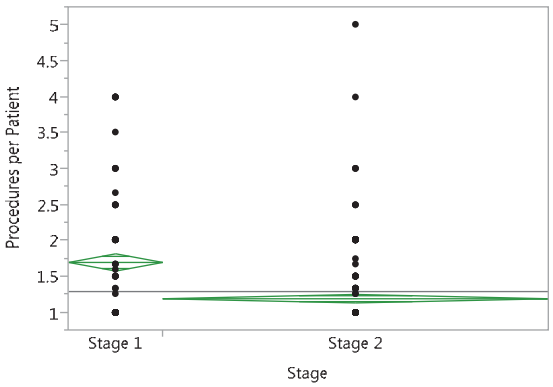



Figure 9: ANOVA graph for BMP for asthma and cerebral salt wasting patients.

## 6 FUTURE WORK

We would like to do more data analysis and research to quantify the benefits of the program including:

- Reduction in hospital-acquired infections such as Central Line-associated Bloodstream Infection (CLABSI)
- Reduction in cost
- Improved patient satisfaction

Additionally, patient satisfaction and cost information are not currently in our data warehouse. So, additional future work is to bring in these data points. This will not only allow us to more accurately monitor performance improvement, it will also enable greater cost transparency to the provider, patient and guarantor. We plan to integrate patient (and parent) experience survey data to quantify improved customer satisfaction and the cost benefits of our results to the hospital, the guarantor and the payor.

Currently, the POKE-R evaluation process is facilitated through printed and emailed reports even though the data is in the Enterprise Data Warehouse.

This is to eliminate the provider from having to access multiple systems. We plan to have the POKE-R details for a patient be directly linked to the patient's electronic health record. This way they can simply view the needed POKE-R information when they are already reviewing the patient's chart.

Currently, we have the number of laboratory tests but not the volume of specimen taken for the tests. Another enhancement we want to make is to interface with our laboratory system to get the precise volume of blood collected. This will give us more accuracy in measuring POKEs and associated risk for anemia.

We are also planning to roll out our POKE-R analytics and process to more departments throughout the hospital in the coming year.

# 7 CONCLUSION

We have implemented a comprehensive and configurable analytics solution to give providers the information they need to address excessive POKE-R events in patients. While our project has only gone into production in one hospital unit, we are already seeing considerable evidence of improvement. This project has the opportunity to reduce cost, improve patient outcomes and increase customer satisfaction.

# REFERENCES


Alharfi, I., Stewart T., Helali, I., Daoud, H., Fraser, D, 2014. Infection rates, fevers and associated factors in pediatric severe traumatic brain injury. In *Journal of Neurotrauma*.

Bateman, J., Lacroix, Boven, K., Forbes, P., Barton, et al, 2008. Anemia, Blood Loss, and Blood Transfusions in North American Children in the Intensive Care Unit". In *American Journal of Respiratory and Critical Care Medicine*.

Salisbury, A.C., Reid, K.J., Alexander, K.P., Masoudi, F.A., Lai, S.M., Chan, P.S., Bach, R.G., Wang, T.Y., Spertus, J.A. and Kosiborod, M., 2011. Diagnostic blood loss from phlebotomy and hospital-acquired anemia during acute myocardial infarction. *Archives of internal medicine*, 171(18), pp.1646-1653.

Foster, C.B. and Sabella, C., 2011. Health care–associated infections in children. *JAMA*, 305(14), pp.1480-1481.

CT Radiation in Kids: How Much of a Risk, Really? *Medscape*. 2014.

Grissinger, M., 2011. Capping intravenous tubing and disinfecting intravenous ports reduce risks of infection. In *Pharmacy and Therapeutics*.

McVoy, M., Shandler, A, 2013. Anemia, Bleeding, and Blood Transfusion in the Intensive Care Unit: Causes, Risks, Costs, and New Strategies. In *Am J Crit Care*.

Ridout R., 2014. Prevent Pain and Organisms from sKin and catheter Entry (POKE): Getting to the Why of Care [Internet]. Harvard Business School Open Forum: Dixie Regional Medical Center. Available from: https://openforum.hbs.org/challenge/hbs-hms-health-acceleration-challenge/innovations/prevent-pain-and-organisms-from-skin-and-catheter-entry-poke-getting-to-the-why-of-care.

Slovis, T., 2002. CT and Computed Radiography: The Pictures Are Great, But Is the Radiation Dose Greater Than Required?. In *American Journal of Roentgenology*.

Brenner D., 2002. Estimating cancer risks from pediatric CT: going from the qualitative to the quantitative. In *Pediatric Radiology*.

Axt-Adam, P., van der Wouden, P., van der Does, E., 1993. Influencing Behavior of Physicians Ordering Laboratory Tests: A Literature Study. In *Medical Care*.

Sengupta, A., Lehmann, C., Diener-West, M., Perl, T., Milstone, A., 2010. Catheter Duration and Risk of CLA-BSI in Neonates With PICCs. In *Pediatrics*.

Kelly, M., Conway, M., Wirth, K., Potter-Bynoe, G., Billett, A.L. and Sandora, T.J., 2011. Moving CLABSI prevention beyond the intensive care unit: risk factors in pediatric oncology patients. In *Infection Control & Hospital Epidemiology*, 32(11), pp.1079-1085.

Sherwood, G., Adams-McNeill, J., Starck, P., Nieto, B., Thompson, C.. 2000. Qualitative assessment of hospitalized patients' satisfaction with pain management. *Research in Nursing Health.*

Stein, J., Murphy, D., Payne, C., Clark, D., Bronstein, W., Tong, D., Castle, B., Shapiro, S, 2013. A Remedy for Fragmented Hospital Care. In *Harvard Business Review*.

Atchison, J., Wachendorfb, J., Haddockb, D., Mysiwb, S., Gribbleb, M., & Corriganb, J. 1993. Hyponatremia-associated cognitive impairment in traumatic brain injury. *Brain Injury*.

Algaze, C. et al., 2016. Use of a Checklist and Clinical Decision Support Tool Reduces Laboratory Use and Improves Cost. In *Pediatrics*.

Bae, J.. 2015. Value-based medicine: concepts and application. In *Epidemiology and Health.*.

Cooke, M., 2010, Cost Consciousness in Patient Care - What Is Medical Education's Responsibility?. In *New England Journal of Medicine.*

Porter, M., 2010. What Is Value in Health Care?. In *New England Journal of Medicine.*

Weinberger, S., 2011. Providing High-Value, Cost-Conscious Care: A Critical Seventh General Competency for Physicians. In *Annals of Internal Medicine.*

Schuh, S., Parkin, P., et al., 1989, High-Versus Low-Dose, Frequently Administered, Nebulized Albuterol in Children With Severe, Acute Asthma. In *Pediatrics*.

Brenner, D.J., 2010. Should we be concerned about the rapid increase in CT usage?. *Reviews on environmental health*, 25(1), pp.63-68.

# Augmenting Guideline-based CDSS with Experts' Knowledge

Nekane Larburu[1,2], Naiara Muro[1,2], Iván Macía[1,2], Eider Sánchez[3], Hui Wang[4], John Winder[5], Jacques Boaud[6,7] and Brigitte Séroussi[7,8]

[1]*eHeatlh and Biomedical Applications, Vicomtech-IK4, Donostia-San Sebastian, Spain*
[2]*Biodonostia, Donostia-San Sebastian, Spain*
[3]*NARU, Donostia-San Sebastian, Spain*
[4]*Computer Science Research Institute, Ulster University, Newtownabbey, U.K.*
[5]*Institute of Nursing and Health Research, Ulster University, Newtownabbey, U.K.*
[6]*AP-HP, DRCD, Paris, France*
[7]*Sorbonne Universités, UPMC Univ. Paris 06, INSERM, Université Paris 13, Sorbonne Paris Cité, UMR S 1142, LIMICS, Paris, France*
[8]*AP-HP, Hôpital Tenon, DSP, Paris, France*
*{nlarburu, nmuro, imacia}@vicomtech.org, e.sanchez@naruintelligence.com, {h.wang, rj.winder}@ulster.ac.uk, {jacques.boaud, brigitte.seroussi}@aphp.fr*

Keywords: Evidence Based Medicine, Breast Cancer, Computer Interpretable Clinical Guidelines, CDSS.

Abstract: Over the past years, clinical guidelines have increasingly become part of the clinical daily practice in order to provide best available Evidence-Based-Medicine services. Hence, their formalization as computer interpretable guidelines (CIG) and their implementation in clinical decision support systems (CDSSs) are emerging to support clinicians in their decision making process and potentially improve medical outcomes. However, guideline compliancy in the clinical daily practice is still "low". Some of the reasons for such low compliance rate are (i) lack of a complete guideline to cover special clinical cases (e.g. oncogeriatric cases), (ii) absence of parameters that current guidelines do not consider (e.g. lifestyle) and (iii) absence of up-to-date guidelines due to lengthy validation procedures. In this paper we present a novel method to build a CDSS that, besides integrating CIGs, stores experts' knowledge to enrich the CDSS and provide best support to clinicians. The knowledge includes new evidence collected over time by the systematic usage of CDSSs.

## 1 INTRODUCTION

In order to offer the best available care, medical practice adopts the Evidence-Based-Medicine (EBM) principle, defined as *"the conscientious, explicit and judicious use of current best evidence in making decisions about care of individual patients"* (Sackett et al., 1996). In the 90s clinical practice guidelines (CPGs) start to appear as rigorous evaluations of different clinical activities that improved the clinical practice and developed health care processes (Grimshaw and Russell, 1994), so that clinicians could follow EBM. However, clinicians still found barriers to adhere to CPGs (Cabana MD et al., 1999). Some of these barriers were lack of awareness, lack of familiarity, lack of agreement, lack of outcome expectancy or the inertia to previous practice. These barriers are still valid in the current practice.

In order to overcome some of the main obstacles, during the last decade multiple CPGs have been formalized in an electronic way, i.e. computer interpretable guidelines (CIG), and applied in Clinical Decision Support Systems (CDSSs) (B. Séroussi et al., 2013). Nevertheless, it was discovered that CPGs still have limitations. For instance, in the context of breast cancer (BC) some factors such as elderly patients, multifocal tumours, occurrence of micrometastasis on lymph-node and patient choice are causes of CPGs non-compliance (Chéreau et al., 2011; Landercasper et al., 2006; Lebeau et al., 2011; B. Séroussi et al., 2013).

In this paper we present a method to acquire expert knowledge in order to develop a knowledge-augmented guideline-based CDSS. It results in the development of new tools to support clinicians on their decision making process for cases that have low evidence (e.g. oncogeriatric cases) or where other aspects (e.g. patient preferences) are crucial.



Larburu, N., Muro, N., Macía, I., Sánchez, E., Wang, H., Winder, J., Boaud, J. and Séroussi, B.
Augmenting Guideline-based CDSS with Experts' Knowledge.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 370-376
ISBN: 978-989-758-213-4

The rest of this paper is organized as follows. Section 2 presents the state of the art on CIG compliance. Section 3 presents the method to augment the guideline-based CDSS with expert´s knowledge. Section 4 presents the application of such method in an EU project, DESIREE, developed in the context of breast cancer. Section 5 proposed a short discussion on the presented method and Section 6 concludes the paper and gives some future work lines.

# 2 STATE OF THE ART

## 2.1 Guideline Compliance

Variations in medical practices have been observed for decades, questioning the quality of care (Mercuri and Gafni, 2011). CPGs compliance is one of the primary performance measures to assess the quality of medical practice. McGlynn et al. (McGlynn et al., 2003) reported that 54.9% of the studied patient population received CPGs' based recommended care, which vary from 10% to 78%. In their work they reported that for BC 75.7% were consistent with recommended care, based on 9 quality indicators. Other studies also demonstrated suboptimal guideline compliance levels in BC (Adegboyega et al., 2015; Landercasper et al., 2006; Lebeau et al., 2011; Wöckel et al., 2010). The published levels of guideline compliance range from 12% (Lebeau et al., 2011) to 100% (Adegboyega et al., 2015), depending on the definition of guideline compliance and the level of abstraction of the guideline. For instance, Wöckel et al. (Wöckel et al., 2010) reported 80% of adherence to German-S3-BC guideline for surgery and for hormone therapy, and 71% for chemotherapy, indicating different compliance levels for the different components of the care plan. Similarly, Lebeau et al. (Lebeau et al., 2011) reported high level of guideline compliance, but also said that "management of non-metastatic BC was fully compliant (considering jointly 20 quality criteria)".

## 2.2 Causes Associated with Guideline Non-Compliance

The causes of variations in care delivery are multifactorial. A review by Flottorp et al. (Flottorp et al., 2013) identified a list of 51 determinants of practice grouped in seven domains: guideline factors, individual health professional factors, patient factors, professional interactions, incentives and resources, capacity for organisational change, and social, political, and legal factors.

However, effective guideline-based CDSSs (Beeler et al., 2014; Roshanov et al., 2013) provide a framework for logging non-compliance cases and learn from them. As demonstrated by Séroussi et al. (B. Séroussi et al., 2013), guideline compliance increases by using guideline-based CDSS. Additionally, Bouaud et al. (Bouaud and Séroussi, 2011) determined the main factors related with CPG non-compliance and reported the distribution of non-compliance causes. Here, we list these causes reported in (Bouaud and Séroussi, 2011):

- ***Patient preferences***: When patients receive more complete information about the benefits and risks of different treatment options, the patients made their own active, informed decisions (Leonard et al., 2011). This decision is influenced by their personal preferences.
- ***Evolution of medical knowledge***: CPG knowledge may not consider latest scientific publications and clinical essays, and hence, they may lag behind 'last' evidence (B. Séroussi et al., 2013). This may include that 'new' parameters are not being considered in the applied guidelines.
- ***Specific situations***: Rare situations that require specific clinical research are also a cause of non-compliance. For example, in BC scenarios shown in (Parks et al., 2012; Schnitt, 1998; B Séroussi et al., 2013), microinvasion, neadjuvant situations and oncogeriatry conditions are the main causes that lead into non-compliance situations.
- ***Medical choices***: One of the main cause of non-compliance is a medical decision that is prioritized over the guideline recommendation. For example, in (Bouaud and Séroussi, 2011), the study reported that BC multidisciplinary staff meetings' choice (i.e. breast units choice) is the main reason reported as the cause for CPG non-compliance.
- ***Others***: Finally, it may be other reasons that lead into CPG non-compliancy that do not belong to any of the previously reported causes.

Some studies provide tools to support clinicians in understanding the reasons of non-compliancy (Hussain et al., 2007). Others exploit the stored patient information to predict patient worsening and prevent potential emergencies (Colantonio et al., 2008). Yet, there is no evidence that all the information related to the whole decision making process (such as additional patient data, the decision criteria for giving a specific treatment and patient

outcomes) is stored and exploited over time to enrich the CDSS and provide better decision support to decision makers in prospective cases.

# 3 METHOD TO AUGMENT GUIDELINE-BASED CDSS

Here we present a method that enables the exploitation of the implicit knowledge used in a decision making process. The method is presented in the following subsections: Section 3.1 presents the starting point, which applies the clinical guideline model, Section 3.2 presents the second stage, which describes the acquisition process of experts' knowledge and Section 3.3 presents how such experts' knowledge is exploited.

## 3.1 Clinical Guideline Model

As discussed in Section 1, CPGs are intended to optimize patient care. Therefore, in this initial stage a clinical guideline model is developed. The clinical guideline model incorporates (i) different guidelines based on users' needs, (ii) updated clinical guidelines or studies, so that the provided recommendations correspond to the latest available evidence, and detects (iii) potential inconsistencies that could be reflected on the implemented guidelines.

## 3.2 Experts' Knowledge Acquisition

The second stage of this method focuses on experts' knowledge acquisition and storage.

We developed a flexible solution that enables the storage of each decisional event (Figure 1). Each **decisional event** reflects all the rationality for taking a decision and the consequences of such decision.

Hence, we define the decisional event as $\boldsymbol{D_n} = < \boldsymbol{P_i}, \boldsymbol{R_j}\ \boldsymbol{FD}, \boldsymbol{C_k}, \boldsymbol{E}, \boldsymbol{O(t)} >$, where (i) $\boldsymbol{P_i}$ is a set of patient *parameters* involved in the decision-making process, (ii) $\boldsymbol{R_j}$ is a set of clinical conditions (e.g. rules) wherein such parameters have been analysed, which results in a set of *recommendations* (iii) $\boldsymbol{FD}$ is a *final decision* that is taken by the decision maker, (iv) $\boldsymbol{C_k}$ is a set of *criteria* for which the final decision is made (which could be a patient parameter), (v) $\boldsymbol{E}$ is the executed treatment (usually, same as the final decision) and (vi) $\boldsymbol{O(t)}$ is the *health outcomes* of a patient measured over time t (e.g. ("ICHOM – International Consortium for Health Outcomes Measurement," n.d.)).

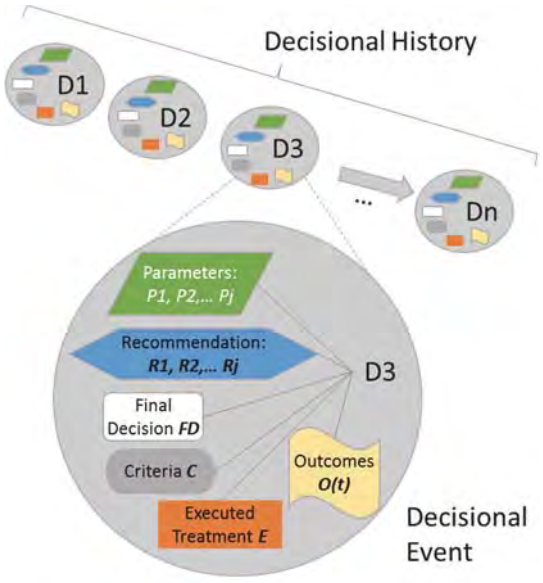



Figure 1: Decisional event and decisional history.

As shown in Figure 1, the storage of decisional events over time lead into a **decisional history**. The decisional history is later used to retrieve conclusions or discover new knowledge (Section 3.3).

## 3.3 Experts Knowledge Exploitation

Here we present the three usages of this decisional history: (i) recommendations assessment, (ii) patient similarity based recommendations and (iii) knowledge discovery to extend the knowledge base.

### 3.3.1 Recommendation Assessment

As presented by (Fox et al., 2009), "the current guideline development lifecycle does not provide appropriate tools to assess their impact on clinical practice". The proposed system is able to evaluate the decisions taken quantitatively (e.g. based on the number of times the recommendation was followed) and qualitatively (e.g. based on the patient outcomes – when the results are successful or match the defined decision criteria). This quantitative and qualitative measurements are presented to clinicians during the clinical decision making process to provide enriched information of the given recommendations.

### 3.3.2 Patients' Similarities

The system also applies similarity features between different patients and their results to support clinicians in their decision making process. For that, the system uses different metrics to determine which (clinical) parameters have higher impact when

determining how similar a patient could be to a retrospective patient (e.g. age range, TNM classification etc.). In cases where the benefits and harms of a specific treatment are not clear, clinicians are able to consult previous similar patient cases and their outcome before taking a decision. The previous patient cases could be specific patient cases, or 'model' cases that summarise *n* past cases.

### 3.3.3 Knowledge Discovery

The experience acquired from the decisional history may enable different type of knowledge acquisition. Here we present the two types of knowledge considered in our research.

Firstly, the information from a large number of cases enables the **adjustment** of CPGs and protocols´ clinical conditions, e.g. in a form of a rule. For that, the scope of the criteria is redefined based on the cases where the given guideline-based recommendation is being followed with successful results. This is implemented using machine learning techniques. For example, if a decision criterion is parameter $a \in [0.5, 1.5]$, after applying machine learning techniques the system recognizes that the recommendation is being successful only when $a \in [0.8, 1.3]$. It also detects when a parameter, not previously included into the clinical condition for the decision making process, is determinant and should be part of the existing decision rule.

Secondly, large number of non-compliant cases with good or better results than the ones that follows the CPGs may lead into an **extension** of the CPGs´ clinical conditions (e.g. rules) by generating 'new' branches. This 'new' branches may include recommendations (treatment actions) that are not considered in the available CPGs (e.g. "clinical trial") or may include recommendations that are in the CPGs, but that are not considered for the given case.

This cases could make the knowledge base either more restrictive when a rule becomes more precise, but also could extend it with further procedures that were not included in the knowledge base.

In both adjustment and extension cases, in order to include the 'new' knowledge into the CDSS, the system verifies if the outcomes are positives and informs clinicians about its potential usage. If approved, this knowledge is included into the knowledge base for the CDSS (Figure 2). Nevertheless, the system provides the information of the recommendation source. This way clinicians are aware if the recommendation is guideline based or created automatically by the system based on the recorded experience or patient similarity properties.

# 4 DESIGN IN DESIREE

This study is being performed in the context of a European H2020 project, named DESIREE. In this section we present DESIREE project (Section 4.1) and the data flow diagram that represents our methodology within DESIREE (Section 4.2).

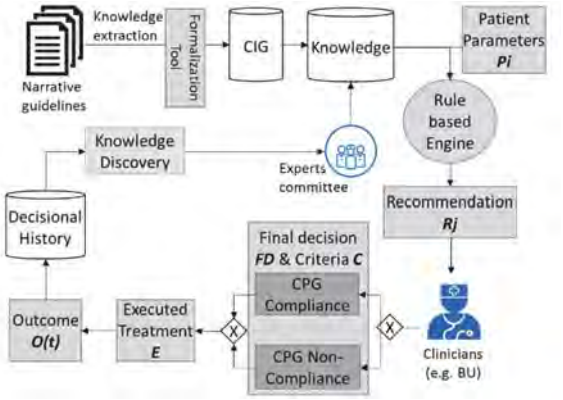



Figure 2: Data flow diagram.

## 4.1 DESIREE

DESIREE aims to provide decision support on the available therapy options by incorporating evidence based guidelines and experience from previous cases and outcomes. Hence, DESIREE goes beyond the limitations of existing guideline-based decision support systems. Such a system targets breast cancer (BC) cases, which is one of the most common and most deadly type of cancer affecting woman in the EU countries, with more than 460,000 new cases and 130,000 deaths in 2012 (Ferlay et al., 2013).

The users of such system are medical domain experts involved on breast units (BU) where patients' diagnosis and treatment decisions are taken. Hence, the system goal is to support BU during their weekly meetings in their multidisciplinary decision making process by providing not only CPGs based decision support, but also additional information extracted from previous cases over time.

## 4.2 DESIREE

The data flow diagram presented in Figure 2 is a high level representation of DESIREE platform. Since DESIREE is developed in the context of BC, in the depicted figure, Breast Units (BU) are the clinical experts that make the final decision. Here, we describe each block presented in Figure 2, omitting the blocks that correspond to the data presented in Section 3.2.

- ***Narrative Guidelines:*** In our methodological approach, the starting point is the analysis of representative and narrative CPGs used in BC care.
- ***Computer Interpretable Guidelines:*** Knowledge engineers extract the relevant information from CPGs and formalize it in a CIG. This covers the recommendations given by guidelines for primary BC in several stages of the whole treatment till the patient is discharged. Hence, the CIG consider the previous treatments and the outcomes of them for the coming decision making action.
- ***Knowledge:*** The knowledge database stores knowledge from the CIG or from the decisional history exploitation´s "new" knowledge.
- ***Rule-based Engine:*** The rule-based engine is able to generate recommendations having as input the structured knowledge. Then, if patient data fulfils the clinical condition, the rules are fired and the engine generates one or more recommendations.
- ***Knowledge Discovery:*** Based on a large set of information stored in the decisional history, the system is capable of retrieving knowledge as discussed in Section 3.3.

# 5 DISCUSSION

The presented method overcomes some limitations of current guideline-based CDSS by providing enriched recommendations and additional information to clinicians in order to support them best in their decision making process. For that, we develop a new information structure based on decisional events. A decisional event stores the whole set of information used in the decision making process, including the consequences of the final decision, such as patient´s outcomes (e.g. quality of life).

Here we present some of the potential benefits and limitations of the proposed method. Firstly, the system promotes the usage of CPGs. Additionally, it assess the impact of the guidelines on clinical practice (Section 3.3.1), which is one of the critical factors detected by (Fox et al., 2009). Secondly, the system flexibility enables the storage of additional valuable information, such as the decision criteria, that could be used to adjust or/and extend the clinical conditions of the given protocols and CPGs over time (Section 3.3.3). This way it helps overcoming some of the limitations of current CPGs presented in Section 2, such as the impact of specific situations. This diverges from the work done in other projects, such as MobiGuide (Larburu et al., 2015), where the guidelines are customized and made context-aware beforehand during the knowledge engineering phase, and not over time depending on previous cases. Neither we focus on the discovery of temporal rules from time-stamped data, like in (Sacchi et al., 2007). Our study aims to discover rules from previous cases tracking each case to assess the outcomes and considering further information often not taken into account in current CPGs, such as the implicit knowledge of clinicians. Finally, the presented method combines both the CPGs and the knowledge generated automatically by the system based on their experience, which overcomes the requirements expressed by clinicians in (Miranda-Mena et al., 2006): "clinicians want a system that combines the protocol (or CPG) and their proper knowledge to suggest treatments".

# 6 CONCLUSIONS & FUTURE WORK

The hypothesis of this research is that such approach is more useful for clinicians, which expect a dynamic system that not only considers available CPGs and protocols, but also a system that is able to learn from the stored information over time to provide enriched decision support system.

In future work we aim to present among others the following points: (i) the tools used to convert the information acquired by experience into knowledge to extend and adjust the CPGs and protocols; (ii) a digital patient model ontology used for the CDSS, and particularly for similarity purposes; (iii) the methodology to assess the recommendations applying different metrics (survival rate, overall well-being, physical functioning etc.); and (iv) the validation of the system in a representative number of patients and the results.

# ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 690238.

# REFERENCES

Adegboyega, T.O., Landercasper, J., Linebarger, J.H., Johnson, J.M., Andersen, J.J., Dietrich, L.L., Driscoll,

C.D., Raghavendra, M., Madadi, A.R., Al-Hamadani, M., Vang, C.A., Marcou, K.A., Hudak, J., Go, R.S., 2015. Institutional review of compliance with NCCN guidelines for breast cancer: lessons learned from real-time multidimensional synoptic reporting. J. Natl. Compr. Cancer Netw. JNCCN 13, 177–183.

Beeler, P.E., Bates, D.W., Hug, B.L., 2014. Clinical decision support systems. Swiss Med. Wkly. 144, w14073.

Bouaud, J., Séroussi, B., 2011. Revisiting the EBM decision model to formalize non-compliance with computerized CPGs: results in the management of breast cancer with OncoDoc2. AMIA. Annu. Symp. Proc. 2011, 125–134.

Cabana MD, Rand CS, Powe NR, et al, 1999. Why don't physicians follow clinical practice guidelines?: A framework for improvement. JAMA 282, 1458–1465.

Chéreau, E., Coutant, C., Gligorov, J., Lesieur, B., Antoine, M., Daraï, E., Uzan, S., Rouzier, R., 2011. Discordance with local guidelines for adjuvant chemotherapy in breast cancer: reasons and effect on survival. Clin. Breast Cancer 11, 46–51.

Colantonio, S., Conforti, D., Martinelli, M., Moroni, D., Perticone, F., Salvetti, O., Sciacqua, A., 2008. An intelligent and integrated platform for supporting the management of chronic heart failure patients, in: 2008 Computers in Cardiology. Presented at the 2008 Computers in Cardiology, pp. 897–900.

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J.W.W., Comber, H., Forman, D., Bray, F., 2013. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012. Eur. J. Cancer 49, 1374–1403.

Flottorp, S.A., Oxman, A.D., Krause, J., Musila, N.R., Wensing, M., Godycki-Cwirko, M., Baker, R., Eccles, M.P., 2013. A checklist for identifying determinants of practice: a systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice. Implement. Sci. IS 8, 35.

Fox, J., Patkar, V., Chronakis, I., Begent, R., 2009. From practice guidelines to clinical decision support: closing the loop. J. R. Soc. Med. 102, 464–473.

Grimshaw, J.M., Russell, I.T., 1994. Achieving health gain through clinical guidelines II: Ensuring guidelines change medical practice. Qual. Health Care 3, 45–52.

Hussain, S., Abidi, S.R., Abidi, S.S.R., 2007. Semantic Web Framework for Knowledge-Centric Clinical Decision Support Systems, in: Artificial Intelligence in Medicine. Presented at the Conference on Artificial Intelligence in Medicine in Europe, Springer, Berlin, Heidelberg, pp. 451–455.

ICHOM – International Consortium for Health Outcomes Measurement [WWW Document], n.d. URL http://www.ichom.org/ (accessed 1.3.17).

Landercasper, J., Dietrich, L.L., Johnson, J.M., 2006. A Breast Center review of compliance with National Comprehensive Cancer Network Breast Cancer guidelines. Am. J. Surg. 192, 525–527.

Larburu, N., Schooten, B. van, Shalom, E., Fung, N., Sinderen, M. van, Hermens, H., Jones, V., 2015. A Quality-of-Data Aware Mobile Decision Support System for Patients with Chronic Illnesses, in Knowledge Representation for Health Care, Lecture Notes in Computer Science. Springer International Publishing, pp. 126–139.

Lebeau, M., Mathoulin-Pélissier, S., Bellera, C., Tunon-de-Lara, C., Daban, A., Lipinski, F., Jaubert, D., Ingrand, P., Migeot, V., REPERES Group, 2011. Breast cancer care compared with clinical Guidelines: an observational study in France. BMC Public Health 11, 45.

Leonard, R., Ballinger, R., Cameron, D., Ellis, P., Fallowfield, L., Gosney, M., Johnson, L., Kilburn, L.S., Makris, A., Mansi, J., Reed, M., Ring, A., Robinson, A., Simmonds, P., Thomas, G., Bliss, J.M., 2011. Adjuvant chemotherapy in older women (ACTION) study - what did we learn from the pilot phase? Br. J. Cancer 105, 1260–1266.

McGlynn, E.A., Asch, S.M., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., Kerr, E.A., 2003. The Quality of Health Care Delivered to Adults in the United States. N. Engl. J. Med. 348, 2635–2645.

Mercuri, M., Gafni, A., 2011. Medical practice variations: what the literature tells us (or does not) about what are warranted and unwarranted variations. J. Eval. Clin. Pract. 17, 671–677.

Miranda-Mena, T.G., Benítez U., S.L., Ochoa, J.L., Martínez-Béjar, R., Fernández-Breis, J.T., Salinas, J., 2006. A knowledge-based approach to assign breast cancer treatments in oncology units. Expert Syst. Appl. 31, 451–457.

Parks, R.M., Lakshmanan, R., Winterbottom, L., AL Morgan, D., Cox, K., Cheung, K.-L., 2012. Comprehensive geriatric assessment for older women with early breast cancer – a systematic review of literature. World J. Surg. Oncol. 10, 88.

Roshanov, P.S., Fernandes, N., Wilczynski, J.M., Hemens, B.J., You, J.J., Handler, S.M., Nieuwlaat, R., Souza, N.M., Beyene, J., Van Spall, H.G.C., Garg, A.X., Haynes, R.B., 2013. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ 346, f657.

Sacchi, L., Larizza, C., Combi, C., Bellazzi, R., 2007. Data mining with Temporal Abstractions: learning rules from time series. Data Min. Knowl. Discov. 15, 217–247.

Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B., Richardson, W.S., 1996. Evidence based medicine: what it is and what it isn't. BMJ 312, 71–72.

Schnitt, S.J., 1998. Microinvasive carcinoma of the breast: a diagnosis in search of a definition. Adv. Anat. Pathol. 5, 367–372.

Séroussi, B., Laouénan, C., Gligorov, J., Uzan, S., Mentré, F., Bouaud, J., 2013. Which breast cancer decisions remain non-compliant with guidelines despite the use of computerised decision support? Br. J. Cancer 109, 1147–1156.

Séroussi, B., Laouénan, C., Gligorov, J., Uzan, S., Mentré, F., Bouaud, J., 2013. Which breast cancer decisions remain non-compliant with guidelines despite the use of computerised decision support? Br. J. Cancer 109, 1147–1156.

Wöckel, A., Kurzeder, C., Geyer, V., Novasphenny, I., Wolters, R., Wischnewsky, M., Kreienberg, R., Varga, D., 2010. Effects of guideline adherence in primary breast cancer--a 5-year multi-center cohort study of 3976 patients. Breast Edinb. Scotl. 19, 120–127.

# Home Monitoring in Portugal
## *An Overview on Current Experiences*

Ana Dias[1], Madalena Vilas Boas[1], Daniel Polónia[1], Alexandra Queirós[2] and Nelson Pacheco Rocha[3]
[1]*Department of Economics, Management, Industrial Engineering and Tourism, University of Aveiro, Campo Universitário de Santiago, 3810-193 Aveiro, Portugal*
[2]*Health Sciences School, University of Aveiro, Campo Universitário de Santiago, 3810-193 Aveiro, Portugal*
[3]*Department of Medical Sciences, University of Aveiro, Campo Universitário de Santiago, 3810-193 Aveiro, Portugal*
*{anadias, mvb, dpolonia, alexandra, npr}@ua.pt*

Keywords: Telemedicine, Teleconsultation, Home Monitoring, Portuguese National Health Service.

Abstract: This paper aims to be a contribution to the discussion on the issue of innovation in healthcare since, in the author's perspective, the health sector, and particularly the Portuguese National Health Service, needs changes in its "business model". There is a need of redirecting care provision to the citizen's natural environment, namely considering the opportunities offered by information and communication technologies. For this purpose the authors surveyed projects already implemented in Portugal, within the Portuguese National Health Service, related to home monitoring, in order to make a critical analysis of the state of the art of ongoing projects. In this study, the authors identified four pilot experiences of home monitoring, all targeted at chronic disease. In spite of some results of these experiments are already known, there is a shortage of available information and scientific evidence, both about the implementation processes themselves and about their clinical, technical and economic evaluation, which, in the opinion of the authors, also hinders their assessment and dissemination.

## 1 INTRODUCTION

The Portuguese National Health Service (Serviço Nacional de Saúde - SNS) presents difficulties in the coordination between levels of care, particularly among primary healthcare and hospital care, which has an impact on patient access to healthcare (Barros and Simões, 2007).

Traditionally, acute diseases were the main concern of health systems, a situation that has changed over the last century, as a result of advances in biomedicine and public health, with a significant impact on the eradication of certain infectious diseases. What is questioned today is whether, with regard to the organizational model of health systems, they fit the current reality where chronic diseases are predominant (Dias, 2015).

Data from the National Health Survey (INS, 2014) are indicative of the new challenges of the Portuguese SNS. In 2014, more than half (52.8%) of the population aged 18 years, was overweight (50.9% a decade ago). The symptoms of depression also worsened, affecting more the retired population (36.5% of the retired population had symptoms of depression, compared to 18,5% of the employed population). Also the percentage of people who reported consuming prescribed drugs increase sharply with age: more than 90% of the population over 65 years. Comparing the results for chronic diseases collected in two surveys (2005-2006 and 2014) it is clearly an increase of the percentage of population affected by these diseases (INS, 2014).

The aging process that the Portuguese population is suffering further enhances this scenario. The National Statistics Institute (Instituto Nacional de Estatística - INE) forecasts that the potential sustainability index (i.e. the ratio between the number of people aged between 15 and 64 and the number of people aged 65 and over) may decrease abruptly: in Portugal, between 2012 and 2060, this index, in one of the most likely scenarios can change from 340 people in working-age for every 100 elderly to 149 people in working-age for every 100 elderly, a value that can decrease to 111 people in working-age for every 100 elderly in the worse-case scenario (INE, 2014a).

We are at a very particular moment of our history in which a "demographic transition" is



Dias, A., Boas, M., Polónia, D., Queirós, A. and Rocha, N.
Home Monitoring in Portugal - An Overview on Current Experiences.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 377-382
ISBN: 978-989-758-213-4

combined with “an epidemiological transition" and "this combination of phenomena is confronting us with a crisis (…) which comes as an opportunity to look critically at what has been done and, based on that, project us in the future with more appropriate tools and skills to new circumstances." (OPSS, 2016, p. 31).

This study thus seeks to contribute to the discussion on the theme of innovation in healthcare in Portugal, taking advantage of the opportunities offered by information and communication technologies. In particular, given the home monitoring potential, which is supported on the progress achieved in mobile technologies, as well as its relevance to chronic diseases, this study aims to analyze the viability of experiences already implemented in Portugal related to home monitoring of patients with chronic diseases.

## 2 RELATED WORK

In Portugal, via the 3571/2013 Order, published in the Official Gazette on March the 6th, 2013, the Ministry of Health, assuming that the use of telemedicine allows the observation, diagnosis, treatment and monitoring in a more convenient place for patients, particularly at home, states that "the services and facilities of the National Health Service (SNS) should increase the use of information and communication technologies in order to promote and ensure the provision of telemedicine services to [its] users" (p.8326).

According to INE, on its Survey on the Use of Information and Communication Technologies in Hospitals (INE, 2014b), one third of the Portuguese hospitals developed telemedicine activities in 2014, an increase of 16 percentage points in ten years (12 percentage points in the last four years) (Figure 1). However, the degree of implementation of telemedicine is quite different in public hospitals (51%) and private hospitals (15%).

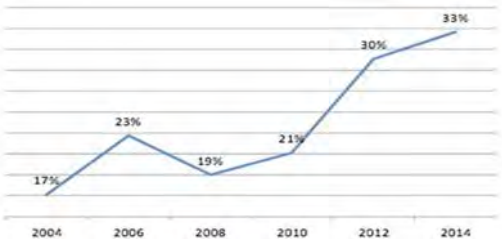



Figure 1: Proportion of hospitals with telemedicine, Portugal, 2004-2014 (Source: INE, 2014b).

Telemedicine activity can take many forms, ranging from remote diagnosis (teleradiology and telepathology) to remote care provision, such as teleconsultation or home monitoring. Within telemedicine activities, the most used was teleradiology (i.e. the exchange of images to discuss cases and for diagnosis), being reported by 84% of hospitals that refer having telemedicine. On the other hand, only 31% of hospitals that report having telemedicine provide teleconsultation (i.e. 10% of the total hospital’s number) (INE, 2014b).

As a result of the telemedicine development, Portugal is involved in several international projects to promote the cooperation between healthcare professionals. In particular, there are some projects with the African Countries with Portuguese as an Official Language (Países Africanos de Língua Oficial Portuguesa - PALOP) (Borja-Santos, 2013). In October 2013 it was launched a telemedicine platform between Portugal and several PALOP (Noronha, 2013) but, previously, there were other projects. For example, in 2012 a project between Portugal and São Tomé and Príncipe allowed an estimated saving of 180,000 euros in the transfer of patients to Portugal and allowed to save one million euros to the Portuguese Ministry of Health (Noronha, 2013). Portugal is also part of a telemedicine network with Angola and the University Hospitals of Geneva that allows technical support for diagnosis and treatment of Angolan patients.

In addition to the experiences within the PALOP, there are others being carried out between Portugal and Spain. Since 2003 the southern region of the Algarve has participated in a telemedicine project in conjunction with the Spanish region of Andalusia, with the aim of creating new communication channels between the Algarve and Andalusia and, inside the Algarve, between the health centers and the hospitals in the region, with the installation of telemedicine equipment in all health centers (Portal da Saúde, 2005).

With regard to teleconsultation, in 2007, it was launched in Portugal the “Linha de Saúde 24”, which provides counselling and referral in a disease situation, accessible through the phone (or chat for people with special needs) as well as therapeutic counselling to clarify particular questions and provide support related to matters as medication (Saúde 24, s.d.).

Within a group of other innovative projects in this area, out of the governmental sphere, the authors highlight two: the "Patient Innovation", a social network for patients who, sharing experiences about

their illness, can develop solutions to their real problems, from therapeutic support to proper medical equipment (Pinho, 2013) and a private service that allows traveler teleconsultations (Consulta do Viajante, 2016).

Specifically, in the region of Alentejo (one of the most aged and sparsely populated regions of the country) several teleconsultation experiments (Oliveira et al., 2014) were reported, dating back to 1998, in order to respond to challenges such as desertification, isolation, low population density, poverty, lack of medical resources in several specialties as well as poor public transports, all of which have acted as barriers to access to healthcare in the region. It should be noted that the Alentejo region represents approximately one-third of Portugal's continental territory, but it is home to only 5% of its population. Teleconsultations are available in 15 medical specialties, ranging from Neurology to Pediatric Surgery. The network includes 20 primary care units and five hospitals, covering almost 30,000 $km^2$ and around 500,000 people. A comprehensive assessment of the costs and consequences of the program is currently underway, since it is stated that there is a lack of evidence of its cost-effectiveness, which, according to Oliveira et al. (2014), hinders the sustainment and realization of the promise of innovative solutions, wherever it is implemented.

Regarding the autonomous islands, the Azores already belongs to several networks, which allows the realization of teleconsultation to determine the clinical need of the patient's displacement to mainland to carry out consultations and exams (Mourato, 2014). In the Azores the use of teleconsultation in Nursing is also frequent, particularly in decision support in the treatment of wounds. Furthermore, there is already teleconsultation in various health centers in the archipelago in the following specialties: Nephrology, Pediatric Cardiology, Neonatology and Endocrinology.

Home monitoring can improve disease prevention, facilitate chronic disease management, including disease self-management, enable personalization of care, and improve productivity in healthcare, thus allowing a more rational use of health services (Queirós et al., 2013; Queirós et al., 2017). In Portugal, according to the Order 8445/2014 of June the 30th, 2014, the Ministry of Health stressed the need to improve the capacity of health monitoring, prevention, detection and treatment of disease in innovative ways, including through models of care in order to maintain people in their homes, promoting their autonomy and encouraging personal responsibility by adopting healthy lifestyles. However, there are no studies reporting the current experiences of home monitoring in Portugal.

# 3 METHODS

Considering the lack of evidence of current Portuguese home monitoring experiences, the present study has the following main objectives:

- To make an inventory of projects already implemented in Portugal, within the SNS in the area of home monitoring, particularly focused on the chronically ill, and preferably projects involving primary healthcare, which is believed that will assume an increasingly central role in the management of chronic disease.
- To make a critical analysis of the state of the art of ongoing projects.

The authors consulted the Central Administration of the Health System (Administração Central do Sistema de Saúde - ACSS) in order to retrieve information on projects related to home monitoring already implemented by the SNS.

Subsequently, an additional survey was conducted to analyze if there were publications that best described these experiences and others in the same area, and possible results already obtained.

Finally, the authors conducted a survey on the web pages of SNS hospitals to search for innovative projects in general, and home monitoring in particular. For this purpose, the web pages of 41 hospitals were identified and analyzed.

# 4 HOME MONITORING EXPERIENCES IN PORTUGAL

With data provided by the ACSS it was possible to identify three projects related to home monitoring already implemented by the Portuguese SNS. In the area of Pulmonology, a home monitoring pilot program of chronic obstructive pulmonary disease was developed, funded by the ACSS and supported by Shared Services of the Health Ministry (Serviços Partilhados do Ministério da Saúde - SPMS), the government agency for eHealth, in partnership with five hospitals, covering a total of 75 patients with severe disease (15 per hospital). These patients were selected in each hospital by their attending

physician, based on their prior history of chronic obstructive pulmonary disease. This home monitoring pilot program began in 2013 and was implemented in a phased manner in the five institutions, namely: Hospital of Faro (Algarve) - beginning in August 2013 (Phase 1); Hospital Pêro da Covilhã (Cova da Beira) - beginning in March 2014; Hospital and University Center of Coimbra - beginning in May 2014; Hospital of Viana do Castelo (Alto Minho) - beginning in October 2014 and Hospital of Portalegre/Elvas (North Alentejo) - beginning in October 2014. Five private companies were also involved in the project and were responsible for the installation of the monitoring devices and their maintenance and for the process of gathering information and transfer it to health professionals. Each patient was assigned the following monitoring devices: blood pressure measuring device, pulse oximeter, thermometer, odometer, device monitoring heart rate and mobile phone. The clinical teams of the hospitals were actively involved in the monitoring of patients integrated in the program and also in their education. According to SPMS (2014), under this pilot program, patients are monitored in their homes. The respective data are then analyzed twice a day by the Pulmonology teams of involved hospitals, trying to reduce the aggravation of their clinical situation, thus avoiding new hospital admissions.

The objectives and results to be achieved in 2016 within the home monitoring program for chronic obstructive pulmonary disease include (ACSS, 2016): raising the quality of services provided to citizens, promoting the continuous monitoring of their health condition; reducing at least one episode of annual hospitalizations as a result of the deterioration of the patients medical condition; reducing at least three episodes of urgency per year; reducing at least two episodes of outpatient consultation per year, and follow, proactively and continuously, the fluctuations of the health conditions of each patient.

In terms of preliminary results of this pilot program, according to Pereira (2016), these have been encouraging, both in terms of satisfaction and health indicators, having already been released some results of an evaluation carried out in the Local Unit of the Alto Minho, namely the reduction in 50% of visits of these patients to emergency services, as well as a decrease in the number of hospitalizations (70%). Pereira (2016) also states that in these five hospitals, 61% of patients considered the quality of the service as "very good or excellent". Although the final overall results are not yet available due to the late start in one of the hospitals, interim evaluations reveal both a reduction in the number of hospital admissions or visits to the urgency services of patients, more evident in some hospitals.

More recently, in November 2015, the "Home Monitoring Plan" was adopted for the definition of sites for the realization of home monitoring and its articulation with the rules of the SNS.

In this context, and for the year 2016, it was contracted activity for the implementation of other two home monitoring pilot programs: a Pilot Program for Home Monitoring of Acute Myocardial Infarction and a Pilot Program for Chronic Heart Failure. Like the pilot program of chronic obstructive pulmonary disease, it was planned that the program would be implemented in five hospitals covering a group of 75 patients (15 per hospital).

In the case of the chronic heart failure, and in the absence of results which can be explained due to the program's newness, it should be noted that from conventional remote monitoring to more recent strategies, using cardiac devices or implantable hemodynamic monitors, this is a topic under active investigation, but, despite previous meta-analyses of small studies have documented the potential benefit of home monitoring, major randomized clinical trials have failed to demonstrate the positive impact of this strategy. In addition, data on the value of the latest monitoring devices are contradictory, since some studies have documented potential prognosis benefit while others cannot confirm it (Sousa et al., 2014).

As a result of the literature review carried out in scientific databases, the authors found a concrete example in Portugal of assessment of a home monitoring experience in cardiac patients, with four hospitals involved. This study, of 2013, indicates that the introduction of home monitoring has the ability to reduce in 25% the costs of monitoring the patients (Costa et al., 2013).

To broaden the scope of this research, and to identify innovative projects, in general, and home monitoring, in particular, the authors decided to search for information on the web pages of the SNS hospitals. From the 41 organizations identified six refer teleconsultation activities on the following areas: Dermatology (referred by three organizations), Pediatric Cardiology (two organizations), Internal Medicine, Endocrinology, Rheumatology, Oncology, Neurosurgery, Pediatrics, Gynecology, Ophthalmology, Genetics, Imaging and Pathology (all referred only by one organization).

However, no organization mentions any home monitoring experiment.

# 5 DISCUSSION

The main purpose of this study was to give visibility to the experiences already implemented in Portugal related to home monitoring of patients with chronic conditions, which the authors believe could permit not to repeat known errors as well as to replicate successes, after being properly evaluated and contextualized. However, it is clear that, although some results of these experiments have been reported in this paper, there is a shortage of scientific evidence regarding, on one hand, the implementation process and, on the other hand, the evaluation of these experiences, what is conflicting to what is defended in the eHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century (Commission of the European Union, 2012), namely: "It is essential to measure and assess the added value of innovative eHealth products and services to achieve wider evidence-based eHealth deployment and create a competitive environment for eHealth solutions." (p.13)

Regarding the home monitoring pilot projects identified in this work the authors would like to stress that the pathologies covered by the pilot projects correspond to chronic diseases and that they are also within the group of priority areas identified by the Portuguese government in 2013.

Concerning these same projects the authors also want to highlight the fact that in the group of institutions covered by these projects, there are no institutions coming from primary healthcare, at least not in an explicit and formal way. As mentioned before, the authors consider that the primary healthcare services must increasingly be involved because of their close proximity to the patient and their informal careers as well as due to their abilities in the management of chronic disease, even because one of the main purposes of home monitoring is to reduce the number of admissions or visits to the hospital emergency services. It is important to emphasize the significance of primary healthcare in the organization of health systems, recalling what in 1978 was stated in the Declaration of Alma Ata (WHO, 1978) on the importance to be given to this level of care: "Primary healthcare is essential healthcare based on practical, scientifically sound and socially acceptable methods and technology made universally accessible to individuals and families in the community through their full participation and at a cost that the community and country can afford (…) It is the first level of contact of individuals, the family and community with the national health system bringing healthcare as close as possible to where people live and work, and constitutes the first element of a continuing healthcare process." (p.1-2)

Concerning the results obtained from the research on the hospital's web pages, the authors would like to highlight the scarce information available with regard to innovative experiences, as well as the complete lack of publicizing information about home monitoring. This, from the author's point of view, can raise questions in particular regarding the opportunities created for patients to participate in these experiences and create obstacles to ensure the equity required in healthcare provision. Despite the fact that, as it has also been demonstrated in this work, home monitoring experiences in Portugal are still small, in number and in size, still the authors would like to discuss the importance of what is (or not) revealed to the public: whether information regarding the experiences as a whole whether information on the criteria for inclusion of patients in these experiments. Another purpose of this advertising is, from the author's point of view, to emphasize the need to make the whole process more transparent, with particular interest to patients and also to other health professionals and institutions, a similar progress that what has been achieved with respect to clinical trials since 2011. Since then, the information on clinical trials with medicines for human use, which are underway in the European Union, is accessible to all European citizens from the portal "EU clinical trials Register" and, more recently, through the international network of clinical trials registers of the World Health Organization. In Portugal, Law 21/2014 of the 16th April, amended by Law 73/2015 of the 27th July, envisages the creation of the National register of clinical trials. Still, and going back to experiences that are developed within the SNS in Portugal, and highlighting once again the difficulties, specifically in this study, on the collection of information on ongoing initiatives, the authors would like to discuss the pertinence of creating a platform for registering, monitoring and disseminating results of these experiences, similar to what is already being done in the clinical trials domain.

# 6 CONCLUSION

In Portugal, within the SNS, the authors identified four pilot experiences of home monitoring, all targeted at chronic disease, but with no direct involvement of primary healthcare, at least explicit

in contracts that were made between the ACSS and the primary healthcare services. This somehow contradicts the need to direct primary care for the prevention, with a view to achieve further gains in health outcomes as well as improvements in terms of efficiency.

The authors would also like to point out the difficulty in getting information related to home monitoring experiences taking place in the SNS, from one source, which, from our point of view, should be either the ACSS or the SPMS. In the author's point of view, if this information is not someway centralized, the evaluation and subsequent dissemination of these experiences will be more difficult to achieve.

It should also be noted that it was not possible to identify evaluation methods with the purpose of, systematically, evaluating experiences, so that the decisions can be based on accurate information, it can be possible to learn from mistakes as well as to innovate by sharing and replicating successful experiences, although this is one of the EU guidelines for the eHealth Action Plan 2012-2020 (Commission of the European Union, 2012).

## REFERENCES


ACSS (2016). Termos de referência para contratualização hospitalar no SNS – Contrato Programa 2016.

Barros, P. P. and Simões, J. D. A. (2007). Health systems in transition: Portugal – Health system review. The European Observatory on Health Systems and Policies.

Borja-Santos, R. (2013). Ministério quer que serviços de saúde apostem mais na telemedicina. Jornal Público. Available in: https://www.publico.pt/sociedade/noticia/ministerio-quer-que-servicos-de-saude-apostem-mais-na-telemedicina-1586809.

Commission of the European Union (2012). eHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century.

Consulta do Viajante (2016). Available in: http://www.consultadoviajante.com/sobre.html.

Costa, P. D., Reis, A. H. and Rodrigues, P. P. (2013). Clinical and economic impact of remote monitoring on the follow-up of patients with implantable electronic cardiovascular devices: An observational study. *Telemedicine and e-health*, 19(2), pp. 71–80.

Dias, A. (2015). Integração de cuidados de saúde primários e hospitalares em Portugal: uma avaliação comparativa do modelo de unidade local de saúde (Tese de Doutoramento). Universidade de Aveiro, Portugal.

INE (2014a). Projeções de População Residente 2012-2060. Lisboa: Instituto Nacional de Estatística.

INE (2014b). Inquérito à utilização das tecnologias de informação e da comunicação nos hospitais 2014. Lisboa: Instituto Nacional de Estatística.

INS (2014). Inquérito nacional de saúde 2014. Available in: http://www.insa.pt/sites/INSA/Portugues/ComInf/Noticias/Documents/2015/Novembro/11INS2014.pdf.

Mourato, P. (2014). Telemedicina generaliza-se até final do ano. Available in: http://www.dn.pt/portugal/acores/interior/telemedicina-generalizase-ate-final-do-ano-3960209.html.

Noronha, N. (2013). Rede de telemedicina entre Portugal e PALOP avança em 2014. Available in: http://lifestyle.sapo.pt/saude/noticias-saude/artigos/rede-de-telemedicina-entre-portugal-e-palop-avanca-em-2014.

Oliveira, T.C., Bayer, S., Gonçalves, L., and Barlow, J. (2014). Brief Communication-Telemedicine in Alentejo. *Telemedicine and e-Health.* 20 (1)

OPSS (2016). Relatório de Primavera 2016 - Saúde: Procuram-se novos caminhos. Observatório Português dos Sistemas de Saúde. Available in: http://www.opss.pt/node/488.

Pereira, J. (2016). Telemedicina em Portugal. NewsPharma MyPneumologia. Available in: http://www.mypneumologia.pt/entrevistas/208-telemedicina-em-portugal.html

Pinho, L. (2013). Patient Innovation é uma rede social para doentes. Jornal Público. Available in: http://p3.publico.pt/actualidade/ciencia/8400/patient-innovation-e-uma-rede-social-para-doentes.

Portal da Saúde (2005). Portugal e Espanha implementam a Telemedicina no Algarve. Available in: http://www2.portaldasaude.pt/portal/conteudos/a+saude+em+portugal/investigacao+e+desenvolvimento/telemedicina.htm.

Queirós, A., Carvalho, S., Pavão, J., Da Rocha, N.P. (2013). AAL information based services and care integration. In *HEALTHINF 2013 - Proceedings of the International Conference on Health Informatics*, pp. 403-406. ISBN: 978-989856537-2.

Queirós, A., Pereira, L., Dias, A. and Rocha, N. (2017). Technologies for Ageing in Place to Support Home Monitoring of Patients with Chronic Diseases. In *HEALTHINF 2017 - Proceedings of the International Conference on Health Informatics*.

Saúde 24 (s.d.). Quem somos. Available in: http://www.saude24.pt/PresentationLayer/ctexto_00.aspx?local=15.

Sousa, C., Leite, S., Lagido, R., Ferreira, L., Silva-Cardoso, J., and Maciel, M.J. (2014). Telemonitoring in heart failure: A state-of-the-art review. *Revista portuguesa de cardiologia*, 33(4), pp.229-239.

SPMS (2014). Telemonitorização DPOC. Available in: http://spms.min-saude.pt/2014/04/telemonitorizacao-dpoc-2/.

WHO (1978). Declaration of Alma-Ata. International Conference on Primary Health Care, Alma-Ata, USSR, 6-12 September 1978.

# Security And Privacy Issues in Healthcare Monitoring Systems: A Case Study

Daniel Tolboe Handler, Lotte Hauge, Angelo Spognardi and Nicola Dragoni
*DTU Compute, Technical University of Denmark, Richard Petersens Plads, 2800, Kongens Lyngby, Denmark*
{*dthandler, lotte.l.hauge*}*@gmail.com*, {*angsp, ndra*}*@dtu.dk*

Keywords: Security, Privacy, Pervasive Healthcare.

Abstract: Security and privacy issues are rarely taken into account in automated systems for monitoring elderly people in their home, exposing inhabitants to a number of threats they are usually not aware of. As a case study to expose the major vulnerabilities these systems are exposed to, this paper reviews a generic example of automated healthcare monitoring system. The security and privacy issues identified in this case study can be easily generalised and regarded as alarm bells for all the pervasive healthcare professionals.

## 1 INTRODUCTION

Health monitoring systems are getting more and more common (Pantelopoulos and Bourbakis, 2010). In particular, elderly patients require systematic and continuous monitoring in order to promptly detect anomalous changes in their health. With the development of new technologies such as mobile computing and wireless sensor network (WSN), many solutions specifically aimed at elder persons have been proposed. Generally, several wireless communication devices are employed and combined with medical sensors, to monitor senior citizens from various points of view (Tsukiyama, 2015; Kotz et al., 2009; Dasios et al., 2015). Surprisingly, many of the proposed systems are not taking into account what security threats the installation provides and which privacy measures are needed. The security risks associated with such systems, indeed, can represent a high concern, because of the sensitive information these services can deal with.

In this paper we want to raise the awareness about the lack of concerns many solution providers show regarding such risks. To do so, we will look at the main weakness of one of these systems, namely a general monitoring system for elderly (Dasios et al., 2015) . This system has been chosen because it is generic enough to provide a good representation of the health monitoring systems available, which are mostly surveillance, only wearable, or mostly environmental sensors. Common for most systems is that none of them have taken security into account, except for adopting encryption for the network protocols. In this paper we want to show which security measures should be investigated in all kind of monitoring systems, by making a risk analysis and threat model for the mentioned system. We also investigate which privacy measures should be stated before implementing the systems. One major contribution will be stating which impact can have the traffic analysis on a health care system, when burglars are planning a break-in. The case study aims at identifying a number of main guidelines to follow in order to propose e-health monitoring systems that should guarantee a reasonable level of security.

*Outline of the Paper.* We describe the case study in Section 2 and then we focus on what impact the health monitoring system can have (Section 3). In Section 4 we present different attacks, as well as vulnerabilities the used network protocols raise. Then we analyse which privacy measures should be taken into account in Section 5. Section 6 concludes the paper.

## 2 ENVIRONMENTAL SENSOR NETWORK

The system we are going to asses is the environmental sensor network described in (Dasios et al., 2015) (sketched in Fig. 1). This monitoring system aims to give an overall health estimation of the elderly at a low cost, without using cameras and microphones, which according to the paper are ”commonly perceived as privacy violators”. They also seek



Handler, D., Hauge, L., Spognardi, A. and Dragoni, N.
Security And Privacy Issues in Healthcare Monitoring Systems: A Case Study.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 383-388
ISBN: 978-989-758-213-4

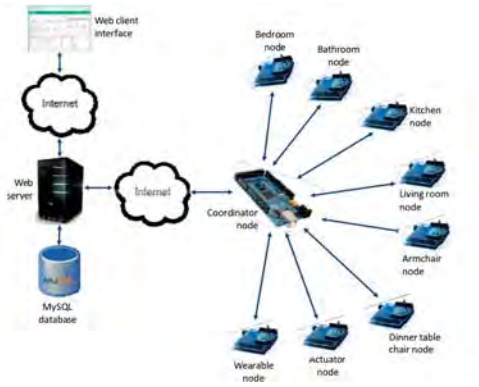



Figure 1: Architecture of the considered environmental sensor network (picture taken from (Dasios et al., 2015)).

a minimal amount of wearable devices. These goals are obtained by a WSN of nine nodes. Four room nodes measure movement, light, temperature, electricity consumption, water flow, and pressure, in order to give an overall assessment of what the senior citizen is doing. Sensors on the chairs detects the presence of a person seating on a chair. A wearable sensor serves as fall detector and as a panic button to call for help. An actuator enables remote control of certain electrical installations such as climate control. Lastly, a coordinator node works as a sink for the WSN, handling data from the network to a web interface and vice versa. The WSN uses ZigBee (see Section 4) to communicate to the sink, that, in turn, uses a wired internet connection sending all the data to a MySQL database and deploying the data on a Web site.

## 3 THREAT MODEL

The threat model is essentially composed by assets, threats, attackers, attacks and vulnerabilities of the system. The final aim is to enable an informed decision-making about application security risk.

**Assets.** The reason for installing monitoring health care systems, is to be able to take care of the elderly. The system reviewed in this paper strives to detect life threatening events and preserving good health, which must be the main assets for the elder. Installing such systems should give the elder a secure feeling of living on their own, while not using an extensive amount of communal resources in salary for health care personnel. Other assets for the elder which are not taken into account in the considered health monitoring system are dignity, possessions of the elder and revealing of activities and location. The environmental sensor network presents a possibility of remotely changing the climate, which they state as an asset for the elder, but with modern technology, this should not be a problem to handle by the elder. In Fig. 2 we list all the considered assets with respect to what we think would be their value for the elder, where we adopted a scale from 1 to 5: very low value (1), low value (2), medium value (3), high value (4), very high value (5).


Figure 2: Considered assets, with priorities related to their value for the elder and attackers, with existence likelihood and technical knowledge to exploit a system.

**Threats.** When looking at the system, the major threats of the assets are false positives, false negatives, or exposure of data. In particular, the exposure of personal health data, location of the elder, and activities. False positives and false negatives are related to the considered assets. A false positive is when the monitoring system misunderstands the status of the elder (for example, raising a false alarm). Conversely, a false negative is when the system fails to detect situations in which the assets are threatened.

**Attackers.** We have defined three main types of attackers (inheritors, criminals and burglars) as summarized in Fig. 2, where also their existence likelihood (from 1: very low, to 5: very high) and related technical knowledge to exploit a system (from 1: very low, to 5: very high) are reported.

*Inheritor.* The most important assets defined, such as life, good health and feeling secure are very dear to the elder, but not something that a lot of others have interest in taking. The only identified attacker for these assets is the inheritor, who would benefit from an early exit for the elder. The inheritor is defined to have less technical (that is, user level) knowledge of the system and very small probability of existing.

*Criminals.* Dignity and indoor climate could be used as blackmailing material and cons for criminals in order to get to the elders possessions. According to Danish police statistic over criminal charges[1] there are about 125 con thefts in inhabitants every year,

[1] https://www.politi.dk/NR/rdonlyres/87ADE11F-C8EE-4249-A64B-7D4B7FDFC7EE/0/Centralenoegletal2014.pdf

and these are mostly committed against elder people. Since it is limited how much value the elder posses, the major criminals are probably not going after the elders homes. Due to this we asses the likelihood of this as small, and the technical knowledge a bit higher than the inheritor, since the criminals have a network where they can distribute malicious tools.

*Burglars.* Possessions are also of great value to burglars. Burglars are becoming more organized, and according to a Danish assurance company they often know what is valuable, and are quick to find and take these items. Knowledge of where values are kept could be fatal in a burglary, which often takes short time. According to Danish police statistic over criminal charges thefts happens more than 20 times as much as robberies, which indicates that burglars tries to avoid people when breaking in. Danish assurance companies also advice their costumers to make their home look inhabited when away, which includes not stating their vacation on Facebook (Alka Ensurance, 2013). This indicates that knowledge of location and activities can have a large impact for burglars when choosing a place to break in. We asses the likelihood of a burglar as high when the burglars are defined to having little technical knowledge (see Fig. 2).

**Attacks.** We have identified five types of attack, which are considered in this paper. These are chosen because they show how serious and diverse can be the attacks available for this types of systems.

*Man in the Middle.* Man-in-the-Middle (MITM) is an attack type where the attacker relay and possibly modify or permutate the communication between two parties (Erickson, 2008). If the attacker is capable of intercepting or modifying the communication into or out of the house, he will be capable of preventing critical system updates, and prevent critical information of going of the house. Additionally is a man in the middle capable of sending fake messages, stating that the elderly has fallen, hence summon the health care company to assist the elderly without reason. If the health care system uses cameras to monitor the elderly, these pictures can be intercepted, and used either for blackmailing the elderly or for locating possession of value in the home of the elderly.

*Denial of Service.* If the health care systems are targeted by a successful denial of service attack, it will have severe consequences. All three health monitoring systems rely on sending information out of the house, if the user has fallen or anomalies is detected. If a denial of service attack prevents this, the whole system is compromised. The environmental sensor network relies on communication between the sensors and the sink. If this is prevented the system will become useless, hence the denial of service would be discovered, since the normal behavior sends steady flow of information, not only when a problem arrives. Compromise of the system in such way can lead to death or injury of the elderly, or a decrease in feeling secure.

*Traffic Analysis.* If the communication flow in the system is analyzed, it might be possible to know whether the elderly is home, is asleep or similar. This information can be useful to a burglar, who doesn't want to be disturbed when breaking in.

*Malware.* Malicious software can be used to modify the system, such that it doesn't fulfill the intended purpose. This can be used to make denial of service or man-in-the-middle attacks. Additionally malware can be used for simple traffic analysis as well.

*Social Engineering.* Attacking the human directly instead of the system has plenty of functions. It can be used to get access to the barebone computer or the sink to install malware, or it can be used to locate the sensors of the network. Elderly people are more vulnerable to social engineering than younger people (Lahtiranta and Kimppa, 2006), hence it is a obvious attack vector for the criminals.

Fig. 3 resumes the relations between threats attacks and attackers we have above detailed.

| Asset | Threat | | | Attacker | Attack |
|---|---|---|---|---|---|
| | False negative | False positive | Exposure of | | |
| Life | X | | | Inheritor | DoS |
| Good Health | X | | | Inheritor | DoS |
| Feeling secure | X | X | Possession | Inheritor | DoS |
| Communal resources | | X | | | MitM |
| | | | Skin | | MitM |
| Dignity | | | Activities | Criminals | SE |
| | | | Possessions | | |
| | | | Activities | Criminals | SE |
| Possessions | | | Location | Burglars | MitM |
| | | | | | MitM |
| Revealing of activities | | | Activities | Burglars | TA |
| Revealing of location | | | Location | Burglars | TA |
| Indoor Climate | | Climate adjustments | | Criminals | MitM |

Figure 3: Threat model for a health monitoring system, linking assets to the different threats, attacks and attackers.

**Vulnerabilities.** The attackers can exploit several vulnerabilities in order to realize the above attacks on the assets, that can be related to the different protocols or to the human being, as the social engineering-based attacks. The considered vulnerabilities are discussed in great detail in the following Section 4.

# 4 SECURITY AND PROTOCOLS

This Section addresses some security issues and vulnerabilities that occur when implementing the system. The network protocol used in the environmental sensor network is ZigBee. This Section reviews how MITM attacks, Denial of Service (DOS) attacks,

and traffic analysis can occur. Then, we generically describe how malware attacks and social engineering can occur. We use the acquired information to provide a risk assessment of the system.

**ZigBee.** ZigBee is widely used in pervasive computing due to its low bandwidth, low cost, and low power consumption (Stelte and Rodosek, 2013). Our vulnerability analysis is based on KillerBee[2], in particular on its *zbtumbler* tool, which detects active ZigBee networks, records and display information about the found devices. This makes it easy to detect whether or not ZigBee is in use.

*MITM Attacks. zbreplay* lounges a replay attack, which is countered in (Stelte and Rodosek, 2013). Another MITM attack is explained in (Choi et al., 2013) and exploits a vulnerability in the key distribution. In ZigBee the key is occasionally updated over the air in standard mode, which makes this attack quite realistic. The counter measures proposed involve using public key encryption instead of symmetric encryption. A greater modification is needed, and public key encryption.

*DOS Attacks. zbassocflood* uses a vulnerability in ZigBee's association method for new nodes joining the network. This floods the coordinator node, creating denial of service from the coordinator note, which is the sink of the sensor network. A counter measure for this problem has been proposed in (Stelte and Rodosek, 2013), by filtering the incoming traffic.

*Traffic Analysis. zbsniff* is a tool that sniffs and analyses the network traffic, thus making network-analysis easy to do. This is very difficult to prevent, due to how easy ZigBee is to detect.

**Malware.** Unintended, malicious software is a problem in all the communication technologies described above. If an attacker is able to inject malware into the system, the attacker will be able to compromise the whole system. Injection of the malware can be done by physically plugging in the malware (e.g. by a USB-stick) or using social engineering to get into the house of the system. Alternatively is it possible to access a wireless router, and through this get access to a computer to inject the malware on. A firewall and a strong security on the wireless network would be a countermeasure to this.

**Social Engineering.** The last attack vector we address is social engineering, covering actions where the human is the target of the attack, rather than the system (Bellovin, 2015). Examples of social engineering attacks are phishing, elicitation and simple phone calls, where the attacker pretends to be someone else (e.g. the IT department) (Hadnagy, 2011). According to Danmarks Statistik[3] the average age of the victims of trick theft (which can be assimilated to social engineering) is 75 years. It is important to keep social engineering in mind when implementing the health care systems, where users primarily are elders. It is impossible to make a comprehensive list of actions to prevent social engineering, however some recommendations can be given:

- Make sure the user of the system (the elderly people) gets information about: system updates, visits from technicians, who in call, if in doubt.
- As far as possible use technicians the user knows, or send a known employee from the health care organization together with the technician.
- If possible configure the system, such that the elder doesn't have to know passwords, etc., such that a social engineer cannot trick the password from the elder.

**Attackers.** Based on the analyses in Sections 3 and 4 we will now define which attackers who are able to carry out which attacks:

*Inheritors.* The computationally skills of the inheritors is limited according to our description in section Section 3. Since the assets the inheritors are targeting are the health and life of the elderly, denial of service attack is their only obvious attack vector. Malware to perform these attacks are considered to hard computationally for the inheritors to do, hence inheritors rely on jamming equipment or similar to perform the attack.

*Criminals.* Social engineering is the first and most obvious option of the criminals. Since their target is to blackmail the elderly, a lot of information is needed. This information can be gathered by man-in-the-middle attacks or traffic analysis. Malware is an option for these attacks.

*Burglars.* The location of the elderly is the main concern of the burglar. Traffic analysis is a good tool for locating the elderly. Man-in-the-middle attacks can be used to intercept pictures of the home of the elderly to locate possessions of value, such that the burglar knows where to look. Malware is one way to conduct a traffic analysis or a man in the middle attack, however it is difficult for the burglar to to able to plant the malware in the system.

[2]https://github.com/riverloopsec/killerbee

[3]http://www.dst.dk/da/presse/Pressemeddelelser/2002

### 4.1 Risk Assessment

As a final evaluation of the above analysis, having defined the risks and threats of the systems in Section 3 and the vulnerabilities of the system in the above sections, we can propose an overview of the five different attack vectors in the system of the considered case study. We start considering easiness and availability of performing the attack, proposing a score for each attack vector. The final outcome is shown in Fig. 4, where, again, we used a scale from 1 to 5, where 1 means very low risk and 5 means very high risk. It is quite clear that the proposed system is most vulnerable to traffic analysis, as well as the social engineering, even if this is a more general problem, not only related to the specific system.

- Man-in-the-middle (4)
- Denial of Service (3)
- Traffic analysis (5)
- Malware (2)
- Social engineering (4)

Figure 4: Easiness and availability of the different attacks, given the vulnerabilities of the system.

The challenges of traffic analysis are that it is very hard to prevent. The use of wireless protocols is always detectable. Even if the data in the communication is encrypted, it is still possible to detect the presence of the data flow.

The full overview of how critical each security issue is, with respect to the assets, can be found in Fig. 5. The scores are calculated as follows.

*Probability Score.* The probability score is a measure of probability of compromising the asset. The value is calculated as the average of the product of the each attackers value (found in Fig. 2) and the linked attack vectors scores (found in Fig. 4) divided by 5 and rounded up. As an example the probability of compromising dignity is:

$$\frac{1}{5}\left\lceil \frac{\textit{Likelihood of criminal} \cdot \textit{ease of MitM}}{2} + \frac{\textit{Likelihood of criminal} \cdot \textit{ease of Social Engineering}}{2} = \right\rceil$$
$$= \frac{1}{5}\left\lceil \frac{2 \cdot 4}{2} + \frac{2 \cdot 4}{2} \right\rceil = 2$$

*Value Score.* The values are defined as the assets value and are scored similar to the score applied to each asset in Fig. 2.

*Final Score.* The final score reveals the overall risk for compromise of this asset the score is calculated as the probability score multiplied by the value score. This reveals a new assessment scale from 1 to 25. Since we do not get above 9 in our assessment we have only colored the scale from 1 to 9, introducing dark green as better than green, dark red as worse than red, and brown as worst in order to easily distinguish between the risks. This is justifiable since everything above 10 could cause serious health problems because of the sensitive subject of health care is, taking care of the already weakened citizens.

| Asset | Prob. | Value | Final |
|---|---|---|---|
| Life | (1) | (5) | (5) |
| Good health | (1) | (1) | (1) |
| Feeling secure | (1) | (4) | (4) |
| Dignity | (2) | (3) | (6) |
| Possessions | (3) | (3) | (9) |
| Revealing of activities | (4) | (2) | (8) |
| Revealing of location | (4) | (2) | (8) |
| Indoor Climate | (2) | (2) | (4) |

Figure 5: Final results of the risk analysis, stating the different risks for the different considered assets.

Fig. 5 reveals that the must critical asset for the elderly is possession, followed by revealing of activities, and revealing of location. The assets most compromised are targeted by burglars doing traffic analysis. It is not impossible to imagine a scenario where a one of the health care systems is widely used, and a burglar will walk around a neighborhood, and use traffic analysis to find out where the systems are used. If the burglar knows in which houses the systems are used, he will be able to determine if the elderly is home, or not, and can then conduct a break-in, without being disturbed.

## 5 PRIVACY AND EQUIPMENT

This Section deals with the privacy issues occurring when monitoring the elderly (*the user* from now on). We use the Common Framework by Markle Foundation (Markle Foundation, 2008) as a base for the analysis, since it has been recommended for health systems in an extensive analysis for privacy frameworks (Kotz et al., 2009).

*Openness and Transparency.* It is important that the user of system knows exactly what data that is collected about her and how this data is used. Furthermore it is important to state who has access to this data and where and how the data is stored.

*Purpose Specification.* It is important that the collected data is used only for its purpose. If the data is intended to be used for new purposes it should be accepted by the user beforehand. If the user is unable to make that decision, it should be clear who can vouch for the decision (children or other next-of-kin).

*Collection Limitation and Data Minimisation.* The stored data should be limited to a minimum. Example: there is no need for storing fall detection data, as long as the user hasn't fallen. It is important to inform the user of the amount of stored data and how long time the data is stored.

*Use Limitation.* The collection of health data is very personal information for the user, and it is important that the integrity of the data is ensured, and that the data in no situation is made available to other than authorized users of the system.

*Individual Participation and Control.* Users should be able to access their own stored data, and should be able to control who else have access to these data.

*Data Quality and Integrity.* Only relevant data should be stored and should be up-to-date. Old and irrelevant data should be removed as soon as possible.

*Security Safeguards and Controls.* All means for protecting the data should be taken, when collection, analyzing, transmitting and storing the data.

*Accountability and Oversight.* The company or people in charge of the health care system must be held accountable for any breach of the security or privacy issues of the system. The system uses many different sensors, and it is important for each type of sensor to tell the user exactly what is stored. The system is used for detecting and analysing the environment the user lives in, hence a certain amount of data will have to be collected and stored for the system to be sufficient. If compromised this system will reveal a great amount of personal data about the user, and it is important that the user is aware of this, and that security measures are launched and maintained.

# 6 CONCLUSION

In this paper we have considered a reference health monitoring system for elderly people as a case study to state the security and privacy risks one should be aware of before implementing such kind of systems. In particular, we have identified attack vectors and groups of attackers, who could compromise the health monitoring system. Moreover, we have raised some privacy issues to address when people are monitored in their own home. Due to the fact that personal data is transmitted and stored on external servers, it should be stated who can access the data and who can be held accountable if data is lost or leaked.

Our major contribution has been the identification of the burglar threat, which could be very prominent if one healthcare system is used in large scale, because traffic analysis is not something a communication protocol can secure, the defence needs to be implemented as a part of what is communicated, which is easy to oversee by the developer.

The aim of the paper is to raise the levels of awareness and understanding of the cyber risks related to home monitoring systems. The hope is that the issues identified in this case study will be regarded as alarm bells for all the pervasive healthcare sector.

# REFERENCES

Alka Ensurance (2013). Her er man mest udsat for indbrud! (retrieved on http://www.alka.dk/).

Bellovin, S. M. (2015). *Thinking Security - Stopping Next Year's Hackers*. Addison-Wesley.

Choi, K., Kim, M., and Chae, K. (2013). Secure and Lightweight Key Distribution with ZigBee Pro for Ubiquitous Sensor Networks. *IJDSN*, 9(7).

Dasios, A., Gavalas, D., Pantziou, G., and Konstantopoulos, C. (2015). Wireless sensor network deployment for remote elderly care monitoring. In *Proc. of PETRA'15*. ACM.

Erickson, J. (2008). *Hacking: The Art of Exploitation*. No Starch Press.

Hadnagy, C. (2011). *Social Engineering: The Art of Human Hacking*. Wiley Publishing, Inc.

Kotz, D., Avancha, S., and Baxi, A. (2009). A privacy framework for mobile health and home-care systems. In *Proc. of SPIMACS'09*. ACM.

Lahtiranta, J. and Kimppa, K. (2006). Elderly people and emerging threats of the internet and new media. In *Proc. of I3E 2006*. Springer.

Markle Foundation (2008). Common framework for networked personal health information: Overview and Principles. Connecting for Health.

Pantelopoulos, A. and Bourbakis, N. G. (2010). A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 40(1):1–12.

Stelte, B. and Rodosek, G. (2013). Thwarting attacks on zigbee - removal of the killerbee stinger. In *Proc. of CNSM'13*.

Tsukiyama, T. (2015). In-home health monitoring system for solitary elderly. In *Proc. of EUSPN'15/ICTH'15*. Procedia Computer Science, Elsevier.

# A Disciplined Innovation Approach to Health Technology Solutions

Noel Carroll[1] and Ita Richardson[1,2]
[1]*ARCH- Applied Research for Connected Health Technology Centre, University of Limerick, Limerick, Ireland*
[2]*Lero- The Irish Software Reseach Centre, University of Limerick, Limerick, Ireland*
*{noel.carroll, ita.richardson}@lero.ie*

Keywords: Connected Health, Innovation, Healthcare Technology, Software Solutions.

Abstract: Despite the potential of innovation-driven healthcare technology services to increase the quality, accessibility and quality of care, the realization and success of such promise has yet to be achieved. This prompted us to explore the barriers towards success for healthcare software companies and examine what frameworks are employed across industry to support their growth in the digital healthcare market. As part of a three-phase study, this article reports on the first phase – to synthesize the literature on the readiness factors for healthcare technology companies. The findings of this research will guide our second phase of this research in surveying industry healthcare software companies. In so doing, we can establish readiness factors for healthcare software companies with a view to offering a more structured and disciplined approach to healthcare innovation.

## 1 INTRODUCTION

We often learn how small and medium-sized enterprises (SMEs) across the national and international service sector must consistently and continuously innovate and adapt to ensure their survival (Gebauer et al. 2012). It is a matter of 'survival of the fittest' to evolve with the dynamic external environment. To do so successfully largely depends on the SME's service innovation capability and competences to deploy resources and improve services. This is a challenge particularly in a healthcare context since technology advancements continue to rapidly grow while concerns around healthcare device safety and regulation continue to surface and challenge innovation (Carroll and Richardson, 2016). Thus, pertinent questions need to be asked such as, *how can an organization continuously evolve and offer a new service to meet healthcare needs*? *Where does the added capability and competencies come from to do so*? From our experience, two key factors here are to 1) identify the unmet healthcare needs and 2) examine how or where the capabilities will come from to address those needs.

We often learn about the growing success of companies breaking new ground in healthcare innovation and dominating market leadership (Carroll, 2016). While, this is very much welcome across the healthcare sector, little is known about why companies, particularly software companies, fail to achieve their business objectives in reaching new markets (Kellermann and Jones, 2013). Thus, uncovering both why companies fail and what we can do to reduce such occurrences, drew our attention towards the concepts of *evaluation*, *organizational readiness* and *capability maturity* to establish a more disciplined view of healthcare technology innovation. Technology has contributed towards a shift within healthcare practice which highlights the growing reliance and trust we now place on software to support healthcare decisions. However, unlike some sectors, for example business, failure to correctly align healthcare needs with software requirements can have devastating consequences on people's health – potentially fatal.

## 2 TOWARDS DISCIPLINED INNOVATION

In recent years, the concept of '*Disciplined Entrepreneurship*[1]' was coined at MIT and offers a comprehensive step-by-step approach to creating solutions. It focuses on the iterative process towards a final solution to meet users' needs. Aulet (2013)

[1] http://disciplinedentrepreneurship.com/



Carroll, N. and Richardson, I.
A Disciplined Innovation Approach to Health Technology Solutions.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 389-395
ISBN: 978-989-758-213-4

attempts to move away from an abstract view of entrepreneurship and proposes a toolkit to guide innovation. This provides a rich insight on entrepreneurship as a skillset. The toolkit provides 24 steps that is described as disciplined entrepreneurship and is a practical step-by-step process to channel the innovation and maximize the chances of success and ultimate impact. Such a formal process is considered beneficial to focus the innovation process. We previously explored how a similar approach could be achieved in a software engineering and healthcare innovation context (Carroll and Richardson, 2016). In this research we employed Design Thinking with a view to aligning healthcare innovation and software requirements and address customer pain points using the *Connected Health Innovation Framework* to a) support software developers in clearly identifying healthcare requirements and b) extend and enrich traditional software requirements gathering techniques. However, we have identified that there is a need to take this a step further and move towards establishing measures of innovation in order to assess risk and the capability to deliver an innovative solution within a process flow. We describe this as 'disciplined innovation'.

## 3 PROBLEM STATEMENT

By services, we often refer to an intangible product, for example, banking, consultancy, healthcare, education and software development. Thus, the service economy is vital across the globe. For the purposes of this paper, we focus on healthcare and software development as an exemplar to support SMEs, i.e. the 'health-tech' market. Yet, despite the significance placed on the service sector, there is a lack of practical measurement or management tools for innovation. Such a gap in literature and practice ensures that the promise of health-tech innovation is never fully realized across SMEs (Kohler et al. 2013). In an attempt to identify a tool to support technology companies to guide SMEs to improve their healthcare innovations, the authors are continuously faced with the need to restart the innovation evaluation process for each company.

In this paper, we propose a decision support tool that will guide organizations to self-assess their current organizational operations. Such a tool would support organizational management practice. It would provides us with a real opportunity to establish a framework to guide organizations through the evolutionary dimensions of healthcare technology innovation.

In a recent article, Christensen et al. (2016) suggest that, "*business model innovation is crowded*" which is driving companies to mount both offensive and defensive initiatives involving new business models. Identifying innovation attributes allows us to have greater control of the innovation flow process and develop predictable business models to drive decision-making tasks, measured performance and accommodate for an efficient innovation process flow. This becomes the primary motivation to introduce a 'disciplined innovation' model.

## 4 HEALTHCARE INNOVATION CAPABILITIES

Healthcare service environments become increasingly complex when technology is implemented to execute specific clinical, technical and business processes to deliver care. This ultimately adds to the complexity of a service environment, making it one of the most difficult environments in which to examine and manage service capabilities. Capabilities are complex, structured, and multi-dimensional. They may be described as fundamental determinants resource utilization to support and sustain organizational performance (Teece, 2009). Managing process maturity has been well documented throughout the business and IT literature. Little research in this area is carried out within a health-tech domain.

In IT management, maturity models play an important and influential role in organizational change (Becker et al., 2009). The availability of service and innovation capabilities has motivated us to review how we conceptualize the health-tech service environment. The success of innovation often relies on a number of contributing factors. For example, according to Birkinshaw et al. (2011, p. 3) the following "conditions" contribute towards sustained innovation: (1) *shared understanding*: cultural understanding of organizational behaviour; (2) *alignment*: aligning systems and processes to achieve desired performance metrics; (3) *tools*: training, concepts, and techniques to innovate; (4) *diversity*: optimizing external influences and insights to offer solutions within a particular domain; (5) *interaction*: establishing platforms to exchange ideas and build networks; and (6) *slack*: providing opportunities to access additional resources to develop ideas. These conditions contribute towards organizations overall competencies and capabilities.

# 5 CAPABILITY MATURITY MODELS

The ultimate goal of an organizational capability is to contribute towards some form of value, e.g. improved healthcare and increased profits. There are a very large number of variables that are dependent on the context and industry which determine the important role capabilities play in value creation. At an abstract level we may identify the attributes of a capability to include (Carroll and Helfert, 2015) value creation, outcome focused, measurable, and maturity driven process. Within an innovation environment, capabilities need to be dynamic. Dynamic capabilities are considered the source of competitive advantage. Teece and Pisano (1994) identify two key aspects in harnessing competitive advantage through dynamic capabilities that may apply to a service innovation environment:

(1) The shifting character of the environment, e.g. healthcare needs;
(2) The importance of strategic management in agility, adaptability and reconfiguring internal resources to meet external demands.

Winter (2003, pp. 4-5) suggests, "*dynamic capabilities typically involve long-term commitments to specialised resources […] [and] […] there must be an ecological demand for the costs of the capability and the use that is actually made for it*". Managing dynamic capabilities requires some form of structure and models – for example, the capability maturity model (CMM) (Paulk, 1995). The CMM assumes progress is made in distinct stages and capture capability maturity at a given time (moving through five progressive stages - initial, repeatable, defined, managed, optimized).

The results of this assessment process supports the organization to position themselves against defined best practices while identifying areas of weakness to drive change (Becker et al., 2009; Carroll and Helfert, 2015). However, existing maturity models tend to focus on large organizations (Blommerde and Lynch, 2016, p.2) and are "*too broad to account for the specificities of service SMEs and fail to reflect their unique characteristics*". We set out to address this gap in a health-tech context.

## 5.1 Health-tech Innovation

Focusing on singular innovation is considered a thing of the past, i.e. developing one solution and forever reaping the rewards. Organizations must continuously innovate and demonstrate their dynamic capabilities to execute effective innovation capability (Blommerde and Lynch, 2016). Blommerde and Lynch (2016, p.2) suggest, "*SMEs are unaware of their service innovation capability or how to improve their innovative maturity mode*". Thus, some form of innovation measurement is required to support SMEs.

Blommerde and Lynch (2016) present the key dimensions of service innovation capability which link to all five stages of capability maturity model using a 'Service Innovation Capability Maturity Index', namely focusing on 1) user involvement; 2) knowledge management; 3) strategizing and 4) networking. In addition, and with a view to focusing on the measurement of innovation, Kohler et al. (2013) introduce a Service Innovation Model that comprises of four layers. From their description, they explain that the top layer (innovation capabilities) is connected with service innovation capability indicators, which are captured in the second layer. Each capability is associated with an indicator that quantitatively captures the implementation of the innovation capability in the company.

Performance is a key factor in innovation and new terms have been introduced over the last decade such as 'disruptive innovation'. While the concept of disruptive innovation stems new terms such as 'value network' which may be described as "*the context within which a firm identifies and responds to customers' needs, solves problems, procures input, reacts to competitors and strives for profit*" (Christensen, 1997; p. 31), we need a systematic approach to manage the innovation process.

In the Service Innovation Model, the indicators are a core focus for the assessment and monitoring of the service innovation capabilities. The indicators are described by Kohler et al. (2013; p. 1350) as being quantitative representation of the innovation capabilities. The indicators are connected to a set of asset categories within sets of assets and assessed on a numerical scale. These assets are categorized into assets, i.e. human, financial, physical, intellectual property rights, information and information technology, and relationship assets. There is also a similar outlook on the dynamic nature of innovation. For example, den Hertog et al. (2010) suggests there are dynamic service innovation capabilities that successful service innovators outperform their competitors in some of the following:

1. Signaling user needs and technological options;
2. Conceptualizing;
3. (Un)bundling;
4. Coproducing and orchestrating;

5. Scaling and stretching.

Thus, we have identified that there is a natural evolutionary process in the innovation process. This process requires an organization to move between specific maturity stages of innovation. Maturity phases are well documented throughout the literature in CMM but may need to be tailored within an innovation context and more specifically, within a health-tech context. For example, Carroll and Helfert (2015) explain how the traditional view of the organizational environment raises concerns regarding the mismatch in the methods used to assess business value and understanding service process maturity. They demonstrate this by unpacking the nature of service capabilities that allows us to understand the primary components of value co-creation and their contribution towards service maturity within an innovation environment to access organizational readiness. This offers a suitable lens to view a disciplined approach to innovation that can be easily adopted by SMEs in health-tech. We also need to examine how organizational readiness aligns with innovation capabilities.

### 5.2 Organizational Readiness

Throughout the literature, organizational readiness is often associated with organizational change management (OCM) (Armenakis et al. 1993; Weiner, 2009). Change is a critical factor for organizational readiness and is a multi-level, multi-faceted construct which healthcare technologies often face to introduce technology innovation. In most cases, such change refers to organizational members' shared resolve to commit towards a change in practice and a collective ability to improve organizational performance. Thus, organizational readiness for change varies as a function of how much organizational members value the change, e.g. within a hospital context.

Value of change must be weighted up against the risk (e.g. cost and investment of resources) associated with innovation. According to Weiner (2009) there are three key determinants of change implementation capability: task demands, resource availability, and situational factors. We argue that *innovation capabilities* are a fourth key determinant of organizational readiness – which needs to be calculated to assess the impact of innovation on organizational readiness. We explain that innovation is the process of introducing new ideas, devices, or methods to bring about some change.

We can begin to uncover the key enablers of innovation by taking a holistic view of change and integrate this with business activity rather than isolated processes. This enables us to develop an innovation model and identify the guiding principles that are grounded in organizational experience – documented throughout literature (phase 1 of our research, as presented in this paper) and captured by surveying industry experiences (phase 2 of our research, future work). Weiner (2009) describes how organizational readiness is "*considered a critical precursor to the successful implementation of complex changes in healthcare setting*s". Weiner also cautions "*most publicly available instruments for measuring organizational readiness for change exhibit limited evidence of reliability or validity*" – hence the motivation for this research.

In the case of healthcare software companies, innovation drives organizational changes to meet new market demands. To ensure that innovation can be successful, metrics must be established to drive such change. Thus, OCM may be described as an approach to transition an organization from their current state to a new desired state. This involves the integration and alignment of people, processes, culture and strategy to innovate.

Before OCM can be successfully implemented, managers must clearly evaluate readiness for change. Armenakis et al. (1993) describes readiness in terms of the organizational members' beliefs, attitudes, and intentions. Thus, there are critical elements of change agents and social dynamics that influence organizational readiness process which may contribute towards the success of healthcare software innovation. Identifying and measuring these elements provides a benchmark on the current organizational state compares with their ideal state to derisk healthcare software innovation. We capture all of these factors of CMM and OCM to present our *Disciplined Innovation Model* for health-tech SMEs.

## 6 DISCIPLINED INNOVATION

Migrating from the current state to the future state of an organization requires a number of key stages to embrace an innovation culture to drive a specific strategy and improve their competitiveness. This enables organizational performance to achieve the desired business goals. Therefore, innovation is not a soft or vague construct, but rather, a critical process to drive organizational performance. Why then is the process of innovation less defined in terms of organizational readiness and process flow measurement? Where are the formal process models to guide SMEs through the innovation process to

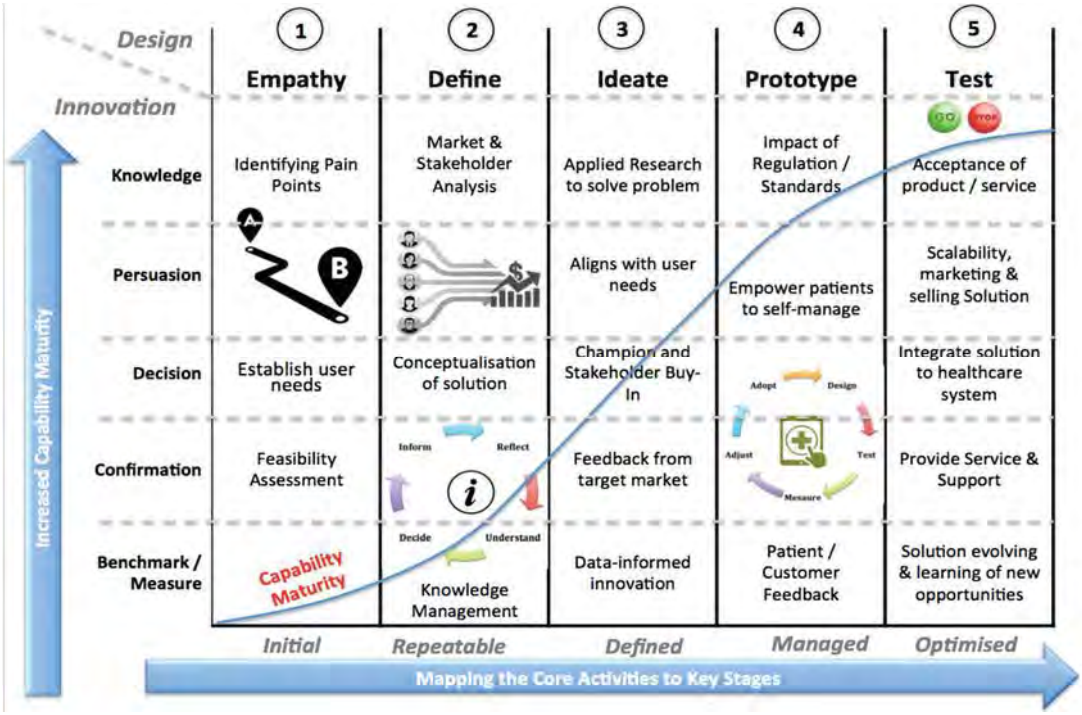



Figure 1: Disciplined Innovation Model.

derisk health-tech initiatives? To begin to address such questions, we need to ask: *Why does a company need to build a new solution, evolve an existing solution to maximize performance*? *How does a strategy cater for such change*? *What specific function(s) of the organization must change to ensure innovation is successful*? *Are there any specific guiding principles to derisk the innovation process*?

Innovation is often linked with creativity and the ability to design solutions for unmet needs in the marketplace. However, it remains unclear whether we can measure innovativeness within organizational readiness. Some attempts were made to measure innovation, open innovation and technological diffusion. For example, Jalles (2010) examines alternative variables such as technological progress (using patents and a Intellectual Property Rights Index) to explain different growth rates of income. In addition, Narayana (2005) suggests the need to measures innovation using a CMM to determine a particular strategic route and whether organizations need to learn of the innovation management process.

## 6.1 Disciplined Innovation Model

Figure 1 illustrates the initial *Disciplined Innovation Model*. It is influenced by:

1. *The key phases of innovation*: knowledge, persuasion, decision, and conformation;
2. *Design Thinking stages*: empathy, define, ideate, prototype, and test;
3. *CMM stages*: initial, repeatable, defined, managed, and optimized.

We also include the need to benchmark each phase to measure the capability maturity as a solution matures through each stage. This captures the essence of our initial development of the *Disciplined Innovation Model* and we have identified the need to establish specific metrics for each stage of the model. As the performance demanded by the customers of a value network increases over time so does the performance provided within a technological paradigm. Within a healthcare technology market, this could include a new set of performance value attributes that are now more relevant than the current paradigm to address healthcare needs.

While there is a strong body of knowledge on innovation as a method of competitive differentiation and as a way to create customer value, less attention has been devoted to developing a measure of innovation (Dobni, 2008). Dobni (2008) identifies innovation culture as an important factor to measure and identifies seven factors: innovation propensity, organizational constituency, organizational learning, creativity and empowerment, market orientation,

value orientation, and implementation context. However, more emphasis needs to be placed on the innovation flow process to support how we can support the innovation process.

## 7 DISCUSSION & CONCLUSION

By embedding data analytics into innovation, organizations can unlock new opportunities if guided through a disciplined process. In healthcare, this can build empathy for users and pave the way to improved experiences to deliver truly user-centered services and improved connectivity of services. We identify that despite the potential of innovation-driven healthcare technology services to increase the quality, accessibility and quality of care, the realization and success of such promise has yet to be achieved.

To address this, we present the initial *Disciplined Innovation Model* as a means to establish a self-assessment toolkit for SMEs to support the advancement of healthcare technology innovations and determine whether they are ready for scaling up their services and targeting innovation opportunities. We also identify the need to evaluate healthcare innovation from a healthcare practitioners perspective (O'Leary et al. 2014) as part of our future research.

While we introduce the initial version of this model, as part of our future research we plan to build on this by identifying specific metrics through industry collaboration and piloting the model through an iterative proves across a number of health-tech SMEs. We anticipate that this model could be tailored to fit other sectors to support SMEs though a disciplined innovation process. We will firstly focus on validating this work with health-tech SME's.

## ACKNOWLEDGEMENTS

This work was supported, in part, by ARCH - Applied Research for Connected Health Technology Centre (www.arch.ie), an initiative jointly funded by Enterprise Ireland and the IDA and Science Foundation Ireland (SFI) Industry Fellowship Grant Number 14/IF/2530.

## REFERENCES

Armenakis, A. A., Harris, S. G., & Mossholder, K. W. (1993). Creating readiness for organizational change. Human relations, 46(6), 681-703.

Aulet, B. (2013). Disciplined entrepreneurship: 24 steps to a successful startup. John Wiley & Sons.

Becker, J., Knackstedt, R., & Pöppelbuß, D. W. I. J. (2009). Developing maturity models for IT management. Business & Information Systems Engineering, 1(3), 213-222.

Birkinshaw, J., Bouquet, C. and Barsoux, J.L. (2011), "The 5 myths of innovation", MIT Sloan Management Review, Vol. 52 No. 2, pp. 43-50.

Blommerde, T., & Lynch, P. (2016). A Maturity Matrix for Assessing Service Innovation Capability. In: Irish Academy of Management Conference 2016, 31 August - 2 September, UCD, Dublin.

Carroll, N. (2016). Key Success Factors for Smart and Connected Health Software Solutions, Computer, Vol. 49, No. 11, pp. 32-38.

Carroll, N., & Helfert, M. (2015). Service capabilities within open innovation: Revisiting the applicability of capability maturity models. Journal of Enterprise Information Management, 28(2), 275-303.

Carroll, N., & Richardson, I. (2016). Software-as-a-Medical Device: Demystifying Connected Health Regulations. Journal of Systems and Information Technology, 18(2), pp.186-215.

Christensen, C.M. (1997). The Innovators Dilemma: when new technologies cause great firms to fail, Harvard Business School Press, Boston, Massachusetts.

Christensen, C.M., Bartman, T., van Bever, D., Strategy, D. and Strategy, E., (2016). The Hard Truth About Business Model Innovation. MIT Sloan: Fall, Research Feature September 13.

den Hertog, P., van der Aa, W., & de Jong, M. W. (2010). Capabilities for managing service innovation: towards a conceptual framework. Journal of Service Management, 21(4), 490-514.

Dobni, C. B. (2008). Measuring innovation culture in organizations: The development of a generalized innovation culture construct using exploratory factor analysis. European Journal of Innovation Management, 11(4), 539-559.

Gebauer, H., Paiola, M. and Edvardsson, B. (2012) 'A capability perspective on service business development in small and medium-sized suppliers', Scandinavian Journal of Management, Vol. 28, No. 4, pp. 321-339.

Jalles, J. T. (2010). How to measure innovation? New evidence of the technology–growth linkage. Research in Economics, 64(2), 81-96.

Kellermann, A. L., & Jones, S. S. (2013). What it will take to achieve the as-yet-unfulfilled promises of health information technology. Health Affairs, 32(1), 63-68. Chicago.

Kohler, M., Feldmann, N., Habryn, F. and Satzger, G. (2013) 'Service Innovation Analytics: Towards Assessment and Monitoring of Innovation Capabilities in Service Firms', 46th International Conference on

System Sciences (HICSS), 7-10 January, Hawaii.

O'Leary, P., Carroll, N., & Richardson, I. (2014). The Practitioner's Perspective on Clinical Pathway Support Systems. In IEEE International Conference on Healthcare Informatics (ICHI), (pp. 194-201). IEEE.

Paulk, M. C. (1995). The capability maturity model: Guidelines for improving the software process. Addison-Wesley Professional.

Narayana, M. G. P. L. (2005). A framework approach to measure innovation maturity. In Proceedings. 2005 IEEE International Engineering Management Conference, 2005. (Vol. 2, pp. 765-769). IEEE.

Teece, D.J. (2009), Dynamic Capabilities and Strategic Management: Organizing for Innovation and Growth, Oxford University Press.

Weiner, B.J., (2009). A theory of organizational readiness for change. Implementation Science, 4(1), p.1. Vancouver.

Winter, S.G. (2003), "Understanding dynamic capabilities", Strategic Management Journal, Vol. 24 No. 10, pp. 991-995.

# Data Format for Storing ANT+ Sensors Data

Petr Ježek and Roman Mouček
*Department of Computer Science and Engineering, New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Plzen, Czech Republic*
{*jezekp, moucek*}*@kiv.zcu.cz*

Keywords: ANT+, NIX, Sensor, Data/metadata Storage, EEGBase, Data Format, eHealth.

Abstract: Medical treatment of sudden and especially chronic diseases has become more expensive. People suffering from a variety of diseases had been traditionally treated in hospitals for a long time. Fortunately, the current situation has been changing also thanks to relatively cheap body sensors and development of systems for home treatment. It brings inconsiderable cost savings and improves patients' comfort. On the other hand, it puts demands on the used technical infrastructure and home treatment system developers who must solve integration of different systems. A crucial point is a definition of unified data formats facilitating transfer and storage of data to/in remote databases. There are standards and APIs such as Zigbee, Bluetooth low energy or ANT+ that define a protocol for data transfer. However, they do not define a suitable format for long term data storing. In this paper, data coming from ANT+ sensors have been studied and metadata related to all kinds of body sensors and raw data and metadata specific to individual sensors have been defined. Then a framework organizing data and metadata obtained from ANT+ sensors into an open and general data format suitable for long term storage of sensor data is introduced. Finally, a sample use-case showing the transfer of data from a sensor into a data storage is presented.

## 1 INTRODUCTION

Medical treatment of sudden and especially chronic diseases has become more expensive, especially with aging population. For instance, there were around 23 million people in the world affected with heart failure in 2011 (Bui et al., 2011). These people had been usually treated in hospitals for a long time. The situation has been changing at present because relatively cheap solutions for home treatment have appeared in the market (Surie et al., 2008), (Kyriacou et al., 2009) and patients can be moved from hospitals to their homes sooner. It brings advantages of a better comfort for patients and makes treatment generally cheaper.

Home treatment systems use a set of wearable sensors, usually powered from batteries, for monitoring of health or fitness level. They have to operate for a long time period without possibility to change batteries frequently. That is why new protocols with low energy consumption such as ZigBee (Farahani, 2008), Bluetooth Low Energy (Heydon, 2012) or ANT (Zaloker, 2014) have been developed. Data from these sensors are transferred to remote servers where they are processed and visualized. Body Area Network (BAN) is an integration of sensors providing a large data collection of body parameters. When the number of sensors connected to BAN increases, requirements for the management, long term storage and sustainability of acquired data also increases.

Although there are some low energy consumption standards for data transfer, these are too fragmented to allow easy manipulation with obtained data. Of course, these standards also do not provide means for long term storage and management of transferred data. As a solution this paper presents how to use a general data format called NIX (Stoewer et al., 2014) for encapsulating and storing ANT+ sensor data.

The paper is organized as follows. Section 2 deals with sensor infrastructure and description of data obtained from sensors. Section 3 describes existing ANT+ profiles; the most suitable profiles for eHealth domain are selected. Section 4 introduces a framework that facilitates conversion of sensor data to the NIX format. Section 5 presents the usage of proposed transformation, a simple use-case is provided. Section 6 summarizes the work and provides an outlook to the future.



Ježek, P. and Mouček, R.
Data Format for Storing ANT+ Sensors Data.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 396-400
ISBN: 978-989-758-213-4

# 2 STATE OF THE ART

There are several approaches aiming to manage Body Area Networks data. These solutions use semantic web technologies or cloud infrastructures. Three layers ontology describing data from different sensors is presented in (Mehmood et al., 2014). This ontology facilitates development of tools used for processing of sensor data. A Sensor-Cloud infrastructure (Yuriyama and Kushida, 2010) represents physical sensors as virtual sensors stored in a cloud infrastructure. Semantic Sensor Web (Sheth et al., 2008) is based on annotation of sensors data by means of Semantic Web. Such annotated data can be distributed on the Internet.

# 3 ANT+ PROFILES DISCUSSION

The ANT protocol is popular because of its low energy consumption and existence of suitable means for description of body parameters. 'Device profiles' (called ANT+) define data over the network in a consistent way (Innovations, 2013) and facilitate development of sensors and management of sensor data in application programs.

ANT+ profiles support a large scale of activities such as cycling, walking, or measurement of body parameters such as heart rate, blood pressure, weight, or muscle oxygen. When browsing individual profiles in a detail we find the attributes that are common for all profiles, for example a device name, device status, manufacturer, signal strength or battery status. Then, there are attributes varying for individual profiles. Individual profile attributes represent raw sensor data or domain specific metadata while common attributes describe general metadata (see Figure 1).

# 4 PROPOSED FRAMEWORK

## 4.1 Prerequisites

Due to the absence of a suitable format/data structure for sensor data representation we have designed a framework for collection and storage of sensor raw data and metadata in a defined structure. The used data structure/format has to be robust, flexible and widely accepted by scientific community to cover a heterogeneous nature of sensor data, provide a long term data sustainability, and ensure its re-usability in third-party systems.

## 4.2 Format Discussion

Within a working group of International Neuroinformatics Coordinating Facility (INCF) (Bjaalie and Grillner, 2007) and its Task Force on Electrophysiology[1] there were introduced two approaches towards defining a standard on electrophysiology data. The first one uses the Hierarchical Data Format (HDF5) (HDF5 Group, 2013). HDF5 is portable and extensible format supporting an unlimited variety of datatypes that is designed for flexible and efficient I/O operations with high volume and complex data. The second approach uses odML (Grewe et al., 2011) as a free form tree-like structure of sections, properties and values suitable for metadata description. This simple, platform-independent and human-readable format also ensures compatibility with other systems developed within the community such as (Zehl et al., 2014), (Le Franc et al., 2014), and (Davison et al., 2013). The next step of the task force, merging of these two approaches (Teeters et al., 2013), resulted in the proposal of the NIX format (Stoewer et al., 2014) that provides a data model for storing experimental data in HDF5 together with its metadata internally organized in the odML format. The NIX format is currently used in Helmholtz (Davison et al., 2013) and EEGBase (Jezek and Moucek, 2012) projects.

Although the NIX format was intended to be used in electrophysiology, its general definition makes it suitable for any time series data.

## 4.3 Proposed Mapping

We selected ANT+ profiles relating to person health and/or fitness level. Figure 1 shows common metadata (see the central circle) and domain specific raw data and metadata (see other circles). The NIX model consists of several main elements: Block, DataArray, Tag, MultiTag, Source, Group, and Dimension. Each element includes a set of attributes (such as id, name, specific attributes) and link to metadata organized in the odML structure. We used a simplified NIX model for mapping ANT+ elements. The Source element represents an ANT+ device, DataArray represents raw data, Dimension represents description of graph axes, and Block wraps a complete record.

[1] http://www.incf.org/programs/datasharing/electrophysiology-task-force

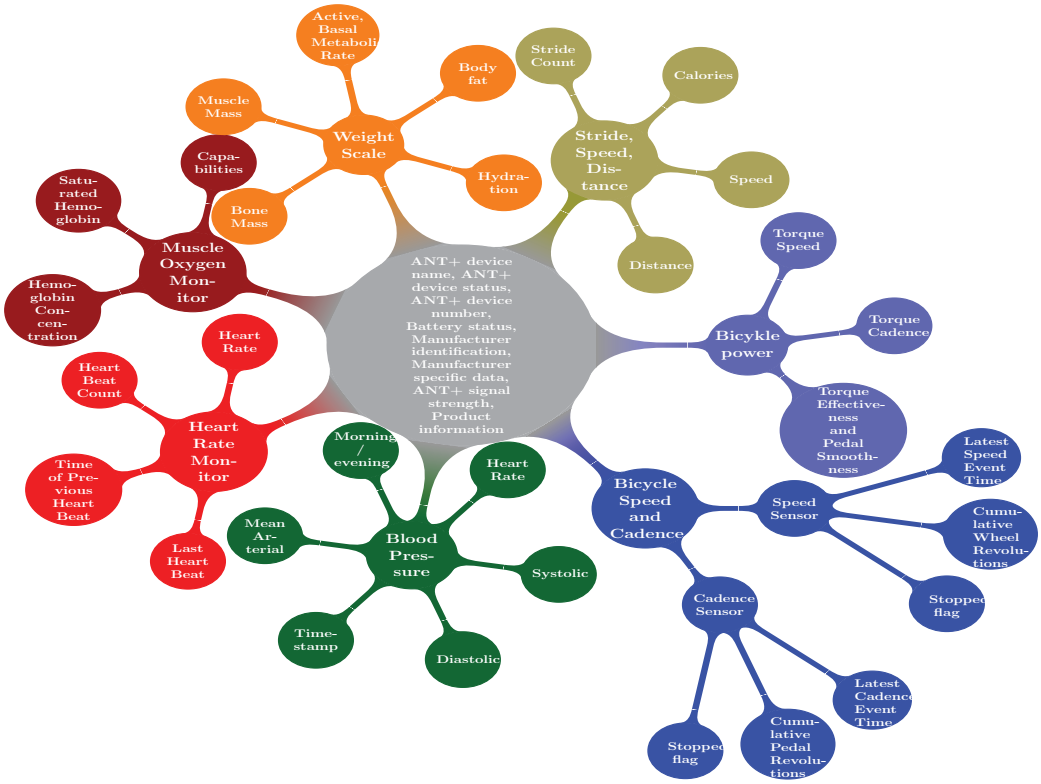



Figure 1: ANT+ Profiles Network.

## 5 USE CASE

The presented framework serves mainly to designers and programmers of the systems for home monitoring. Let's assume the following use-case. A programmer wants to implement a system for heart rate monitoring of elderly people. There are the following system requirements: sensors have to be easy to use and sensors data must be easily transferred to a computer where they are stored and evaluated. Moreover, during regular medical checks a physician uses long term records to check health condition of the patient and eventually starts a treatment. It means that both the patient and physician have access to the infrastructure that ensures data storing and management as well as data security, consistency and sustainability.

In this use-case we used the Garmin Premium Strap Heart Rate Monitor as a representative of ANT+ supporting devices. The Android SDK[2] was used to read ANT+ data into Android smart phones. We integrated this SDK into a custom mobile application MoBio[3] that reads data from ANT+ sensors and store

```
Section
├── name = heart_rate
├── type = metadata
└── Properties
    ├── Device name = Strap heart rate
    │   monitor
    ├── Device number = 1
    └── Product Information = Garmin
        Premium Strap Heart Rate Monitor
```

Figure 2: Metadata from the heart rate sensor in odML structure.

them on a SD card. The user of MoBio can pair available sensors, record data and visualize them. This solution is available to a large number of users due to the existence of cheap Android smart phones and heart rate monitor straps on the market.

Since our framework was also integrated into MoBio, recorded data can be stored in the NIX format. MoBio parses the record, metadata are transferred into a structure with one section and several properties (see an example in Figure 2) and continuously read heart beats data are stored into the DataArray element (see an example in Figure 3). The Source element has an attribute metadata that contains a link to

[2]http://developer.android.com/sdk/

[3]https://github.com/NEUROINFORMATICS-GROUP-FAV-KIV-ZCU/MoBio

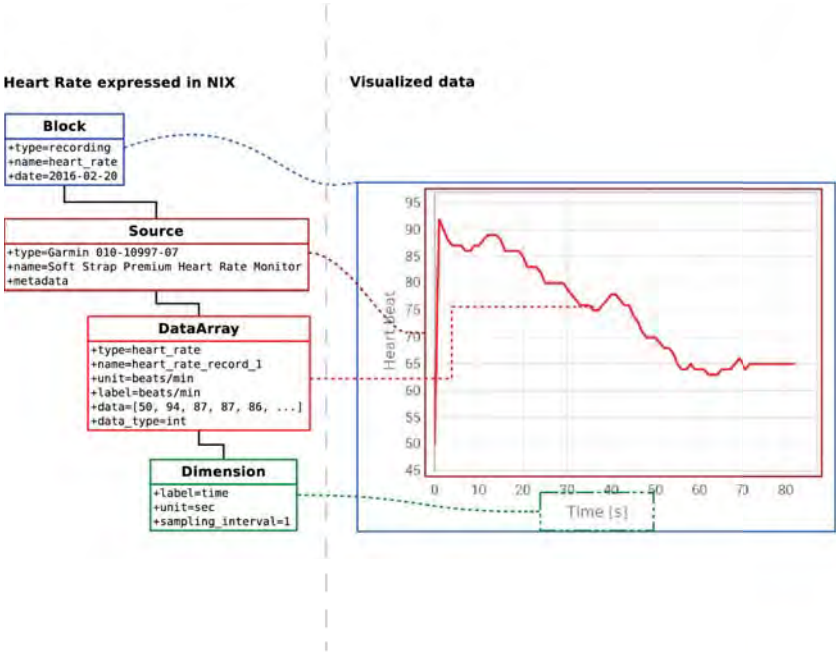



Figure 3: Heart rate record in the NIX format.

Metadata

Experiment

| Property | Value |
|---|---|
| private-experiment | false |
| start-time | 7/3/15 |
| end-time | 7/3/15 |
| temperature | 23 |
| environment-note | - |
| research-group | EEG/ERP group |
| scenario-title | Fitness level |

Subject

| Property | Value |
|---|---|
| gender | M |
| age | 25 |

Weather

| Property | Value |
|---|---|
| title | Sunny |

Artifact

| Property | Value |
|---|---|
| compensation | NotKnown |
| reject-condition | NotKnown |

Heart Rate

metadata

| Property | Value |
|---|---|
| Device name | Strap heart rate |
| Device Number | 1 |
| Device Information | Garmin Premium Strap Heart Rate Monitor |

Figure 4: Metadata stored in EEGBase.

the odML structure.

Once the data and metadata are stored they can be transferred to a suitable database. Figure 4 shows the metadata stored and visualized in EEGBase. A complete description of the experiment contains metadata from the Heart Rate strap. The raw data are stored as well.

# 6 CONCLUSIONS AND FUTURE WORK

Together with raising popularity of sensors for home treatment several low energy standards have been defined. These standards enable data to be transferred from body sensors into common computers where they are processed and visualized. ANT+ supported by significant sensors producers is one of the most used standards. Although transfer protocols and several APIs for working with sensors are defined, an open standard for storing sensors data are not substantially provided. Since home treatment systems use proprietary data formats, they cannot be easily integrated with variety of sensors.

In this paper we overcome these difficulties by designing a framework that maps data from ANT+ sensors into the open and generally applicable NIX format. The format brings advantages of two layers structure, metadata are structured using the flexible odML format and data are organized using the HDF5 format. Two layered organization of ANT+ sensors data is also a significant contribution of this work.

The functionality of the framework is shown on a simple use case. Within our future work the testing of a large collection of sensors followed by data transfer to a few databases is supposed. We also plan to invite developers of home treatment systems to integrate the framework into their solutions. When the framework is fully tested, we start to work on the

transformation of data using the Bluetooth low energy standard into the NIX format as well.

## ACKNOWLEDGEMENTS

This publication was supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports under the program NPU I.

## REFERENCES

Bjaalie, J. G. and Grillner, S. (2007). Global neuroinformatics: the International Neuroinformatics Coordinating Facility. *J Neurosci*, 27(14):3613–5.

Bui, A. L., Horwich, T. B., and Fonarow, G. C. (2011). Epidemiology and risk profile of heart failure. *Nature Reviews Cardiology*, 8(1):30–41.

Davison, A. P., Brizzi, T., Guarino, D., Manette, O. F., Monier, C., Sadoc, G., and Frégnac, Y. (2013). Helmholtz: a customizable framework for neurophysiology data management. *Frontiers in Neuroinformatics*, (25).

Farahani, S. (2008). *ZigBee Wireless Networks and Transceivers*. Newnes, Newton, MA, USA.

Grewe, J., Wachtler, T., and Benda, J. (2011). A bottom-up approach to data annotation in neurophysiology. *Frontiers in Neuroinformatics*, 5(16).

HDF5 Group (2013). Hierarchical data format. http://www.hdfgroup.org/HDF5/.

Heydon, R. (2012). *Bluetooth low energy: the developer's handbook*. Prentice Hall.

Innovations, D. (2013). Ant message protocol and usage.

Jezek, P. and Moucek, R. (2012). SYSTEM FOR EEG/ERP DATA AND METADATA STORAGE AND MANAGEMENT. *NEURAL NETWORK WORLD*, 22(3):277–290.

Kyriacou, E., Pattichis, C., and Pattichis, M. (2009). An overview of recent health care support systems for eemergency and mhealth applications. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 1246–1249.

Le Franc, Y., Gonzalez, D., Mylyanyk, I., Grewe, J., Jezek, P., Moucek, R., and Wachtler, T. (2014). Mobile metadata: bringing neuroinformatics tools to the bench. *Frontiers in Neuroinformatics*, (53).

Mehmood, N. Q., Culmone, R., and Mostarda, L. (2014). An ontology driven software framework for the healthcare applications based on ant+ protocol. In *Advanced Information Networking and Applications Workshops (WAINA), 2014 28th International Conference on*, pages 245–250. IEEE.

Sheth, A., Henson, C., and Sahoo, S. (2008). Semantic sensor web. *Internet Computing, IEEE*, 12(4):78–83.

Stoewer, A., Kellner, C., Benda, J., Wachtler, T., and Grewe, J. (2014). File format and library for neuroscience data and metadata. *Front. Neuroinform. Conference Abstract: Neuroinformatics 2014*.

Surie, D., Laguionie, O., and Pederson, T. (2008). Wireless sensor networking of everyday objects in a smart home environment. In *Intelligent Sensors, Sensor Networks and Information Processing, 2008. ISSNIP 2008. International Conference on*, pages 189–194.

Teeters, J. L., Benda, J., Davison, A. P., Eglen, S., Gerhard, S., Gerkin, R. C., Grewe, J., Harris, K., Jackson, T., Moucek, R., Pröpper, R., Sessions, H. L., Smith, L. S., Sobolev, A., Sommer, F. T., Stoewer, A., and Wachtler, T. (2013). Considerations for developing a standard for storing electrophysiology data in hdf5. *Frontiers in Neuroinformatics*, (69).

Yuriyama, M. and Kushida, T. (2010). Sensor-cloud infrastructure - physical sensor management with virtualized sensors on cloud computing. In *Network-Based Information Systems (NBiS), 2010 13th International Conference on*, pages 1–8.

Zaloker, J. (2014). Ant/ant+. *Arrow M2M representative*.

Zehl, L., Denker, M., Stoewer, A., Jaillet, F., Brochier, T., Riehle, A., Wachtler, T., and Grün, S. (2014). Handling complex metadata in neurophysiological experiments. *Frontiers in Neuroinformatics*, (29).

# Towards a Multi-level Approach for the Maintenance of Semantic Annotations

Silvio Domingos Cardoso[1,2], Chantal Reynaud-Delaître[2], Marcos Da Silveira[1], Ying-Chi Lin[3], Anika Groß[3], Erhard Rahm[3] and Cédric Pruski[1]

[1]*LIST, Luxembourg Institute of Science and Technology, 5, avenue des Hauts-Fourneaux, L-4362 Esch-sur-Alzette, Luxembourg*

[2]*LRI, University of Paris-Sud XI, Orsay, France*

[3]*Institute of Computer Science, Universität Leipzig, P.O. Box 100920, 04009 Leipzig, Germany*

{*silvio.cardoso, cedric.pruski, marcos.dasilveira*}*@list.lu*, {*lin, gross, rahm*}*@informatik.uni-leipzig.de, chantal.reynaud@lri.fr*

Keywords: Semantic Annotation, Annotation Maintenance, Ontology Evolution, Life Sciences.

Abstract: Semantic annotations are often used to enrich documents as clinical trials and electronic health records. However, the usability of these annotations tends to decrease over time due to the evolution of the domain ontologies. The maintenance of these annotations is critical for tools that exploit them (e.g., search engines and decision support systems) in order to assure an acceptable level of performance. Despite the recent advances in ontology evolution systems, the maintenance of semantic annotations remains an open problem. In this paper, we introduce, based on previous experiments, the main components of a multi-level approach towards the automatic maintenance of semantic annotations. We further provide examples for strengthening our proposal.

## 1 INTRODUCTION

The use of Knowledge Organization Systems (KOS) (Hodge, 2000), such as classification schemes, controlled terminologies, thesauri or ontologies in the medical field to annotate medical data is gaining interest over the last years (Gimenez et al., 2012; Funk et al., 2014; Yimam et al., 2016). Usually, KOS elements are used to annotate documents such as clinical reports or medical images in order to make their semantics explicit for humans and software applications. The KOS entities (concepts, properties, relationships, etc.) are associated with documents producing semantic annotations (Da Silveira et al., 2015). This process is commonly made by humans or automatic annotators and brings many benefits for end users such as, enhancing the retrieval of relevant information for decision support or improving semantic interoperability between systems (Uren et al., 2006).

However, the dynamic nature of medical knowledge forces to continuously revise KOS content. Thus, semantic annotations based on previous versions of the KOS can be impacted and loose their validity. Therefore, mechanisms to adapt these impacted annotations to the new version of KOS are required. In our previous work (Groß et al., 2012; Cardoso et al., 2016), we have shown a strong correlation between the modification of KOS elements and the modification of semantic annotations. We also manage to categorize the various evolution that can affect KOS and associate these changes with modifications of elements defining annotations.

In the literature, three families of approaches dealing with annotation maintenance can be found. The first one addresses the problem of automatic detection of inconsistent annotations (Eilbeck et al., 2009; Qin and Atluri, 2009; Köpke and Eder, 2011; Zavalina et al., 2015). However, mechanisms to support the correction of impacted annotations are not proposed. The second family of approach put the focus on the automatic detection and manual correction of invalid annotations (Maynard et al., 2007; Auer and Herre, 2007; Burger et al., 2010; Abgaz, 2013). However, these approaches only consider basic ontology changes, e.g., deletion and addition of concepts in ontology while more complex changes are important to consider and requires human intervention to perform the maintenance which is hardly applicable in the medical domain by virtue of the huge amount of annotations to adapt. Last, the most advance works implement an automatic correction of the annotation (Luong and Dieng-Kuntz, 2006; Tissaoui et al., 2011;



Cardoso, S., Reynaud-Delaître, C., Silveira, M., Lin, Y-C., Groß, A., Rahm, E. and Pruski, C.
Towards a Multi-level Approach for the Maintenance of Semantic Annotations.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 401-406
ISBN: 978-989-758-213-4

Park et al., 2011; Frost and Moore, 2014). This is mostly done based on reasoning techniques which rely of the logic formalism of the KOS. However, as medical KOS are expressed using lightweight description logic, these technique must be adapted.

The literature review highlights that there is no annotation maintenance/adaptation framework able to cope with the specificity of the medical domain e.g., size of the KOS, amount of annotations. In this paper we discuss the foundation of a (semi-)automatic approach to manage semantic annotations when their underlying KOS evolve over time without re-annotating the documents. We further justify our ideas based on previous experiments and examples.

We structure the remainder of this paper as follows: In Section 2 we define the annotation model used throughout this paper. Section 3 introduces our ideas towards the (semi-)automatic maintenance of semantic annotations. In Section 4 we conclude the paper by outlining future work.

## 2 ANNOTATION MODEL

The development of novel annotation maintenance approaches requires an appropriate annotation model covering evolution and quality aspects for annotations. We refer to our previously described annotation model (Groß et al., 2012; Cardoso et al., 2016) and give a brief overview of the main aspects.

A single annotation is defined as $a = (i, c, \{q\})$ where an instance item $i \in I_u$ is annotated with an ontology concept $c \in ON_v$, and a set of quality indicators $\{q\} \in Q$. An instance might be an electronic health record (EHR) or a question item from a case report form (CRFs) as used within clinical trials. In general, a concept can be used to annotate many items and an item might be annotated with several concepts. Different quality indicators can be used to retain quality, reliability and provenance information for each annotation, e.g. by attaching numerical confidence values, categorical ratings or evidence codes (Groß et al., 2009). Note that the quality of automatically generated annotations can vary significantly depending on the used methods, tools and their configurations (Funk et al., 2014).

Both, instance data and ontologies, underlie continuous changes. Hence, we denote $I_u$ as an instance in the version $u$ and $ON_v$ is an ontology in the version $v$. In this proposition paper, we focus on maintaining annotations due to evolution of ontology. We include further elements in the annotation model to better trace KOS changes and to correctly update the annotations. For instance, we retain the position of an annotation within an instance item ($offset$) since items can cover several concepts. The $offset$ can be useful, e.g. to link concepts from different versions with the same part of an item. We further consider the semantic relationship between a KOS concept and an item or the annotated part of an item. For instance, one item can be annotated as equivalent to a concept, more/less specific, partial match, etc. The semantic type of an annotation is useful to update outdated annotations. For instance, instead of removing an impacted annotation after concept deletion, one could preserve the annotated item by linking it to the superclass of the removed concept and changing its semantic type to "less specific". As additional provenance information, our annotation model includes an element to indicate which concept attribute (e.g., title, synonym, preferred terms, etc.) has mainly been used to produce an annotation. This can be valuable during the maintenance process, for example, to decide whether a basic attribute change is relevant and might entail an annotation modification.

## 3 FOUNDATION FOR SEMANTIC ANNOTATION MAINTENANCE

As discussed in Section 1, our long term objective is to design a (semi-)automatic approach for maintaining semantic annotations valid over time if the underlying KOS is evolving without a complete re-annotation of the document and by guaranteeing a high quality in the annotation after maintenance. We have analyzed the evolution of several KOS of the medical domain and we identified the behavior of annotations under different scenarios. We rely on these findings to derive different aspects to take into consideration for the maintaining semantic annotations. It can be seen as a multi-level approach that can be split according to inputs, process and outputs. It allows us to optimize the annotation maintenance task by considering at each step more information of different nature to maintain annotation that remain invalid after the previous step.

### 3.1 Maintenance Process

The different maintenance processes we have identified consist in: i) Automatically detecting inconsistent annotations caused by the evolution of the underlying KOS; ii) Using information gained from the evolution of the KOS only to adapt impacted annotations; iii) Using information of external KOS to maintain annotations that could not be maintained by considering local resource; iv) Using change patterns to finalize

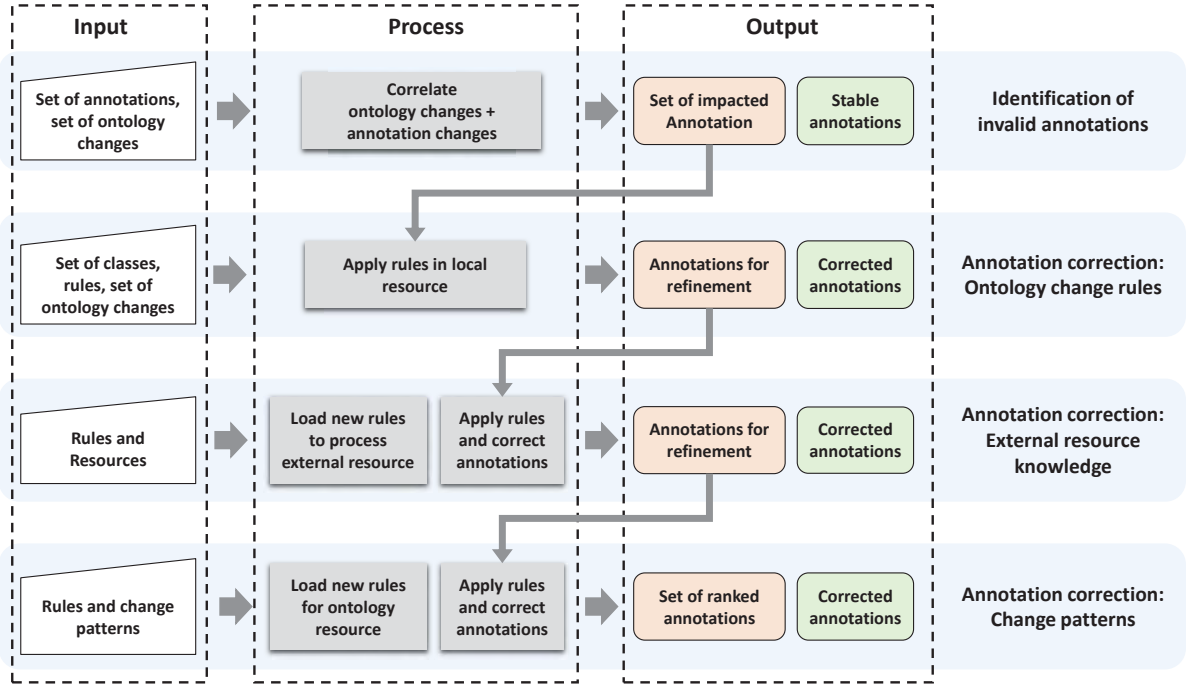



Figure 1: A multi-level framework for supporting annotation maintenance.

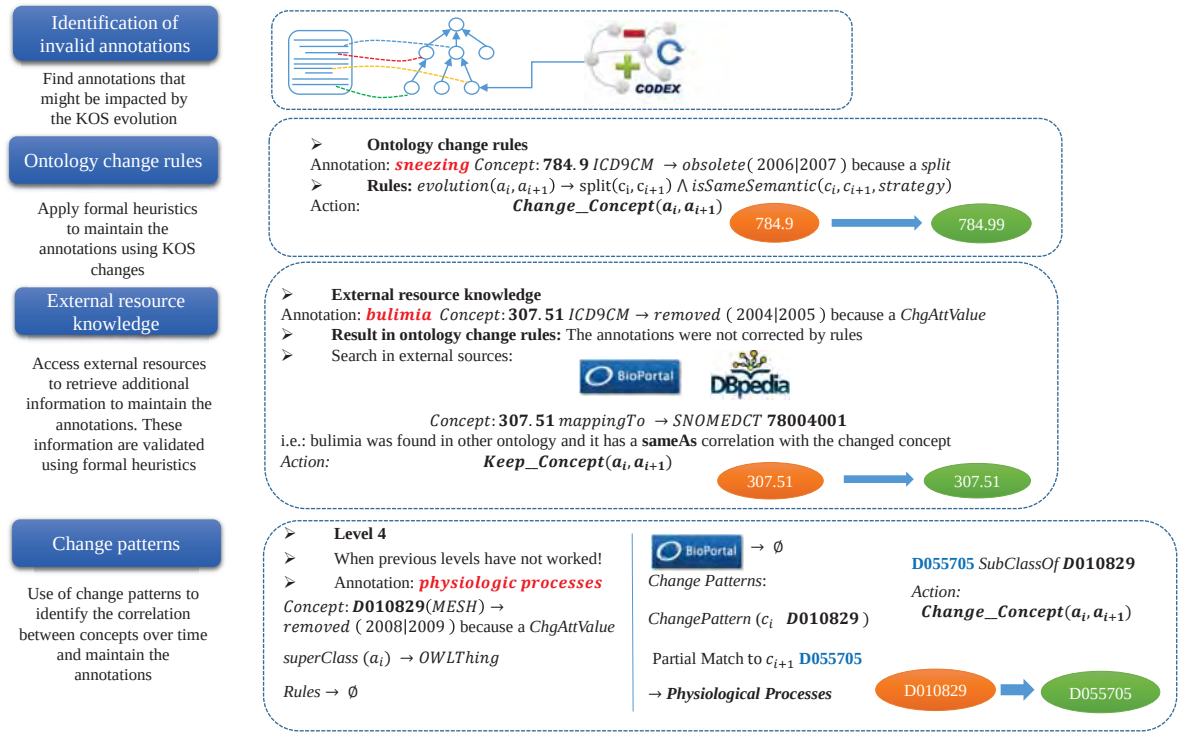



Figure 2: Examples illustrating the behaviour of the framework at each level.

the maintenance and optimize the quality of the set of adapted annotations.

- **Identification of invalid annotations**: It consists in identifying invalid annotations by analyzing the evolution of the associated KOS. To this end, it takes as input a set of annotations and two successive versions of the used KOS namely $K_n$ and $K_{n+1}$. The identification of concepts that have

changed between $K_n$ and $K_{n+1}$ can be obtained using an ontology *Diff* tool (Hartung et al., 2013; Noy et al., 2002) as well as additional information specifying the type of changes that have affected these concepts. As it is the case for ontology mapping adaptation (Groß et al., 2013), such information plays a key role in the maintenance task because it will determine the type of correction to apply to the annotations in the next levels. For instance, the deletion of a concept attribute can lead to the deletion of annotations but the deletion of the same attribute value in the context of a split of concept can lead to the migration of the annotation to the evolved version of the concept (i.e. the result of the split). It is therefore crucial to consider not only basic ontological changes (i.e. addition/deletion of concept) as it is the case in existing approaches for annotation maintenance but complex changes (i.e. split/merge of concepts) to optimize both the maintenance process and the quality of the adapted annotations.

- **Annotation correction using ontology change rules**: It consists in using information derived from the set of annotations itself as well as the data of the *Diff* between the two KOS versions $K_n$ and $K_{n+1}$ coming from the previous level to adapt the identified invalid annotations. At this level, the correction of annotations can be specified in rules that combine the context of evolution of the KOS and the status of the annotations. Under these conditions, the rules must specify the maintenance action to perform. For instance, we observed cases where an annotation was impacted because in its new version the label of concept associated with this particular annotation adopted the plural form. Therefore the corresponding rule can look like:

  *"If the label of concept is set to its plural form then do not change the annotation"*

  The type of ontological change contained in the *Diff* allows to propose more elaborated correction rules acting directly on the element of the annotation model like the $offset$. For instance, if the attribute used to annotate the text contained in an EHR is modified (e.g., a new word was added at the end of the label), then we check if we can find this modification in the text of the document to annotate (e.g., if the new word is also adjacent to the old text) by checking the information located at the $offset$ position. The corresponding rule is:

  *"If the label of concept increase and the data located beside the offset of the annotation is equal to the added word then increase the offset"*

  Another example is depicted in Figure 2. The annotation *sneezing* associated with the concept having as code 784.09 in ICD-9-CM version 2006 is no more valid in 2007. It is the direct consequence of the split of concept 784.9 between 2006 and 2007.

  In the example of Figure 2 the depicted rule checks if a concept was split. Basically, it specifies if the concept 784.99 from the new version of ICD-9-CM was engendered by a split of concept and whether it has a label (or an attribute value) which is equal to the same text of the annotation. It also verifies that the new version of the concept 784.9 (if it still exists) has no label of concept that fully match the text of the annotation. As a result, the action to maintain this annotation is to change the source concept of the annotation to 784.99.

- **Annotation correction using external resource knowledge**: It consists in using information inferred from external knowledge sources to maintain the annotations that could not be corrected using local resources of the previous level. Actually, in many cases the drift of ontological concepts can be characterized only by considering the semantic relationships provided by other ontologies (Pruski et al., 2016). Often labels of concept are completely different, from the syntactic point of view, before and after evolution. Therefore, considering local resources only does not allow to characterize their evolution and, in turn, cannot be reused for annotation maintenance purpose. The example depicted in Figure 2 about the evolution of the label of concept 307.51 of ICD-9-CM "Bulimia" in 2004 to "Bulimia nervosa" in 2005 shows another use case that requires external knowledge source. Applying existing approaches on annotation associated with this concept would simply lead to the deletion of the annotations. But the consideration of external resource (here mappings between ICD-9-CM and SNOMED CT provided by Bioportal) tells that these two terms are synonyms therefore the annotation can be kept. Nevertheless, the nature of the external knowledge resources can vary. Whether RDF datasets like BIO2RDF (Belleau et al., 2008) or expressive OWL ontologies contained in Bioportal (Noy et al., 2009) are considered, the inferred information can be of different quality and can affect the quality of the maintenance process.

- **Annotation correction using change patterns**: At this stage, information provided by the *Diff* and the use of external resources are not sufficient to maintain invalid annotations. The analysis of the morphosyntactic form of concept labels can reveal information to take decision about the maintenance of annotation. This technique has

already been explored in the context of ontology mapping adaptation (Dos Reis et al., 2015) but remains less relevant in terms of quality in the resulting maintenance decisions. Change Patterns are modifications observed in attribute values of a concept using linguistic-based features to identify the correlation between concepts over time. For instance, a *Partial Copy* between concepts is computed if and only if there exists a partial overlap between words from an attribute present in the KOS version $K_n$ and an attribute in the new KOS version $K_{n+1}$ (i.e., the attribute $a_0$ becomes $a_1$).

For instance, the annotation "Physiologic processes", shown at the bottom of Figure 2 produced using MeSH in period 2008/2009 was removed. This is due to a change in the attribute value in the definition of the concept D010829 leading to "Physiological Phenomena". Assuming the following conditions: i) we do not have information inside the ontology to handle with this change, ii) the super class from concept D010829 is *Thing* iii) external resources do not provide the necessary information to make decision, the application of four change patterns (*total copy, total transfer, partial copy, partial transfer*) considering only the attributes in the same sub-ontology e.g., the sub classes from concept D010829 allow to change the concept associated to this annotation from D010829 to D055705.

### 3.2 Output

Our approach was designed to process the annotations according to different levels of granularity, but the outputs only contain three kinds of data.

The first one refers to the nature of the annotations. It makes the distinction between annotations impacted by the evolution of the underlying KOS and non impacted annotations. We described in details these annotations in (Cardoso et al., 2016).

At the levels dealing with the correction of the annotations, the outputs are: i) the corrected annotations and ii) the set of annotations that need further investigation. Once corrected, the annotations are also enriched with evolution information making future modifications easier and enhancing their quality.

If invalid annotations remain, the definition of another levels exploiting different kind of information for maintenance purpose need to be implemented. The complexity of the evolution affecting KOS, the nature of the annotation, the specificities of the kind of object to annotate need to be taken into account in the definition of the additional levels. The rules that are used at each level also need to be defined by considering the quality of the adapted annotations.

## 4 CONCLUSION

We have presented a multi-level approach towards the (semi-)automatic maintenance of the annotations turned invalid after the evolution of their associated KOS. Our proposal is based on literature review as well as experimentations and consists in the progressive integration of complex information of different nature and various sources for correcting invalid semantic annotations without re-annotating documents. As future work, we will put the stress in the definition and validation of such a framework. Since our annotation framework will be used to (semi-)automatically correct outdated annotations, the used methods will need careful evaluation according to the quality of the produced results. For future work, we also plan to evaluate the different maintenance approaches using several annotation datasets from the biomedical domain such as annotated CRFs or EHR.

## ACKNOWLEDGEMENTS

The National Research Fund (FNR) of Luxembourg and Deutsche Forschungsgemeinschaft (DFG) support this work under the ELISA research project.

## REFERENCES

Abgaz, Y. M. (2013). *Change impact analysis for evolving ontology-based content management*. PhD thesis, Dublin City University.

Auer, S. and Herre, H. (2007). *A Versioning and Evolution Framework for RDF Knowledge Bases*, pages 55–69. Springer Berlin Heidelberg, Berlin, Heidelberg.

Belleau, F., Nolin, M.-A., Tourigny, N., Rigault, P., and Morissette, J. (2008). Bio2rdf: towards a mashup to build bioinformatics knowledge systems. *Journal of biomedical informatics*, 41(5):706–716.

Burger, T., Morozova, O., Zaihrayeu, I., Andrews, P., and Pane, J. (2010). Report on methods and algorithms for linking user-generated semantic annotations to semantic web and supporting their evolution in time.

Cardoso, S. D., Pruski, C., Silveira, M. D., Lin, Y., Groß, A., Rahm, E., and Reynaud-Delaître, C. (2016). Leveraging the impact of ontology evolution on semantic annotations. In *Knowledge Engineering and Knowledge Management - 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings*, pages 68–82.

Da Silveira, M., Dos Reis, J. C., and Pruski, C. (2015). Management of dynamic biomedical terminologies:

Current status and future challenges. *Yearbook of Medical informatics*, 10(1):125–133.

Dos Reis, J. C., Dinh, D., Da Silveira, M., Pruski, C., and Reynaud-Delaître, C. (2015). Recognizing lexical and semantic change patterns in evolving life science ontologies to inform mapping adaptation. *Artificial intelligence in medicine*, 63(3):153–170.

Eilbeck, K., Moore, B., Holt, C., and Yandell, M. (2009). Quantitative measures for the management and comparison of annotated genomes. *BMC Bioinformatics*, 10(1):67.

Frost, H. R. and Moore, J. H. (2014). Optimization of gene set annotations via entropy minimization over variable clusters (emvc). *Bioinformatics (Oxford, England)*, 30(12):1698–1706.

Funk, C., Baumgartner, W., Garcia, B., Roeder, C., Bada, M., Cohen, K. B., Hunter, L. E., and Verspoor, K. (2014). Large-scale biomedical concept recognition: an evaluation of current automatic annotators and their parameters. *BMC Bioinformatics*, 15(1):1–29.

Gimenez, F., Xu, J., Liu, Y., Liu, T. T., Beaulieu, C. F., Rubin, D. L., and Napel, S. (2012). Automatic annotation of radiological observations in liver CT images. In *AMIA 2012, American Medical Informatics Association Annual Symposium, Chicago, Illinois, USA, November 3-7, 2012*.

Groß, A., Dos Reis, J. C., Hartung, M., Pruski, C., and Rahm, E. (2013). Semi-automatic adaptation of mappings between life science ontologies. In *International Conference on Data Integration in the Life Sciences*, pages 90–104. Springer.

Groß, A., Hartung, M., Kirsten, T., and Rahm, E. (2009). Estimating the quality of ontology-based annotations by considering evolutionary changes. In *International Workshop on Data Integration in the Life Sciences*, pages 71–87. Springer.

Groß, A., Hartung, M., Prufer, K., Kelso, J., and Rahm, E. (2012). Impact of ontology evolution on functional analyses. *Bioinformatics*, 28(20):2671–2677.

Hartung, M., Groß, A., and Rahm, E. (2013). Conto–diff: generation of complex evolution mappings for life science ontologies. *Journal of biomedical informatics*, 46(1):15–32.

Hodge, G. (2000). Systems of knowledge organization for digital libraries: Beyond traditional authority files. Reports - Descriptive.

Köpke, J. and Eder, J. (2011). Semantic invalidation of annotations due to ontology evolution. In Meersman, R., Dillon, T., Herrero, P., Kumar, A., Reichert, M., Qing, L., Ooi, B.-C., Damiani, E., Schmidt, D., White, J., Hauswirth, M., Hitzler, P., and Mohania, M., editors, *On the Move to Meaningful Internet Systems: OTM 2011*, volume 7045 of *Lecture Notes in Computer Science*, pages 763–780. Springer Berlin Heidelberg.

Luong, P.-H. and Dieng-Kuntz, R. (2006). A rule-based approach for semantic annotation evolution in the coswem system. In *Canadian Semantic Web*, volume 2 of *Semantic Web and Beyond*, pages 103–120. Springer US.

Maynard, D., Peters, W., and Sabou, M. (2007). Change management for metadata evolution.

Noy, N. F., Musen, M. A., et al. (2002). Promptdiff: A fixed-point algorithm for comparing ontology versions. *AAAI/IAAI*, 2002:744–750.

Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D. L., Storey, M.-A., Chute, C. G., et al. (2009). Bioportal: ontologies and integrated data resources at the click of a mouse. *Nucleic acids research*, page gkp440.

Park, Y. R., Kim, J., Lee, H. W., Yoon, Y. J., and Kim, J. H. (2011). Gochase-ii: correcting semantic inconsistencies from gene ontology-based annotations for gene products. *BMC Bioinformatics*, 12(1):1–7.

Pruski, C., Dos Reis, J. C., and Da Silveira, M. (2016). Capturing the relationship between evolving biomedical concepts via background knowledge. In *the 9th Semantic Web Applications and Tools for Life Sciences International Conference*.

Qin, L. and Atluri, V. (2009). Evaluating the validity of data instances against ontology evolution over the semantic web. *Information and Software Technology*, 51(1):83 – 97.

Tissaoui, A., Aussenac-Gilles, N., Hernandez, N., and Laublet, P. (2011). Evonto - joint evolution of ontologies and semantic annotations. In Dietz, J., editor, *International Conference on Knowledge Engineering and Ontology Development (KEOD), Paris, 26/10/2011-29/10/2011*, pages 226–231.

Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., and Ciravegna, F. (2006). Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(1):14 – 28.

Yimam, S. M., Biemann, C., Majnaric, L., Šabanović, Š., and Holzinger, A. (2016). An adaptive annotation approach for biomedical entity and relation recognition. *Brain Informatics*, 3(3):157–168.

Zavalina, O. L., Kizhakkethil, P., Alemneh, D. G., Phillips, M. E., and Tarver, H. (2015). Building a framework of metadata change to support knowledge management. *Journal of Information & Knowledge Management*, 14(01):1550005.

# Diabetes Among Children (DAC)
## *Project - Exploring Opportunities with Support from Mobile Applications in a Cross Cultural Indo-Swedish Study*

Jenny Lundberg[1], Soniya Billore[2] and Clara Axelsson[3]
[1]*Department of Mediatechnology, Linnaeus University, PG Vejdes väg, Växjö, Sweden*
[2]*Department of Marketing, Linnaeus University, PG Vejdes väg, Växjö, Sweden*
[3]*eHealth Institute, Linnaeus University, Bredbandet, Kalmar, Sweden*
*{jenny.lundberg, soniya.billore, clara.axelsson}@lnu.se*

Keywords: Diabetes, Indo-Swedish, Global Health, Quality Assurance Systems, Metronics, Open Systems.

Abstract: In this paper we present opportunities and challenges to meet the worldwide challenge of diabetes. Diabetes has devastating long-term complications that cause very great personal suffering and social costs locally and globally. The prevalence of diabetes is increasing globally as an epidemic and affects 415 million people today, which is expected to increase to 642 million in 2040. In this paper we explore possibilities to join in Indo-Swedish R&D collaboration. We present and motivate the research purpose. Furthermore we present a research framework for mobile application development between Sweden and India. The scientific framework is elaborated and this paper ends with specific challenges and further work.

## 1 INTRODUCTION

According to the International Diabetes Federation (2015), 415 million in the world have got diabetes and it is estimated that by 2040, 642 million will be diagnosed with diabetes. Probably no other condition constitutes a challenge to the patient as diabetic, as glucose levels is dependent on the content of each meal, and the physical activity before and after that time, all the time, and every day. Level of emotional stress also affects sugar levels via the hypothalamic-pituitary-adrenals (HPA) cortisol-producing systems (Melin, 2015). Too little insulin is a threat in the long term, due to the risk of complications, and too much can lead to serious symptoms, mental and physical impact, in the worst cases, coma or death within minutes-hours. Patients should preferably combine information on food intake, especially the amount of carbohydrates, and previously scheduled and physical activity, and the current stress level, to calculate the optimal dose of insulin before each meal, several times per day.

Diabetes has two main types. In type 1 patients are dependent on insulin injections several times a day, to survive; in type 2 are many dependent on insulin for control, and to minimize the risk of complications (Forbes, 2016) (Thunander et al., 2012). Diabetes means risk of death in the short term, if not cared for, and the high risk of long term complications such as macrovascular (myocardial infarction, heart failure, stroke and peripheral arterial insufficiency, if not heal foot ulcers and amputations) and microvascular (retinopathy / blindness, kidney failure with dialysis or transplantation, and nerve problems).

Diabetes puts high demands on the individual when it comes to individual care, and in the case of children it puts higher demands on the family and especially the parents. Complications can occur and it is commonly known that this creates difficult health conditions and high social costs. Potentially this can be handled with support from not only medical science but also the latest developments in mobile devices and sensor technology. The measurements obtained from the close interaction of the mobile applications with the patients, here the children, and the consequent big data can be used to handle some of the issues in this area. Given that diabetes in children is a growing concern all across the world, this pilot project aims to look at the development and use of mobile applications designed specifically for them. As part of the background motivation for this study a focus group was conducted in India among paediatricians and diabetologists to inquire about the knowledge and usage of mobile applications for Diabetes. It was observed that the doctors were aware



Lundberg, J., Billore, S. and Axelsson, C.
Diabetes Among Children (DAC) - Project - Exploring Opportunities with Support from Mobile Applications in a Cross Cultural Indo-Swedish Study.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 407-412
ISBN: 978-989-758-213-4

of such applications on a global level but these were not used commonly by the diabetic children in India. Also it was not an established practice yet for the doctors to recommend any mobile application to the parents or the children to support their lifestyle in dealing with this challenge.

## 2 BACKGROUND

In this background we present the situation related to diabetes from the two countries specific perspectives. This presented within the three subsections India, Sweden and motivation for the Indo-Swedish research study.

### 2.1 India

The presence of Diabetes among children is a growing concern in India. According to a recent report (The Times of India, 2014) there are almost 70,000 cases of children with Type I and nearly 40,000 cases with type 2 diabetes. Obesity and the fast food fad is a common factor coupled with the fact that nearly 68% of urban children do not engage in regular exercise (type 2). Praveen and Tandon (2016) discuss the incidences of type II diabetes among children in India and underline that obesity is the leading cause of the situation. It is also of concern that parents, caretakers and schools emphasise the need for a healthy lifestyle among the children so that type II incidences can be curbed and brought under control. One of the possible influencers could be the digital support systems such as mobile technologies. Forbes, 2016 reports that the mobile penetration is India has reached nearly a billion subscribers and the market has the third largest smartphone users in the world. This opens immense opportunities for using the mobile technology for health related benefits. The first diabetes app named Diabeto was launched only recently in 2015. This application connects to glucometers of 30 different types and helps to keep track of the sugar levels, insulin levels including Bolus and Basal. Further it keeps the data stored in the cloud and presents visual representations of the patient's data for further, more efficient diabetes management. There is growing awareness about this application for Diabetes management in India and there is also growing interest among health and technology entrepreneurs to develop more applications for Diabetes.


Figure 1: The Diabeto application. Photo credit: www.startupexplore.com.

Figure 2: The synchronisation of the app with the glucometer. Photo credit: http://www.medgadget.com/2015/01/diabeto-to-sync-glucometers-with-smartphones-video.html.

In the next section, the situation related to diabetes in Sweden is presented.

### 2.2 Sweden

Sweden has the second highest presence of diabetes among children in the world approximately 7-8000 individuals and 800 are diagnosed every year. But there are some popularly used applications such as Triabetes, a CE marked medical software (iTunes, Triabetes). Research, development & innovation related to diabetes are ongoing and mature enough with rich data material to do further analysis on. Different quality assurance systems where devices related to diabetes are included exists. CE markings (Läkemedelsverket, 2016) and other initiatives as for example Diabetesappar.se are a collection of apps related to diabetes developed in cooperation with the Swedish health care professionals and Swedish application providers to facilitate and improve everyday life for patients and healthcare. Modeling of

diabetes (Marmarelis and Mitis, 2014) are considered in essence in regular technical issue, with glucose - insulin regulations. Furthermore, there are national approaches towards open quality assured databases such as the National Diabetes Registry, NDR with a specific open registry for children, Swediabkids. In the pictures below, figure 2, an example of a diabetes kit for children in Sweden is presented. Furthermore, continuous monitoring (CGM) like for example the automatic insulin pumps are examples of additional products in use.

Figure 3: Example of children's "diabetes kit" in Sweden. Photo credit: Jenny Lundberg.

### 2.3 Motivation for the Indo-Swedish Study of the DAC Project

For children and adolescents diagnosed with diabetes lifelong lifestyle changes are inevitable to ensure quality of life. Dietary regulations, sound exercise and monitoring as well as medication need to be obtained regularly every day (Barndiabetesfonden, 2016). This is an immense task for the young and their family who all need to be highly motivated to succeed long term. Cafazzo et al., 2012 have shown that motivation for self-monitoring can increase with the help of specialised mobile applications.

It is believed that the experience of healthcare and researchers Sweden could be used to explore the use of similar applications for the diabetic children in India. An Indo-Swedish cross cultural study of the lifestyles of diabetes among children can provide sufficient data and enable a higher level of generalization that can be translated to create more efficient and open mobile applications for the young generation.

## 3 RESEARCH PURPOSE

This research will conduct the following exploratory steps to have a deeper insight of the situation regarding e-initiatives and specifically mobile applications for Diabetes among children.

1. Explore the possibilities for designing e-initiatives with a study of the child patients and the level of adjustments and adaptations needed to deal with Diabetes as a health condition
2. Explore the personal needs of the patients in order to obtain a personalized and microscopic look to their needs and wants for the e-initiative
3. Utilise information portals and big data sources to study the possibilities of building an interactive mobile based empowerment platform for virtual coordination
4. Design mobile based solutions to empower and assist the children and their parents to deal with the presence of this diagnose.
5. Connect the results of the above to health industry entrepreneurs for mobile app. creation and development. This includes connecting Swedish entrepreneurs to the ones in India so that collaborative efforts on technology and user needs can be optimally utilised for creating market specific mobile applications.

## 4 SCIENTIFIC FRAMEWORK

As seen from the framework, the research attempts to explore in depth the situation of the children suffering from Diabetes type I and II in the countries of Sweden and India. The research will be done for mainly 2 aspect, A: the lifestyle of the children and the requirements that the children have given their health condition. By requirements it can mean a number of parameters such as access to health and medical advice, infrastructure support in terms of access to sports and other facilities that can help keep their body healthy and active, emotional support from families and friends, and access to other children who suffer from the same health situation through a network or group either online or otherwise where they and their families could meet and support each other. The other parameter is B: systems that support diabetes management in children such as the quality of awareness regarding the condition among the families, schools and local society and the relationship between them and the local health care units. Also this will explore the technological support that children have access to as they themselves want

to know more of their health condition and how best they could manage it. For example , the use of mobile applications that can be used to bring the knowledge , awareness and management of the diabetes related health situation closer to the children suffering from it as well as making the platform user friendly.

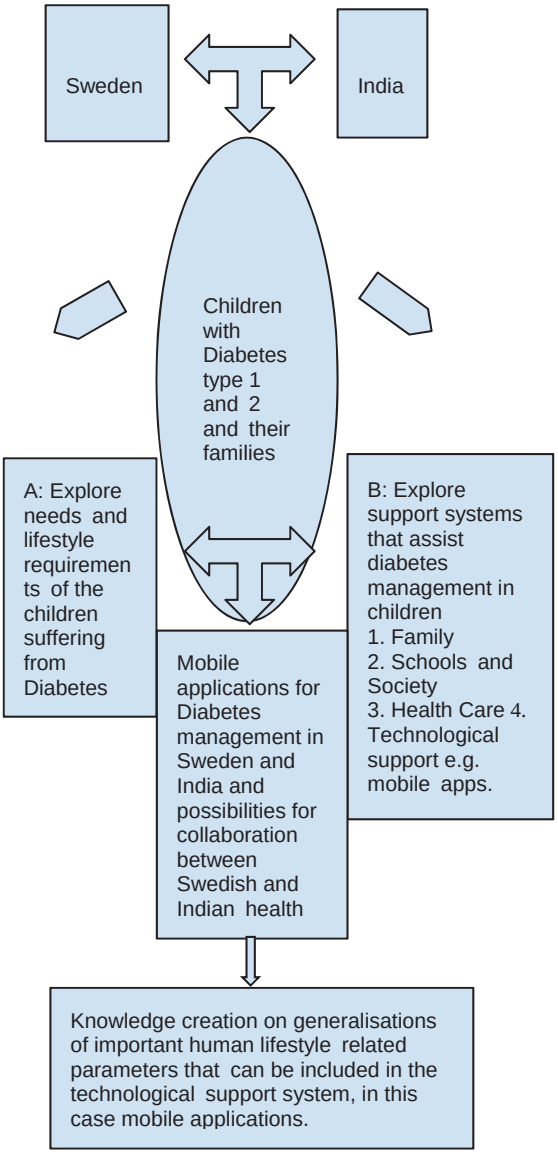



Figure 4: Research framework for mobile application development between Sweden and India.

It is known by our initial secondary research on the topic that Sweden has been using a number of mobile applications for Diabetes management but the situation is not the same for India. It will be interesting to further see how the applications in Sweden are customised to children patients if any.

By knowing the situation for parameters A and B described above we can have a list of factors that need to be tested against the presently available applications and explore the possibility of including new ones in them. Further on, the research will take the next step of bringing these observations to the world of entrepreneurship by building a link between the mobile app developers in the health industry of Sweden and India. It is believed that since Sweden and India have increasing number of diabetes affected children in very varied social and cultural settings, the knowledge created through the study will help generalise some important human and disease related elements that could be used on a wider context for Diabetes management among children on a global level.

## 5 SPECIFIC OPPORTUNITIES AND CHALLENGES

Novel technology with sensors / actuators, Internet of Things (IoT), micro sensors and open socio-technical systems, e.g. Internet-of-Things have unique opportunities to produce scientific data on e.g. diabetic patient's situation occurred. From here, and directly from the electronic medical record can be models based on real data and large data sets over time be developed and on contributing to a more secure basis for informed decisions in everyday life, such as about insulin doses, for patients and caregivers. The second area of this project concerns the so-called Big Data management. Much information is available today in the electronic medical record, but not available. Output data must be retrieved by experts in analysis departments, in e.g. Excel. Basic facilities required, for example. standard reports, for everyday recurring purposes. Most diabetic patients are and will be in developing countries, with type 2 diabetes, with more sophisticated technology the so-called artificial pancreas will not be an option, but there techniques that can be used with smart phones are more accessible. A combination with the automatic biological data, as pulse rate, as visualized in parallel with glucose levels and insulin doses, would provide additional information, e.g. about symptoms associated with hypoglycaemic events. Hypoglycaemia is still the biggest obstacle to achieving ideal blood glucose control, incl. fear of hypoglycaemia. Many factors influence the outcome of doses of insulin in diabetes, such as age, sex, duration of diabetes, type of insulin used (short- and

long-acting types), the patient's body composition, physical activity, why automatic data collection via a "bio-bracelet", which such as Smart Band, and practical visualization of these complex tasks, and their interactions, both can be helpful for patients and their advisers or assistants in clinical settings, and provide information to be analysed at the graduate level, can provide new information that can be converted into new clinical counselling. Analyses of data for glucose levels and insulin doses with details of HbA1c levels, over time, longitudinally, can provide both new insights into the relationship between glucose variability (i.e., frequency and amplitude of the high and low values) and the level of HbA1c (and over time relation to the development of diabetic complications), and in the patterns of glucose levels, frequency of episodes of hypoglycaemia, relationships actually taken insulin doses, etc. are not available today, and not for larger groups of real patients outside of clinical trials, and put them in the with bias (s) involved. The complex situation of diabetes care is a major global challenge (Guariguata et al., 2014) and is a good example of the area where the development and utilization of techniques to facilitate better target fulfilment can spread easily benefit both individuals and society (Sundström, 2016), (Hu et al., 2015), (Lundberg, et al., 2015), (Eriksén, 2015).

# 6 FUTURE WORK

Collaboration between different expert domains are of interest to battle the challenges related to diabetes. Closer collaboration between stakeholders, with care institutions and related partners for a healthy lifestyle among children are desired.

# ACKNOWLEDGEMENTS

To the system-cross organisation at Linnaeus University giving us this unique opportunity to initiate a small scale R&D cooperation. Furthermore to the paediatric clinics, endocrinologists in Växjö and Kalmar hospitals for support.

# REFERENCES

International Diabetes Federation, *About Diabetes*, Available at http://www.idf.org/about-diabetes/facts-figures, accessed 22 September 2016

*Nearly 70000 children suffer from Diabetes*, Available at http://timesofindia.indiatimes.com/life-style/health-fitness/health-news/Nearly-70000-children-suffer-from-diabetes/articleshow/45133397.cms, accessed 26 September 2016

India's first diabetes app, available at: http://timesofindia.indiatimes.com/life-style/health-fitness/health-news/Indias-first-diabetes-app-launched-to-ease-diabetic-lives/articleshow/49980670.cms, accessed 24 September 2016.

India's mobile subscribers, available at: http:// www.forbes.com/sites/sariatharia/2016/01/06 india-just-crossed-1-billion-mobile-subscribers- milestone-and-the-excitements-just-beginning/#783308a55ac2

http://www.diabetesappar.se/

Barndiabetesfonden, available at http://www.barndiabetesfonden.se/Om-diabetes/, accessed 28 September 2016

http://www.forbes.com/sites/sarithara i/2016/01/06/india-just-crossed-1-billion-mobile-subscribers-milestone-and-the-excitements-just-beginning/#783308a55ac2

http://timesofindia.indiatimes.com/life-style/health-fitness /health-news/Indias-first-diabetes-app-launched-to-eas e-diabetic-lives/articleshow/49980670.cms

https://lakemedelsverket.se/Alla-nyheter/NYHETER-2015 /Medicinska-appar-ska-folja-regelverkets-krav-for-att-vara-sakra/

https://itunes.apple.com/se/app/triabetes-diabetes-allt-i-ett/ id657352839?mt=8

https://www.ndr.nu/#/

Cafazzo, J. A., Casselman, M., Hamming, N., Katzman, D. K., & Palmert, M. R. (2012). *Design of a mHealth App for the self-management of adolescent type 1 diabetes: a pilot study. Journal of medical Internet research*, *14*(3), e70.

Praveen, P. A., & Tandon, N. *Childhood obesity* and type 2 diabetes in India. *WHO South-East Asia J Public Health*, *5*(1), 17-21., 2016

https://swediabkids.ndr.nu/

Marmarelis, V. & Mitsis, G. *Data-driven modeling for diabetes - diagnosis and treatments*, Springer Verlag, 2014

Diabeto application

Melin, E., *Psychosomatic aspects on diabetes and chronic pain, Alexithymia, depression and salivary cortisol; The affect School and Script Analysis Therapy*, in *Department of Clinical sciences.* 2014, Lund University, Faculty of Medicine: Lund

Thunander, M., Torn, C., Petersson, C., Ossiansson, B., Fornander, J., Landin-Olsson, M., *Levels of C-peptide, body mass index and age, and their usefulness in classification of diabetes in relation to autoimmunity, in adults with newly diagnosed diabetes in Kronoberg, Sweden.* Eur J Endocrinol, 2012. 166(6): p. 1021-9.

Guariguata, L., et al., *Global estimates of diabetes prevalence for 2013 and projections for 2035.* Diabetes Res Clin Pract, 2014. 103(2): p. 137-49.

Sundström, P., *Överenskommelse mellan staten och Sveriges Kommuner och Landsting; vision e-hälsa 2025* Avdelningen för digitalisering, 2016.

Hu, Y., Bai, G., Lundberg, J., & Eriksén, S. (2015). *A Virtual Community Design for Home-based Chronic Disease Healthcare.* In 2nd International Workshop on Usability and Accessibility focused Requirements Engineering. Springer book chapter.

Lundberg, Jenny, Jhunjhunwala, Ashok, Prashant, Suma, Ravindran, Deapica , Eriksén, Sara, Hofflander, Malin, Nilsson, Lina *Health in Hand - Indo-Swedish R&D collaboration around innovative mobile technologies for health promotion and disease prevention*, poster, Womencourage, 24-26 sept., Uppsala, Sweden, 2015.

Eriksén Sara, Jhungjhungwala, Ashok, Lundberg Jenny, Prashant Suma, Ravindran Deapica, *Perzonalized mobile interventions for health promotion – locating participatory design in emergent situated practices of global healthcare infrastructuring*, Infrastructures for healthcare 5th International workshop, Trento 18-19 June 2015.

# Understanding Users' Perception on the Trustworthiness of Online Health Information

Shijuan Li, Beidi Yu, Jiexi Liu and Biting Huang
*Department of Information Management, Peking University, Beijing, China*
*{shijuan.li, ybdyu, 1400093002, 1400016609}@pku.edu.cn*

Keywords: Information Credibility, Information Trustworthiness, Information Literacy, Youth, Health Information Seeking.

Abstract: The evaluation of online health information, i.e. its reliability, credibility, trustworthiness, etc., plays a significant role in the users' judging and health behaviour change process. Related works haven't come to a consensus on the framework of trust formation process. Nor much attention has been paid to one of the most active groups of online information users, the youth. This study, from the young people's perspective, examines the perception and judgment on the trustworthiness towards online health information. To test the design and reveal deficiencies of the study and procedure before time and resources are expended on large-scale studies, a pilot study was designed and conducted. Then semi-structured interviews were employed involving students from two groups: university freshman and the seniors respectively. The preliminary results cover: the exploration of their health information seeking process, factor analysis towards trustworthiness of online health information, the perceptions on HON measures and related health information literacy.

## 1 INTRODUCTION

The assessment of online health information, i.e. its reliability, credibility, trustworthiness, etc., influences greatly the users' judging and health behavior change process. Related works haven't come to a consensus on the framework of trust formation process. It has been revealed that credibility is a key impact factor of trust formation; influencing users' trust judgement and health behaviour change eventually (Everard and Galletta, 2005, Rowley et al., 2015). Other work indicated that trust and credibility are two concepts overlapped, since they share same sub-domains, i.e. brand, content, usefulness, style, while considering constructing an assessment framework (Fogg et al., 2003, Rowley et al., 2015).

Nor much attention has been paid to one of the most active groups of online information users, the youth. Young people have been recognized as one of the most active groups of the Internet users in China (China Internet Network Information Center, 2016). According to China Internet Network Information Center (CNNIC), there were approximately 688 million Internet users in China by 2015, among which 36.4% (around 224 million) were young people (aged between 10 to 24). Though probably they are supposed to have comprehensive surfing skills and have extensive Internet experience, they still have difficulty in evaluating the quality of online information, especially health information (Gray et al., 2005, Dobransky and Hargittai, 2012).

This study, from the young people's perspective, particularly examines the perception and judgment towards trustworthiness of online health information, which then will help to explore their health information behaviour. The core questions are designed include: If the youth are capable of judging the quality of online health information? What are their processes of trust formation towards online health information? What factors influence positively or negatively their acceptance and use of online health information? The rest of the paper is organized as follows: explanations of methods and materials utilized during data collection and data analysis process; preliminary analysis results from the qualitative study, consisting of a pilot study and semi-structured interviews; end with a discussion.



Li, S., Yu, B., Liu, J. and Huang, B.
Understanding Users' Perception on the Trustworthiness of Online Health Information.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 413-417
ISBN: 978-989-758-213-4

# 2 METHODS AND MATERIALS

The methods of the ongoing work are designed including of a pilot study and semi-structured interviews. Young people's perception and judgment towards the trustworthiness of online health information are investigated utilizing searching experiment and follow-up interviews. Volunteered participants, aged between 18 to 24, were employed in the Peking University (China) to take part in the investigation. The participants were asked to firstly attend 30 minutes long session to search for information within a given health field. Each participant should provide concurrent protocols as they search the Internet for health information. Meanwhile, a numerical scale, from one to ten, will be provided to them to score the trustworthiness of the searching results. Then a brief interview is followed to explore the rational on how they allocate scores for the trustworthiness of each result (Yu, 2016).

## 2.1 Related Measures

Evaluation tools are available to assess online health information, and previous studies have used generic tools, i.e., HON code, to examine the quality of web-based healthcare information (Harland, 2007; Hsu W-C, 2008; Kim, 1999). The Health On the Net Foundation (HON) (www.healthonnet.org) promotes and guides the deployment of useful and reliable online health information, and its appropriate and efficient use. The purpose of HON code is to protect all from incorrect, indeed misleading medical and health information. With this intention, HON accredits web sites according to eight ethical principles: 1) from which authority the information is coming from; 2) what is the purpose of the site; 3) the confidentiality policy; 4) the origin of the sources used; 5) the justification about benefits and performance of a specific treatment; 6) the transparency of the authorship; 7) the transparency of sponsorship as well as honesty in advertising; 8) editorial policies. The user's perceptions on those selected measures are investigated in this study as well.

## 2.2 Data Collection

The qualitative data are collected from both the pilot study and interviews. A pilot study was conducted to reveal deficiencies in the design of the study and procedure before time and resources are expended on large-scale studies. The participant approached was a senior university student who was interviewed about basic information about online health information search and trustworthiness judgement. Questions that were not answered as expected were re-worded or re-scaled; time that taken to complete the session was recorded to decide whether it is reasonable; ambiguities in questions were identified and clarified.

Then semi-structured interviews were conducted involving students from two groups: university freshman and the seniors respectively. The recorded interviews were transcribed by one the research team and then crosschecked by another member. The preliminary results were analysed utilizing content-based analysis that will be discussed in the next section in greater details.

## 2.3 Data Analysis

The data collected from both the pilot study and the interviews were recorded and transcribed with permission from the participants for further analysis.

One concern is that of contamination of data. This may arise from either where data from the pilot study are included in the main results or where pilot participants are included in the main study, but new data are collected from these people (van Teijlingen and Hundley, 2001). It is believed that contamination is less of a concern in qualitative research, where researchers often use some or all of their pilot data as part of the main study. Some have therefore argued that in qualitative approaches separate pilot studies are not necessary (Holloway, 1997). Frankland and Bloor (Frankland and Bloor, 1999) indicated that piloting provides the qualitative researcher with a "clear definition of the focus of the study" which in turn helps the researcher to concentrate data collection on a narrow spectrum of projected analytical topics. Therefore, data from the pilot study was included in the overall analysis.

The interpretation and analysis of the data employed content-based thematic analysis that involved classifying and coding the content into themes. An explanation of the codebook used is as shown in Table 1, including the code, theme and their corresponding explanations. The main themes formed from the data include: the health information seeking process, perception on HON measures, health information literacy, factor analysis on trustworthiness etc.

Table 1: Codebook.

| Code | Theme | Explanation |
|---|---|---|
| IS | Credible information source | Health websites, hospital websites, question and answer community |
| IP | Credible information provider | Large public hospitals, doctors from famous hospitals |
| CI | Credible information | Disease symptoms, appointment information, hospital and department ranks, misunderstands of information |
| NI | Not credible information | Recommended medicine, recommended therapy |
| NF | Not credible factor | Advertisement |
| AA | Active acquisition | Hospitals and departments |
| PA | Passive acquisition | Pushing articles |
| TF | Information trust formation | Information from authority consistently |

# 3 PRELIMINARY RESULTS

Basing on both related theories and the previous findings, the preliminary results from the qualitative data are categorized into the following themes:

- Exploration of their health information seeking process;
- Their perception on HON measures and related health information literacy;
- Factor analysis towards trustworthiness of online health information.

## 3.1 Health Information Seeking Process

As shown in Figure 1, the participant actively seeks for health information when a health related issue occurred, or they passively received pushed information from well-known information sources, such as top websites or official organizations. The take-for-granted trust on experts and large public hospitals arguably affects young people's assessment of credibility of online health information, and excludes other information sources such as private sectors.

They tend to believe the health information immediately as long as they are from large hospitals with good reputation (i.e. Three grade hospital in China), or the online health information were provided by the physicians who have been working in those large hospitals.

*"The most credible website is hospital website."*

*"Large hospitals hired these doctors have verified their professionalism."*

*"That information is provided by professional doctors and it can't impact my health."*

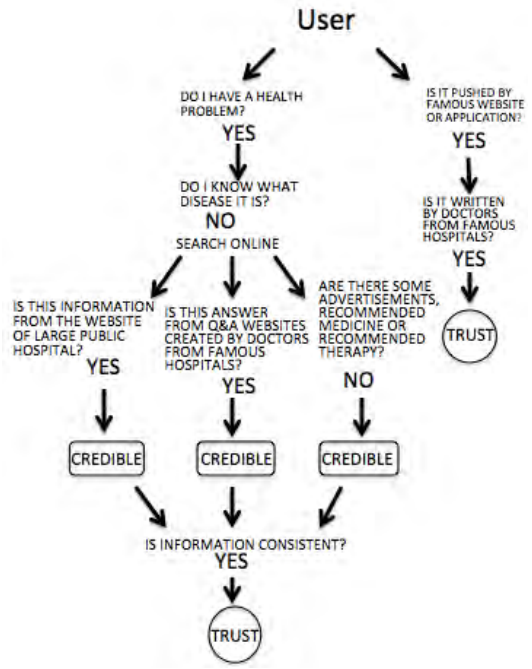



Figure 1: Health information seeking process.

## 3.2 Perceptions on HON Measures

Not very surprisingly, nearly all of the participants have not heard of the HON code measures previously due to their health information literacy limitations. On the other hand, the HON code has been well established in the western world prior its translation into Chinese languages. Regarding to the detailed eight measures that forming the HON's ethical principles; the authority, the purpose, confidentiality and the frequency of the occurrence of advertisement could have affected their judgement and perception towards the online health information's credibility. Moreover, senior students showed capability of more sophisticated information literacy and willingness to explore complicated measures for online health information.

*"If I suffer from health problems which I can't solve with common sense, I must go to the hospital. I don't trust commercial recommendations."*

*"Heavy advertisements are for profit only. They may exaggerate the issues."*

### 3.3 Factor Analysis on Trustworthiness

Young people pay more attention to well-known healthcare providers. They tend to neglect the content or design of health information. Therefore, their take-for-granted trust on the experts affects the youth's choices greatly.

Figure 2 and Figure 3 summarized the subdomains of two important factors, "trust" and "credibility", from related studies.

"Trust/trustworthiness" and "credibility" have been identified as two factors that affect each other, which therefore confused the formation of trustworthiness towards online health information and the define of those two concepts. Nevertheless, other factors, such as brand, content, verification, style, design look, authority, expertise, real-world feel, are identified as factors for the youth to decide the trustworthiness towards online health information.


Figure 2: Subdomains of "Trust".

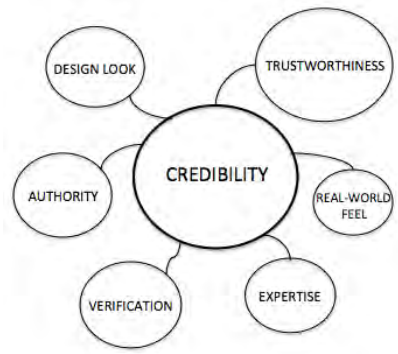



Figure 3: Subdomains of "credibility".

*"I read these articles carefully because they are filtered by editors who should be responsible."*

*"I use Internet to search what the disease is and which hospital and department I should go to.*

*"When I browse the information I get from search engine, all of them are consistent."*

## 4 DISCUSSION

The preliminary findings from our work is summarized as following:

- The participants paid more attention to the trustworthy information providers, i.e. leading hospitals, doctors and web editors, which contribute to the credibility of health information.
- The take-for-granted trust on experts and large public hospitals arguably affects young people's assessment of credibility of online health information, and excludes other information sources such as private sectors.
- The lack of provision of health information from well-known organizations and consumers' deficiency in health information literacy restricts the health information seeking and utilizing behaviour.

Sillence et al. (2007) noted that the design and content of information were major impact factors within the dimension that users utilized to assess the credibility of online health information. The reason behind this difference between previous study and ours will be investigated in future large-scale studies. In addition, the result of pilot study indicated that when conducting searching experiment, the task should be set about a common symptom which young people are more likely to suffer from. The active acquisition and passive acquisition should also be taken into consideration to augment the research methodologies.

## REFERENCES

China Internet Network Information Center, 2016. Statistical Report on Internet Development in China.< http://www1.cnnic.cn/IDR/ReportDownloads/201604/P020160419390562421055.pdf>.

Dobransky, K. Hargittai, E., 2012. *Inquiring Minds Acquiring Wellness: Uses of Online and Offline Sources for Health Information*. Health Communication, Vol. 27 No. 4, pp. 331-343.

Everard, A., Galletta, D. F., 2005. *How Presentation Flaws Affect Perceived Site Quality, Trust, and Intention to Purchase from an Online Store*. Journal of Management Information Systems, Vol. 22 No. 3, pp. 55-95.

Fogg, B. J. et al, 2003. *How Do Users Evaluate the Credibility of Web Sites? A Study with Over 2,500*

*Participants.* In proceedings of the 2003 conference on designing for user experiences, San Francisco, California, pp. 1-15.

Frankland, J., Bloor, M., 1999. *Some issues arising in the systematic analysis of focus group material*, IN Barbour, R., Kitzinger, J., (Eds.) *Developing focus group research: politics, theory & practice.* London, Sage, 154.

Gray, N. J. et al, 2005. *Health information-seeking behaviour in adolescence: the place of the internet.* Social Science & Medicine, Vol. 60 No. 7, pp. 1467-1478.

Harland, J, Bath, PA., 2007. *Assessing the quality of web sites providing information on multiple sclerosis: evaluating tools and comparing sites.* Health Information Journal, 13 (3), 207-221.

Holloway, I., 1997. *Basic concepts for qualitative research*, Oxford, Blackwell Science, 121.

Hsu, W-C, Bath, PA., 2008. *Development of a patient-oriented tool for evaluating the quality of breast cancer information on the Internet.* In: eHealth Beyond the Horizon – Get IT There, Proceedings of MIE2008. Volume 136 Studies in Health Technology and Informatics. Editors: Andersen, S.K., et. al. IOS Press, pp.297-302.

Kim, P, Eng, TR, Deering, MJ, Maxfield, A., 1999. *Published criteria for evaluating health related web sites: review.* British Medical Journal, 318 (7184), 647-649.

Rowley, J. et al, 2015. *Students' trust judgements in online health information seeking.* Health Informatics Journal, Vol. 21 No. 4, pp. 316-327.

Sillence, E., et al, 2007. *How do patients evaluate and make use of online health information?* Social Science & Medicine, Vol. 64 No. 9, pp. 1853-1862.

Van Teijlingen, E. R., Hundley, V., 2001. *The importance of pilot studies.* social research UPDATE.

Yu, B. D., Li, S., Liu, TY., 2016. *Exploring young people's perception on the assessment of the credibility of online health information: a case study on university students in china,* In proceedings of the 15th International Conference on WWW/Internet 2016, Mannheim, Germany, pp. 312-314.

# Making Environments Work for People
## *A Novel Approach towards Personal Lifestyle Management Informatics*

Eleni Kaldoudi
*School of Medicine, Democritus University of Thrace, Dragana, Alexandroupoli, Greece*
*kaldoudi@med.duth.gr*

Keywords: eHealth, Personal Health Systems, Healthy Behaviour, Lifestyle Related Disease, Self-engagement, Self-management.

Abstract: This paper introduces a new paradigm for personalized systems used by the citizen for self-management of health and disease: using smart technologies to exploit the health potential of surrounding environments and to support the citizen in decisions related to a healthy lifestyle. This approach proposes meshing the citizen's specific needs for healthier behaviours with what is available to meet these needs in the surrounding environment. Rather than focusing on health monitoring alone, the aim should be to create a healthy digital envelope – a Healthy Place – around the citizen as they move through their everyday lives. The implementation of this approach needs to integrate existing systems for health assessment and environmental predictions; collect personal private data from mobile personal sensors and public data on health content of the environment; design dynamic testable models of behaviour change, that situate the individual within their environment; develop advanced analytics for context understanding and situational awareness that will couple the current goals of the person with what his or her environment can offer; and create personalized decision support services for behaviour change that exploit the current match between a person's needs and the opportunities offered by his or her environment.

## 1 INTRODUCTION

Information and communication technologies have conventionally been used to support disease management. A second generation of interventions addresses personal patient informatics, building the 'quantified self' to increase self-knowledge and autonomy via (large scale) personal data collection. The current trend is for personal devices and applications whose primary purpose is less to enlighten users with information than to urge them to change (Singer, 2015). However, broadcasting generic health messages (e.g. 'do this, don't do that') has relatively modest effects unless the context/environment makes the advocated changes very easy to carry out. Technology today cannot significantly alter physical environments in this respect, but it can alter something equally or even more important: the perceived environment.

In this paper, we propose an innovative approach of meshing the citizen's specific needs and goals for healthier behaviours with what is available to meet these needs in the surrounding environment. Rather than focusing on health monitoring alone or individual medical and behaviour change plans, the proposed approach aims to create a healthy digital envelope – a *Healthy Place* – around the citizens as they move through their everyday lives.

## 2 BACKGROUND

Lifestyle-related diseases are defined as non-communicable diseases and are caused by non-physiological lifestyle factors such as unhealthy diet, physical inactivity, tobacco use, excessive use of alcohol and psychosocial factors e.g. chronic stress and depression, are leading causes of death globally.

Chronic non-communicable diseases such as cardiovascular disease, cancer, diabetes and chronic respiratory disease, were responsible for 36 million deaths (67% of all deaths) in a single year (WHO, 2010). All these diseases are profoundly impacted by lifestyle options including dietary intake, exercise, stress, sleep, and use of alcohol.

Lifestyle changes in these patients can prevent the progress of the disease more successfully than



Kaldoudi, E.
Making Environments Work for People - A Novel Approach towards Personal Lifestyle Management Informatics.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 418-423
ISBN: 978-989-758-213-4

any drug treatment (Spruijt-Metz, 2014; Wilson, 2014; Williamson, 2005; Li, 2014)

During the last decade, different forms of health care delivery have also been considered, based on the specific needs of patients with chronic diseases, such as patient empowerment. Environmental information as provided by smart cities infrastructure has been proposed to augment personal health applications and help citizens choose wisely their whereabouts in the urban environment (Solanas, 2014). Furthermore, the WHO, among others, has promoted the idea of preventing disease through healthy environments (Pruss-Ustun, 2006). To ensure sustainable healthy behavioural change, we must foster and promote environments that can support people in healthier lifestyle habits.

As it is difficult to radically change the environments of developed societies, we propose to change how the environment is perceived. This will be achieved by highlighting, on an individual basis, those aspects of the environment that are most conducive to encouraging and maintaining personalized healthy behaviours. Thus, the aim is to 'change' the place around the citizen into a 'perceived' *Healthy Place*. For the rest of this document, therefore, the term *Healthy Place* is used to refer to a place whose health-related aspects have been highlighted by the proposed technological framework.

# 3 HEALTHY PLACES CONCEPT

For any place of everyday life to turn into a perceived *Healthy Place*, its objects and concepts must be semantically described and linked to healthy habits and values, thus creating an augmented environment for individuals to manage health, lifestyle and disease. To be relevant, this needs to relate to the individual's every day, real world context and be coupled to the personal lifestyle and the medical/behavioural goals of the individual in question.

There are three different types of data that pertain to this approach: private personal data, public data on the ambient surroundings, and health related scientific evidence including predictive models and risk assessment (Figure 1).

Private, individual-level data includes: (a) personal information on health issues, e.g. demographics, allergies, risk factors, etc. as available from a personal health record; (b) real-time information on physical activity, location, and dietary choices; (c) lifestyle related information from the person's social media presence, including attitudes, intention and relations to the community; and (d) via analysis of the above, information on the motivational and emotional state of the person.

Public data includes data on healthy aspects of resources and activities of a place, incorporating commerce, retail, leisure, workplace and community aspects, or the ambient environment. Aggregating public data on life contexts can be driven by stakeholders in commerce, retail, leisure, workplace, community who will publish key data; it can also use participatory sensing approaches.

Major lifestyle related contexts include:

- Food: the ingredients and calorific content of food items are an obvious and important source of data when considering one's health. This data can be used in relation to: managing weight, ensuring an appropriate balance of nutrients is consumed and coping with food allergies. Additional information includes meal preparation processes, e.g. the type of fat used for frying, whether menu items have been in the proximity of nuts and other common allergens.
- Recreational activities: rich description of activities in terms of their work load, difficulty, special requirements, indications and contra-indications for the healthy person at different ages and capacities and when suffering from different health conditions and disease.
- Public Transport: with the emergence of 'Smart Cities' and public Open Data there is a growing tendency for data concerned with transport routes, stop and station locations, timetables and vehicle locations to be publicly available. This data can be used to help with maintaining a certain level of exercise and energy consumption.
- Ambient environment: temperature, meteorological conditions, noise level, air pollution, airborne allergens, etc. pollen counts, as well as meteorological conditions can affect health.

Ground medical evidence and health prediction models can serve as the basis on which the current environment is analysed for the opportunities it offers and threats it presents for the individual. The goal is to dynamically highlight the most suitable attribute of each environment for the individual and deter from any threats this environment may hold, based on the needs of each person. Thus, the same environment is presented differently according to the health condition and requirements of each individual.

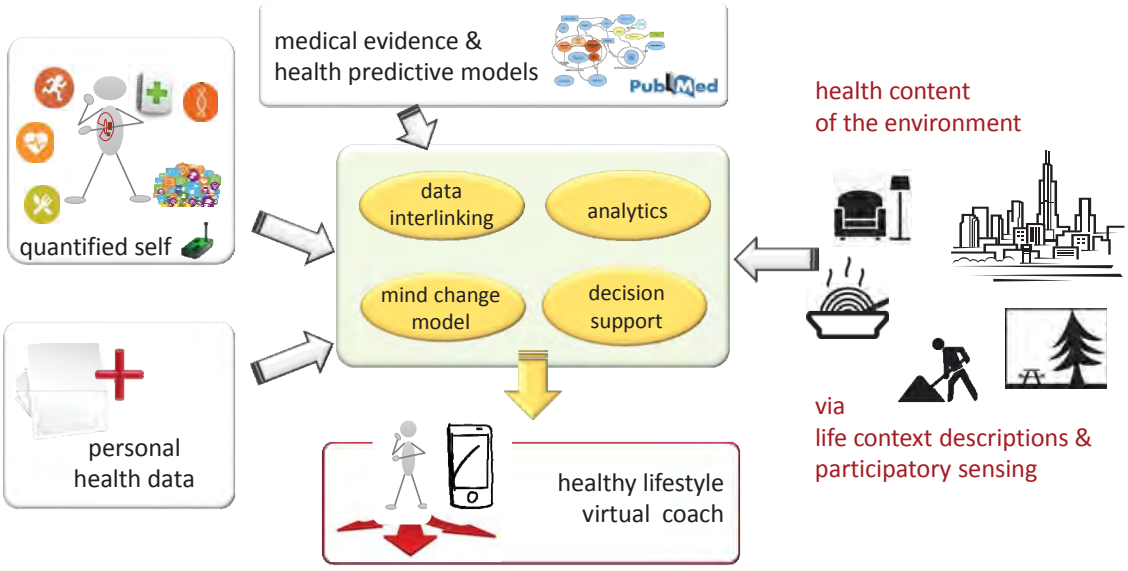



Figure 1: Data aggregation to realize the concept of a *Healthy Place*.

The coupling of public and personal data projected against a behaviour and behaviour change model to create personalized recommendation services for the citizen via a personal application. This can virtually coach individuals by supporting: (a) advanced behaviour and environment awareness; (b) self-monitoring, goal setting, and action planning; and (c) motivational and sustained behaviour change. Thus, individuals can explore personal motives, competences, life goals, preferences, social connections, and other internally-driven motivational elements around physical activity, healthy eating and healthy behaviours in general. These can be coupled with opportunities offered by the current environment. Social intelligence tools can also be used to tap into participants' sense of volition and ownership (as opposed to merely external pressure), confidence and competence (as opposed to self-perceptions of unpreparedness and even failure), and positive social support for their activities and goals.

## 4 OPEN RESEARCH ISSUES

The proposed concept of combining health related content with personal information to create an optimum perceived *Healthy Place* for each individual can be approached by a modular architecture as shown in Figure 2.

Starting from top to bottom, data and knowledge acquisition components acquire personal and public data. Personal data is nowadays readily available via a variety of commercially available wearable and other personal sensors and systems (e.g. personal health records). However, describing the health content of the environment may prove challenging. Participatory sensing can be employed to exploit the crowds and their mobile devices to harvest and enrich the information about the environment and living spaces (e.g. the noise level and the temperature of a place, photos of food items for optical nutritional recognition, etc.).

Information a person creates on social networks may prove to be a significant determinant of behaviour: We are all individuals who are interconnected with other individuals by personal, social, economic and workplace relationships. These connections facilitate communication, can motivate us to different behaviours, and have the potential to support shifts toward healthier lifestyles. Further research is expected to adapt and develop the necessary natural language processing, information retrieval and machine learning methods, combining data mining with semantic technologies and expertise in social media sentiment analysis.

Data enriching and interlinking is of outmost importance to bring up rich data relationships that would help couple personal health requirements with the opportunities in the environment. Thus, graph data repositories (Angles, 2012) are chosen to establish the middle layers of the proposed architecture: a public one for the health context of living spaces and a private for the personal

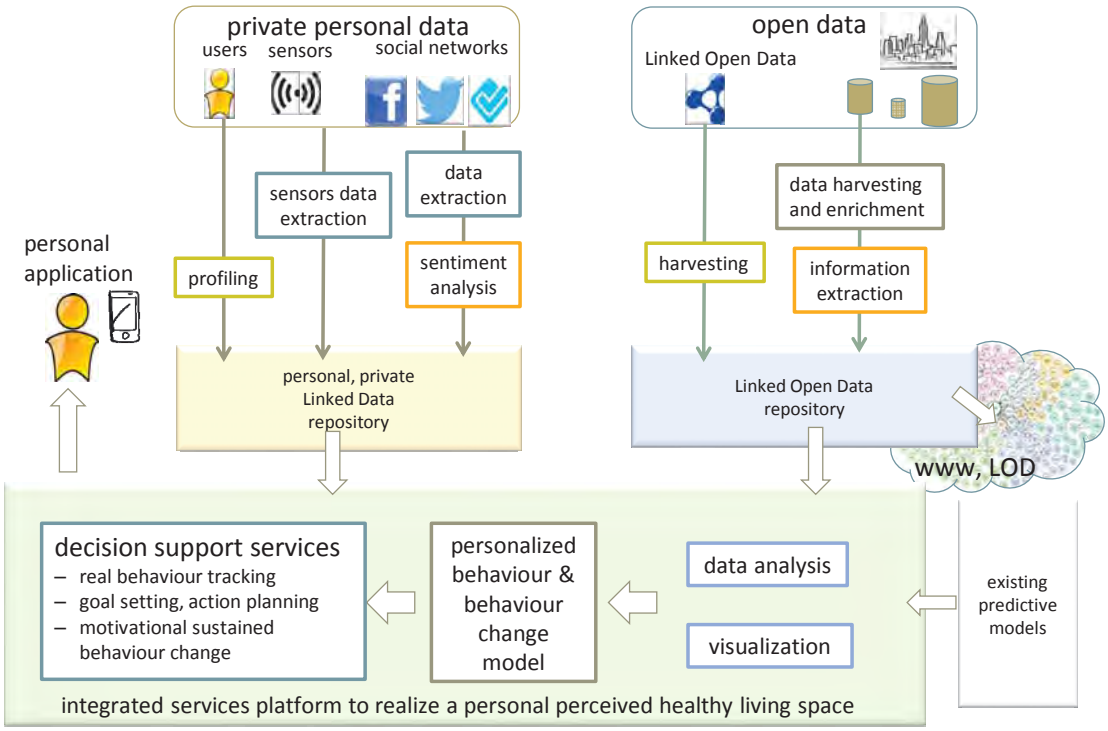



Figure 2: Overview of an abstract framework to realize personal perceived *Healthy Places*.

information. Although a lot of work has been put on releasing semantically rich Open Data, contemporary solutions often fall short of fully exploiting the Semantic Web's potential (d'Aquin, 2008). This shortcoming owes mainly to a) the shortage of adequate knowledge acquisition mechanisms, b) the lack of an environment-based, life and health related semantically integrated approach and c) the slow progress concerning the linkage of user data with the Web of Data.

Despite the proliferation of semantic web data, most published data remains semantically poor (e.g. XHTML, XML, CSV files). To leverage this wealth, research should focus on developing solutions that enable knowledge acquisition, which can be accommodated using different approaches, such as data and information extraction, and sentiment analysis. Applications that are based on traditional database modelling principles suffer from difficulty in capturing evolution of the data model, high software maintenance cost and low reusability. The semantic web achieves the translation of data across boundaries that separate different domains and overcomes these limitations (Feigenbaum, 2007). Further research is required to define and analyse the model and corresponding ontology/schema to describe health context of living spaces. This entails detailed examination and exploitation of existing health and life vocabularies and development of the appropriate Linked Services and service information (Pedrinaci, 2010).

The emergence of social web has led to the generation of user interaction and preferences traces that are often distributed, fragmented and detached (Rowe, 2009). This limitation not only complicates efforts to gather relevant user information, but also weakens the ties between personal data and the web of data. However, it fails to safeguard privacy, given the potential for a determined data-gatherer to integrate diverse data sources to form a detailed picture of an individual's actions and preferences. Concerning the linkage between personal data and the web of data further research is required to integrate personal data and web of data through a single services interface, while ensuring the highest level of privacy protection for individuals.

Support for analysis and visualization of large data sets can in principle be done by aggregation performed in either data space (data reduction) or in visual space (visual aggregation). Scalability is a key challenge in visual analytics as it determines the ability to process large datasets by means of computational overhead as well as appropriate rendering techniques. Often, the huge amount of

data that must be visualized exceeds the limited number of pixels on a display by several orders of magnitude. Currently existing techniques typically focus on a single given data type, e.g., time series or text data, so further research is required to address multiple data perspectives simultaneously.

The wealth of available personal health devices and applications, should be amended by novel applications tracking real-time personalized lifestyle to deduce the person's current real behaviour and how much this deviates from what is a healthy behaviour for this person (especially as specified by the individual in terms of his or her personal behavioural goals) and to determine mind changing actions (i.e., behaviour change through cognitive and emotional determinants). Novel tools are also required to support individuals to engage with self-monitoring, goal setting, personal projects and coping plans. New motivational and sustained behaviour change decision support applications should be devised to allow individuals to explore personal motives, competences, life goals, preferences, social connections, and other internally-driven motivational elements around physical activity, healthy eating and healthy behaviours in general.

A major enabling factor for realizing the personal healthy space via messing public and private health related data lies in the challenge to preserve privacy (Vayena, 2015). Although healthcare data are customarily anonymised to ensure a certain level anonymity (Gkoulalas-Divanis, 2014), they remain susceptible to threats caused by data linkage (e.g., with publicly available data sources) or by background knowledge. Thus, effective measures for preserving privacy must be developed (Viceconti, 2015). Also, patient consent (and its revocation) is recognized as a major limitation in broadly re-using available healthcare datasets for novel big data analytics (Barash, 2015). Additionally, the recent agreement on Commission's EU Data Protection Reform (EU Regulation 2016/679) recognizes that practices have to respect the citizens' rights to (1) easily access their own data, (2) transfer data among providers; (3) have their data deleted when no longer needed; (4) know when their own data have been hacked. Thus, new research is required to extend anonymization algorithms to work on a distributed setting, where multiple parties hold different parts of the data that cannot or are not willing to share in raw form. develop and validate computationally efficient algorithms for detecting complex events in healthcare data streams. This research should also be complemented by novel privacy preserving consent management mechanisms and cryptography enabled techniques for anonymous and unlikable feedback and reward mechanisms to return useful service output to the citizen (e.g. a health prediction or personal health status) or reward the citizen for contributing personal data.

# 5 DISCUSSION

In a radical departure from traditional eHealth, this paper introduces a new paradigm for personalized systems used by the citizen for self-management of health and disease: smart technologies based on existing predictive systems are used to exploit the health potential of the surrounding environment and support the citizen in his/her decisions related to a healthy lifestyle. To achieve this, novel research should address the following:

- integrate existing predictive systems from different domains, namely (a) health risk assessment models, calculating risks based on current health condition; (b) health predictive systems based on environmental factors; (c) behavioural models; and (d) predictive models of environmental parameters, e.g. weather conditions, air pollution and noise levels;
- collect personal data from mobile personal sensors but also collect data on health content of the environment via citizen participatory sensing;
- design dynamic testable models of behaviour change, that situate the individual within their environment and take full account of cognitive, social and emotional aspects;
- develop technology that allows the semantic description of health-related aspects of an environment as well as of health-related aspects of a person's behaviour;
- perform advanced analytics for context understanding and situational awareness that will couple the current goals of the person with what the current environment can offer; and
- deploy personalized recommendation services for behaviour change that, based on personalized predictions, exploit the current match between a person's needs and the opportunities offered by his or her environment.

The goal is to: (a) help citizens manage actively health and eventually adopt and maintain a healthy behaviour, and thus prevent lifestyle related diseases; and (b) make stakeholders in food, commerce, retail, leisure, workplace and community level aware of the healthy (or non-healthy) aspects of the goods, opportunities and premises they offer

to the public and provide them with technology to promote what is healthier for each citizen.

Health cannot be successfully promoted and sustained by health care systems alone – these are naturally focussed on treatment much more than prevention. At the same time, changing environments with health-related goals in mind is extremely difficult – it requires political will and sometimes costly investments; health is a goal that needs to be balanced alongside other priorities, such as prosperity or efficiency.

Changing citizens' awareness of their everyday environment, in the light of their own priorities and goals, creates a new possibility for the prevention of lifestyle related diseases and, indeed, for the co-production of health and reduction of potential chronic, life quality reducing and costly health conditions and complications.

## ACKNOWLEDGEMENTS

The concept presented in this paper was inspired mainly by work done in the FP7-ICT project CARRE (No. 611140), funded by the European Commission

## REFERENCES

Angles, R., 2012. A comparison of current graph database models. In *28th International Conference on Data Engineering Workshops (ICDEW)*, IEEE, 171-177.

d'Aquin, M., Motta, E., Sabou, M., Angeletou, S., Gridinoc. L., Lopez, V., Guidi, D., 2008. Toward a new generation of semantic web applications. *IEEE Intelligent Systems*, 23(3), 20-28.

Feigenbaum, L., Herman, I., Hongsermeier, T., Neumann, E., Stephens, S., 2007. The semantic web in action. *Scientific American*, 297(6), 90–97.

Gkoulalas-Divanis, A., Loukides, G., Sun, J., 2014. Publishing data from electronic health records while preserving privacy: a survey of algorithms. *J Biomed Inf*, 50, 4-19.

Li, D., 2014. Effect of the vegetarian diet on non-communicable diseases. *J Sci Food Agric*, 94(2), 169-173.

Pedrinaci, C., Liu, D., Maleshkova, M., Lambert, D., Kopecky, J., Domingue, J., 2010. iServe: a linked services publishing platform. In *CEUR workshop Proceedings*, 596, 10.

Prüss-Üstün, A., Wolf, J., Corvalan, C., Bos, R., Neira, M., 2006. Preventing disease through healthy environments. World Health Organization WHO, Geneva.

Rowe, M., 2009. Interlinking distributed social graphs. In *Linked Data on the Web Workshop, WWW2009.*

Singer, N., 2015. From knowing yourself, to prodding yourself, *The New York Times*, April 19, BU3.

Solanas, A., Patsakis, C., Conti, M., Vlachos, I.S., Ramos, V., Falcone, F., Postolache, O., Perez-Martinez, P.A., Di Pietro, R., Perrea, D.N., Martinz-Balleste, A., 2014. Smart Health: A Context-Aware Health Paradigm within Smart Cities. *IEEE Comm Mag*, 52(8), 74-81.

Spruijt-Metz, D., O'Reilly, G.A., Cook, L., Page, K.A. Quinn, C., 2014. Behavioral contributions to the pathogenesis of type 2 diabetes. *Curr Diab Rep*, 14(4), 475.

Vayena, E., Salathe, M., Madoff, L.C., Brownstain, J.S., 2015. Ethical challenges of big data in public health. *PLOS Comp Biol*, 11(2), e1003904.

Viceconti, M., Hunter, P., Hose, R., 2015. Big data, big knowledge: big data for personalized healthcare. *IEEE J Biomed Health Inf*, 19(4), 1209-1215.

WHO, 2010. Global status report on non-communicable Diseases.

Williamson, D.A., Stewart, T.M., 2005. Behavior and lifestyle: approaches to treatment of obesity. *J La State Med Soc*, 157(1), S50-S55.

Wilson, D.E., van Vlack, T., Schievink, B.P., Doak, E.B., Shane, J.S., Dean, E., 2014. Lifestyle factors in hypertension drug research: systematic analysis of articles in a leading Cochrane report. *Int J Hypertens*, 2014, 835716.

# Health Information Systems: Background and Trends of Development Worldwide and in Russia

Elena Vaganova[1,2], Tatiana Ishchuk[1], Anatoly Zemtsov[1] and Dmitry Zhdanov[1,2]
[1]*National Research Tomsk State University, 36, Lenin Ave., Tomsk, Russia*
[2]*Diagnostika + LLC ,of.47, 8, Razvitiya Ave., Tomsk, Russia*
*hailun@mail.ru*

Keywords: Medical Informatics, Health Information Systems (HIS), Special Aspects of Development, Opportunities and Challenges, Trends.

Abstract: The paper is to study the background, opportunities, challenges, and trends of development of health information systems in Russia and worldwide. There are two main types of HIS: electronic medical records and clinical decision support. The key areas of their application include patient management, clinical management, diagnostics and treatment, research and education. The development of economic efficiency of HIS is considered to be one of the future research field in medical informatics.

## 1 INTRODUCTION

Health information systems (HIS) belong to IT-industry, which contributes to the global economy providing jobs for IT-specialists and increasing tax revenues from the activities related. The development of medical engineering and technologies in general and HIS in particular is related to changes in the needs of health care industry including steady increase of knowledge in medical field, complexity of the examination, diagnostic, and treatment methods.

## 2 BACKGROUND OF THE DEVELOPMENT OF HEALTH INFORMATION SYSTEMS

Medical informatics as a discipline is still young, in particular when compared with other medical disciplines However approaches to the data processing in medicine and health care have over 50 years of history.

A historical analysis shows major milestones of the development of global medical informatics and HIS:

1959, Robert Ledley and Lee B. Lusted published a widely read paper on diagnostic decision-making appeared in *Science*, in which the authors expressed hope that by using computers, much of physicians' work would become automated and that many human errors could therefore be avoided.

1965 – one of the first clinically-oriented health care Information Systems Technicon Medical Information System was developed as a collaborative project between Lockheed and El Camino Hospital in California.

1967 – Health Evaluation through Logical Processing (HELP) was the first hospital information system to integrate clinical data accumulation and clinical decision support.

1967 – International Medical Informatics Association (IMIA) was established. It has close ties with the World Health Organization (WHO) as a Non Government Organization, and with the International Federation of Health Information Management (IFHIMA).

1968 – COmputer STored Ambulatory Record (COSTAR), an electronic medical record, was developed by the Laboratory of Computer Science at Massachusetts General Hospital between for Harvard Community Health Plan by Octo Barnett and Jerome Grossman.

1960s – first hospital information systems were first introduced. The staff used them primarily for managing billing and hospital inventory. Major work on: signal analysis, laboratory applications, modeling and simulation of some biological processes, databases; first attempts on decision support (diagnosis).



Vaganova, E., Ishchuk, T., Zemtsov, A. and Zhdanov, D.
Health Information Systems: Background and Trends of Development Worldwide and in Russia.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 424-428
ISBN: 978-989-758-213-4

1973 - in the Netherlands at the Free University in Amsterdam the department of Medical Informatics started under the chairmanship of Jan van Bemmel.

1974 - the department of Medical Cybernetics and Informatics was established in the Soviet Union, headed by S.A. Gasparyan.

1976 - The Problem-Oriented Medical Information System, or PROMIS, was designed for maintaining health care records at the University of Vermont by Jan Schultz and Dr. Lawrence Weed, M.D.

1980 - Edward H. Shortliffe founded one of the earliest formal degree programs in biomedical informatics at Stanford University, emphasizing a rigorous and experimentalist approach.

1986 - European Society for Artificial Intelligence in Medicine (AIME) was established.

1970s - 1980s - a shift from a paper-based to computer-based records system; founding most national and international organizations, conferences; attempts to systematize major areas of medical informatics; first specialized schools and courses; principles of clinical and hospital information systems, security and medical data protection; advanced decision support systems – expert systems.

1990s–2000s - medical Informatics consolidates its position as an independent discipline and is mandatory in most medical schools; hospital information systems are implemented in some hospitals, mainly for management; first e-health and telemedicine research; notable progress in data bases, medical imaging; more visible importance and complexity of electronic health record (HER), including confidentiality, data protection, standards etc.

2000 – 2010 - clearer understanding of e-health potential as a specialized industry and business; hidden gaps and difficulties in real implementation: integration and interoperability, modest rate of user acceptance, quality assessment. Clear contour of sub disciplines: bioinformatics, neuroinformatics etc.

## 3 KEY CONCEPTS OF HEALTH INFORMATION SYSTEMS

In recent decades medicine and health care have changed significantly. Major companies (IBM, Cisco, Microsoft, AGFA, GE et al.) are involved in the development of hardware and software solutions for health care. Special attention is paid to standards of digital medicine, HIS and their components. For Russian conditions (a lot of remote parts of the country) implementation of e-health programs including telemedicine systems, networks and data banks is of great current interest.

Intended use and functional options of HIS depend on the territorial level of health care, as well as the special features of a particular health care organization. The main objectives of HIS usage are enhancement of efficiency of treatment (reducing of medical errors), and optimization of diagnosis and treatment expenses including health and clinical management and patient records. The most urgent and challenging task is considered to develop computer-based medical decision-support.

Healthcare information systems, health information systems and hospital information systems are often used today to refer to the same concept. A series of terms such as *computerized patient records*, *electronic medical records*, and *electronic health records, have been* have been mentioned in scientific papers in the evolution of this phenomenon from its early foundations in the 1960s. They are commonly used almost interchangeably.

Thus there are two main types of HIS: *electronic medical records (EMR)* и *clinical decision support (CDS)*.

Electronic medical records maintain patient information and physician notes in a computerized data base. Electronic records allow the provider to track the patient's health over time, read the input of other consulting physicians, or recall his own clinical assessment from a previous day or hospital visit. Clinical decision support provides timely reminders and suggestions to medical practitioners. Decision support may recommend screening tests based on a patient's age and medical conditions, and drug allergy information. Electronic medical records and clinical decision support systems together form the backbone of the hospital information system.

The main application fields and functions of HIS consist of:

*Patient management* (patient registry, scheduling of appointments, admittance and bed control; emergency care; in-patient/out-patient system);

*Clinical management* (hospital releases; medical reports, electronic prescriptions; surgery appointments);

*Diagnostics and treatment* ( lab exams);

*Supplies management* (stockroom; ordering of supplies; pharmacy; current assets);

*Financial management* (accounts payable and receivable; banking control);

*Support services* (hospital infection controls; assets maintenance; vaccine control);

*Research and education* (library; convention center scheduling, recruiting and personnel).

# 4 SPECIAL ASPECTS OF THE DEVELOPMENT OF HEALTH INFORMATION SYSTEMS

HIS is a set of software, hardware, and data for automation of health care processes in medical institutions health and recreation resorts. Besides the above mentioned main application fields and functions a corporate HIS carries out the following tasks:

- maintaining of common information space, intended for immediate access to data;
- improvement of the quality of medical records;
- control of health care quality and reduction of medical errors;
- increasing transparency of a medical institutions;
- constant analysis of economic aspects of health care;
- reduction of time of examination and treatment.

The developers of health information systems have to deal with a constantly changing subject area. The most important sources of these changes are:

- development of social and economic spheres;
- development of medical science;
- the influence of information technologies on patients' behavior (they become more informed) and health management in general.

Unlike most industries, in medicine there are three sides of financial and economic relations: the party that receives services (patient), the party that provides services (medical organization), and the party who pays for services (patient, insurance company, government).

Another special feature of health information which must be always considered is *privacy*.

# 5 OPPORTUNITIES AND CHALLENGES OF HEALTH INFORMATION SYSTEMS

## 5.1 Opportunities

*Economic and administrative efficiency of a hospital:*

- cost savings due to reducing of paper work and errors in billing;
- cost savings on medication due to instant access to comparison of drugs consumed);
- cost savings on laboratory studies (due to access to comparison of total annual costs of laboratories);
- standardization of hospital administration;
- improvement of management decisions due to an integrated information system;
- access to a more complete, accurate and structured documentation of clinical data;
- automatic sorting of data;
- direct access to instant updates, including remote access to a patient's medical history;
- reduction in medical errors due to more accurate data entry;
- continuous remote monitoring of patients;
- access to analysis and interpretation of data that can be used for the study of diseases and preventive measuresin clinical practice.

- *Improvement of patient care quality:*

- access to patient data from other hospitals;
- simplification of administrative procedures;
- processing of medical records and quick results.

*Remote data processing and transmission* - cloud technologies.

## 5.2 Challenges

The development of HIS is a complex socio-technical process, characterized by a *high level of uncertainty*.

*Different needs* of practical health care representatives (regional authorities, chief physicians, doctors, nurses etc.).

*State medical institutions* which purchase devices and equipment have to adhere to very serious *limitations.*

*Long-term implementation* of HIS in the context of constantly changing healthcare conditions and obligations of staff.

*My baby syndrome:* most of the innovation products in the field of healthcare are promoted by developers who are too confident in their project. They may not be able to abandon their project, ignoring its economic efficiency.

*Implementation* of information technologies is *time-consuming*. It is impossible to immediately estimate the efficiency of a freshly introduced HIS. Users need to get used to new tools and new opportunities provided by software or hardware modifications. The value and applicability a particular HIS are constantly changing due to different economic and management reasons.

Each HIS is to some extend unique. Some information technologies may be similar in various aspects, but the functionality may be different.

The success / failure of a HIS depends on compatibility between the developed device and existing realities of current medical institutions.

Two key stakeholders of HIS development, developers and users, may have different versions of a reality. Developers see HIS exclusively from its technical feasibility.

Hospitals are concerned about the costs, payback period, and even the interests of external stakeholders, e.g. the government.

Creation of large regional and national health information systems for the exchange of data regarding patients and specialized medical centers is one of the key direction in health IT. The problem of such projects is they often *do not involve actual participants of the process*. Obtaining and sharing medical data stored in isolated systems require a carefully specified product developed by professionals. At the same time, different regions and countries apply *very different requirements to such complex systems*. The product must have easily and quickly adaptable to these demands and provide additional developmental options.

# 6 FACTORS OF THE DEVELOPMENT OF HEALTH INFORMATION SYSTEMS

The main *aspects* of the development of medical informatics in general are progress in information and communication technologies (including data processing methods), improvement of public healthcare, and constant changes in the needs, requirements and expectations of the society.

In Russia and worldwide there are *factors* which are more likely to influence the development and implementation of health information systems at local and regional levels.

The *state policy* in the field of IT-based management is one of them. In this regard attention must be paid to the problem of *different levels of knowledge in the field of information technologies*. Computer literacy of medical professionals, software and hardware, unification and standardization of primary data, methods of processing and transmission, the possibility of treatment in remote data centers (cloud architecture), data availability for physicians and patients are required.

The worldwide popularization of *cloud computing* and gradual integration of health information systems in *web-applications and mobile devices*, requiring support of international standards, must be noted as another impact factor.

Financial capacity of healthcare institutions directly affects the future of HIS. No matter what potentially interesting and attractive ideas and opportunities are offered in the industry, available financing of each health facility in particular and a region in general must be properly assessed. The costs of the proposed solutions and possible financial support must be considered. Thus, another *important factor* of the development of HIS is the *cost - efficiency ratio* of information technologies.

# 7 CONCLUSIONS: TRENDS OF THE DEVELOPMENT OF HIS

The globalization of IT business is one of the main tendencies. At present, any person (or company) is a potential data consumer. Therefore, even considering tough competition of major manufacturers, the possibilities of IT market are still boundless. The main market players include producers form the US, Japan, France, Britain and Germany, South Korea, Taiwan, Singapore and others.

Implementation of software products in medical institutions for clinical examination and treatment includes mobile networks, virtualization technologies, and telemedicine systems which provide medical professionals with a fast and secure access to necessary data.

The main trends of the decade 2010 – 2020 include: big data approach; cloud computing; social networks on health; certification of educational programs in medical informatics; full interoperability - communication, devices, semantic interoperability; integration of molecular & genetic

data; patient empowerment, involvement of Personal medical record; further steps towards *personalized medicine*, increase of patient safety, reduction of medical accidents and errors; improvement of preventive medicine and reduction of curative medicine; use of portable/wearable devices for monitoring, prediction & prevention; deployment of home monitoring systems and tele-assistance; deeper penetration of IT tools in medical research (modeling & simulation, digital patient etc); advanced decision support systems.

At present in Russia Medical Cybernetics is a full-blown health care specialty. Education is provided by the Siberian State Medical University, Penza State University and the Voyno-Yasenetsky Krasnoyarsk State Medical University, in addition to the education provided by the Pirogov Russian National Research Medical University. In 2000 the Russian Ministry of Health approved a program for this discipline, prepared by the Second Pirogov Moscow Medical Institute.

In Russia, one of the key trends is creation of a united national electronic information system including telemedicine, providing a hot link between medical organizations of various levels aimed at remote consultations of physicians. Distance learning courses and continuous educational programs for health care workers are to be developed.

In recent years, in Russia a growing number of health information systems for automation of health care institutions have been developed and implemented. Russia is a comparatively young participant in the market discovering the achievements of heath informatics, reproduction and restoration of human resources on the basis of new technologies.

## REFERENCES


Reinhold Haux, 2010. International Journal of Medical Informatics 79, p. 599–610

Morris F. Collen, 2015. Chapter The History of Medical Informatics in the United States Part of the series Health Informatics p. 123-206

Nazarenko G. I., Guliev Ja. I., Ermakov D. E. 2005. Book. MedicalInformation systems: theory and practice (Medicinskie informacionnye sistemy: teorija i praktika), 320p.

M. Ngafeeson, 2014. Chapter. Healthcare Information Systems: Opportunities and Challenges (IGI Global, Hershey)

Sonia Valle W. et al., 2011. Journal of Information Systems and Technology Management Vol. 8, No. 1, p. 155-178

Jorgen P. Banslera et al., 2010. International Journal of Medical Informatics № 7 9, p. 637–648

Kossova E. V., Muhin Ju. Ju. 2010-2011 Information systems in healthcare (Informacionnye tehnologii v medicine), p. 124-148.

Peter Littlejohns et al., 2003. British Medical Journal № 326, p. 860–863

Elske Ammenwerth et al., 2003. International Journal of Medical Informatics № 71, p. 125—135

Lebedev G.S. 2011-2012. Information systems in healthcare (Informacionnye tehnologii v medicine) p. 42–62.

Gesev A. V. 2012. Radiotechnics (Radiotehnika)

Zarubina T. V. 2013. Data-measuring and operating systems (Informacionno-izmeritel'nye i upravljajushhie sistemy) vol. 11, № 10. P. 4–8.

Gusev A.V. 2006. Doctor and information technologies (Vrach i informacionnye tehnologii) № 5. p. 32–43.

International Medical Informatics Association. URL: http://imia-medinfo.org/wp/welcome-to-imia-2/

World Health Organization. URL: www.who.int

Agha L. 2014. Journal of Health Economics. № 34. – P. 19–30.

N. S. Fleming et.al. 2011. Health Affairs. – Vol. 30, № 3. – P.481–489.

S. V. W. de Oliveira et.al. // Journal of Information Systems and Technology Management. – 2011. – Vol. 8, № 1. – P. 155–178.

George I. Mihalas, 2014. Acta Informatica Medica 22(1): 37–43.

Hasman A., Mantas J., and Zarubina T. 2014. Acta Informatica Medica. 22(1) p. 25–36.

# A Telecare System on Smart Watches that Communicate with Wireless Bio Sensors

Akio Sashima, Mitsuru Kawamoto and Koichi Kurumatani
*Human Informatics Research Institute, National Institute of Advanced Industrial Science and Technology (AIST),*
*2-3-26 Aomi, Koto-ku, Tokyo, 135-0064, Japan*
{*sashima-akio, m.kawamoto, k.kurumatani*}*@aist.go.jp*

Keywords: Telecare System, Wireless Bio-sensor, Wristwatch-type Device, Smart Watch, Peer-to-peer, Tactile Interface.

Abstract: In this paper, we propose a prototype of telecare system by using wristwatch-type devices, so called smart watches. In the prototype system, the smart watch receives physiological data of a user, such as heart beats and body movements, by communicating with a wireless bio-sensor worn by the user. The smart watch sends the physiological data to the other users' smart watches connected to the Internet. The sensed data are shared among family members in a peer-to-peer manner so as to remotely monitor the physical health status of the other members. We have designed a user interface which visually shows the remote user's current body posture and enables the others to tactilely feel his/her heart beats. We describe the overview of this prototype system and its user interface implemented with the smart watches. We show experimental results of communication performances of the system.

## 1 INTRODUCTION

Recently, wristwatch-type information devices, so called "smart watches", have drawn attention of industrial and research communities. In general, the wristwatch-type information device is regarded as a peripheral device of smartphone because it is a small sensing device attached with the user's wrist to sense her/his activity and shows alerts of the connected smartphone, such as when a mail is received. However, some of them can be used as stand alone devices which do not need to work with smartphones because they have wireless communication modules, such as Wi-Fi and LTE (4G), to independently communicate with other devices on the Internet.

In addition, as a user interface device, wristwatch-type devices have distinguished features from the smartphones. For example, a feature is intimacy of the devices: they are attached with the users wrists and stay close with the users at all times; they can directly provide tactile information for the users by using their vibration. By utilizing this feature, we can create a new service that directly and successively provides the telecare information with care givers. An example of such intimate service is "Mediated social touch," (Haans and IJsselsteijn, 2006). It is an important issue in teleacare researches how we can mediate social touch and tactile information between family members.

In this paper, we describe a prototype of peer-to-peer telecare system by using wristwatch-type information devices instead of smartphones. The system is implemented with small wireless bio-sensors and wristwatch-type devices. Physiological information of a service user, such as electrocardiogram and body movements, are sensed by the bio-sensor and shared with the others' devices in a peer-to-peer manner. Utilizing the above features of wristwatch-type device, we have developed a user interface which visually shows the remote user's current body movements and enables the others to tactilely feel his/her heart beats. The user interface has been designed to realize intuitive communication in family members. First, we describe overview of the system and its user interface. Then experimental results to investigate its communication performances are shown.

## 2 BACKGROUND

Mobile sensor devices have been used in telecare services to remotely monitor statuses of older persons who live alone (Lin, 2012)(Triantafyllidis et al., 2015)(Sashima et al., 2008). The services provide their health and activity statuses, which are obtained by the wearable devices, for their trustworthy persons



Sashima, A., Kawamoto, M. and Kurumatani, K.
A Telecare System on Smart Watches that Communicate with Wireless Bio Sensors.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 429-434
ISBN: 978-989-758-213-4

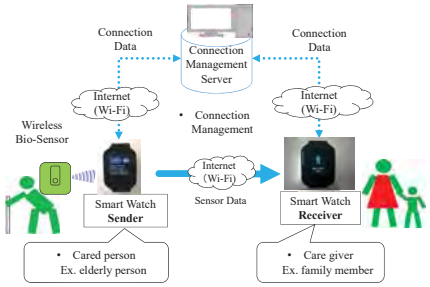



Figure 1: Overview of telecare service for elderly person.

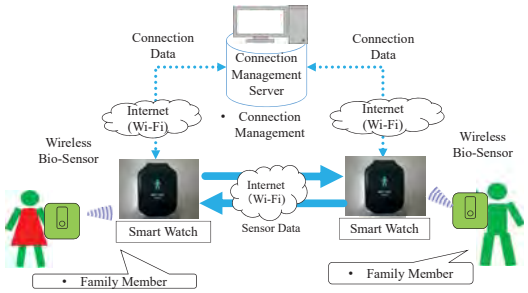



Figure 2: Overview of P2P telecare service for family members.

who take care about them, such as caregivers, family members, doctors, by using smart phones.

So far, in order to implement such telecare services, most of them have been built based on *server-client communication model*; all of sensing data of the older person are sent to a central data server and the server manages and provides the data of the users. However, the approach has a drawback to start up and maintain a practical service in reality. It is about management costs of the physiological sensing data in a central data server; the management costs rise as increasing personal health data in the server because it requires careful handling.

In order to prevent the management issue arising from the server-client model, we have proposed *peer-to-peer communication model (P2P communication model)* for telecare services (Sashima and Kurumatani, 2016). The P2P communication model assumes that the smartphones of the users directly communicate with each other in a peer-to-peer manner. The sensing data are directly sent to a smartphone of a caregiver not going through a central data server so that the management cost is less.

In this paper we apply the P2P communication model to a telecare system on wristwatch-type devices. We have implemented a prototype of telecare system based on the P2P communication model and evaluated the validity of the system. In addition, we propose a user interface that fits for wristwatch-type devices.

# 3 TELECARE SERVICE MODELS

We propose two service models using the peer-to-peer telecare system implemented on wristwatch-type devices;one is for elderly person is shown in the figure 1; the other is for family members is shown in the figure 2.

## 3.1 Telecare Service for Elderly Person

In "telecare service for elderly person" model, we assume that the service consists of a wireless bio-sensor, a connection management server, a wristwatch-type device of sensed user and a wristwatch-type device of remote user. The service user, who wears a wristwatch-type device and a wireless bio-sensor, is an elderly person living in the nursing care facilities which has Wi-FI access points. The other remote users are care givers (e.g., family member, medical staffs). The bio-sensor senses the user's physiological data, such as electrocardiograph data, and wirelessly sends the data to the his or her wristwatch-type device. The data which represents the user's physical status are sent to the other remote wristwatch-type devices connected to the Internet. The network address of the remote devices, which is used for establishing a communication channel among devices, are sent by the connection management server. Hence, it is possible to share the latest physical status of the user by using the wristwatch-type device.

## 3.2 P2P Telecare Service for Family Members

In "P2P telecare service for family members" model, we assume that the service consists of a connection management server, two wireless bio-sensors, and two wristwatch-type devices. Each service user has a wristwatch-type device and a bio-sensor. This model is a symmetrical telecare service model for family members; each member can be take care of other members. The difference between "telecare service for elderly person" model and this model is that the data flows in both directions between the wristwatch-type devices.

# 4 IMPLEMENTATION

## 4.1 System Components

The system consists of the following components.

### 4.1.1 Wireless Bio-sensor Device

A wireless bio-sensor device is shown at the left side in Figure 3. The sensor device is attached to a user's chest by sticking electrodes with a peel-off sticker. The device consists of 5 kinds of sensors (electrocardiograph, 3-axis accelerometer, barometer, thermometer, hygrometer), a lithium ion battery, and a Bluetooth[1] module which communicates with a smart watch. In the prototype, the device continuously senses electrocardiographic data and 3-axes acceleration data of the user and wirelessly sends the data to the smart watch.

### 4.1.2 Wristwatch-type Information Device (Smart Watch)

We have implemented the system on the Android Wear smart watches[2]. The wristwatch-type information device is shown at the right side in Figure 3. It commutates with a connection management server to know the network address of remote user's device. By knowing the latest address of the device, it relays the physiological data obtained by the bio-sensor to the remote user's device in a peer-to-peer manner. Using a graphical user interface of the system, the other user can know the his or her physical status visually. We can configure the system settings e.g., specifying a opponent service user, by the user interface. The system can inform the user's status by control the inner vibrator. In this research, we use the vibrator to provide remote user's heart beats.

### 4.1.3 Connection Management Server

A connection management server stores the network address (IP address and port number) of each user's smart watch, and dynamically updates them when the smart watch connects from a different network environment.

The network address of wristwatch-type information device is dynamically changed according to its network environments. To enable two devices communicate with each other, the network address of the peer is required.

[1]https://www.bluetooth.com/

[2]https://www.android.com/wear/


Figure 3: Wireless bio-sensor (left) and wristwatch-type information device (right).

To solve this issue, we have introduced a connection management server in the system. A connection management server stores the network address of each user's watch devices, and updates them when the device starts to connect to the server. It also tells the updated address to other devices. By knowing the latest address, the devices can communicate each other even if the addresses are dynamically changed.

Notice that the server does not handle sensing data but just initiates the communication of users.

## 4.2 Communication

The current wristwatch-type devices have Wi-Fi communication facilities and directly communicate with the other devices using Internet Protocols, such as UDP. We use the communication facilities to implement the telecare system that can send and receive UDP packets between the devices.

### 4.2.1 Communication Protocol

We adopt a protocol which has been proposed for fast communications of mobile devices (Sashima and Kurumatani, 2016). Figure 4 shows an outline of the protocol. It is a right-weight protocol, which does not have a mechanism to handle lost packets, so as to be suitable for the limited computational resources of wristwatch-type devices. In addition, as it does not wait to receive acknowledge packets, it can immediately send the packet that represents the latest user status. However, as losing packets is a weak point of this protocol, we describe experimental results of communication performances of the system in a later section.

To establish a communication channel between a wristwatch-type information device in a private network and the device in the outside of the network, we apply the NAT traversal technique (Ford et al., 2005) that enable peer to peer communications over a NAT router.

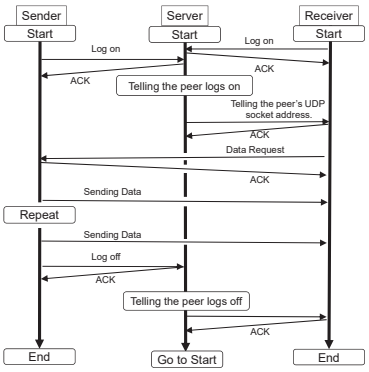



Figure 4: Overview of the communication protocol (Sashima and Kurumatani, 2016).

## 4.3 Data Representation

To share the sensing data between the devices in the service, the data has two data representations for the service models: a) "raw data representation" which includes raw sensing data and b) "abstract data representation" which includes the data analyzed at a wristwatch device connecting to wireless bio-sensor.

### 4.3.1 Raw Data Representation

This representation includes raw data obtained by the bio-sensor for the application that requires to know detailed data of the sensed user. We assume that typical usage of this representation is for medical applications, such as remote monitoring of electrocardiograph. Thus the communication is one way from a sensed user to a care giver, such as medical staff. This raw representation data is synchronously sent at a constant sensing interval. For example, if the sampling period of the sensor is 5 msec, the data is sent at each 5 msec interval.

### 4.3.2 Abstract Data Representation

two way communication assumed in "P2P telecare service for family members" requires more computational power for handling traffics. We have developed the abstract representation to reduce the traffics in the two way communication. The representation includes an abstract data summarized by a data sequence of raw sensing data: a "posture" data of the abstract representation is derived by analyzing a sequence of 3-axes acceleration data; and a "heart beat" data is derived by analyzing a sequence of electrocardiographic data. In addition, the abstract representation data is asynchronously sent when the system detects a change of the status. For example, the data is sent as soon as possible when the posture of the user is changed. Hence, using this representation, the amount of traffic data becomes 10–100 times smaller than using the raw data representation in our prototype system.

## 4.4 User Interface

The wristwatch type device is tactilely attached with the user's wrist and stay close with the user at all times. To utilize the features of the device, we have designed a user interface which directly and successively provide telecare service to the user. The user interface visually shows the users' current body movements and enables the user to tactilely feel other user's heart beats.

### 4.4.1 Graphical User Interface for Showing User's Status

Because the wristwatch-type device has a small display, we have designed a simple graphical user interface showing the status of the remote user. Figure 5 shows display images of the user interface. In the figure, each image shows a physical status of the remote user. Currently, the system classifies the following 6 statuses: standing, moving, laying-up, laying-down, walking, unclassified. The system automatically classifies the status based on the 3-axis acceleration data obtained by a wireless bio-sensor. The sampling rate of the data is 100 Hz. For the classification, we have used a rule-based algorithm using predefined threshold values.

To share the classification with the raw data representations, the devices share the raw acceleration data and do the classification based on the data independently. Because the classification process is done by each device, the analysis results can be slightly different between the devices.

To share the classification with the abstract data representations, the devices connected to the bio-sensor analyses the raw acceleration data and send the abstract posture data to other devices. Because the classification process is done by a device, the same analysis results are shared by the devices.

### 4.4.2 Tactile User Interface for Feeling Heart Beats

Based on electrocardiographic data obtained by the wireless bio-sensor, the system automatically detects the pase of heart beats of sensed user. To detect the pase of heart beats, the system monitors peak locations of the R-wave by calculating differentiation values of the electrocardiographic data, and detect sharply changing points as the peaks of the R-wave.


Figure 5: Graphical user interface of the prototype system. Each image shows a physical status of a remote user: standing (left), moving (center), laying up (right). HR field shows the pase of heart beats of the user.

The sampling rate of the data is 200 Hz. It calculates "beats per minute (BPM)," which stands for the pase of heart beats. The BPM is based on the latest R-R period, and shown at the heart rate (HR) field in the graphical user interface (see Figure 5).

To share the pase of heart beats with the abstract data representations, the device is connected to a wireless bio-sensor. It detects an R-peak from the electrocardiographic data and repeatedly sends a packet to remote user's wristwatch device. We call the packet corresponding to the detected R-peaks R-peak packet. When the remote user's device receives the R-peak packet from sensed user's device, it immediately vibrates for a short time (30 msec). Hence, it vibrates in synchronism with the heart beat of sensed user so as to others can tactilely feel his/her heart beat by the vibrations. Strictly speaking, as there are delay times to receive the R-peak packets and the times fluctuate with network conditions, others can tactilely feel "his/her heart beat like pattern" by the vibrations.

To share the pase of heart beats with the raw data representations, the devices sends the raw electrocardiographic data to the other users. The others calculate the BPM based on the raw data independently.

# 5 EVALUATIONS

We have experimentally evaluated communication performances by measuring the packet loss and end-to-end delays of the prototype system.

The experimental network environments are a Wi-Fi network (LAN) in our laboratory. We have measured the communication performance of two devices which connect to the same Wi-Fi access point that might be used by the other users. To prevent the effects of the sleep mode of the devices, we have applied power to USB ports of them.

We have evaluated communication performance in the following two conditions assuming the different service models: *one-way communication* used for "telecare service for elderly person," and *two-way communication* used for "P2P telecare service for family members." For the above two conditions, we have compared the performance between the abstract data representation and raw data representation. We have analyzed the performance based on the data recorded by each device in each condition. We have used the data recorded within ten minutes from the starting time of the communication.

## 5.1 Packet Losses

Experimental results about packet losses between the wristwatch-type devices are shown in Table 1. Total samples means a total number of the data sent by the devices. The "condition" column of Table 1 shows combinations of communication styles and data representations. For example "Two Way/Raw" stands for two way communication with raw data representation. Using the abstract representation, the total numbers of data became small in this experiment because the data is sent asynchronously when the user changed his or her posture.

While many packets were lost by using the raw representation especially in the two way communication, few packets were lost by using the abstract representation. Because the wristwatch-type device has limited computational resources, the abstract representation with asynchronous communication is more suitable for the prototype than the raw representation.

## 5.2 End-to-end Delay

We have measured delay times of the communication in the conditions. The results are shown in Figure 6. Before the experiments clocks of the devices are synchronized based on a clock of the connection server. The average delay time of "Two Way/Raw" condition is omitted in the graph because it took for a long time

Table 1: Experimental results of packet losses.

| condition | total samples | lost samples | loss rate (%) |
|---|---|---|---|
| One Way/Raw | 180000 | 677 | 0.4 |
| One Way/Abstract | 870 | 0 | 0 |
| Two Way/Raw | 360000 | 149845 | 41.6 |
| Two Way/Abstract | 3267 | 4 | 0.1 |

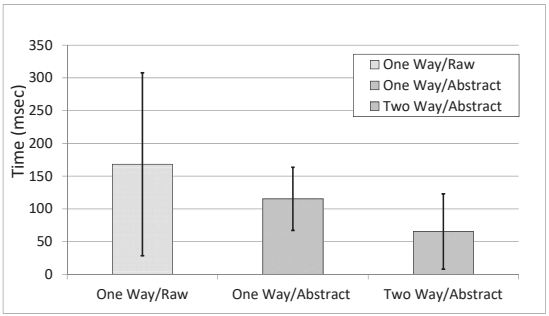



Figure 6: End-to-end delay times (msec).

(over 30sec). Average delay times of the other conditions are under 200 msec. Although more careful investigations are required in various network conditions, we believe that the result shows the feasibility of the telecare services by using wristwatch-type devices.

## 6 CONCLUSIONS

We have described a prototype of peer-to-peer telecare system by using wristwatch-type information devices not using smartphones. Physiological information of service user, such as electrocardiogram and body movements, sensed by the wireless bio-sensor are shared among the service users' wristwatch-type information devices in a peer-to-peer manner. To realize intuitive communication between family members, we have developed a user interface which visually shows the sensed user's current body movement and enables the others to tactilely feel his or her heart beats. We believe that the system opens up a new design space in which all users in a social group can take care of each others at the same time.

## ACKNOWLEDGEMENTS

This work was supported in part by JSPS KAKENHI Grant Number 26330125.

## REFERENCES

Ford, B., Srisuresh, P., and Kegel, D. (2005). Peer-to-peer communication across network address translators. In *Proceedings of the Annual Conference on USENIX Annual Technical Conference*, ATEC '05, pages 13–13, Berkeley, CA, USA. USENIX Association.

Haans, A. and IJsselsteijn, W. (2006). Mediated social touch: a review of current research and future directions. *Virtual Reality*, 9(2-3):149–159.

Lin, C.-F. (2012). Mobile telemedicine: A survey study. *J. Med. Syst.*, 36(2):511–520.

Sashima, A., Inoue, Y., Ikeda, T., Yamashita, T., Ohta, M., and Kurumatani, K. (2008). Toward mobile healthcare services by using everyday mobile phones. In *Proceedings of the First International Conference on Health Informatics, HEALTHINF 2008, Funchal, Madeira, Portugal, January 28-31, 2008, Volume 1*, pages 242–245.

Sashima, A. and Kurumatani, K. (2016). Towards a peer-to-peer communication model for mobile telecare services. In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2016) - Volume 5: HEALTHINF, Rome, Italy, February 21-23, 2016.*, pages 542–549.

Triantafyllidis, A., Velardo, C., Salvi, D., Shah, S., Koutkias, V., and Tarassenko, L. (2015). A survey of mobile phone sensing, self-reporting and social sharing for pervasive healthcare. *Biomedical and Health Informatics, IEEE Journal of*, PP(99):1–1.

# Designing a Social Machine for the Heart Manual Service

Vanessa Hanschke[1], Areti Manataki[2], Cristina Adriana Alexandru[2], Petros Papapanagiotou[2], Carolyn Deighan[3], Louise Taylor[3] and Dave Robertson[2]

[1]*Microsoft, Milan, Italy*

[2]*School of Informatics, The University of Edinburgh, Edinburgh, U.K.*

[3]*The Heart Manual Department, NHS Lothian, Edinburgh, U.K.*

*vhanschke@aol.com, {a.manataki, pe.p}@ed.ac.uk, calexan4@staffmail.ed.ac.uk,*
*{carolyn.deighan, louise.taylor}@nhslothian.scot.nhs.uk, dr@inf.ed.ac.uk*

Keywords: Social Machine, Heart Manual, Self-management Rehabilitation, Participatory Design.

Abstract: Social machines are emerging as a focus of research within the field of informatics as they begin to become the central administrator of our everyday communications. The difficulty of applying such systems to specialised contexts, such as healthcare, calls for guidelines on how to design them, so that they become truly useful. In collaboration with the Heart Manual Department, this project is an attempt at finding suitable methods for designing social machines in a healthcare context. It suggests that adopting a participatory approach where stakeholders are active, equal participants throughout the design process leads to a more usable, likeable, and thus more successful social machine. We describe the process of designing a social machine for the Heart Manual service, in which requirements were elicited through various participatory design methods and a proof of concept evaluation was carried out with a prototype. The prototype was received largely positively and scored highly on the System Usability Scale, indicating the success of the proposed methodology.

## 1 INTRODUCTION

The Heart Manual (The Heart Manual Department, 2016) is the UK's leading home-based supported self-management programme for individuals recovering from acute myocardial infarction and/or revascularisation. With the help of a team of facilitators (typically nurses, psychologists or general practitioners), who are based in different locations nationally and internationally, it guides patients through a series of sessions which empower them to improve their lifestyle. Apart from the initial training that facilitators receive, there is no infrastructure for community building, and knowledge exchange takes place in an ad-hoc fashion, with individual facilitators contacting the Heart Manual (HM) team to clarify their questions.

Social machines (Hendler and Berners-Lee, 2010) can provide an appropriate infrastructure for such interactions. Through a closer interaction between humans and machines, such systems can support communication and the sharing of experience, and contribute to a sense of community. Despite recent research interest in this topic, there are still no best practices for designing successful social machines.

In this paper, we propose a participatory approach to the design of healthcare social machines. We posit that the involvement of users throughout the design process can address usability and likeability aspects, which are key success factors for health technologies. In particular, we describe our experience in employing participatory design methods to develop a social machine for the HM facilitators. Background information is provided in Section 2 and our methodology is introduced in Section 3, followed by an account of requirement elicitation in Section 4. We next present a first prototype created to collect quick feedback (Section 5) and how this was further adapted and evaluated by HM facilitators (Section 6). We conclude with an overview of lessons learnt and future work.

## 2 BACKGROUND

The Heart Manual (The Heart Manual Department, 2016) is a home based cardiac rehabilitation programme, supported by trained facilitators and evidenced by three randomized control trials (Clark et al., 2011). It consists of six weekly sessions that include education, exercise, relaxation and stress management. The programme was digitised in 2015



Hanschke, V., Manataki, A., Alexandru, C., Papapanagiotou, P., Deighan, C., Taylor, L. and Robertson, D.
Designing a Social Machine for the Heart Manual Service.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 435-440
ISBN: 978-989-758-213-4

(Deighan et al., 2015), and there is now the opportunity to extend it with recent advances in large and distributed open systems, such as social machines.

Social machines are technology-enabled social systems, seen as computational entities governed by both computational and social processes (Hendler and Berners-Lee, 2010). They are defined as "Web-based socio-technical systems in which the human and technological elements play the role of participant machinery with respect to the mechanistic realization of system-level processes" (Smart et al., 2014). Social machines are a recent research theme, with related work focusing mostly on analysing existing examples, such as Facebook, Wikipedia and Stack Overflow (Shadbolt et al., 2013; Smart et al., 2014). Literature on designing social machines is relatively sparse. Donath (Donath, 2014) approaches this topic in terms of conceptualising communities and strengthening ties, while Murray-Rust and Robertson (Murray-Rust and Robertson, 2015) suggest using existing social machines that bring together designers and developers to generate new social machines. Both approaches push for an understanding of the users and a close tailoring of the social machine to its participants. Yet we are still lacking guidelines for practical requirement elicitation for new social machines.

In healthcare, a number of social machines aim to bring patients together to help them live healthier lives. PatientsLikeMe.com, for instance, connects patients with similar conditions, allowing for peer support and knowledge sharing, while the Fitbit and Nike+ FuelBand online communities operate as behavioural interventions that complement activity trackers. However, the majority of such machines lack a clinical evaluation, and there are concerns around confidentiality and patient self-diagnosis.

Similarly to the proposed HM social machine, other machines connect clinicians rather than patients (e.g. doc2doc.bmj.com). A recent review (Rolls et al., 2016) explores how social media have helped health professionals worldwide create virtual communities where they exchange knowledge and network. However, the paper does not give a comprehensive review of design methodologies. The closest example in this literature were the trials for the Midwifery Forum (Brooks et al., 2004). However, the participatory design mentioned in the paper was limited to prototype evaluation, rather than collaborative design.

# 3 METHODOLOGY

The lack of guidelines for designing social machines is a considerable gap, especially in the healthcare context, where technology uptake is relatively slow. Including the capabilities of the users in the interface design can have a significant effect on its usability and, thus, the success of a healthcare social machine.

We thus adopt a participatory approach to the design of a social machine for the HM facilitators. Participatory design is a user-centred design methodology, in which the end-user is made a full participant of the design process, typically by interacting with mock-ups, prototypes and other tools that represent developing systems (Schuler and Namioka, 1993; Simonsen and Robertson, 2012). We hypothesise that *participatory design is an effective method for the design of a social machine of health services with respect to system likeability and usability, as measured by a small-scale evaluation of a web-based prototype*.

An iterative development process was followed, consisting of three main phases: requirements elicitation, low-tech paper-based prototype design and high-tech wireframe prototype development. Inspired by the spiral model of software development, each phase included a repetition of steps, namely planning objectives, collecting information from users, analysing their feedback and creating a next level prototype (or list of requirements, in the case of Phase 1).

# 4 ESTABLISHING REQUIREMENTS

We utilised Shadbolt's constructs for classifying social machines (Shadbolt et al., 2013; Smart et al., 2014) to establish a set of questions which provide a framework for gathering requirements for a social machine. The questions devised cover the tasks and purpose of participation, participants and their roles, as well as motivation and incentives (see Figure 1).

Three main methods were used for establishing requirements, described hereafter:

**Training session:** We first attended a two-day training session that prepares healthcare professionals to become facilitators of the HM programme. This allowed us to gain a clearer understanding of their role and the challenges they face (e.g. cultural issues that might impede patients from following dietary restrictions). To address such challenges, the session also made apparent the usefulness of a space for facilitators to discuss their opinions and experiences, share good practice and provide peer support, especially in the case of facilitators working in remote areas.

**Brainstorming session:** This was organised in the form of a "Future Workshop" towards idea generation (Simonsen and Robertson, 2012). Future workshops are common in participatory design and consist

| Questions | Answers for HM social machine |
|---|---|
| **Tasks, purpose and context of participation** | |
| What creative content is going to be produced? | sharing of good practice; peer support; updates |
| How visible is every individual contribution? | individual, named posts |
| Is there going to be some form of rating of the content? | yes: favourites; contributions marked as helpful; reporting |
| Will tasks be domain specific? | yes: directly related to or surrounding the HM |
| Does the machine owner (NHS, HM Programme) benefit from participation? How? | yes: touching base with facilitators; circulating surveys |
| Are activities pre-defined or participant-defined? | facilitators start discussions; HM team hold chaired live sessions |
| Is there variation in tasks? | yes, but it should be easy to find and access tasks |
| What media is used for participation? (Web, desktop, mobile devices...) | desktop computer |
| Is the participation done in a mobile context or in a physical location relevant to the service (i.e. at work/hospital)? Are the facilitators engaging with this tool during working hours or in their spare time or both? | at work during working hours |
| **Participants and roles** | |
| Who are the participants? How many are expected to join? How many will be active constantly? | HM facilitators (especially isolated ones); 935 currently active; weekly or monthly use |
| Can users be solely passive? | possibly, but preferably not |
| How autonomous are they? Or is their participation dependent on others? | activity and interest may depend on other members' activity |
| Are they anonymous? | yes, if desired, for patient confidentiality purposes |
| What is the hierarchical structure? | HM team: supervise and mediate; HM facilitators: participate |
| Are roles clearly separated? | yes: admin have more power |
| **Motivation and incentives** | |
| What is the intrinsic motivation of participants? Gaining/sharing knowledge? Benefiting the Heart Manual? Widening their expertise? Solving problematic situations? | revalidation and understanding how other facilitators work with the manual; ask and answer questions; |
| Will there be extrinsic rewards? (payment, status) | mixed opinions about a badges system |
| Will the tool also have autotelic elements? | yes: enjoyable aesthetics; feeling part of the community |

Figure 1: Questions to guide requirement gathering for a social machine, and answers for the HM social machine.

of three stages: i) critiquing the present, ii) envisioning the future and iii) implementing - moving from present to the future. Four clinicians participated in the workshop: a health psychologist, an assistant psychologist, a specialist nurse who is also an HM facilitator, and the lead of the HM team. Following the Future Workshop structure, we first focused on existing communication between clinicians, and then we discussed the goals of the HM social machine. Based on these goals, the participants were then asked to devise dream solutions, unrestricted by technical knowledge or possibilities. Three examples of social machines were then presented to them: Facebook, doc2doc and Stack Overflow. These were carefully chosen, so as to prompt discussions around different forms of communication, privacy and confidentiality regulations, as well as user reputation, respectively. This presentation initiated a new discussion about the dream solutions, leading to a jointly ordered list of requirements.

**Telephone interviews:** Three semi-structured interviews with additional facilitators were organised so as to verify some of the ideas collected previously, as well as to clarify points where participants had differing opinions. For example, participants gave their opinion about the motivation for a social machine and the associated concerns, the need for private groups or private messaging, anonymous posting, etc.

The brainstorming session and telephone interviews were recorded (with the written consent of the participants), transcribed and analysed following top-down thematic analysis, as guided by the questions in Figure 1. Outputs included explicit answers to these questions (presented in the same figure), leading to a list of functional and non-functional requirements. Among functional requirements, we distinguish the following elements: discussion forum, quizzes, surveys, blogs, events and notepad. Posting anonymously and tagging content as helpful were deemed desirable, but setting up groups or private messaging were not. The list of non-functional requirements included high levels of security and privacy, low maintainability and the option to report forum abuse.

# 5 FIRST PROTOTYPE AND ITS EVALUATION

Following best practices in participatory design (Simonsen and Robertson, 2012) and based on the requirements gathered, we created a low-fidelity, paper-based website prototype, so as to quickly obtain feedback and generate ideas at an early stage of the design process, before committing to design decisions that would be harder to change later on. A set of pages were designed to capture the main functionality of the website: i) homepage, ii) discussions, iii) submitting a discussion (see Figure 2) and iv) user profile.

A **prototyping session** was organised to collect feedback, with the partaking of the four participants from the brainstorming session. Each participant was questioned individually around two activities. In the first activity, they were given a set of cut out items (e.g. buttons and taskbars) which they could arrange

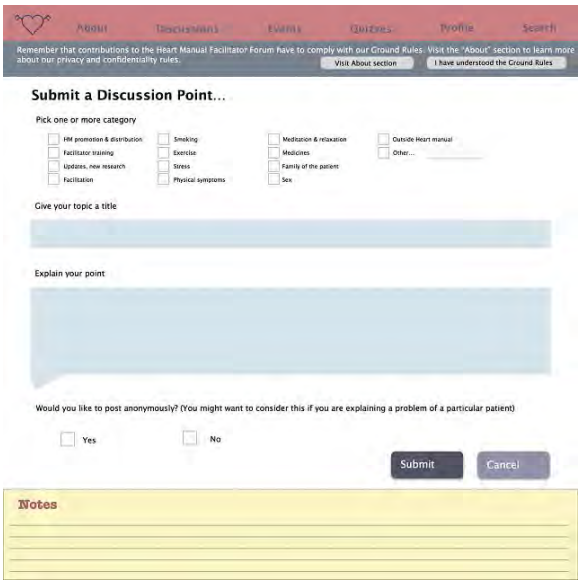



Figure 2: Paper-based prototype for submitting a discussion.


Figure 3: The home page in the final prototype.

on a sheet of paper to reflect their idea of the homepage. In the second activity, they were presented suggestions for the design of the screens to elicit opinions and given screen print-outs to annotate. Topics discussed included the notes facility and the presentation of ground rules for participating in the discussions, so as to avoid misinformation and breach of confidentiality. The prototyping session was audio-recorded, handwritten notes were taken and any visual material created was documented with photographs. These resources were analysed qualitatively using a combination of top-down and bottom-up thematic analysis. In general, it was found that participants were pleased with the layout that was presented to them. Opinions were evenly split on the presentation of the ground rules and on a simple versus busy look of the homepage. A pop-up window and a busy feel, respectively, were chosen by the designer for the next version of the prototype, which would be further evaluated, thus reducing the associated risk.

## 6 FINAL PROTOTYPE AND ITS EVALUATION

A hi-fi web-based prototype was developed by adapting the first prototype given the feedback gathered. It was presented on a browser and made from linked screens to generate a feel of the interactions that would take place (see Figure 3). Its main difference to the first prototype was that users could interact with the system and see its dynamics in operation. The prototype was not fully functional, but a proof of concept. It was designed to work for certain scenarios that were set up for the evaluation: 1) starting a discussion, 2) entering a discussion point, 3) marking a contribution as helpful, 4) reporting a contribution as harmful, 5) entering an event, 6) adding items to favourites, and 7) accessing favourites.

The prototype was tested with seven potential users (five members of the HM team and two facilitators) for usability and likeability. Three of the HM team members had participated in the first prototyping session, while the remaining two had not been involved in any of the prior activities. The two additional facilitators had been interviewed for establishing requirements. The evaluation was divided in two main activities (i.e. think aloud and questionnaire), while a survey was set up for carrying it out remotely.

**Think Aloud and Short Structured Interviews:** For this part of the evaluation, participants were presented with the prototype on a laptop. They were given four sets of tasks: i) submit a new discussion, ii) find a discussion, mark it as a favourite, mark and report a contribution, iii) find an event and mark it as a favourite and iv) revisit favourites. The participants were asked to "think aloud" while completing these tasks, i.e. explain their thought process, so as to make explicit where they were stuck and why, what they were looking for, what their expectations were, etc. Following their think aloud, participants gave their opinion about the presented features and their intuitiveness.

**Questionnaire:** Participants were next presented with a questionnaire around usability and likeability, which included three sets of questions: i) SUS scale (Brooke, 2013), ii) how often (in a scale from 1,

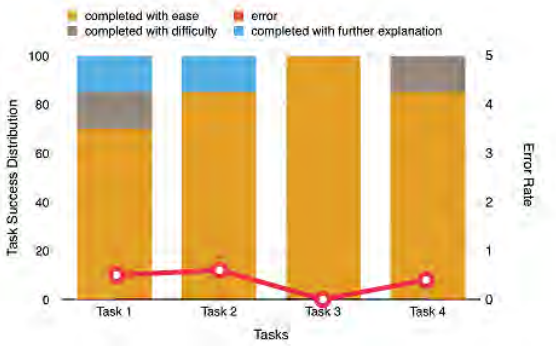



Figure 4: Accuracy and completeness in Think Aloud tasks.

i.e. never, to 5, i.e. very often) participants thought they would use certain features and iii) whether they thought certain aspects were particularly enjoyable, unenjoyable or concerning and why.

**Survey:** Participants who performed the evaluation remotely were emailed an online survey containing the same questionnaire and Think Aloud tasks, and a link to the prototype. They were also asked to put remarks in comments if they were stuck at any point, especially if they were unable to complete the task.

The five participants of the HM team carried out the different evaluation activities in the presence of the researcher. One of the facilitators completed the evaluation using the survey, while the other completed it remotely, with the researcher on the phone and watching the screen navigation in real time.

The Think Aloud sessions and structured interviews were audio recorded (with the participants' permission), transcribed and, together with the survey data, analysed qualitatively using a combination of top-down and bottom-up thematic analysis. Interview or on-line survey replies on intuitiveness were attributed a ranking from 1 to 5. During the Think Aloud session, the time for completing each task was recorded, and notes were taken about whether it was completed successfully and the number of errors. These data were analysed quantitatively with regards to efficiency, task accuracy and completeness. Questionnaires were also analysed quantitatively.

Figure 4 presents the accuracy and completeness measured during the Think Aloud session. All participants were able to complete all tasks independently, except for one participant who required further explanations on two tasks. The errors observed had to do with misunderstanding instructions and with the fact that the search bar lacked functionality at this stage.

Regarding efficiency, the average time in seconds for each of Tasks 1-4 was: 57.4 (SD 57.4), 67.4 (SD 43.5), 21 (SD 8) and 35 (SD 19.4), respectively. Note that the standard deviations here are inevitably large, given the small number of participants recruited.

Based on participant remarks and answers during the Think Aloud session and the short structured interviews, all tasks scored highly on intuitiveness: the average intuitiveness score (on a scale from 1 to 5) for Tasks 1-4 was 4.6, 4.4, 5 and 3.7, respectively.

The average SUS score was 86.4 (SD 18.02), indicating "excellent" satisfaction, according to published grade rankings of SUS scores (Bangor et al., 2009). With the exception of one participant who gave a score of 50 (i.e. "ok"), the rest of the scores ranged from 80 (i.e. "Good") to 100 (i.e."Best imaginable").

The results to the question of how often participants would use certain features ranged from 3 (i.e. every now and then) to 4.43 (i.e. between regularly and very often). The former was assigned to the use of the notepad for keeping personal notes, while the latter to reading discussions and looking up events.

A variety of features were found to be enjoyable. Discussions, in particular, were mentioned by five out of the seven participants. For instance, a facilitator mentioned: "Starting a discussion. I liked that. And I liked it because if I know I've got a thought in my head and I can put it out there and see what my colleagues and counterparts have. So it can validate the ideas I might have". Three facilitators mentioned that the favourites feature also seemed very useful. No feature was pointed out as unenjoyable or concerning.

## 7 CONCLUSIONS

Based on the quantitative measures presented in Section 6, the hi-fi prototype was a great success, performing well on both satisfaction and effectiveness. The SUS score was very positive, corresponding to an "excellent" usability grade. Only one user was unable to complete two tasks independently, totalling an overall average of 92% success per user. The overall error rate was low, with no errors on many tasks and less than one error on average for tasks with errors. Some of the errors, such as accessing the search bar (which was not yet functional), are easily resolvable. The efficiency of the tasks can only be compared between each other, as we do not have an absolute measure of how long these should take. A learning effect between Tasks 1 and 3 can be noted, as they take a similar amount of clicks, but the latter is faster.

The system seemed to have a high level of acceptance, as most features were marked to be used between "every now and then" and "very often". This shows that the features proposed in the prototype were in line with the expectations of the participants. This is also demonstrated by the fact that discussions were highlighted as a particularly enjoyable feature.

This interdisciplinary project highlighted several themes which we expect to be recurring when designing social machines for health professionals, given the general characteristics of the domain. Firstly, designing for time constraints is crucial, given that clinicians are very busy. Hence, an easily learnable interface is recommended, where all features are directly accessible on display. Security is another important theme, and access control aspects should be carefully thought out. Furthermore, social machines that support knowledge sharing should allow users to identify misinformation, which is critical in healthcare.

The choice of participatory design allowed us to identify these themes and concerns, which might not have been addressed without input from the users. Our overall experience of using participatory design in healthcare was greatly positive. Participants were very motivated and gave thorough and careful feedback. On the other hand, recruiting participants was the biggest issue, as healthcare professionals are generally very busy and communication channels tend to be controlled. We worked around this issue by ensuring that different types of stakeholders were involved in the study, from experienced to new facilitators and from isolated workplaces to busy hospitals.

The range of participatory design methods used provided us with a wealth of both quantitative and qualitative data and allowed for a good level of user engagement and interaction. The two-hour brainstorming session, which resulted in 26 pages of transcript, allowed participants to exchange and visualise ideas, and thus served as a good basis for establishing requirements. The phone interviews did not provide the same level of idea generation, but carrying them out separately with each facilitator provided a safe space for them to express their personal views on a HM machine. Furthermore, the paper-based prototyping session was found to be very useful, as it allowed us to concretise ideas at an early stage.

The questions presented in Figure 1 were particularly useful for designing the HM social machine. They guided the requirement elicitation process, from structuring the brainstorming session to analysing the qualitative data obtained. We would, hence, recommend their use to other social machine designers.

In the future, we wish to continue using the participatory design methodology and implement the feedback received as part of the current study. In order to create a truly usable social machine for the Heart Manual service, we plan to include more participants and to investigate new topics, such as the moderating role of the HM team. Applying the methodology presented in this paper to the design of other social machines is another exciting avenue for future work.

## ACKNOWLEDGEMENTS

This research was supported by the EPSRC SociaM project under grant EP/J017728/1.

## REFERENCES

Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *J. Usability Studies*, 4(3):114–123.

Brooke, J. (2013). SUS: a retrospective. *Journal of usability studies*, 8(2):29–40.

Brooks, F., Pospopa, C., and Schott, P. (2004). Midwifery on the net: new communication technology. *British Journal of Midwifery*, 12(2):107–110.

Clark, M., Kelly, T., and Deighan, C. (2011). A systematic review of the heart manual literature. *European Journal of Cardiovascular Nursing*, 10(1):3–13.

Deighan, C., Pagliari, C., Michalova, L., Elliott, J., Brookmann, F., and Taylor, L. (2015). Digital heart manual: making technology accessible with a user-friendly resource. In *Health Informatics Scotland Conference*. British Computer Society.

Donath, J. (2014). *The social machine: designs for living online*. MIT Press.

Hendler, J. and Berners-Lee, T. (2010). From the semantic web to social machines: A research challenge for AI on the world wide web. *Artificial Intelligence*, 174(2):156–161.

Murray-Rust, D. and Robertson, D. (2015). Bootstrapping the next generation of social machines. In *Crowdsourcing*, pages 53–71. Springer.

Rolls, K., Hansen, M., Jackson, D., and Elliott, D. (2016). How health care professionals use social media to create virtual communities: An integrative review. *Journal of Medical Internet Research*, 18(6):e166.

Schuler, D. and Namioka, A., editors (1993). *Participatory Design: Principles and Practices*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA.

Shadbolt, N. R., Smith, D. A., Simperl, E., Van Kleek, M., Yang, Y., and Hall, W. (2013). Towards a classification framework for social machines. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 905–912. ACM.

Simonsen, J. and Robertson, T. (2012). *Routledge International Handbook of Participatory Design*. Routledge.

Smart, P., Simperl, E., and Shadbolt, N. (2014). A taxonomic framework for social machines. In *Social Collective Intelligence*, pages 51–85. Springer.

The Heart Manual Department (2016). The Heart Manual. http://www.theheartmanual.com.

# Experimental Design and Collection of Brain and Respiratory Data for Detection of Driver's Attention

Roman Mouček[1,2], Lukáš Hnojský[1], Lukáš Vařeka[1,2], Tomáš Prokop[1,2] and Petr Brůha[1,2]
[1]*Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, Pilsen, Czech Republic*
[2]*NTIS – New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, Pilsen, Czech Republic*
*moucek@kiv.zcu.cz, hnojsky.l@gmail.com, {lvareka, prokop, pbruha}@kiv.zcu.cz*

Keywords: Neuroinformatics, Brain Activity, Electroencephalography, Event Related Potentials, Respiration Rate, Driver's Attention, Simulated Drive, Data Validation, Deep Learning, Stacked Autoencoder.

Abstract: Attention of drivers is very important for road safety and it is worth observing even in laboratory conditions during a simulated drive. This paper deals with design of an experiment investigating driver's attention, validation of collected data, and first preprocessing and processing steps used within data analysis. Brain activity is considered as a primary biosignal and is measured and analyzed using the techniques and methods of electroencephalography and event related potentials. Respiration is considered as a secondary biosignal that is captured together with brain activity. Validation of collected data using a stacked autoencoder is emphasized as an important step preceding data analysis.

## 1 INTRODUCTION

Attention of drivers is a very important factor of road safety. Inattentive drivers are dangerous to their surroundings and cause a considerable number of accidents. Since decline of attention, especially during long rides, is natural, it is worthwhile to investigate it even in laboratory conditions during a simulated drive. Results from laboratory experiments can be then used in real environment, e.g. for development of devices maintaining driver's attention or development of autonomous driving systems used at first when the driver is tired or inattentive.

In this paper we follow experiments and studies previously provided by our neuroinformatics research group and published in (Mouček and Řeřicha, 2012), (Mouček and Řondík, 2012), and (Mouček and Košař, 2014). A pilot experiment that is in more detail presented further in this paper partly shares the same assumption as the already published experiments. However, it differs in design and extends it by collecting an additional biosignal - data from respiration, and also by validation of collected data using a stacked autoencoder. Within the presented experiment the experimental design is proposed and the data and metadata suitable for investigation of influence of monotonous drive on driver's attention during simulated drive (a car simulator is used) are collected and evaluated. Then the basic preprocessing and processing steps used within data analysis are described.

When creating the experimental design the methods and techniques of electroencephalography (EEG) and event related potentials (ERP) are used to monitor and analyze brain activity of participating drivers. Event related potentials measures use stimulation techniques to investigate brain responses, so attention of a driver is not only affected by driving on a monotonous track, but also tested and influenced using auditory stimulation in our case. It is considered that the peak latency (peak latency represents a level of driver's attention) of the P3 component (the brain cognitive response described in Section 2) increases in time as the driver is more tired from monotonous drive. However, this component has to be first detected in the collected data. Therefore, the data validation step has to be done. Besides the brain activity the respiratory rate is also captured and its changes (most probably its decrease) are anticipated.

University students in the role of tested subjects participated in the experiment; the captured data were analyzed and partially interpreted observing particular trends in them. Due to an assigned page limit, a more detailed analysis including statistical evaluation and more detailed discussion of the results is not



Mouček, R., Hnojský, L., Vařeka, L., Prokop, T. and Brůha, P.
Experimental Design and Collection of Brain and Respiratory Data for Detection of Driver's Attention.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 441-450
ISBN: 978-989-758-213-4

provided. Moreover, since we publicly provide the collected data, the analysis can be provided by independent experts. We also realize that the number of subjects participated in the experiment (15 participants) is low for more precise statistical evaluation, but this number seems to be sufficient for a pilot experiment and for the decision about meaningfulness of the used experimental design and usability of the collected data.

The paper is organized as follows. Section 2 gives a short overview of basic principles of the ERP technique and assumptions related to amplitude and latency of P3 component. Then some recent experiments dealing with attention of drivers are briefly presented with respect to the experiments already presented in papers (Mouček and Řeřicha, 2012), (Mouček and Řondík, 2012), and (Mouček and Košař, 2014). The objectives of the proposed experiment are given in Section 3. Section 4 deals with experimental design, used hardware devices and software tools, participants, environment, and course of the experiment. The collected data and metadata are described in Subsection 4.5. The validation of the collected data is provided in Section 5. Sections 6 and 7 describe data preprocessing and processing. Experimental results together with final discussion are provided in Section 8. The last section contains concluding remarks.

## 2 STATE OF THE ART

This section provides a necessary short description of the ERP technique, the P3 component and the relation of P3 amplitude and P3 latency to attention. Then an assumption dealing with respiratory data is given. Finally, a short overview of EEG/ERP experiments dealing with driver's attention is presented.

Electroencephalography (EEG) and event related potentials (ERP) techniques are used for monitoring brain activity by measuring voltage changes on a scalp surface. ERPs have two advantages compared to classic behavioral methods: they are useful for determining which stage or stages of processing are influenced by a given experimental manipulation (a detailed set of examples is in (Luck et al., 2000)) and they provide an online measure of the processing of stimuli even when there is no behavioral response (Luck, 2005).

The P3 component (also referred to as the P300 component) as one of the event related components (waveforms) depends entirely on the task performed by the subject and is not directly influenced by the physical properties of the stimulus. It is sensitive to a variety of global factors, such as time since the last meal, weather, body temperature, and even day time or the time of year (Luck, 2005). It is not known exactly what the P3 component really means, it is probably related to a process called context updating (Luck, 2005).

However, there are known factors which influence the amplitude and the latency of the P3 component. The P3 component is sensitive to the probability of a target stimulus. P3 amplitude increases when the probability of the target stimulus class decreases or when it is preceded by a greater number of non-target stimuli. P3 amplitude is also larger when the subject pays more attention to the task. However, it is smaller if the subject does not know whether a given stimulus is / is not a target.

P3 latency is associated with stimulus categorization; if stimulus categorization is postponed, P3 latency increases. However, P3 latency does not depend on consequent processes (e.g. response selection). Thus P3 latency can be used to determine if a performed experiment influences the processes of stimulus categorization or processes related to a response (Luck, 2005). More detailed information is available in the literature provided above.

In our case we suppose that stimulus categorization is influenced by driver fatigue that is related to inattention and the time required for low-level sensory processing of incoming stimuli increases with the level of fatigue. We also consider that respiratory rate decreases with the level of fatigue.

When omitting behavioral studies, not many experiments dealing with driver's attention during simulated drive were performed using the techniques of electroencephalography and especially event related potentials.

Suitability of EEG-based techniques is described in (Schier, 2000); drivers' activity during a driving simulation task was recorded. As the result, an increase in alpha activity was interpreted as less attentional activity and a decrease as more attentional activity. EEG data as an effective indicator to evaluate driver fatigue are presented in (Li et al., 2012). The impact of a surrogate Forward Collision Warning System and its reliability according to the driver's attentional state by recording both behavioral and electrophysiological data was presented in (Bueno et al., 2012).

A systematic framework for measuring and understanding cognitive distraction in the automobile was presented in (Strayer et al., 2015). Primary, secondary, subjective, and physiological measures were collected and integrated into a cognitive distraction scale. Simultaneous recording of EEG and

eye-tracking for investigating situation awareness and working memory load in distracted driving was introduced in (Ichiki et al., 2015).

The ERP technique was used in (Wester et al., 2008) where the impact of secondary task performance (an auditory oddball task) on a primary driving task (lane keeping) was investigated. The study showed that when performing a simple secondary task during driving, performance of the driving task and this secondary task are both unaffected (Wester et al., 2008).

Amplitude of the P3 component reflecting individual differences of navigation performance in a driving task was investigated in (Bo et al., 2012). Two groups of navigators with good and poor navigation performance participated in a driving task; P3 amplitude was measured while two types of triggers were presented (intersections and street signs). Poor navigators showed larger P3 amplitude than good navigators on the left hemisphere, right hemisphere, the temporal, parietal and occipital sites when intersection triggers were presented, and on the occipital site when street sign triggers were presented, reflecting different levels of mental resource needed to process the spatial information between these two groups.

# 3 OBJECTIVES OF EXPERIMENT

The presented experiment was designed and conducted to investigate attention of drivers during simulated drive. The assumptions described in Section 2 were considered during designing and performing the experiment and the following objectives were set:

- to construct and implement a monotonous track where a substantial decrease of attention is supposed,
- to use a car simulator located in our neuroinformatics laboratory,
- to capture the brain activity of drivers during a simulated drive,
- to use auditory stimuli during experiment to evaluate brain activity,
- to capture respiration rate together with drivers' brain activity monitoring,
- to perform a pilot experiment on a group of at least ten participants,
- to validate the collected data,
- to annotate, store and made the collected data public,
- to compare the latency of averaged P3 components and evaluate the respiration waveforms to give a preliminary view if the results follow the considerations given in Section 1 and should be further elaborated.

# 4 DESIGN OF EXPERIMENT

The design of the experiment is a variant on the classic odd-ball paradigm in which presentations of sequences of frequent (non-target) audio/visual stimuli are interrupted by infrequent (target) stimuli. Infrequent stimuli usually elicit a much larger P3 component than frequent stimuli.

Two auditory stimuli were used to elicit the brain activity of the participants:

- non-target stimulus S1 - the sound of car wipers, duration time 900 ms, probability of occurrence p = 0.8,
- target stimulus S2 - the sound of thunder, duration time 900 ms, probability of occurrence p = 0.2,

Both these stimuli were played from the headphones worn by the participants during the whole simulated drive. The participant had to press the button to react to each target stimulus. The response button was located under the steering wheel. Although this response event required some movement from the participant, it did not cause undesired artifacts. The stimulus onset asynchrony (SOA) was set to 3900 ms (it means that the interstimulus interval was 3000 ms). The following rules were applied within the experimental scenario:

- at least two first stimuli are non-targets,
- each target stimulus appears randomly with respect to its probability of occurrence,
- two target stimuli cannot be sequential.

The background sound of drizzling was played from speakers to imitate the real environment. The speakers were located inside the car simulator behind the driver's seat.

The overall length of the experiment was 60 minutes. This length was experimentally verified as a maximum time frame during which the participant did not have difficulties (bad feelings) with an EEG cap placed on his/her head. It was also expected that fatigue would increase approximately after 30 minutes of driving.

The experiment was divided into three driving sessions, each session lasted 15 minutes and was followed by five minutes break. The breaks served both for relaxation of the participant (from driving and

watching the simulation scene) and for preventing the participant from familiarity with presented stimuli (habituation to stimuli was thus limited).

### 4.1 Hardware Equipment

The experiment was performed in the neuoroinformatics laboratory of the University of West Bohemia, Czech Republic that was equipped with all necessary hardware infrastructure for EEG/ERPs and respiratory rate recordings. The experimental car simulator (a front part of a real Skoda Octavia car) was equipped with the Logitech G27 wheel, accelerator, and brake. These were connected to the computer via the USB port.

Three computers were used: the first one for presentation of auditory stimuli, the second one for storing recorded data, and the third one for presentation of the track. The track was projected on the wall in front of the car simulator. V-Amp produced by the Brain Products company was used as an EEG amplifier. It served also as an input of the sensor capturing respiratory rate.

### 4.2 Software Tools

The stimulation protocol was implemented in the Presentation software tool produced by Neurobehavioral Systems, Inc (Neuro Behavioral Systems, 2014). The sequence of stimuli was generated randomly, but it always contained the same number of target and non-target stimuli. All the sounds (including background drizzling) were recorded using the Audacity software tool and stored in the .wav format. The track was prepared using the World Racing 2 game produced by the Synetic Company (SYNETIC GmbH, 2014). The same track as in (Mouček and Řeřicha, 2012) was used. The BrainVision Recorder (Brain Products, 2014) was used for recording and storing EEG/ERP data and respiration rate. MATLAB, EEGLAB, and ERPLAB software tools were used for processing and analysis of experimental data.

### 4.3 Recording System

Common EEG caps (the 10-20 system defining locations of scalp electrodes) were used depending on the size of the participants' heads. The reference electrode was placed approximately 0,5 cm - 1 cm above the nose and the ground electrode was placed on the ear. The respiratory rate sensor (produced by the Brain Products company) as well as the EEG cap were connected to the V-Amp amplifier.

### 4.4 Participants and Course of Experiment

A group of 15 volunteers, university students (thirteen men, two women), aged 21-28, participated in the experiment. The participants got necessary information about the experiment in a written form in advance. Then informed consent was obtained from all of them.

Before starting the experiment, each participant was familiarized with basic behavioral rules during an EEG/ERP experiment (e.g. not to use cosmetic products before the experimental session, or reduce eye blinking and unnecessary movements to decrease the number of artifacts). Then the participant was familiarized with all sounds played during the experiment, with car simulator controls, and with the track. Subsequently the participants were allowed to drive around to get accustomed to the car simulator and simulated drive.

During the experiment the experimenter was controlling data recording and checking the correct behavior of the stimulation program. When the experimental session finished, the participant left the car simulator. Then the experimenter asked him/her to fill in the questionnaire containing questions related to his/her feeling of fatigue during/after the drive.

### 4.5 Data and Metadata

EEG/ERP data were recorded with the sampling frequency of 1 kHz; no filters were used during data recording. The resulting signal was stored into three files:

- .eeg file containing raw data,
- .vhdr file containing metadata that describe raw data in the related .eeg file,
- .avg file containing the averaged signal around the used stimuli.

All recorded data and collected metadata were stored into the EEG/ERP portal (experiments No. 205-209, 225-236) (EEG/ERP Portal, 2016). These data are publicly available for registered users (registration is free).

## 5 DATA VALIDATION

The data validation was based on the main objective of P3-based experiments: target and non-target trials are expected to be associated with differently shaped ERP components, especially P2, N2, and

P3 (Blankertz et al., 2011). To validate this objective, dichotomous machine learning was used. If classification of a specific dataset from one subject yields low error rates (defined later), the objective of the odd-ball paradigm is considered to be fulfilled.

The classifier was trained on a randomly selected data subset. The training subset contained 730 ERP trials (described in detail in (Vařeka et al., 2014a)) with equal numbers of targets and non-targets. The trained classifier was subsequently applied to the data of individual subjects.

The Matlab scripts available in (Vařeka et al., 2014b) and using EEGLAB and BCILAB functions were used for the implementation. Feature extraction follows the Windowed Means Method proposed in (Blankertz et al., 2011). This method includes feature extraction: low pass filtering and spatial filtering, and machine learning technique based on one of the deep learning models - stacked autoencoders. Feature extraction was described in detail in (Vařeka et al., 2014a). Machine learning was designed as follows.

The Matlab implementation of stacked autoencoders was used. The parameters (including number of layers, number of neurons in each layer, etc.) were empirically optimized. The experimentation started with two layers, then either new neurons were added into the layer, or a new layer was added until the performance of the classifier stopped increasing.

Finally, the following procedure was used to train the network:

1. The first autoencoder with 100 hidden neurons was trained. The maximum number of training epochs was limited to 500.
2. The second autoencoder with 75 hidden neurons was connected with the first autoencoder to form a 133-100-75-133 neural network, and trained. The maximum number of training epochs was limited to 300.
3. The third autoencoder with 60 hidden neurons was connected with second first autoencoder to form a 133-100-75-60-133 neural network, and trained. The maximum number of training epochs was limited to 200.
4. The fourth autoencoder with 30 hidden neurons was connected with third autoencoder to form a 133-100-75-60-30-133 neural network, and trained. The maximum number of training epochs was limited to 200.

Furthermore, the following parameters were set for the network globally: L2WeightRegularization was set to 0.004, SparsityRegularization was set to 4, and SparsityProportion was set to 0.18.

After the training of each autoencoder, the input feature vectors were encoded using that autoencoder to form input vectors of the next autoencoder.

Using the output of the last autoencoder, softmax supervised classifier was trained with 200 training iterations. Finally, the whole pre-trained 133-100-75-60-30-2 network was fine-tuned using backpropagation.

The learned model was first verified on other P300-based data (Vařeka et al., 2014a). Then, for each subject, error rates depicted by red bars were obtained by applying the classifier in the testing mode. Let us suppose that we have $t_p$ - number of correctly classified targets, $t_n$ - number of correctly classified non-targets, $f_p$ - number of misclassified non-targets, $f_n$ - number of misclassified targets. The error rate was calculated according to Equation 1.

$$ERR = \frac{fp + fn}{tp + tn + fp + fn} \tag{1}$$

As a result, error rates indicate the extent to which the classifier was unable to separate target and non-target single trials. The classification results may slightly differ with each run because of the indeterministic training process.

# 6 DATA PREPROCESSING

The recorded EEG/ERP data as well as the data obtained from the respiratory sensor were processed using the following workflow:

- Channel selection: The following channels capturing brain data were selected for the initial processing: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, Fz, Cz, and Pz.
- Driving session selection: Data for each driving session were processed separately.
- Data filtering: IIR Butterworth filter (frequency range 0,01 Hz - 20 Hz) was applied to the data.
- Data segmentation: The epochs were extracted from datasets, data corresponding to each target and non-target stimulus were selected in the time interval (-100 ms before the stimulus, 1000 ms after the stimulus) in the area of occurrence of the target or non-target stimulus.
- Application of the filter for automatic artifacts detection: The segmented data exceeding the range (-100 microV, 100 microV) were denoted as possible artifacts and provided for manual inspection.
- Rejection of corrupted data: The data automatically denoted as artifacts were manually inspected

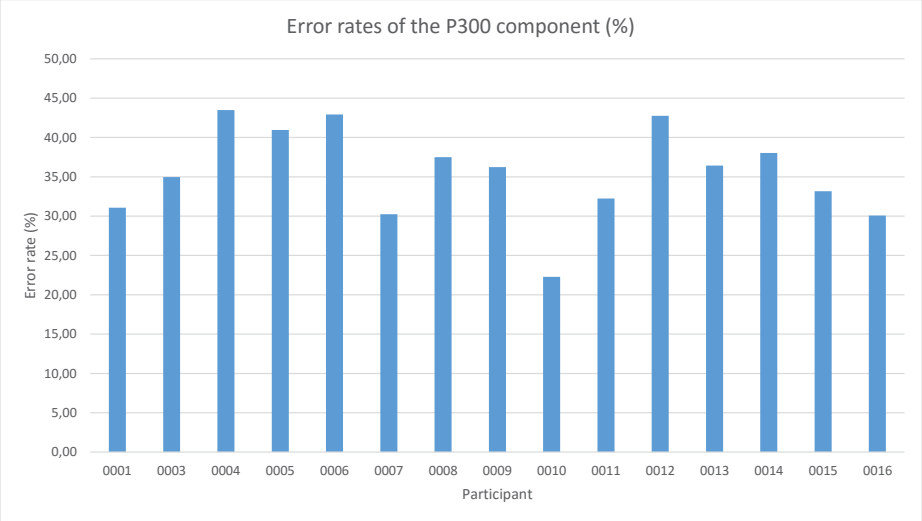



Figure 1: Results of validation. The error rates for each subject are depicted in bars. Higher error rates mean lower amplitudes of P3s and/or more distortion in the EEG/ERP signal.

and most of them rejected. Moreover, each dataset was manually inspected and in case of suspected artifacts the related epochs were rejected. Artifacts were usually caused by eyes blinking, swallowing or movements.

- Baseline correction: The baseline was corrected using the interval (-100ms, 0ms) before occurrence of each target or non-target stimulus.
- Data averaging: The accepted epochs for each participant and each session were averaged and stored separately for target and non-target stimuli.
- The data captured by the respiratory sensor were processed using the following workflow: only the .vhdr file was read, filtering was applied, and the scale for respiration visualization was adjusted.

## 7 DATA PROCESSING

For the next analysis only the channels P3, P4, Fz, Cz, and Pz (since the occurrence of the P3 component is more significant on these channels) were selected. Then grand averages for each driving session and each participant were computed (separately for target and non-target stimuli). The latency of the P3 component was determined using the technique of peak latency. It is the simplest way to determine the latency of the P3 component when the maximum amplitude in the time frame of possible occurrence of the P3 component is searched for. The P3 component time frame was set to (300 ms, 450 ms) reflecting the expected location of peak latency values occurrence in case of auditory stimulation. Finally, the peak latency was determined manually from computed grand averages.

## 8 RESULTS AND DISCUSSION

The results from performed experiments are summarized in the figures and tables presented further. The P3 component for the participant 0010 and the values of his/her peak latency for the first driving session and for channels P3, P4, Fz, Cz, and Pz are shown in Figure 2.

It can be seen that the component P3 is clearly identifiable and there is a substantial difference between reaction to target and non-target stimuli. However, the identification of the P3 component was not so evident for all participants. This is shown in Table 1 where peak latencies for selected channels and each driving session are available.

The data collected from participant 0002 were rejected, the resulting values were out of reasonable range, most probably because of technical failure during the measurement. The data from participants 0005, 0006, 0008, and 0009 were also not further interpreted since the N2 component (a repetitive, non-target stimulus elicits N2 deflection that can be thought of as the basic N2 component) had a very high amplitude while the component P3 was not clearly evident (an example for the participant 0006 is available in Figure 3). It can also easily observed that the amplitude of the N2 component decreased in time; the difference in peak amplitudes of this component is

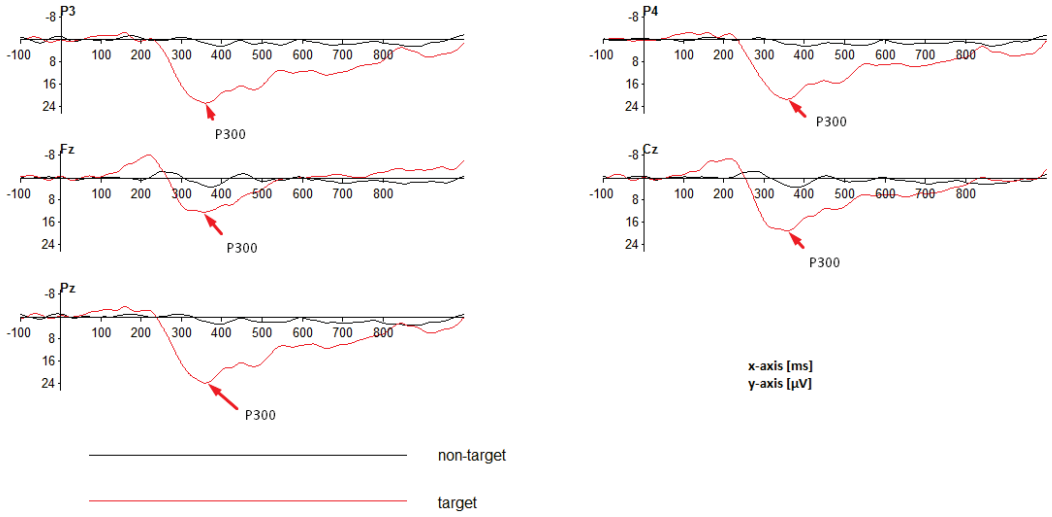



Figure 2: Grand average for participant 0010, the first driving session, channels P3, P4, Fz, Cz, and Pz, peak latency of the P3 component for target and non-target stimuli.

noticeable comparing the waveforms for each driving session. It was probably caused by habituation of participants during the course of the experiment.

However, it does not mean that the data containing high N2 amplitude and low P3 amplitude are incorrect. Since we were unsure with their correct interpretation, they were rejected only for this preliminary analysis. It is reasonably possible that they will be considered for later analysis after reviewing them by another expert in the field. The data from participants 0012 and 0013 were not interpreted because of low amplitude of the P3 component.

It is also evident that peak latencies differ among the participants in average. This is a natural phenomenon that reveals that the cognitive processing of stimuli is different for each individual in the considered time frame. It can be also seen that the average latency for all participants for each selected channel increased as the experiment continued to the next driving sessions.

However, different results we got by computing grand averages for all epochs containing the target stimulus and for all participants. It is not an average of peak latencies as computed above, all waveforms related to the target stimulus within driving sessions were averaged. The resulting waveforms are given in Figure 4. The determined values of peak latency for averaged waveforms are clearly shown in Table 2. It can be seen that grand averages for target stimuli, selected channels and all participants differ between the first and second driving session, while there is no evident difference between the second and third driving session.

The respiration rate was computed for each participant and for each driving session as it can be seen in Table 3. The average respiratory rate decreased in time. Participants 0013 and 0014 had an increased respiratory rate in all driving sessions in general, but it also decreased within the course of the experiment. The values from participants 0002 and 0015 were not captured correctly due to technical difficulties with the sensor.

## 9 CONCLUSIONS

This paper described the experiment dealing with attention of drivers during a simulated drive. Brain activity and respiratory rate of the participants of the experiment were measured and investigated by using mainly the methods of electroencephalography and event related potentials. Drivers, university students, were stimulated by simple auditory stimuli while driving a car simulator on a monotonous track. The collected data were annotated, stored, preprocessed, validated, and partly analyzed. The peak latency of the P3 component was derived from the data and grand averages for each participant and driving session as well as grand averages for all participants and each driving session were computed.

Experimental results showed that the P3 component had been identified at most participants during all driving sessions. However, some experimental results were not interpreted in this article because of the high amplitude of the N2 component compared to the amplitude of the P3 component. Prolongation of the peak latency of the P3 component was evident in case of most participants and in case of simple averaging of

Table 1: Grand averages for driving sessions, participants and selected channels.

| Participant | 1st driving session | | | | | 2nd driving session | | | | | 3rd driving session | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Peak latency of P3 component [ms] on channels | | | | | Peak latency of P3 component [ms] on channels | | | | | Peak latency of P3 component [ms] on channels | | | | |
| | **P3** | **P4** | **Fz** | **Cz** | **Pz** | **P3** | **P4** | **Fz** | **Cz** | **Pz** | **P3** | **P4** | **Fz** | **Cz** | **Pz** |
| 0001 | 423 | 421 | 336 | 410 | 418 | 452 | 450 | 404 | 406 | 486 | 497 | 507 | 402 | 400 | 497 |
| 0002 | rejected because of technical failure | | | | | | | | | | | | | | |
| 0003 | 383 | 384 | 382 | 386 | 383 | 390 | 416 | 394 | 399 | 378 | 398 | 431 | 387 | 398 | 346 |
| 0004 | 459 | 443 | 422 | 435 | 445 | 464 | 462 | 409 | 456 | 455 | 473 | 476 | 454 | 465 | 469 |
| 0005 | too high amplitude of N200 component | | | | | | | | | | | | | | |
| 0006 | too high amplitude of N200 component | | | | | | | | | | | | | | |
| 0007 | 345 | 334 | 323 | 343 | 341 | 356 | 353 | 357 | 356 | 354 | 357 | 355 | 359 | 358 | 356 |
| 0008 | too high amplitude of N200 component | | | | | | | | | | | | | | |
| 0009 | too high amplitude of N200 component | | | | | | | | | | | | | | |
| 0010 | 361 | 356 | 357 | 357 | 359 | 389 | 389 | 391 | 389 | 389 | 399 | 397 | 394 | 398 | 396 |
| 0011 | 317 | 318 | 347 | 305 | 322 | 319 | 311 | 309 | 303 | 327 | 363 | 410 | 307 | 304 | 407 |
| 0012 | Low amplitude of P3 component | | | | | | | | | | | | | | |
| 0013 | Low amplitude of P3 component | | | | | | | | | | | | | | |
| 0014 | 351 | 348 | 358 | 347 | 350 | 399 | 336 | 353 | 337 | 335 | 448 | 359 | 399 | 363 | 365 |
| 0015 | 444 | 413 | 404 | 406 | 411 | 428 | 418 | 411 | 415 | 418 | 498 | 505 | 448 | 456 | 449 |
| 0016 | 405 | 404 | 392 | 387 | 404 | 418 | 416 | 404 | 406 | 415 | 404 | 389 | 402 | 393 | 399 |
| **Avg** | **388** | **380** | **369** | **375** | **381** | **402** | **395** | **381** | **385** | **395** | **426** | **425** | **395** | **393** | **409** |

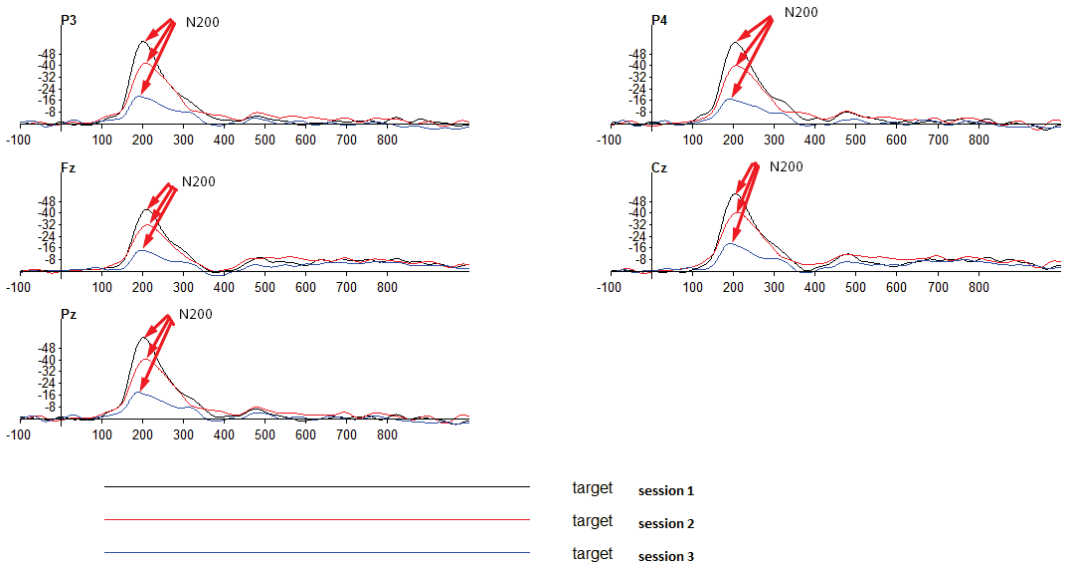



Figure 3: Grand average for the participant 0006, channels P3, P4, Fz, Cz, and Pz, peak latency of the N2 component for target stimulus.

their peak latencies. Despite expectations, prolongation of peak latency in time was not clearly observed when grand averages for all participants were investigated. This prolongation is evident only as a difference between grand averages of the first and second driving sessions. We also supposed that when the driver expected his/her drive to be almost completed (during the third driving session), his/her attention increased. The average respiration rate and respiration rates for most participants showed a decreasing trend during the course of the experiment.

The results in this article were not statistically evaluated. However, the trend of increasing latency and decreasing respiratory rate is clearly visible. The experimental results were naturally affected by different brain reactions of participated drivers and sensibility of captured data to the environmental noise, participants' overall mental conditions, and their movements that caused occurrences of artifacts. Although there was a big effort to eliminate these circumstances by experimental design, setting of experimental conditions, and usage of data preprocessing and process-

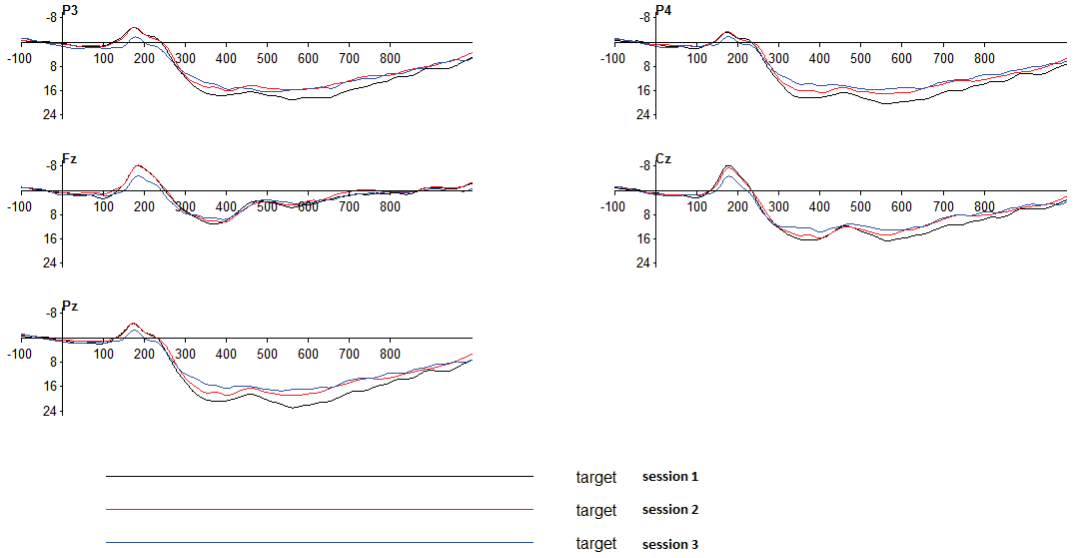



Figure 4: Grand averages for all participants, channels P3, P4, Fz, Cz, and Pz, peak latency of the P3 component for the target stimulus.

Table 2: Grand averages of peak latencies of the P3 component in [ms] for selected channels, target stimulus.

| **Grand average** | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st driving session | | | | | 2nd driving session | | | | | 3rd driving session | | | | | |
| P3 | P4 | Fz | Cz | Pz | P3 | P4 | Fz | Cz | Pz | P3 | P4 | Fz | Cz | Pz | |
| 391 | 391 | 369 | 386 | 389 | 407 | 407 | 395 | 402 | 405 | 406 | 405 | 395 | 402 | 402 | |

Table 3: Respiration rate for each participant and driving session.

| | **Respiratory rate** | | |
|---|---|---|---|
| **Participant** | **1st driving session** | **2nd driving session** | **3rd driving session** |
| 0001 | 18 | 16 | 16 |
| 0002 | rejected | | |
| 0003 | 16 | 15 | 14 |
| 0004 | 15 | 14 | 14 |
| 0005 | 19 | 18 | 18 |
| 0006 | 16 | 16 | 15 |
| 0007 | 16 | 15 | 15 |
| 0008 | 18 | 17 | 16 |
| 0009 | 17 | 17 | 15 |
| 0010 | 16 | 15 | 15 |
| 0011 | 14 | 13 | 13 |
| 0012 | 18 | 18 | 18 |
| 0013 | 23 | 22 | 22 |
| 0014 | 26 | 25 | 25 |
| 0015 | rejected | | |
| 0016 | 16 | 16 | 16 |
| **Avg** | **17,69** | **17,00** | **16,62** |

ing methods, they could not be completely removed. That is why we also validated the collected data using the stacked autoencoder.

The observations collected by experimenters from the participants were also summarized. The participants (they filled in the questionnaire) did not report a highly increased level of fatigue after having finished their experiment. Although the participants complained about the tedious ride, the stimulation kept them relatively attentive. One third of participants pointed out that they would like to drive longer without any brake. On the other hand, most participants complained about unpleasant feelings caused by the ground electrode placed on their ear. This feedback have been used for the design of further ongoing experiments.

We believe that by using the results of ongoing research and technological innovations it will possible to capture biosignals more easily in the future. It would facilitate recognition of the human attention level and decrease the number of accidents not only in transport but also during activities that are directly influenced by human attention.

## ACKNOWLEDGMENTS

This publication was supported by the UWB grant SGS-2016-018 Data and Software Engineering for Advanced Applications and the project LO1506 of the Czech Ministry of Education, Youth and Sports under the program NPU I.

# REFERENCES


Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of erp components – a tutorial. *NeuroImage*, 56(2):814 – 825. Multivariate Decoding and Brain Reading.

Bo, O., Changxu, W., Guozhen, Z., and Jianhui, W. (2012). P300 amplitude reflects individual differences of navigation performance in a driving task. *International Journal of Industrial Ergonomics*, 42(1):8–16.

Brain Products (2014). Brain vision recorder.

Bueno, M., Fabrigoule, C., Deleurence, P., Ndiaye, D., and Fort, A. (2012). An electrophysiological study of the impact of a forward collision warning system in a simulator driving task. *Brain Research*, 1470:69–79.

EEG/ERP Portal (2008-2016). EEG/ERP Portal.

Ichiki, M., Ai, G., and Wagatsuma, H. (2015). Simultaneous recording of eegs and eye-tracking for investigating situation awareness and working memory load in distracted driving: A prospective analysis toward the neuro-driving framework. *Frontiers in Neuroscience*, (10).

Li, W., He, Q.-C., Fan, X.-M., and Fei, Z.-M. (2012). Evaluation of driver fatigue on two channels of eeg data. *Neuroscience Letters*, 506(2):235–239.

Luck, S., Woodman, G., and Vogel, E. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, 4(11).

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique (Cognitive Neuroscience)*. A Bradford Book, 1 edition.

Mouček, R. and Košař, V. (2014). Attention of driver during simulated drive. *HEALTHINF 2014 - 7th International Conference on Health Informatics, Proceedings; Part of 7th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2014*, pages 543–550.

Mouček, R. and Řeřicha, J. (2012). Driver's attention during monotonous driving. *2012 5th International Conference on Biomedical Engineering and Informatics, BMEI 2012*, pages 486–490.

Mouček, R. and Řondík, T. (2012). Influence of mental load on driver's attention. *Transaction on Transport Sciences*, 5(1):21–26.

Neuro Behavioral Systems (2014). Home page.

Schier, M. (2000). Changes in eeg alpha power during simulated driving: A demonstration. *International Journal of Psychophysiology*, 37(2):155–162.

Strayer, D., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., and F., B. (2015). Assessing cognitive distraction in the automobile. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(8):1300–?1324.

SYNETIC GmbH (2014). World racing 2.

Vařeka, L., Brůha, P., and Mouček, R. (2014a). Event-related potential datasets based on a three-stimulus paradigm. *GigaScience*, 3(1):35.

Vařeka, L., Brůha, P., and Mouček, R. (2014b). P3-validator.

Wester, A., Böcker, K., Volkerts, E., Verster, J., and Kenemans, J. (2008). Event-related potentials and secondary task performance during simulated driving. *Accident Analysis and Prevention*, 40(1):1–7.

# The Need for Trustworthiness Models in Healthcare Software Solutions

Raja Manzar Abbas[1], Noel Carroll[1,2], Ita Richardson[1,2] and Sarah Beecham[1]
[1]*Lero- The Irish Software Research Centre, University of Limerick, Limerick, Ireland*
[2]*ARCH- Applied Research for Connected Health Technology Centre, University of Limerick, Limerick, Ireland*
*{raja.abbas, noel.carroll, ita.richardson, sarah.beecham}@lero.ie*

Keywords: Trustworthiness, Healthcare Software, Trustworthiness Attributes.

Abstract: Trustworthiness in software is of vital importance to technology users, especially in health, where lives may depend on the correct application of software that is fit for purpose. Despite the risk posed by improper use of technology in the health domain, there is evidence to suggest that stakeholders often trust the software without fully appreciating the possible consequences. In this paper, we explore what determines trustworthiness in healthcare software solutions. While there are often claims of improved quality of care, increased safety and improved patient outcomes using healthcare technology – the scientific basis for such claims appear to be uncritically accepted. Ultimately, this can lead to a surge in healthcare software solutions, some of which may be misaligned with healthcare needs and potentially lead to fatal outcomes. To support health technology stakeholders, we propose a 'trustworthiness healthcare software model' that can be employed to assess the level of trustworthiness associated with healthcare software solutions.

## 1 INTRODUCTION

The United Nation's International Standard Industrial Classification, (2016) categorizes healthcare as generally consisting of hospital activities, medical and dental practice activities. Implementations of potentially transformative healthcare technologies are currently underway internationally, often with significant impact on national expenditure. For example, Ireland has invested approximately €900 million in its e-health while the UK has invested at least £12.8 billion in a National Programme for Information Technology (NPfIT) for the National Health Service. Similarly, the Obama administration in the United States has committed to a US$38 billion Healthcare investment (Catwel et al., 2009).

Such large-scale expenditure has been justified on the grounds that electronic health records (EHRs), picture archiving and communication systems (PACS), electronic prescribing (ePrescribing) and associated computerised provider (or physician) order entry systems (CPOE), and computerised decision support systems (CDSSs) will help address the problems of variable quality, safety and trust in the modern health care. However, the scientific basis of achieved quality and trust – which are repeatedly made and are seemingly uncritically accepted – remains to be established (Huckvale et al., 2012; Institute of Medicine, 2007).

### 1.1 Problem Statement

The ultimate goal of software is to help end-users to accomplish their tasks in a convenient and efficient manner. However, the literature suggests that software technology advancements in healthcare often failed to ease the lives of the healthcare professionals. Instead, healthcare professionals often report a loss of productivity while using healthcare software. This leads to a lack of trust in the healthcare software (Velsen et al., 2016).

### 1.2 Research Question

In this paper we examine the literature on trustworthiness in healthcare and look particularly at the associated attributes. We also explore the need for a healthcare software model of trustworthiness. Considering the broad and vast nature of software technology use in healthcare, we argue that stakeholders need to have a set of criteria by which they can assess the level of trustworthiness of a given technology.



Abbas, R., Carroll, N., Richardson, I. and Beecham, S.
The Need for Trustworthiness Models in Healthcare Software Solutions.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 451-456
ISBN: 978-989-758-213-4

There is an apparent lack of insight into what a trustworthiness healthcare software model should capture, and how it should be applied. To address these gaps, we formulate the following research questions:

- *RQ1*. What are the key attributes that define trustworthiness of healthcare software?
- *RQ2*. What are the current models or frameworks that capture trustworthiness of healthcare technology?

### 1.3 Methodology

To explore these questions, we undertook a structured literature review. A structured literature review may be described as appraisals of past studies conducted systematically, purposefully and methodologically (Armitage and Keeble-Allen, 2008; Petticrew, 2001).

In the research discussed in this article, a literature search was completed in the bibliographic databases ACM Digital Library, IEEE, Springer LINK and Google Scholar, using the keyword search phrases '*trustworthiness'*, '*healthcare software*', '*healthcare trust*', '*trustworthiness models'*, *trustworthiness frameworks'*, '*trust attributes'*, '*trustworthy attributes'*, '*software trustworthiness*' and '*healthcare software trustworthiness*'. *2894* initial reference sources were found. From these, after screening titles and abstracts , *2224* were deemed not eligible. Out of remaining *670* research articles, *536* articles were screened out after applying the exclusion criteria on the titles and abstracts - *238* were not relevant to software engineering, *107* research articles had no specific intervention about software trustworthiness (trustworthy, trust), *187* articles did not mention software attributes and/or models and *4* research articles were not written in English. After reviewing the full text of the remaining *134* studies, *83* more studies were excluded due to lack of relevance to the topic and *51* studies were selected as primary studies.

## 2 IMPACT OF HEALTHCARE TECHNOLOGY

Due to the growth in population and shift in demographics, there is a significant pressure on the global healthcare system. Shojania et al. (2016), attribute a toll to medical error of 251,454 deaths in US hospitals per year, making, they say, medical error the third-leading cause of death in the USA. The Institute of Medicine study estimated the cost of nonfatal medical errors is between $17 billion and $19 billion each year, and that between 2.9% and 3.7% of all patients admitted suffer some type of injury because of medical mismanagement. As a result, there is a growing focus on healthcare technology to offer greater service efficiency and it has given rise to a comprehensive sociotechnical model for managing healthcare through software solutions.

Technological advances have encouraged the development of new technologies that drive connectivity across the healthcare sector—apps, gadgets, and systems that personalize, track, and manage care using just-in- time information exchanged through various patient and community connections (Leroy et al., 2014).

This paradigm shift heavily emphasizes the process of software development in healthcare systems. It has also contributed to a shift in healthcare practice, highlighting our growing reliance and trustworthiness of software to support healthcare decisions. However, trusting the healthcare software solution without validating can have serious and potentially fatal consequences (Carroll, 2016).

## 3 TRUSTWORTHINESS – WHO CARES?

Trustworthiness in healthcare software is the sum of trust in different factors. The composition these factors can differ for different healthcare users. For example, for patients, trustworthiness in software consists of, mostly, a perceived level of control and privacy, while for healthcare professionals, a larger and different set of issues play a role, including reliability and a transparent data storage policy. The set of factors that affect trustworthiness in a healthcare portal are different from the sets that have exist for general software domain. There is a need to study trustworthiness in healthcare software as a separate subject to inform the design of reliable interventions.

### 3.1 Need for Trustworthiness Healthcare Software Model

With significant growth in healthcare software solutions, software is having an increasing impact on clinical decisions and diagnosis. However, there is little evidence as to the trustworthiness of software. For example, a glitch in St. Mary's Mercy Medical Centre's (Cork, Ireland) patient management system "killed" 8,500 patients on paper (National Computer Security Center, 1985). When St. Mary's upgraded its

patient management software, there was a mapping error in the alteration process that triggered the program to notify 8,500 patients of their incorrect death.

While unregulated medical devices rarely find their way to patients, the same cannot be said about the largely unregulated market for health applications and software. As such, in reality, there exists a considerable gap between the potential benefits that software's could provide, and what healthcare professionals are currently likely delivering in practice. Recent reviews in the therapeutic areas of bulimia (Nicholas et al., 2015), asthma (Huckvale et al., 2012), Post-traumatic stress disorder (PTSD) (Olff, 2015), insulin dosing (Huckvale, 2015) and suicide prevention (Larsen et al., 2016) have yielded worrying conclusions regarding the quality, scientific basis and often blatant disregard for safety (The Daily Mail, 2014),

These errors go some way to illustrating the need for trustworthy healthcare software. Although different researchers have tried to address some attributes of trustworthiness not all attributes have been identified. We discuss some of the models and standards and the attributes they cover in an effort to understand what areas they are lacking in.

### 3.2 Impact of Trustworthiness on Quality

The number of medical errors caused by devices (which have embedded software) and software applications naturally leads to the questions:

- How can healthcare software be made trustworthy?
- What process/mechanism would achieve this?
- How do we inform users and healthcare providers which healthcare software solution can be trusted and why?

We first identified the attributes of trustworthy healthcare software and why there is a need for a trustworthy healthcare software model.

## 4 DEFINING TRUSTWORTHINESS

According to Merriam–Webster Dictionary (2004), trustworthy means '*worthy of confidence'*. For software products, researchers and practitioners have a slightly different understanding of '*trustworthy software systems'*, since we need to view trustworthiness over time.

Trustworthiness is defined by Amoroso et al. (1994) as a "*level of confidence or degree of confidence*" and software trustworthiness is defined as a "*degree of confidence that the software satisfies its requirements*". Since the definition is expressed as a "*degree of confidence*", Amoroso and Taylor illustrates trustworthiness is dependent upon management and technical decisions made by individuals or groups of individuals evaluating the software. Software trustworthiness is expressed in terms of a set of requirements, where the 'set' is variable. For example, trustworthiness may be dependent on the set of functional requirements, or may be a critical subset of functional requirements, or it may be some set of requirements that include non-functional assurance requirements like safety or security (Amoroso et al., 1994).

In his ICSE 2006 Keynote speech, Boehm (2006) pointed out the increasing trend of software criticality and dependability as one of the key software trends. Over the past 50 years, different strategies such as formal methods, security assurance techniques, defect prediction, failure mode and effects analysis, testing methods, and software assurance techniques have been proposed to address different aspects of software trustworthy challenges. Based on these studies, numerous quality categories and attributes have been studied as major factors influencing on software trustworthiness. Among them are included functionality, reliability, safety, usability, security, portability, and maintainability, etc.

In addition, Zhang et al. (2012) reviews the appropriateness of the software attributes summarized by Yang et al. (2009), and suggests that the trustworthiness of software is related to the following set of properties which they redefined to address the trustworthiness context as:

- Safety
- Validity
- Reliability
- Reusability
- Scalability
- Maintainability
- Performance

Carbone et al. (2013) defined trustworthiness from information and communication technology (ICT) systems where security challenges include both confidentiality (or privacy) and in integrity (or trust) of the data. In particular, the notion of trustworthiness seems relevant for tagging databases and electronic patient records with information about the extent to which test results, diagnoses and treatments can be trusted.

Initial findings from literature suggest a lack of understanding of trustworthiness in the healthcare software domain and a need for a trustworthiness process model that define standards or best practices about the trustworthiness of healthcare software.

In this position paper, we propose the key attributes that define trustworthiness of healthcare software and current models that capture trustworthiness of healthcare technology.

# 5 THEORETICAL FOUNDATIONS

Though there is a growing consensus that trustworthiness is characterized as one that satisfies a collection of critical quality attributes, yet, there is a lack of common understanding of healthcare software trustworthiness – particularly in a healthcare context. Based on the literature and the sample of definitions introduced here, we identify that the key factors of trustworthiness for healthcare software should be regulation, confidence of the users and meet its requirements and objectives in a satisfactorily manner.

## 5.1 Theoretical Influences on Developing a Trustworthiness Healthcare Software Model

The Capability Maturity Model Integration (CMMI) guides organisations with a view to ensuring that the correct software is being developed throughout each stage in the development cycle and conforms to specification. However, it is important to realise that software process models, such as CMMI, do not cover healthcare regulations, and that they need to be used in conjunction with the regulations (Burton et al., 2006).

There have been some new developments in this area, for example the development of MDevSPICE (formally known as MediSPICE) (Clarke et al., 2014; McCaffery et al., 2010). The MDevSPICE framework is one of the first attempts to address the safety concerns faced by healthcare software producers and presents a software safety assessment process. Verification and validation activities are very important in software development and can consume much of a project's costs and effort. While verification and validation are addressed by process models and standards for both generic and safety critical software development, there are still challenges in undertaking its successful implementation as part of the software development process. The process of verification and validation requires a clear understanding of how each activity is undertaken and related to each other, which is important in a healthcare environment (Carroll and Richardson, 2016).

For example, the development of an international software process improvement (SPI) framework for the medical device industry acts as a key enabler of best practice for the healthcare sector. SPI techniques offer a continuous cycle of performing an assessment and restarting the cycle (McHugh et al., 2012) with the aim of reducing defective software. Software may also be vulnerable to outside attack. Many hospitals and healthcare facilities use various threat management software and firewalls to monitor their mobile device applications to ensure that they are secure and safe. In most cases, within the USA, this is a requirement of Health Insurance Portability and Accountability Act (HIPPA).

HIPAA is a framework which is followed by number of organisations for maintaining the security and privacy of the health information. HIPAA came into force in 1996 to address a number of concerns, most notably the need for increased protection of the medical records of the patients against unauthorised access (Wu et al., 2012). HIPAA provides a national standard for electronic healthcare transactions. It also provides regulations regarding healthcare information security and privacy (Jepsen, T, 2003). HIPAA covers entities such as healthcare providers, insurers and providers of health plan. Healthcare organisations are now required to individually assess their security and privacy requirements using various auditing tools. Healthcare technology systems have access to personal identifiable information.

Our traditional view of privacy protection methods through various anonymization techniques does not provide an efficient way to deal with the privacy of technological healthcare software solutions. For example, in response to growing concerns on privacy and data security, in 2014, the European Commission published a Green Paper on mHealth (European Commission, 2014). Through wide stakeholder consultation, the paper discusses the main barriers and issues related to mHealth deployment. They highlight a number of key topics including data protection, security of health data, informed consent, big data management, patient safety and transparency of information across the EU and, ultimately, on the need to regulate mHealth applications.

One of the main concerns across industry is the lack of a unified model which can incorporate all of

the best practices for healthcare software development. There is a need to formulate a healthcare software model that can accurately propagate trustworthiness throughout the process.

### 5.2 Towards Developing a Trustworthiness Healthcare Software Model

By carefully reviewing the appropriateness of the attributes summarized through a software lens, we suggest that the trustworthiness of healthcare software model is related to the following set of properties:

- **Security:** inclusion of security mechanisms in the model with respect to access control processes.
- **Efficiency:** effectiveness of the model construction that is able to give a quick response or reaction with minimal resources and/or time taken.
- **Safety:** inclusion of semantics that represent process requirements related to safety, and the ability to highlight inconsistencies in the process model with respect to safety-related processes.
- **Functionality:** the functions at the level expressed in functional requirements of the model, emphasizing at the level of final user functionality
- **Reliability:** the probability of the process model delivering results that is consistent with the model assumptions.
- **Regulation:** decision support reference model that will ensure that healthcare software products are safe and effective to protect and promote public health through various standards and regulations.
- **Validity:** the ability of the process model to reflect the assumptions and constraints about the software process specified by process stakeholders.
- **Accuracy:** the measurement tolerance, or transmission of the process model that defines or removes the limits of the errors.

## 6 FUTURE RESEARCH

Having established a foundation for the Trustworthiness Healthcare Software Model by identifying the key attributes of trustworthiness from a healthcare software perspective, we will continue to build on this to establish key processes and metrics within the model.

As part of our future research, we will examine and modify existing trustworthiness models. The subsequent focus will be on extending and modifying existing techniques based on our identified attributes for the analysis. Then our next step will be to take the trustworthiness model that we develop, to test and refine it on a large scale with healthcare software sector. This way, we will also be able to say which attributes are the most important.

## 7 CONCLUSIONS

The primary goal for adopting healthcare software is to provide patients the best service possible by gathering and interpreting accurate information. This should help them to take correct and timely decisions which reduces cost, time and effort, thereby resulting in the timely treatment of the patient. But, there are apparent concerns regarding whether we can trust healthcare software solutions.

We have identified that there is a gap, and therefore, a consequent need to introduce a Trustworthiness Healthcare Software Model. In this study, we have focused on an initial definition of trustworthiness attributes from literature. We highlight our next steps towards the development of the Trustworthiness Healthcare Software Model and its validation across the healthcare software sector.

## ACKNOWLEDGEMENTS

“This work was supported with the financial support of the Science Foundation Ireland grant 13/RC/2094 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero - the Irish Software Research Centre (www.lero.ie)”.

## REFERENCES

Amoroso, E., Taylor, C., Watson, J. and Weiss, J., 1994, November. A process-oriented methodology for assessing and improving software trustworthiness. *In Proceedings of the 2nd ACM Conference on Computer and communications security pp. 39-50.*

Armitage, A and Keeble-Allen, D. 2008."Undertaking a structured literature review or structuring a literature review: Tales from the field", Electronic Journal of

Business Research Methods 6 (2), pp. 103-114. Available www.ejbrm.com.

Boehm, B., 2006, May. A view of 20th and 21st century software engineering. *In Proceedings of the 28th international conference on Software engineering pp. 12-29*. ACM.

Burton, J., McCaffery, F. and Richardson, I., 2006, May. A risk management capability model for use in medical device companies. *In Proceedings of the 2006 international workshop on Software quality pp. 3-8.*ACM.

Carbone, M., Christensen, A.S., Nielson, F., Nielson, H.R., Hildebrandt, T. and Sølvkjær, M., 2013, August. ICT-powered Health Care Processes. *In International Symposium on Foundations of Health Informatics Engineering and Systems pp. 59-68*. Springer Berlin Heidelberg.

Carroll, N. 2016. *Key Success Factors for Smart and Connected Health Software Solutions, Computer, Vol. 49, No. 11, pp. 32-38.*

Carroll, N. and Richardson, I., 2016. Software-as-a-Medical Device: Demystifying Connected Health Regulations. *Journal of Systems and Information Technology, 18(2).*

Catwell, L. and Sheikh, A., 2009. Evaluating eHealth interventions: the need for continuous systemic evaluation. *PLoS Med, 6(8)*, p.e1000126.

Clarke, P., Lepmets, M., McCaffery, F., Finnegan, A., Dorling, A. and Flood, D., 2014, November. MDevSPICE-a comprehensive solution for manufacturers and assessors of safety-critical medical device software. *In International Conference on Software Process Improvement and Capability Determination pp. 274-278*. Springer International Publishing.

European Commission (2014), "Green paper on mobile health ('mHealth')", COM 219 final, SWD 135 Final, available at: http://ec.europa.eu/digital-agenda/en/news/green-paper-mobilehealth- *mhealth (accessed November, 2016).*

Huckvale, K., Adomaviciute, S., Prieto, J.T., Leow, M.K.S. and Car, J., 2015. Smartphone apps for calculating insulin dose: a systematic assessment. *BMC medicine, 13(1*), p.1.

Huckvale, K., Car, M., Morrison, C. and Car, J., 2012. Apps for asthma self-management: a systematic assessment of content and tools. *BMC medicine, 10(1), p.1*.

Huckvale, K., Car, M., Morrison, C. and Car, J., 2012. Apps for asthma self-management: a systematic assessment of content and tools. *BMC medicine, 10(1), p.1*.

Institute of Medicine (2007) *Preventing medication errors*. Washington (D.C.): National Academy Press.

Jepsen, T., 2003. IT in healthcare: progress report. IT professional, 5(1), pp.8-14.

Larsen, M.E., Nicholas, J. and Christensen, H., 2016. A systematic assessment of smartphone tools for suicide prevention. *PloS one, 11(4)*, p.e0152285.

Leroy, G., Chen, H. and Rindflesch, T.C., 2014. Smart and connected health. *IEEE Intelligent Systems, 29(3), pp.2-5.*

McCaffery, F., Dorling, A. and Casey, V. (2010), *"Medi SPICE: an update"*, available at: http://eprints.dkit.ie/48/ (accessed 14 November 2016).

McHugh, M., McCaffery, F. and Casey, V., 2012. Software process improvement to assist medical device software development organisations to comply with the amendments to the medical device directive. *IET software, 6(5), pp.431-437.*

Merriam-Webster, 2004. *Merriam-Webster's Collegiate Dictionary*. Merriam-Webster.

Nicholas, J., Larsen, M.E., Proudfoot, J. and Christensen, H., 2015. Mobile apps for bipolar disorder: a systematic review of features and content quality. *Journal of medical Internet research, 17(8).*

Olff, M., 2015. Mobile mental health: A challenging research agenda. *European journal of psychotraumatology, 6.*

Petticrew, M.A. (2001)."Systematic literature reviews from astronomy to zoology: Myths and misconceptions", *British Medical Journal 322 (7278), pp. 98-101*. Available www.bmj.com/cgi/contact/full/322/7278/98.

Shojania, K.G. and Dixon-Woods, M., 2016. Estimating deaths due to medical error: the ongoing controversy and why it matters. *BMJ Quality & Safety, pp.bmjqs-2016.*

The Daily Mail: *Health warning over blood pressure monitoring apps as doctors warn they are 'untested, inaccurate and potentially dangerous'*. Available at: http://www.dailymail.co.uk/sciencetech/article-2887791/Health-warning-blood-pressure-apps-doctors-warn-untested-inaccurate-potentially-dangerous.html.

United Nations. *International Standard Industrial Classification of All Economic Activities, Rev.3.* New York.

Van Velsen, L., Wildevuur, S., Flierman, I., Van Schooten, B., Tabak, M. and Hermens, H., 2016. Trust in telemedicine portals for rehabilitation care: an exploratory focus group study with patients and healthcare professionals. *BMC medical informatics and decision making, 16(1), p.1*.

Wu, R., Ahn, G.J. and Hu, H., 2012, January. Towards HIPAA-compliant healthcare systems. *In Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium pp. 593-602.* ACM.

Yang, Y., Wang, Q. and Li, M., 2009, May. Process trustworthiness as a capability indicator for measuring and improving software trustworthiness. In *International Conference on Software Process pp. 389-401*. Springer Berlin Heidelberg.

Zhang, H., Kitchenham, B. and Jeffery, R., 2012. Toward trustworthy software process models: an exploratory study on transformable process modeling. *Journal of Software: Evolution and Process, 24(7), pp.741-763*.

# A Step Towards the Standardisation of HIV Care Practices

Cristina-Adriana Alexandru[1], Daniel Clutterbuck[2], Petros Papapanagiotou[1], Jacques Fleuriot[1] and Areti Manataki[1]

[1]*School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh, EH8 9AB, U.K.*

[2]*Genitourinary & HIV Medicine, Chalmers Sexual Health Centre, NHS Lothian, 2A Chalmers St, Edinburgh, EH3 9ES, U.K.*

{*calexan4, ppapapan, jdf, amanatak*}*@inf.ed.ac.uk, daniel.clutterbuck@nhs.net*

Keywords: HIV, Integrated Care Pathways, Formal Workflow Modelling.

Abstract: Recent improvements to HIV care at the NHS Lothian Board have concentrated on a re-mapping of the processes involved in their existing Integrated Care Pathway (ICP), in order to incorporate improvements identified during the ICP implementation and consider new advances in care. Our work aims to extend and enhance this mapping by formalising care workflows using our logic-based tool WorkflowFM. This paper presents our progress to date in terms of methodology and initial findings concerning actors, resources and workflows involved in the first 3 months of HIV care for the Chalmers Sexual Health Centre. We argue that the resulting models and analysis could address some of the difficulties faced by units providing HIV outpatient care.

## 1 INTRODUCTION

HIV infection is a major global concern. In 2015, 36.7 million people were living with HIV, and it was the cause of 1.1 million deaths worldwide (UNAIDS, 2016). HIV care is complex and lifelong, which makes the efficient collaboration of multidisciplinary teams at local, regional and national levels important (HIS, 2011). In Scotland, a recognition of the inconsistency of existing HIV services has led in 2011 to the consideration of Integrated Care Pathways (ICPs) as a key priority area within HIV standards (Standard 9 (HIS, 2011)). ICPs are recognised as valuable in the long term management of HIV as a chronic disease, but the particular complexity of the first 3 months of care following diagnosis resulted in a prioritisation of support for developing ICPs for the first 3 months of HIV care across the NHS Boards (HIS, 2013).

An ICP is defined as "a structured chronological, multidisciplinary clinical record, developed by local development groups, to suit a local situation" (SPA, 2016) and NHS Scotland uses the one proposed by the Scottish Pathways Association (SPA). The SPA also points out that an ICP could contain a chronological plan of a patient's care or treatment, usage protocols, information on guidelines or instructions, sign off sections for allocating responsibilities for entries and variance tracking sections for recording cases when the patient does not follow the usual pathway.

The NHS Lothian Board has been using an ICP for the first 3 months of HIV care since April 2012 (Wielding et al., 2013). The ICP acts as a map of the patient's journey through care and was devised through a development process following a recognised methodology (ICPUS, 2007) that involved multidisciplinary HIV care teams – consultants, nurses, pharmacists, other health professionals – and patient representatives. Its creation proceeded mainly through an adhoc identification of the various activities involved in the provision of care and resulted in documents intended to capture a multidisciplinary, chronological and structured case-record for each patient. The developed ICP pathway documents were first used in paper, and since 2014, electronic form.

The development and use of the NHS Lothian HIV ICP helped not only improve and maintain a consistent standard of care amongst patients, but also identify a range of potential improvements to care delivery. These included, for example, routinely checking vaccination status prior to clinic visits, checking results prior to patient attendance, cutting down on redundant documentation and instigating some self-completion of social and mental health status information. Implementing such changes requires an ongoing re-thinking and recording of the flow of processes. This is not uncommon, as ICPs are regarded by the research community as continuous care process improvement interventions (Vanhaecht et al., 2010).

As part of the continuous improvement process, we started a detailed, post-hoc mapping of the work-



Alexandru, C-A., Clutterbuck, D., Papapanagiotou, P., Fleuriot, J. and Manataki, A.
A Step Towards the Standardisation of HIV Care Practices.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 457-462
ISBN: 978-989-758-213-4

flows[1] involved in the ICP in collaboration with the NHS Lothian HIV ICP Development Group. We aim to create and document an unambiguous model of care that is shareable among the HIV team and across NHS boards and can be modified as new policies, guidelines and systems are introduced in HIV care. Ultimately, this model could be the precursor to the systematic design of an adaptable IT system that facilitates documentation, integrates the existing medical record systems, tracks each patient's journey, flags up the tasks that need to be done, and routinely records outputs such as completion rates, timescales etc.

The process of mapping the care processes in the initial ICP in Lothian was iterative and relatively informal: using a map of the pathway and asking multidisciplinary team members to add detail individually or in groups, then reviewing and revising the map and the resulting ICP document over multiple cycles. The aim of the current work is to extend and improve these mapping activities with the use of the process modelling framework WorkflowFM (Papapanagiotou and Fleuriot, 2014). This provides a diagrammatic interface to build process workflows in a formal, systematic way, similar in some respects to other common process modelling languages such as BPMN (Object Management Group, 2011). In addition, though, its logical core helps reduce design errors by providing mathematical guarantees of correctness, systematic accounting of resources and freedom from deadlocks. The resulting workflows clearly depict the modelled practices using a simple graphical notation, can be easily modified to reflect policy changes and can be executed to simulate patient journeys.

The capabilities of WorkflowFM fit well with the process mapping needs of the HIV ICP Development group, as it not only allows a structured, shareable representation of their deployed ICP, but can also help analyse and evaluate potential changes in the ICP before they are applied in practice. We note, in passing (due to space limitations), WorkflowFM's growing track record in the analysis of various healthcare processes. For a more in-depth look, we refer the interested reader to our work on collaboration patterns in healthcare (Papapanagiotou and Fleuriot, 2014) and intra-hospital patient transfers (Manataki et al., 2016).

This paper presents our methodology and progress so far for the Chalmers Sexual Health Centre, one of the largest HIV care providers in Scotland.

[1]In this paper, we define a workflow as a set of well-defined processes, sequential or concurrent, which take a work task from initiation to completion. Processes can be seen at different levels of abstraction – from single steps, to aggregates of several steps, to even a whole workflow.

# 2 METHODOLOGY

In our approach, we adopt a 3-stage methodology:

1. **Gain an in-depth understanding of HIV care in the organisation** by consulting available documentation, shadowing and interviewing the different involved stakeholders, including clinicians, nurses, administrators and patients.
2. **Use the results of the first step to formally specify the workflows for the existing Lothian ICP, suggest improvements** (e.g. with respect to efficiency or new approaches to patient care) **and propose new formalised ICPs for areas not covered by the current one** using WorkflowFM.
3. **Evaluate the formal ICPs** by re-approaching the participants from Step 1 with a questionnaire about the key processes, exceptions, resources and information used in the workflow diagrams.

At the end of the project, we will thus have a set of validated formal ICPs that can be shared with different HIV care providers to help them re-think their workflows and agree on an improved, up-to-date ICP.

# 3 DATA COLLECTION

The data for this work was collected by going through the existing documentation, consisting of preliminary flowcharts, checklists and tables devised by the NHS Lothian's HIV ICP Development Group, chaired by D. Clutterbuck, a Consultant in Genitourinary and HIV medicine and one of the authors of this article.

In order to gain a range of perspectives on the care processes at play, we interviewed a variety of stakeholders including (other) consultants, nurses, pharmacists and coders. We focused on breaking down the resources (paper or electronic) used, the flow of information, how these depend on the patient and time in their care, and whether the performed activities actually fit the existing records. Interviews were semi-structured for one hour. With the participants' approval in a consent form, interviews were audio recorded, and notes were taken. Once the interviews were transcribed, the data was analysed qualitatively.

# 4 INITIAL RESULTS

## 4.1 The Actors Involved in HIV Care

The main actors involved in HIV care in the Chalmers Sexual Health Centre are nurses, consultants, pharmacists and clinical coders. The responsibilities of

**nurses** vary from doing basic clinical tasks such as taking bloods or doing urine analysis (Band 2, also known as clinical support workers), to giving prescribed vaccines and treatments (Band 5), deciding on treatment and seeing patients independently (Band 6, the majority in Chalmers) and having all of the previous responsibilities but also a leadership role in the management of HIV care (Band 7). **Senior nurses** (Bands 6 and 7) are usually the first point of contact for new patients. They also have **health advisor** roles within their team, and thus can discuss partner notification with patients. Additionally, during the first 3 months of care they act as **ICP Leads**, making sure that the ICP documents are completed within this time. Senior nurses have their own appointments with patients for completing the required elements of care and filling in the ICP documents. **Consultants** are specialist HIV clinicians who see patients for a medical review and decide on their treatment. They work during Monday or Thursday afternoon clinics to see their own patients (6-8 appointments). They may also be contacted about them outside of these clinics. On a rotational basis, they occupy the role of **senior genitourinary medicine (GUM) consultants** covering the clinic throughout the day, which involves being available for consultation with the senior nurses, seeing patients in person for urgent cases (e.g. a patient being unwell), or liaising with the pharmacist regarding emergency medication requests. **Pharmacists** are responsible for medicine reconciliation (i.e. checking what other medication patients are already on, what drug allergies they have and checking for interaction with their ARV regimen), advising patients on treatment options and dispensing drugs. Clinical pharmacists also take part in formal and informal meetings with doctors and senior nurses to decide on HIV treatment initiation and treatment changes, providing expertise on appropriate drug choices, interactions and suitability. During the first 3 months of HIV care, the **clinical coders** are responsible for keeping the computer records up-to-date and consistent and moving information on blood results from electronic form to a paper care plan for pharmacists for weekly meetings.

## 4.2 Resources Used

Chalmers currently use two main computer systems to support HIV care: an **Access database**, initially developed for infectious diseases in the Western General Regional Infectious Disease Unit (RIDU), which now also contains forms for the ICP checklists, and the **National Sexual Health System (NaSH)**, a comprehensive patient management system for sexual and reproductive health which was introduced for clinical care in the Lothian Sexual and Reproductive Health Service in 2011 and adopted for use in HIV care in 2014. Both of the systems are important for HIV care in Chalmers. NaSH is used in all clinics in Scotland for sexual and reproductive health care and facilitates communication with other sexual health clinics, replacing paper records. The HIV database facilitates prescribing and helps ensure that all of the necessary steps for the first 3 months of care have been performed. However, the two systems are not integrated and they contain numerous common fields, thus leading to a lot of duplication. Keeping information up-to-date and consistent between them is time consuming and an ongoing concern. Common fields are often in different formats. Moreover, each of the systems has additional fields and functionality, making information very scattered and requiring clinicians to frequently move between the two systems. To speed this up at regular meetings discussing the management of each patient who is to be seen that week, workarounds have evolved. **Standard patient letters** developed on NaSH have been introduced, containing an overview of the patient's health status, medication, immune status and serology and highlighting any issues or outstanding tasks and gaps in the ICP documentation. These are updated by clinicians after each patient appointment with the primary purpose of forwarding them to the general practitioners (GPs) or using as transfer letters, but also serving as a concise summary of the ICP and its completion during meetings. After each appointment, consultants and nurses also write **clinical note summaries on NaSH**, which are also used in meetings. Moreover, pharmacists use a **paper care plan** when discussing prescriptions, which is pre-populated with blood results and clinical issues from the two systems.

HIV clinicians may sometimes need to consult other computer systems that do not currently communicate with NaSH or the HIV database: **Trak** for demographics, lab results, ordering other tests including imaging, and for onward referral to other specialities (Trak is the main patient management system for hospital services in Lothian, so patients will have a record there if they are seen for other conditions), **Apex** for lab results and **SCCR** for cervical smear test results.

## 4.3 First 3 Months of HIV Care

A patient may enter the HIV care pathway through different means: he/she may be newly diagnosed with HIV – either in Chalmers or in one of the numerous Outreach clinics doing point of care testing in Lothian – or transferred from another HIV health centre, or referred from their GP. Once he/she enters the pathway,

the **ideal workflow** involves the following processes:

1. No matter how he/she entered the pathway, the patient is seen for a **one hour first visit by one of the senior nurses**. Patients who were recently diagnosed in Chalmers are given their diagnosis during this visit. Moreover, all patients have a confirmatory HIV antibody test and other baseline investigations (depending on whether they have done some of them previously) and have an initial discussion with the senior nurse which varies very much case-by-case, but usually is about how they feel, what HIV means to them and how they are going to cope. Very importantly, during the first visit with a patient the senior nurse fills in as much as possible of his/her NaSH record, registers the patient on the database and fills in as much as possible of the ICP checklist from the database together with the patient. The senior nurse must also discuss with the patient about his/her availability and preference for a certain consultant (e.g. some patients may wish to be seen by the same consultant as their partners). This, together with the availability of the consultants and the day of the patient's first visit, will inform the senior nurse's decision regarding the team (Monday or Thursday) and consultant to whom to allocate the patient. A diagrammatic representation of the workflow for the first visit for new patients after they have received their diagnosis is presented in Fig. 1, and will be discussed in subsection 4.4. After the first visit, the senior nurse books an appointment for the patient with his/her consultant (the medical review visit) within the following 2 weeks, when investigation results are back.
2. The patient attends the **first medical review visit**, during which the consultant usually discusses with him/her presenting issues, past medical history, medications, sometimes family history, and decides on a management plan. Depending on the patient, antiretroviral treatment options may also be discussed at this time. The consultant ideally fills in medical fields on NaSH and the database during this discussion. After the visit, the consultant summarizes the discussion in a NaSH clinical note and fills in the standard patient letter from NaSH which, with the patient's approval, is forwarded to his/her GP.
3. If an antiretroviral treatment regimen has been decided, the consultant contacts the pharmacists about the treatment decision or proposed options. One of them will meet with the patient (**pharmacist visit**) within the same clinic to collect information on medications and drug allergies and make a decision on medicine reconciliation. The pharmacist may discuss treatment options with the patient, help him/her make a decision and provide the drugs.
4. For most patients, the baseline assessment and ICP are completed within 3 months of entering the pathway. Appointment frequency then decreases once patients become more stable and their viral load is undetectable (usually within 6 months). For patients who have an undetectable viral load on therapy and do not have significant psychological or social problems or other physical comorbidities, visits then occur every 6 months as part of the organised programme of routine care.

These processes seem mostly sequential and quite straightforward. However, we have marked them as ideal because they only apply to patients who do not require a lot of support, who are at a stage in their disease when things can progress at a normal pace, not needing an urgent medical review, and for cases where emergencies never occur. Moreover, they rely on the availability of staff and patients, and in particular limited numbers of staff and increasing numbers of patients in Chalmers are an issue. In reality, the ideal workflow occurs rarely. More often than not the workflow has variances such as exceptions, optional processes, repeated processes, or processes happening in any order, making it very complex. Such variances are due to the patients' state, need for support and how this evolves over time:

1. Sometimes a senior nurse must make a decision to bypass a patient's first visit in order to have him/her see a consultant sooner. In such situations, the items from the first visit usually need to be deferred to later, one or more, regular (30 minute) appointments with the senior nurse. The most important cases are the following:
   - If a new transferred in or referred patient is found during an initial prioritization before the first visit (by checking the transfer/referral letter or calling the patient on the phone) to be unwell or already taking antiretrovirals and not having enough medication left, the senior nurse will set up an early medical review appointment for him/her with an available consultant.
   - When a patient seems unwell and needs an urgent review when coming in for his/her first visit, the senior nurse contacts the senior GUM consultant to see him/her immediately.
2. Especially if newly diagnosed, a patient may feel distressed and need additional support, and so the senior nurse may need to meet with him/her repeatedly (regular nurse 30-minute visits) after the

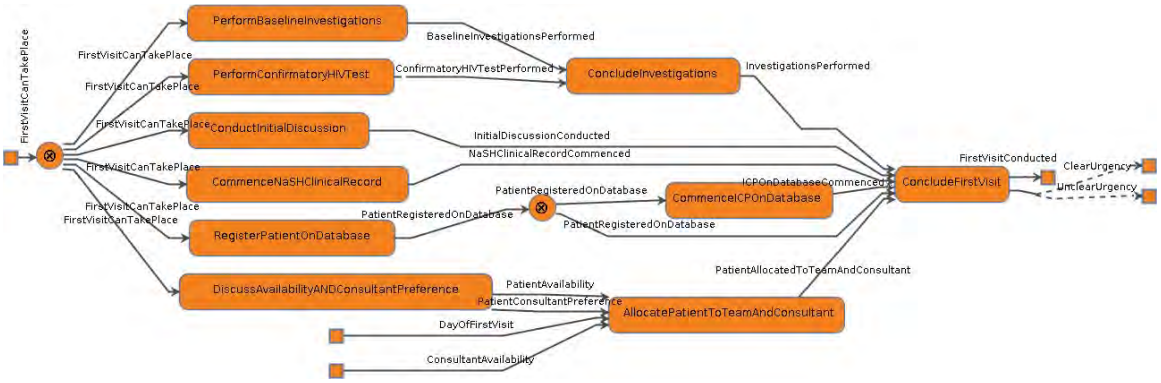



Figure 1: Workflow for the first visit after diagnosis for new patients.

first visit. The nurse will usually try to fit such subsequent appointments in the same day with the patient's medical review visit with the consultant, but this is not always possible.

3. Following a senior GUM consultant visit, or the first visit with the senior nurse, depending on his/her state a patient may require an early review appointment (occurring earlier than the default 2 weeks) to be booked with a consultant.
4. Following a medical review visit, a consultant may decide that early follow-up appointments are necessary, e.g. if the patient has abnormal results, medical issues, mental issues, is vulnerable, has comorbidities that affect his/her immune status or will require an STI screen or vaccination.

Another important exception is given by the fact that senior nurses often do not manage to finish filling in all of the relevant parts of the NaSH and ICP checklists from the database during first visits. In this case, they need to, sometimes repeatedly, book additional routine nurse appointments with the patients.

The steps required for filling in information on the systems are, actually, often performed in an order which depends on the patient's state and needs, and earlier or later within the actual visits, which makes them optional at different times. Although there is an expectation for consultants to fill in some information on the systems, we have found that this is not done consistently, as they prioritise patient care. Some have expressed their concern about the unclear sharing of responsibility for filling in the ICP checklists, and the lack of time to do this during an appointment.

Another less frequent exception is given by the case of patients whose level of urgency for being seen for a medical review is not completely clear to the senior nurse after their first visit. In such circumstance, the senior nurse first discusses with a consultant before booking subsequent appointments for the patient.

Apart from the work directly involving patients, there is also important background work happening in Chalmers during the first 3 months of care. In particular, the two Chalmers HIV teams, the Monday and the Thursday team, meet every week to discuss the management of the patients who are scheduled in clinic that week, and to check progress with the ICP. Patients are discussed in turn, using the NaSH summary page (projected onto a screen from the system), the pharmacists' care plan with pre-filled bloods and clinical details and their printed provisional prescriptions. If a patient's status is not clear or more up to date information is needed, the team may also browse through clinical notes or other pages from NaSH, the HIV database and even other systems such as Apex. Once a week, clinical coders move any additional information from NaSH onto the ICP checklists from the database. The ICP leads (senior nurses) check progress with the ICP checklists on the database for each patient and flag up missing information before the meeting during their admin time. They may use clinical notes to remind consultants to fill it in. Once information of a form is complete, they sign it off.

## 4.4 Rigorous Workflow Modelling

As mentioned previously, we model and compose the workflows involved in the first 3 months ICP using WorkflowFM, a graphical tool built on top of logical foundations. Processes in WorkflowFM are represented visually as rectangles, and their inputs and outputs as edges. An example diagram of the workflow for the first visit for new patients after they have received their diagnosis is provided in Fig. 1.

In this, one can clearly identify processes that are independent and can therefore be performed concurrently, in any order. For example, once the patient attends for the first visit, the senior nurse can perform different investigations, have an initial discussion with the patient, commence the NaSH clinical record, register him/her on the database or discuss

his/her availability and consultant preference in any order. The priority of these steps is influenced by the patient's particular state and needs.

Other processes must be performed sequentially, as they are dependent on the ones preceding them. For example, only once the senior nurse registers the patient on the database can he/she commence the ICP on the database together with the patient. Also, the patient's availability and preference for a certain consultant partly influences the senior nurse's decision to allocate the patient to a certain team and consultant.

The final output of the workflow is the completion of the first visit and the evaluation of the level of urgency of the patient's state. The senior nurse may not be able to clearly determine this, in which case (marked as `UnclearUrgency`) he/she will need to discuss the case with a consultant before booking any other appointments (as described in subsection 4.3).

WorkflowFM allows an executable deployment of the developed models for both simulation and use in practice. Simulation enables further analysis of the workflows with respect to the available resources on site (including clinical workload), acceptable completion rates, costs, conflicts across multiple pathways, etc. The models can also be deployed to support the ICP through rigorous tracking of information and resources and automatically generated checklists for each step. This provides a skeleton for a larger system that can also integrate with the NaSH clinical record system, provide decision support, and send notifications and reminders to guide users through the ICP. We view this as a promising way of taking our work further once we conclude the evaluation of the ICPs (Step 3 of our methodology – see Section 2).

# 5 CONCLUSION

This paper described our progress towards formally re-mapping the integrated care pathways for the first 3 months of HIV care at the Chalmers Sexual Health Centre. Our study showed that the most important difficulty faced by HIV specialists is the requirement to continuously check and update information on several non integrated systems. This leads to inefficiency, frustration and burdens them with the need to ensure consistency in order to avoid errors. We believe the formal models we have developed using our logic-based tool WorkflowFM are not only a key first step towards addressing these core issues and optimizing current practices, but also offer a coherent, readily adaptable blueprint for designing and implementing an effective IT support system for HIV care at Chalmers and beyond.

# ACKNOWLEDGEMENTS

This research is supported by EPSRC grant EP/N014758/1. We would like to thank the team of senior nurses, consultants, pharmacists and clinical coders from the Chalmers Sexual Health Centre for their valuable insights into their work practices.

# REFERENCES

HIS (2011). Human immunodeficiency virus (HIV) services - standards. Technical report, Healthcare Improvement Scotland.

HIS (2013). HIV implementation and improvement programme. http://www.healthcareimprovementscotland.org/our_work/long_term_conditions/hiv_treatment_and_care/hiv_icp_implementation.aspx.

ICPUS (2007). Integrated care pathway users in Scotland – A workbook for people starting to develop integrated care pathways. Technical report.

Manataki, A., Fleuriot, J., and Papapanagiotou, P. (2016). A workflow-driven, formal methods approach to the generation of structured checklists for intra-hospital patient transfers. *Journal of Biomedical and Health Informatics.*

Object Management Group (2011). Business Process Model and Notation (BPMN), version 2.0. http://www.omg.org/spec/BPMN/2.0/PDF.

Papapanagiotou, P. and Fleuriot, J. D. (2014). Formal verification of collaboration patterns in healthcare. *Behaviour & Information Technology*, 33(12):1278–1293.

SPA (2016). What is an integrated care pathway? Technical report, Scottish Pathways Association.

UNAIDS (2016). AIDS update. Technical report, Joint United Nations Programme on HIV/AIDS.

Vanhaecht, K., Panella, M., Van Zelm, R., and Sermeus, W. (2010). An overview on the history and concept of care pathways as complex interventions. *International Journal of Care Pathways*, 14(3):117–123.

Wielding, S., Clutterbuck, D. J., Wilks, D., and Panton, L. (2013). Development of an ICP for HIV outpatient care in Lothian, Scotland. In *European Care Pathways Conference Glasgow*.

# Sharing of Big Data in Healthcare: Public Opinion, Trust, and Privacy Considerations for Health Informatics Researchers

Laura Moss[1,2], Martin Shaw[2], Ian Piper[2], Christopher Hawthorne[3] and John Kinsella[1]
[1]*Department of Anaesthesia, Pain & Critical Care, School of Medicine, University of Glasgow, Glasgow, U.K.*
[2]*Department of Clinical Physics, NHS Greater Glasgow & Clyde, Glasgow, U.K.*
[3]*Department of Neuroanaesthesia, NHS Greater Glasgow & Clyde, Glasgow, U.K.*
{*laura.moss, john.kinsella*}*@glasgow.ac.uk, martin.shaw@nhs.net, ian.piper@brainit.org, cwhawthorne@doctors.org*

Keywords: Big Data, Privacy, Critical Care, Data Analysis, Trust, Data Security.

Abstract: Advances in technology has transformed clinical medicine; electronic patient records routinely store clinical notes, internet-enabled mobile apps support self-management of chronic diseases, point-of-care testing enables laboratory tests to be performed outside of hospital environments, patient treatment can be delivered over wide geographic areas and wireless sensor networks are able to collect and send physiological data. Increasingly, this technology leads to the development of large databases of sensitive electronic patient information. There is public interest into the secondary use of this data; many concerns are voiced about the involvement of private companies and the security and privacy of this data, but at the same time, these databases present a valuable source of clinical information which can drive health informatics and clinical research, leading to improved patient treatment. In this position paper, we argue that for health informatics projects to be successful, public concerns over the secondary use of patient data need to be addressed in the design and implementation of the technology and conduct of the research project.

## 1 INTRODUCTION

Healthcare is rapidly changing and advanced technology is enabling the collection of vast amounts of patient data. Consequently healthcare is experiencing a Big Data phenomenon. Whilst health informatics research is focused on the development of novel approaches to enable the intelligent analysis of this data, the use of electronic patient data to advance these approaches and the implementation of these technologies within real world healthcare environments raises many ethical questions and divides public opinion.

In this paper we explore some of the issues regarding the use of electronic patient data for secondary purposes, specifically trust, security and patient confidentiality. The paper is organised as follows: section 2 describes the prominence of Big Data within healthcare and challenges faced; section 3 discusses issues raised in the sharing of data, in particular the involvement of private companies and views of the general public; section 4 illustrates real-world issues faced whilst implementing a big data analysis platform into a healthcare environment; section 5 discusses considerations and opportunities for health informatics researchers; finally, section 6 concludes the discussion.

## 2 BACKGROUND

As technology advances, data is increasingly collected through a variety of mechanisms. It is estimated that 2.5 quintillion bytes of data is currently generated each day (IBM a, 2016). Big Data is a term which is now commonly used to describe such large and complex datasets. Advanced analytics are often applied to Big Data to extract meaning, insights and discovery of new knowledge. However, using these large datasets presents challenges due to the data's volume (many sources e.g. sensors, social networking), variety (many formats, e.g. videos, text) and velocity (speed at which data is produced and requirements for near-real time processing). Other attributes can include veracity (noisy, messy data), variability (meaning of the data can be constantly shifting) and fine-grained (Kitchen and McArdle, 2016).

Within healthcare, the increasing use of technology is creating large volumes of clinical data; in 2011, the global size of healthcare data was estimated to be 161 billion gigabytes (IBM b, 2016). This influx of data is from a variety of technical advances: e.g. enhanced clinical imaging, electronic medical records, and physiological sensors. Additionally, the growth



Moss, L., Shaw, M., Piper, I., Hawthorne, C. and Kinsella, J.
Sharing of Big Data in Healthcare: Public Opinion, Trust, and Privacy Considerations for Health Informatics Researchers.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 463-468
ISBN: 978-989-758-213-4

of wearable sensor devices (e.g. smart watches) is leading to patients themselves being able to generate large amounts of healthcare data.

The potential clinical benefit of using this data is well established. Analysis of healthcare data can drive improvement in many areas, including: clinical and organisation processes, optimisation of treatments, and reduction of healthcare costs through intervention at an earlier point and more proactive and targeted care. Patient data can be used to predict clinical risk, targeting resources where they are needed most and identifying problems that would benefit from early intervention. Technologies that aid clinical decision-making and help clinicians to manage the exponential growth in medical knowledge offer substantial opportunities to reduce variation, improve the quality of care, and possibly lower costs (Jaspers et al., 2011), (Fillmore et al., 2013).

Whilst a thriving health informatics community is focused on the development of novel approaches and tools to enable the intelligent and sophisticated analysis and use of healthcare data, the transfer of these technologies and ideas to real world clinical environments faces many challenges. Healthcare providers often do not have access themselves to data science expertise, required technology infrastructure and funding to do this themselves. In many cases, to enable the potential for big data analysis to improve healthcare, partnerships have to be formed between healthcare providers, research organisations and private companies. High profile examples include DeepMind and IBM Watson technology. IBM's Watson is being used within oncology in US and Canadian hospitals to assess tumours (IBM c, 2016) and DeepMind is using patient data from UK hospitals to develop diagnostic tools in areas such as acute kidney injury (DeepMind a, 2016) and ophthalmology (DeepMind b, 2016).

# 3 TRUST, SECURITY AND PRIVACY

Trust and confidentiality between a clinical provider and patient is not new: it is central to the practice of healthcare and has been focused on since Hippocrates. Whilst the concept of patient confidentiality has endured as an ideal throughout history, the precise nature of it has changed with the sociohistoric context (for a detailed review see (A.Ferguson, 2012)). In the digital age, patient confidentiality is often framed within the context of electronic patient records and the potential involvement of third parties. Whilst the involvement of private organisations and other research organisations can resolve many practical issues for healthcare providers, it often involves the transfer of sensitive patient data to these organisations for research development and this can raise many questions. For example:

- Should electronic patient data be used for secondary research purposes?
- Who does the electronic patient data belong to?
- Who should be able to use the electronic patient data? Public sector healthcare organisations (e.g. NHS) and/or commercial companies?
- Who is collecting the data and where is it stored?
- What safeguards should be put in place to protect patient confidentiality?
- How do patients and relatives feel about the collection and use of this data?

Most countries have some form of regulation or processes which have been legally put in place to resolve some of the above issues. Additionally, institutions often have internal procedures and practices in place to protect patient data. However, public perception of the adequacy of these frameworks can be divided and concerns raised about whether private companies, in particular, can be trusted with patient data. For example, the public sector can be perceived as more trustworthy than profit-making organisations when using patient data (Focus Groups et al., 2013). Recent publicity in cases such as DeepMind's arrangement with the Royal Free Hospital, London, has been controversial and highlighted the requirement for more robust safeguards to be put in place to ensure patient data is adequately secured (New Scientist, 2016). Additionally, there are constant reports in the media about data leaks and unsuccessful, large-scale, health I.T projects (Presser et al., 2015).

It also raises issues about whether patient consent and public awareness of healthcare data sharing is adequate. A number of studies have identified that there is low awareness by the general public of electronic patient record systems and how and why healthcare data might be used (BMA, 2016) (Riordan et al., 2015). Riordan et al (Riordan et al., 2015) found that most people would prefer to opt-in before their identifiable records were used and half of participants would share their de-identified records under implicit consent. A recent consultation with the general public in the UK on this issue identified that they had little confidence in the safeguards put in place to protect data. Additionally, it was felt that there was a lack of accountability within the system, and malicious use of data by private companies (e.g. pharmaceutical and insurance) was a concern for many

people (BMA, 2016). Lessons learnt from the failure of large scale patient data sharing projects show the requirements for clear communication to the public, easy to understand consent rules and strong oversight and communication regarding distribution and use of patient data (Presser et al., 2015). For a more detailed systematic review of the literature regarding public responses to sharing of health data see (Aitken et al., 2016b).

To investigate the opinions of the general public further, in a small study we asked the audience at a science festival their thoughts on the topic of secondary use of patient data (Kinsella et al., 2017). Questions covered included: whether the participants were aware of the potential of using their medical data for secondary research purposes; whether patient data should be used for research purposes and how likely would they be themselves to share their own personal data for research; if they trust clinicians with their data; and their opinions on the role of private/commercial companies in supporting and/or carrying out research on their own medical data. 39 out of 41 adults responded to the survey (from which we have full results for 37 adults). Table 1 shows results from the survey. The vast majority of respondents felt that their medical data should be used for research purposes and would be happy to share their data. This is in keeping with a number of other studies (as summarised in (Aitken et al., 2016b). Additionally, most respondents trusted clinicians, but when it came to private companies, the response was mixed. This difference in perceived trustworthiness between clinicians and private companies has been found previously (e.g.(Aitken et al., 2016a),(BMA, 2016)).

We also looked for any age divide in the participant's responses. We asked which age category participants fell into (21 and under, 22-34, 35-44, 45-54, 55-64, 65 and over) and three questions regarding the involvement of private companies: 1) Would you trust a private company to do research with your medical data? 2) Would you trust a private company to do research with anonymous medical data? 3) Would you be comfortable with a private company providing the support to medical researchers to enable them to do medical research? Tables 2 and 3 display the results of this analysis. In both age groups, more people had positive responses (agreed or strongly agreed) than negative ones (disagree or strongly disagree). Although younger people are often thought to be more confident with technology and data sharing, no sizable difference between the two age groups was found in this study. A larger sample size would be required to show statistical significance.

Public opinion and perception of the use of healthcare data can be divided and it is clear from these studies and the opinions of other researchers (e.g. (van Staa et al., 2016)), that for health informatics projects involving transfer and analysis of patient data by third parties to be a success, the trust of the general public needs to be earned and respected by all involved.

# 4 CHART-ADAPT CASE STUDY

To illustrate some of the issues which may need to be considered in collaborative health informatics research projects, in this section we discuss some of the actions which the CHART-ADAPT project instigated to try and overcome data sharing concerns (CHART-ADAPT, 2016).

The CHART-ADAPT platform allows the fast analysis of high and low frequency data collected from a critical care unit; enabling the creation and assessment of novel, closed loop, diagnostic or therapeutic models and algorithms. Routinely recorded patient data is automatically transferred from the electronic patient record system in the critical care unit to a high performance computing platform implementing a Spark (Apache a, 2016), Scala (Scala, 2016) and Hadoop (Apache b, 2016) technology stack. Complex physiological algorithms are then applied to the data to derive clinically useful variables which are returned back into the clinical environment and integrated with the existing electronic patient record system.

A lack of the required technical infrastructure within the hospital to process the patient data within clinically meaningful timescales meant it was essential for the healthcare provider and academic researchers to form a collaborative team which included commercial partners to provide the required high performance computing infrastructure.

Due to the nature of the collaboration and the requirement to transfer patient data, the project team was aware of the need to maintain patient confidentiality, the governance of patient data, and the need to gain the confidence of patients and unit staff. Several concerns were identified: 1) failure of the de-identification software and subsequent transfer of identifiable patient data outside the healthcare provider's network, 2) secure handling of the anonymised patient data by the commercial partner, 3) correct re-identification of patient data when it reentered the healthcare provider's network, and 4) public perception of a commercial partner supporting the patient data analysis. These concerns were considered from the start of the project and all the collaborating organisations worked together to integrate the following activities into the project plan:

Table 1: General Public Opinions on Secondary Healthcare Use.

| **Question** | **Yes** | **No** | |
|---|---|---|---|
| Are you aware that medical data could be used for research? | 30 (81.2%) | 7 (18.9%) | |
| **Question** | **Strongly agree/agree** | **Undecided** | **Disagree/Strongly disagree** |
| Medical data should be used for research | 33 (89.2%) | 4 (10.8%) | 0 |
| Would you be happy to share your healthcare data? | 31 (83.8%) | 4 (10.8%) | 2 (5.4%) |
| Do you trust clinicians with your healthcare data? | 27 (73%) | 7 (18.9%) | 3 (8.1%) |
| Do you trust private companies to use your medical data for research purposes? | 8 (21.6%) | 20 (54%) | 9 (24.3%) |

Table 2: Responses $<$ 35 years old, Total = 21 respondents.

| **Question** | **Strongly Agree** | **Agree** | **Undecided** | **Disagree** | **Strongly Disagree** |
|---|---|---|---|---|---|
| 1 | 1 (4.8%) | 7 (33.3%) | 10 (47.6%) | 2 (9.5%) | 1 (4.8%) |
| 2 | 4 (19%) | 8 (38.1%) | 7 (33.3%) | 1 (4.8%) | 1 (4.8%) |
| 3 | 5 (23.8%) | 10 (47.6%) | 5 (23.8%) | 0 (0%) | 1 (4.8%) |

Table 3: Responses $>=$ 35 years old, Total = 16 respondents

| **Question** | **Strongly Agree** | **Agree** | **Undecided** | **Disagree** | **Strongly Disagree** |
|---|---|---|---|---|---|
| 1 | 1 (6.25%) | 3 (18.8%) | 8 (50%) | 2 (12.5%) | 2 (12.5%) |
| 2 | 2 (12.5%) | 7 (43.8%) | 5 (31.3%) | 2 (12.5%) | 0 (0%) |
| 3 | 4 (25%) | 5 (31.3%) | 6 (37.5%) | 1 (6.25%) | 0 (0%) |

- Regulatory approval (beyond minimum requirements) was acquired for the transfer of patient data (e.g. NHS Research Ethics, Caldicott Guardian approval)
- The project developed software to automatically anonymise the patient data before it left the healthcare environment. A rigorous testing plan was followed and repeated at regular intervals to ensure confidentiality was maintained.
- The commercial partner responsible for technical support of the data analysis (Aridhia) developed an Information Governance Strategy for the project which made explicit the data handling and security procedures put in place. Close communication was also maintained between the personnel responsible for Information Governance in both organisations (healthcare and commercial).
- Public engagement initiatives were implemented. For example, posters and leaflets were made available in the unit, staff were briefed and updated on project progress, and a public event was hosted to discuss patient data sharing within critical care.
- Attendance at relevant academic and healthcare events was scheduled into the project. This gave the team the opportunity to discuss the platform and gather feedback which was fed into the development of the project.

Although some activities were time-consuming and beyond the usual scope of a research project, it was beneficial not only for development of the platform, but also to make sure, to aid future acceptance of the technology, that we took the clinical staff and general public with the project, rather than exclude them and present the technology as a fait accompli.

# 5 OPPORTUNITIES FOR HEALTH INFORMATICS RESEARCH

Whilst public opinion on trust, security and privacy of patient data needs be carefully considered in research projects, there is also an opportunity for the health informatics community to develop tools and technologies to address these concerns. Below are some suggestions (although this is not exclusive) and comments on how the health informatics community may be able to contribute:

- **Communication of Health Informatics Projects** - There is a need to develop clear, concise, up-to-

date summaries of health informatics projects to aid transparency, in particular regarding the use of patient data. Researchers should consider how they will engage the public when designing and implementing the health informatics project.

- **Dynamic Consent** - Current mechanisms of informed consent for patient data sharing are static, paper-based and designed around legal frameworks and regulation. They are also specific to individual research studies and have to be repeated for subsequent studies. There is a growing awareness that this is inadequate and future policies are moving towards a more sophisticated form of consent (e.g. the proposed EU General Data Protection Regulation (GDPR, 2016)). Dynamic consent provides patients with a personalised interface, placing them in control of how their healthcare data is used; data sharing preferences can be stated and often they can view how their data is being used by researchers (Kaye et al., 2015), (Williams et al., 2015). Once consent has been specified by patients, new tools and technologies are required which enable their preferences to be dynamically and automatically applied across multiple clinical databases and research studies.
- **Safe Havens** - To control how electronic patient data is used by researchers, many healthcare providers are making it accessible through Safe Havens (i.e. it doesn't leave an authorised environment) (Caldicott, 2016). Safe Havens pull together data from multiple healthcare sources and links made between the datasets whilst maintaining patient confidentiality. Safe Havens require a suite of software tools to: ensure security of the centrally stored data (e.g. defend against cyber attacks), enforce data access controls, and audit the use of the patient data. Whilst basic tools have been implemented, there is still potential for more sophisticated software to support these activities.
- **De-identification of Patient Data** - Generally, there is public support for the sharing of de-identified data for research purposes. National and international guidelines specify methods for de-identification and can include the removal or generalisation of certain attributes. Experts can also be asked to identify attributes with an associated risk leading to patient identification. As removal of data can lead to a lack of quality of the dataset overall, there is a balance to be struck between usability and patient confidentiality. This is a non-trivial optimisation problem which computing and artificial intelligence fields are well placed to contribute towards workable solutions.
- **Re-identification of Patient Data** - Even when patient data has been de-identified, there is still a possibility that it can be re-identified through the use of other, publicly available, datasets. This is likely to be a growing concern, especially with initiatives to make more data available and machine readable (e.g. Semantic Web). Some solutions to reduce the chances of this happening include: removal of high risk variables from a dataset (e.g. features which are available in multiple documents and publicly available); and generalisation of patient data into 'bins' of data (e.g. values are generalised over 5 patients). Again, computing and artificial intelligence fields are well placed to develop tools which enable the automatic identification of high risk attributes.

# 6 CONCLUSION

The health informatics community has an important role to play in the development of novel technology and algorithms to enable advances in clinical knowledge and the quality of patient care. This type of research requires access to sufficient volumes of patient data which raises important issues by the general public regarding ethics, trust and security of patient data, especially if private companies are involved in the research activities. Our position is that, despite these concerns, it is necessary for private companies, research institutions and healthcare providers to work together to successfully transition technology projects from research to real-world environments. However, it is vital that patient confidentiality is maintained during all stages of development. There is a role for policy makers to ensure that existing legislation and procedures are adequate for a fast moving technology industry and that there is clear accountability. Additionally, there needs to be greater public engagement on health informatics projects and open communication regarding the potential use of their data.

# ACKNOWLEDGEMENTS

The CHART-ADAPT project is a collaboration between the University of Glasgow, Aridhia, NHS Greater Glasgow & Clyde, and Philips Healthcare. It has been co-funded by Innovate UK (ref:102113). Approval for the CHART-ADAPT work was granted by the West of Scotland Research Ethics Committee (REC ref: 15/WS/0222) and local Caldicott Guardian approval has been provided. Approval for the survey (section 4) was provided by MVLS, University of

Glasgow. The CHART-ADAPT project we would like to acknowledge the staff and patients of the Neurointensive care unit, Neurosciences Institute, Glasgow.

## REFERENCES

A.Ferguson (2012). The evolution of confidentiality in the united kingdom and the west. *AMA Journal of Ethics*, 14(9):738–742.

Aitken, M., Cunningham-Burley, S., and Pagliari, C. (2016a). Moving from trust to trustworthiness: Experiences of public engagement in the scottish health informatics programme. *Science and Public Policy*, 43(5):713–723.

Aitken, M., de St Jorre, J., Pagliari, C., Jepson, R., and Cunningham-Burley, S. (2016b). Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Medical Ethics*, 17(73).

Apache a (2016). Apache Spark. https://spark.apache.org/. Accessed: Nov 2016.

Apache b (2016). Apache Hadoop. https://hadoop.apache.org/. Accessed: Nov 2016.

BMA (2016). Secondary Uses of Data, Public Workshop. https://www.bma.org.uk/collective-voice/policy-and-research/ethics/secondary-uses-of-data/public-workshop. Accessed: Nov 2016.

Caldicott (2016). Information: To share or not to share? The Information Governance Review. https://www.gov.uk/government/uploads/system/uploads /attachment_data/file/192572/2900774_InfoGovernance_accv2.pdf. Accessed: Nov 2016.

CHART-ADAPT (2016). CHART-ADAPT. http://www.chartadapt.org. Accessed: Nov 2016.

DeepMind a (2016). DeepMind Acute Kidney Injury. Royal Free London. Google DeepMind: Q&A. https://www.royalfree.nhs.uk/news-media/news/google-deepmind-qa/. Accessed: Nov 2016.

DeepMind b (2016). DeepMind Moorfields Eye Hospital. Moorfields announces research partnership. http://www.moorfields.nhs.uk/news/moorfields-announces-research-partnership. Accessed: Nov 2016.

Fillmore, C., Braye, C., and Kawamoto, K. (2013). Systematic review of clinical decision support interventions with potential for inpatient cost reduction. *BMC Med Inform Decis Mak*, 13(135).

Focus Groups, Stevenson, F., Lloyd, N., Harrington, L., and Wallace, P. (2013). Use of electronic patient records for research: views of patients and staff in general practice. *Fam Pract*, 30(2):227–23.

GDPR (2016). GDPR: Regulation (EU) 2016/679. http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf. Accessed: Nov 2016.

IBM a (2016). IBM Big Data. Extracting business value from the 4 V's of big data. http://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data. Accessed: Nov 2016.

IBM b (2016). The 4 V's of big data. http://www.ibmbigdatahub.com/infographic/four-vs-big-data. Accessed: Nov 2016.

IBM c (2016). IBM's Watson supercomputer to speed up cancer care. http://www.bbc.co.uk/news/technology-32607688. Accessed: Nov 2016.

Jaspers, M., Smeulers, M., Vermeulen, H., and Peute, L. (2011). Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc*, 18(3):327–34.

Kaye, J., Whitley, E., Lund, D., Morrison, M., Teare, H., and Melham, K. (2015). Dynamic consent: a patient interface for twenty-first century research networks. *Eur J Hum Genet*, 23(2):141–6.

Kinsella, J., Hawthorne, C., Shaw, M., Piper, I., Healthcare, P., Aridhia, and L.Moss (2017). Public perception of the collection and use of critical care patient data beyond treatment: a pilot study. In *Proceedings of the Society of Critical Care Medicine Congress (SCCM)*. SCCM.

Kitchen, R. and McArdle, G. (2016). What makes big data, big data? exploring the ontological characteristics of 26 datasets. *Big Data & Society*, Jan-June 2016(3):1–10.

New Scientist (2016). Revealed: Google AI has access to huge haul of NHS patient data. New Scientist 2016 Apr 29. https://www.newscientist.com/article/\2086454-revealed-google-\ai-has-access-to-\huge-haul-of-nhs-patient-data/. Accessed: Nov 2016.

Presser, L., Hruskova, M., Rowbottom, H., and Kancir, J. (2015). Care.data and access to uk health records: patient privacy and public trust. *Technology Science*, 2015081103(Aug 11).

Riordan, F., Papoutsi, C., Reed, J., Marston, C., Bell, D., and Majeed, A. (2015). Patient and public attitudes towards informed consent models and levels of awareness of electronic health records in the uk. *Int J Med Inform*, 84(4):237–347.

Scala (2016). Scala Programming Language. http://www.scala-lang.org/. Accessed: Nov 2016.

van Staa, T.-P., Goldacre, B., Buchan, I., and Smeeth, L. (2016). Big health data: the need to earn public trust. *BMJ*, 354:i3636.

Williams, H., Spencer, K., Sanders, K., Lund, D., Whitley, E., Kaye, J., and Dixon, W. (2015). Dynamic consent: A possible solution to improve patient confidence and trust in how electronic patient records are used in medical research. *JMIR Med Inform*, 3(1).

# The Virtual Enterprise Data Warehouse for Healthcare

James P. McGlothlin, Amar Madugula and Ilija Stojic
*Fusion Consulting Inc, Irving, TX, U.S.A.*
*(jmcglothlin, amadugula, istojic}@fusionconsultinginc.com*

Keywords: Data Warehousing, Business Intelligence, Data Virtualization.

Abstract: Healthcare organizations have access to more data than ever before. Healthcare analytics is a vital tool for healthcare organizations and hospitals to analyze performance, identify opportunities to improve, make informed decisions, and comply with government and payor regulations. However, the field of medicine and the political and regulatory landscape are constantly changing, thus these requirements and opportunities rapidly evolve. The traditional best practice solution for business analytics is to organize and consolidate the data into a dimensional data warehouse for analytics purposes. Due to the size of the data, the number of disparate sources and the volume of analytics needs, the overhead to create and maintain such a data warehouse is becoming prohibitive. In this paper, we introduce a virtual data warehouse solution that combines the design and modelling principles of traditional dimensional modelling with data virtualization and in-memory database architectures to create a system which is more agile, flexible and scalable.

## 1 INTRODUCTION

In the healthcare industry in the United States, there has been a rapid and transformational move to electronic medical records (EMRs). The result of these technological advancements is that much more data is available. The challenge every hospital faces is how to use this vast supply of data to improve and make better decisions. This problem is magnified by the ever changing quality metrics, regulatory requirements, payment and incentive programs, political programs and environment. Healthcare organizations must be able to support different analytics and even operational processes for different patient populations and payors.

The amount of data available is staggering. Not only do modern EMRs allow digital access to every medication administration, order and test result, but personalized medicine is allowing the use of specific gene and DNA information to improve patient care. Additionally, personal electronic sensors and wearables are allowing healthcare organizations to analyze patient data even outside of the office or hospital. The volume of healthcare data is growing at a rate of 48% annually (Leventhal, 2014).

In addition to the exponential growth of healthcare data, there is also an exponential growth of healthcare costs. This is being magnified by increased life expectancy and a large aging population. Payors are pushing down these costs through changing payment models such as pay for performance, managed care, full risk plans, value based purchasing and more. With each of these programs comes different analytics needs and different requirements for compliance, reimbursement and incentives.

The traditional best practice for analytics has been to create a dimensional model data warehouse which organizes the most important enterprise data for analytics. Sets of business intelligence tools, reports and dashboards can then utilize these data warehouses to provide the analytics needs of the organization. However, this approach is becoming less sustainable for large organizations in the healthcare industry. The needs and requirements change too quickly and are too specialized to allow for development of custom extract/transform/load (ETL) processes for each need. The number of data sources is too diverse and the data varies too much in availability, quality and format to allow for complete daily extraction into the data warehouse. The sheer volume of data overloads the data warehouse and makes the storage, memory and scalability requirements untenable. In a recent survey, healthcare data scientists reported that 49% were having difficulty fitting data into relational databases, and that data variety was an even greater challenge (Miliard, 2014).

In this paper, we introduce a solution that combines the design and advantages of a traditional



McGlothlin, J., Madugula, A. and Stojic, I.
The Virtual Enterprise Data Warehouse for Healthcare.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 469-476
ISBN: 978-989-758-213-4

data warehouse with the latest advances in data virtualization technology. Additionally, we leverage in-memory databases and column stores to further accelerate performance and agility. We will describe our solution, how we are using it to integrate data from many different sources, and analyze the benefits of this approach.

## 2 THE TECHNICAL SOLUTION

Data virtualization is an approach and technology for integrating multiple sources of data. Our goal with data virtualization is to abstract the logic of the data model from the specifics of the data location and source formatting. This means that applications and users consuming the data do not need to be aware of how or where the data is physically stored. This allows us extreme agility, because we can choose at any point to consolidate data, move data, transform data or cache data without any effect on the tools and users consuming the data upstream.

We implemented our virtual enterprise data warehouse using the Cisco Data Virtualization (Cisco DV) platform. Cisco DV supplies data federation to many types of sources including relational databases, files, cloud and big data technology solutions such as Hadoop (Zikopoulos and Eaton, 2011), web services, and multi-dimensional sources. These sources are accessed and integrated using advanced query planning and optimization, parallelization and distributed joins. However, this data virtualization platform is for more than just data federation. Our goal with a DV platform is to create a true single version of the truth for the enterprise. We chose Cisco DV because it provides a development environment to create a logical data model and then map it to the source systems. Also, it provides a business directory allowing the data points to be defined and made available in business terms. This provides the foundation for a data governance data dictionary for the enterprise. Furthermore, Cisco DV maintains and persists a metadata repository that defines the data model as views and the technical details to map the information view to the underlying data source system. Since this metadata is persisted with history and version control, it provides an excellent solution for data lineage. Our experience is that data lineage is an absolute requirement to achieve user trust in the data and user adoption. Figure 1 shows the architectural diagram of the Cisco DV Suite.

Data virtualization provides some solutions for performance issues including query optimization and caching. However, we found that most of the benefits

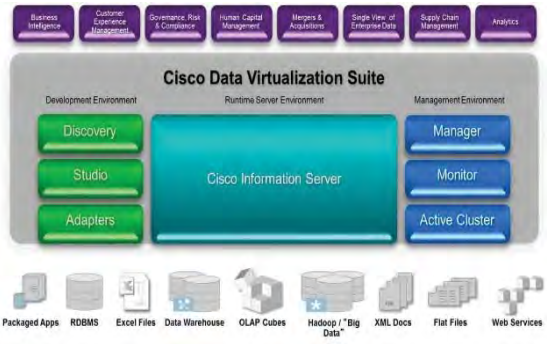



Figure 1: Cisco Data Virtualization Suite.

of data virtualizations were reduction in ETL development level of effort, reduction in the time to market on new projects, improved data management and governance, and reduction of ETL daily execution time. These features are important but they do not address the issue of performance for analytics to the end user. In fact, depending on the source system, it is possible that the traditional consolidated data warehouse, which is designed for analytics queries, will outperform a virtualized approach. We consider this a very important problem to solve so we introduced additional technology to accelerate performance.

SAP HANA is an in-memory, column store database appliance designed for analytics data warehouses (Sikka et al, 2013). Column store databases perform especially well for analytics because they optimize read-only access of the data, whereas traditional database optimize single row transactions. Because columns are stored together, there is significantly less local data variety and therefore more opportunity for data compression. Also, column stores only retrieve data requested in the query. In dimensional modelling, generally the analytics user chooses specific dimensions for constraints or analysis. Because column stores only retrieve the information requested, they are especially well-suited for analytics data warehouse queries (Stonebraker et al, 2005). This is even more magnified with self-service reporting, where there is no way to optimize the report ahead of time because the user has the option to change the query. Finally, and most importantly, HANA is completely in-memory. Therefore, queries are extremely fast. Because of the column store architecture and advanced compression technologies, we have found compression rates ranging from 5x to 47x depending on the type and sparsity of the data. Figure 2 shows the architecture of the SAP HANA Platform.

As we stated earlier, data virtualization hides from the consumer and upstream applications the

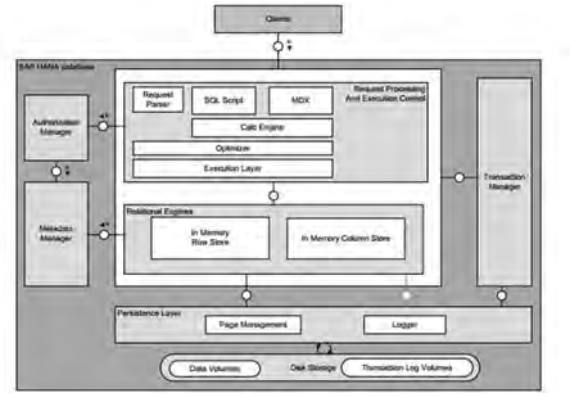


Figure 2: SAP HANA Platform.

physical source of the data. This allows us to move the most important data to SAP HANA and to adjust which data is stored in HANA based on optimization needs. This has been shown to improve some queries' performance by over 100x. There is no impact or change required to the tools, reports or dashboards. We are terming our use of HANA as physical cache. Cisco DV handles moving the data from the original source into HANA so no extra development effort is required.

We continue to use industry standard Kimball dimensional modelling design for our virtual enterprise data warehouse (Kimball, 2011). All of our data models are defined using facts, dimensions, bridges and other standard data warehouse design techniques. We implemented algorithms needed for data integration such as patient matching, provider attribution, cross walk tables, and standard code sets. We created flexible, source-agnostic business model for healthcare using dimensional modelling. The primary difference is that this is a logical model, we are not always physically populating tables that match the model schema. Instead, we are using data virtualization views as appropriate. Figure 3 shows the solution architecture.

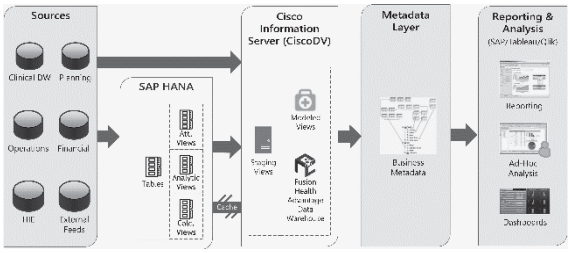



Figure 3: Cisco DV/SAP HANA Solution.

# 3 IMPLEMENTATION

## 3.1 EMR Data

For a hospital, the most important data source is the hospital electronic medical record (EMR). Many EMRs now supply data warehouse and analytics solutions. Our goal is certainly to leverage these solutions. However, we have found many instances where we had to add custom extension tables because of different processes at our hospital or different analytics needs. Here are some of many examples:

a. Blood pressure on an office visit to be lowest rather than last
b. Discharge provider on a hospital visit to be based on the bill rather than the treatment team log
c. Provider attribution
d. Quality metrics that look for clinical events in both clinical documentation and the bill and claim
e. DRGs to include the secondary DRG coded on the bill
f. Cancellation reasons for cancelled appointments or surgeries
g. Different documentation data points for expected discharge delay reasons

Our challenge is that the vendor does not allow us to change their tables. We can create our own tables but now extra logic and table joins is needed when doing analysis and reports.

We have defined a pure data model and metadata layer in our virtual data warehouse. In accordance with traditional Kimball dimensional modelling, our model matches the business model and analytics needs, rather than the source (Kimball, 2011). So even though three or four tables from the EMR vendor data warehouse and extensions may be required, it will look like a single table in the virtual enterprise data warehouse. This allowed us to cover all of the information in the vendor data warehouse with 40 less tables and to considerably reduce the complexity of the queries used by reports and dashboards.

For example, the vendor data warehouse has fact tables for hospital visits, billing accounts, and services. We wish to know the discharge provider and last service for the hospital visit. For our hospital, the discharge provider is inaccurate on the hospital visit fact, but correct as the attending provider on the hospital account fact. The last service is not the hospital service on the hospital visit fact, but can be determined by determining the last service for the

patient chronologically. This logic is complex for a report writer and is very likely to create reporting errors. Specifically, the discharge provider on the source table is not the correct discharge provider. We were able to use data virtualization to create a single hospital visit fact with the correct values for these columns for our business. This allows our data governance team to choose the correct business definition and us to expose it to the entire enterprise. The complex logic and the inaccurate columns from the EMR vendor data warehouse are not exposed to the user. However, the EMR vendor data warehouse is still utilized to source the data. This allows us to create a much better data warehouse for our clinical EMR data and our end users.

## 3.2 Other Clinical Sources

With the current focus on preventive care and population health, it is becoming more imperative to have all information related to a patient's health. This can include data from outside of the hospital's EMR including claims, pharmacy and lab data. This can also include clinical data from independent providers or Health Information Exchange(s). Furthermore, hospital networks continue to consolidate, and often the different hospitals and clinics are using different EMR systems. One key challenge health care business intelligence teams face is integrating clinical and operational data from multiple sources. Integrating data allows a provider or care coordinator to be aware of patient office visits, diagnoses, lab results, prescriptions, images and hospital visits which occur outside of their primary EMR. This improves care management and risk assessment, allows gaps in care to be addressed and makes it possible to do quality metrics with complete information. Also, outside data can be used to better stratify patient risk.

For example, if we have pharmaceutical claims information, we can know if the patient received their flu vaccine at the local grocery store, and we can assess their adherence to medication orders. If we have information from an affiliated ophthalmologist's EMR, we can know whether the patient received their diabetic eye exam. If we have claims information, we can know about hospital admissions while the patient was on vacation. We can connect with risk stratification engines to know what potential events the patient is most at risk for, and what preventive care measures might help avoid these issues. We can use benchmarks to see how our admission rates, length of stay, supply cost and other information compare to others in the industry.

Bringing in these data sources is challenging. We have to match the patients and providers with those already in our enterprise data warehouse. We have to maintain the original source system identifiers, so we will be able to process updates or additional patient information in the future. This information comes in at various times which we do not control, so we cannot perform a daily extract as easily as our process for our EMR extraction. The data comes in many different formats and uses different code sets. So, the logic needed to conform the data can vary depending on the source.

We have brought in claims data both from payors and from network affiliate providers. We have used custom extracts to bring in specific clinical information from affiliate providers EMRs. In the future, we plan to bring in lab and pharmacy data.

We developed logic for patient matching and persisted the patient matching results and a crosswalk to the source system in our data warehouse. We then virtualized all of the other data. The end result was that we created quality dashboards that examined patients' entire health across all of the clinical source systems. This dashboard only accessed the virtual metadata abstract layer so the reports did not need any information about the source systems or formats. However, we did include metadata about the source system, so that users could know the data lineage of the information. This allows a physician at our hospital to know that his patient had a lab result from an outside provider.

## 3.3 Non-clinical Systems

Our hospital has many sources of data which are not clinical. However, all of these systems provide increased value when analytics which includes the clinical data can be provided.

For example, decision support costing systems allow us to determine the costs associate to a billing transaction, a surgery, an order or a medication. This can include fixed and variable costs in many different accounting buckets such as labor, supply and overhead. Integrating this data with the clinical data warehouse lets us analyze costs related to specific diseases, patient cohorts, locations, providers, procedures, etc. Because this data is managed in a different system and is quite large, we do not want to physically consolidate this data so we are using our data virtualization platform.

We also have materials management and supply chain information. This allows us to evaluate inventory and purchasing contracts. This information feeds our cost algorithms. There is significant value

in making this data available in our data warehouse for analytic purposes.

Another example is HR information. This information often involves many different systems and forms including position information, salary and benefits information, provider credentialing and time and attendance. Including time and attendance with the clinical events performed by the employee allows us to evaluate productivity. We can analyze wages and overtime to determine opportunities for improved resource management, training information and cost.

Other examples of peripheral non-clinical data include accounts receivable collections information and budgeting information.

## 3.4 Clinical Support Systems

There is a vast amount of clinical information available in hospitals which many not be in the central EMR. This includes case management systems which monitor physician reviews, expected discharges, avoidable days, etc., statistical systems which are used for clinical details such as Apache (Knaus et al, 1981) and Prism (Murray et al, 1988) critical care evaluation techniques, lab systems which have more detailed information about specimens collected or blood units supplied, radiology systems which have detailed information about images, and clinical engineering systems for oncology, pathology, cath labs, etc. These systems vary for each hospital we have worked with.

Generally, we have found it is not necessary to bring in all of the data from these ancillary systems. However, often specific key data points are very important to our data warehouse. We have used data virtualization to target and pull out specific data elements which augment data structures we already have in our data warehouse.

## 3.5 Benchmarks

Every hospital and healthcare organization wants to know how it is performing relative to its peers. This provides valuable insight identifying opportunities for achievable improvement. There are hundreds of sources for benchmarks of all different varieties. Examples include quality benchmarks like Medicare Stars ratings and Pay for Performance percentiles, financial benchmarks like supply cost for OR in the region, benchmarks like Centers for Medicare and Medicaid Services (CMS) length of stay by DRG. These are simple benchmarks but there are much more complicated clinical benchmarks and whole companies which special in providing benchmark information. We plan to use data virtualization to integrate these benchmarks into the enterprise data warehouse so we can show opportunities, targets and concerns in our dashboards and visualizations. We have brought in many of the simple ones, and plan to bring in more comprehensive and detailed benchmarks in the future such as critical care length of stay by service and comorbidity.

## 3.6 Patient Experience

It is important for a hospital to monitor patient satisfaction. Patient satisfaction is measured through customer surveys. Generally, these surveys are outsourced so they can be objective, fair and consistent. Analyzing the results of this external information can provide the hospital valuable insight into improvement opportunities.

## 3.7 Precision Medicine

Precision medicine uses patient information to tailor personalized treatment. For example, analysing patients' genomes can allow the most effective cancer treatment medication and therapy to be chosen. There is considerable research funding being applied to precision medicine and it is considered a very significant development for improving healthcare treatment. (Jameson and Longo, 2015)

Clinical information such as medications administered, medication reactions, diagnoses, pathology results, and precise imaging information is vital to properly tailor a personalized medicine approach. So, important information exists in the enterprise data warehouse to identify the appropriate patient cohorts and monitor the progress of treatment.

However, precision medicine almost always involves gene analysis. Clearly, genome databases are huge and cannot be consolidated physically into our data warehouse. Thus, the data virtualization approach is absolutely vital to implementing precision medicine.

## 3.8 FHIR

Fast Healthcare Interoperability Resources (FHIR) is a framework for next generation intercommunication between healthcare data systems. FHIR uses RESTful (representational state transfer) application programming interfaces (APIs) and defined data points and elements (resources) to exchanging information electronically. FHIR is a standard managed by the HL7 organization, the major standardization organization for healthcare data

(Bender, 2013).

Cisco DV supports web services as a source for the enterprise data warehouse include RESTful APIs and XML resources. As such, we can integrate data platforms which support FHIR using this standard.

# 4 BENEFITS

In addition to allowing us to integrate so many different sources, our virtual enterprise data warehouse approach solves many problems we have encountered in our traditional healthcare data warehouses.

## 4.1 ETL Performance

Because of the complexity of the healthcare data and EMR, we have found the daily process of extracting the data time-consuming. Best practice requires us to have multiple data warehouses for production, development and user acceptance testing. Generally, they source from the same operational data store for the EMR. It has been a constant challenge to have this ETL finish in a timely manner. If we were to increase the logic in this data warehouse transformation, the ETL time would grow. If we were to bring other sources into the physical data warehouse, the ETL time would definitely grow. Data virtualization allows us to avoid bringing other data sources into our physical data warehouse. It also allows us to move some of the logic out of the physical data warehouse and into the abstraction layer.

## 4.2 Scalability

Healthcare information is very detailed. A week long hospital stay can have as much as 10,000 separate data points documented. There is a challenge both on disk space and ETL load time to get all this data into the data warehouse. This problem is magnified when data outside the organization such as claims information and affiliate provider data is brought in and integrated. The growth in this data can be hard to predict as can the additional data needs of the organizations which are constantly evolving.

Clearly, the virtual data warehouse reduces the physical disk space requirement by leaving some data in place. Moreover, it is inherently scalable. Because the data transformation is not tied to the data storage and consumers of the data are not connected to the data storage, we can easily move where the data is stored. This allows us the flexibility to integrate cloud solutions or to choose new technologies at a future time without needing to make final decisions now. The organization is given the flexibility to change databases or use big data technologies in the future without impacting the architecture or the data consumers.

## 4.3 Tool Agnostic

Many business intelligence tools such as SAP provide a metadata layer. However, our experience is different tools are required for different purposes. Many hospitals use both SAP tools and Tableau, Qlik or other visualization tools. In the past, it was necessary to recreate the metadata layers and security for each tool set or risk inconsistencies between applications. In our virtual data warehouse solution, the metadata is persisted in the data virtualization layer and consumed by all of our business intelligence tools.

## 4.4 Security

Few things are more important to healthcare organizations than security. Compliance and privacy regulations are very strict. The organization must define who can see each piece of data. This includes object level security (can the user see this type of data at all based on their role) and row level security (can the user view this specific data element based on the value - such as location, patient, provider). The data virtualization layer provides a single place to define and enforce security which will then be consumed consistently across the organization.

## 4.5 Data Availability

Our source data becomes available at different times. Some data such as census we have enabled for almost realtime access. Much of the EMR data is available daily. Some external data such as claims may only be provided monthly. Some calculated data is only updated quarterly. By disconnecting the source from the abstraction layer, we can have greater control over when data is refreshed and can support on-demand access, custom extracts, and pipeline push consumption.

Additionally, it is important to make the data always available and consistent to the user. We want to avoid restricting access during loads, but we never want to provide partial or inconsistent information. The data virtualization layer gives us a place to manage this. Generally, we can provide stale data or cached data during updates.

### 4.6 Data Governance

Data governance is a methodology and process for managing information as an asset. As part of the data governance program, the hospital chooses which data points are important, a standard name and definition for that data point, a correct source of truth, and who should be allowed to see the data. Data governance and metadata management is vital to obtaining "a single version of the truth", which is a important yet difficult goal. The virtual data warehouse gives all analytics and reporting users a single place to go to obtain data. The data and logic can be defined in an organized manner. The data dictionary provides the definition in business terms and the data lineage in technical terms. Users and data stewards can search the data dictionary so that data is used consistently rather than extracted repeatedly. All business intelligence tools can source the data from the data virtualization layer allowing the logic and naming to be consistent across the organization.

## 5 RESULTS AND CONCLUSION

We have implemented our data virtualization approach at several major hospitals and continue to expand these projects. We have been able to successfully deploy the virtual data warehouse and enable access to the physical EMR data warehouse quite quickly. Then, we grow and adjust this model to bring in the other sources important to the enterprise analytics. All of our projects are still growing but we have seen very encouraging early results including faster project development times, user adoption of the metadata, improved data governance implementation and significant reduction in model complexity.

With the growth in healthcare data in both volume and variety, and the growth in analytics needs, the traditional data warehouse and analytics approach is simply not agile enough to scale for the needs of the healthcare industry. By introducing data virtualization and in-memory persistent caching, and by preserving the dimensional model foundation of the data warehouse approach, we assert that we have created a solution that is sufficiently agile to scale and grow with the needs of the modern hospital.

## REFERENCES


Leventhal, R, 2014. Report: Healthcare Data is Growing Exponentially, Needs Protection. In *Healthcare Informatics.*

Miliard, M, 2014. Data variety bigger hurdle than volume. In *Healthcare IT News.*

Sikka, V., Färber, F., Goel, A., Lehner, W., 2013. SAP HANA: the evolution from a modern main-memory data platform to an enterprise application platform. In *Very Large Databases.*

Stonebraker, M., Abadi, D, Batkin, A. et al, 2005. C-Store: A Column-oriented DBMS. In *Very Large Databases.*

Kimball, R, 2011. *The data warehouse toolkit: the complete guide to dimensional modelling.* Wiley Computer Publishing.

Bender, D, 2013. HL7 FHIR:An agile and RESTful approach to healthcare information exchange. In *CBMS.*

Zikopoulos, P, Eaton, C, 2011. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill Osborne Media.

Ellisman, M. and Peltier, S., 2003, December. Medical data federation: The biomedical informatics research network. In The Grid (Vol. 2).

Bloom, K. and Cms Collaboration, 2014. CMS Use of a Data Federation. In Journal of Physics: Conference Series (Vol. 513, No. 4, p. 042005). IOP Publishing.

Kahn, B.K., Strong, D.M. and Wang, R.Y., 2002. Information quality benchmarks: product and service performance. Communications of the ACM, 45(4), pp.184-192.

Tesch, T. and Levy, A., 2008. Measuring service line success: the new model for benchmarking: the service line model benefits nearly all stakeholders involved in healthcare delivery. But how is its success measured?. Healthcare Financial Management, 62(7), pp.68-75.

Schneider, Polly. "How Do You Measure Success?." In *Healthcare Informatics* 15.3 (1998): 45-56.

Raghupathi, W. and Raghupathi, V., 2014. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), p.1.

Goth, G., 2007. Virtualization: Old technology offers huge new potential. IEEE Distributed Systems Online, 8(2), p.3.

Feldman, B., Martin, E.M. and Skotnes, T., 2012. Big Data in Healthcare Hype and Hope. October 2012. Dr. Bonnie, 360.

Hopkins, B., Cullen, A., Gilpin, M., Evelson, B., Leganza, G. and Cahill, M., 2011. Data virtualization reaches the critical mass. Forrester Report.

Lupşe, O.S., Vida, M.M. and Tivadar, L., 2012. Cloud computing and interoperability in healthcare information systems. In *The First International Conference on Intelligent Systems and Applications* (pp. 81-85).

Koufi, V. and Vassilacopoulos, G., 2008. Context-aware access control for pervasive access to process-based healthcare systems. Studies in health technology and informatics, 136, p.679.

Knaus, W.A., Zimmerman, J.E., Wagner, D.P., Draper, E.A. and Lawrence, D.E., 1981. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. Critical care medicine, 9(8), pp.591-597.

Pollack, Murray M., Urs E. Ruttimann, and Pamela R. Getson. "Pediatric risk of mortality (PRISM) score." Critical care medicine 16.11 (1988): 1110-1116.

Jameson, J.L. and Longo, D.L., 2015. Precision medicine—personalized, problematic, and promising. Obstetrical & Gynecological Survey, 70(10), pp.612-614.

# Exercise and Wellness Health Strategy Framework
## *Software Prototype for Rapid Collection and Storage of Heterogeneous Health Related Data*

Petr Brůha[1,2], Roman Mouček[1,2], Pavel Šnejdar[1], David Bohmann[1], Václav Kraft[1] and Peter Rehor[1]
[1]*Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, Pilsen, Czech Republic*
[2]*NTIS - New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, Pilsen, Czech Republic*
{*pbruha, moucek*}*@kiv.zcu.cz,* {*snejdar, bohmannd, vkraft*}*@students.zcu.cz, rehorp@rek.zcu.cz*

Keywords: Exercise and Wellness, Chronic Disease, Health Related Data, Health Informatics Systems, Software Prototype.

Abstract: Unwillingness of many people to assume responsibilities for a personal health, fitness and wellness seems to be widespread. This can be partially remedied by individualized exercise and wellness program that integrates the basic knowledge domains: lifestyle, sports and fitness, and nutrition and personal/environmental health. However, collection, management and analysis of data and metadata related to these domains is demanding and time consuming task. Moreover, the appropriate annotation of raw data is crucial for their next processing. To promote such a program a software infrastructure for collection, storage, management, analysis and interpretation of health related data and metadata has been proposed and part of this infrastructure has been developed and tested outside laboratory conditions. This software prototype allows experimenters to collect various heterogeneous health related data in a highly organized and efficient way. Data are then evaluated and users can view relevant information related to their health and fitness.

## 1 INTRODUCTION

In the countries of the European Union (EU) deaths due to chronic diseases are projected to increase. It was estimated that out of the projected millions of people who would die by 2015, 64% of them die of a chronic disease – unless urgent action is taken (Tunstall-Pedoe, 2006). Given that both physical inactivity and obesity are strong independent causes and predictors of chronic illness and disability, it has been estimated that they impose a significant economic burden on the health-care system in EU. The most recent research in EU indicates that only 7.5% of children and 15% of adults are physically active for at least 150 minutes per week, while over 18% are obese and over 46% are overweight (Busse, 2010; Organization et al., 2010; Webber et al., 2014).

It is apparent that physical activity is essential in the prevention of chronic disease and premature death (Lee and Skerrett, 2001). Chronic diseases develop over ones lifetime, with clinical symptoms occurring many years after the underlying origins of the disease have occurred. As we move ahead in the 21st century, cardiovascular diseases, i.e. coronary artery disease (CAD) hypertension, stroke, and heart failure, type 2 diabetes, metabolic syndrome, and cancer are the leading killers in westernized society and are increasing dramatically in developing nations. Physical inactivity is a modifiable risk factor for cardiovascular disease and a widening variety of other chronic diseases, including diabetes mellitus, cancer (colon and breast), obesity, hypertension, bone and joint diseases (osteoporosis and osteoarthritis), and depression 1-14 (Taylor et al., 2004; Blair et al., 2001; Shephard, 2001).

The onset of progression of chronic diseases is mediated in the vast majority of cases by an interaction between genetic factors and their interaction with environmental factors. These environmental factors are largely lifestyle factors, namely physical activity and dietary patterns, but also include other influences, such as smoking, alcohol consumption, stress, and hazardous environmental compounds. These factors are modifiable, and, as such, disease manifestations from these factors are largely preventable.

To cope with these modifiable factors exercise and wellness intervention programs have been introduced, mainly in the United, States, Canada and Australia. This paper presents the first steps that have been done



Brůha, P., Mouček, R., Šnejdar, P., Bohmann, D., Kraft, V. and Rehor, P.
Exercise and Wellness Health Strategy Framework - Software Prototype for Rapid Collection and Storage of Heterogeneous Health Related Data.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 477-483
ISBN: 978-989-758-213-4

to introduce such exercise and wellness health strategy framework and related academic program within the European Union. Its attempt is to create professionals who are capable influence self-care practices positively, reinforce healthy habits and prepare responsible citizens for the future. This innovative, applied and unique initiative combines three traditionally exclusive disciplines (Etiology of Chronic Disease, Exercise Science and Behavioral Psychology) into one comprehensive curriculum that addresses two major front page challenges in Europe: chronic disease management and sedentary lifestyle. It is being applied through extensive clinical and fieldwork experiences, the degree will provide students with the theoretical knowledge, practical skills, and prerequisites that are necessary to provide a professional guidance during lifestyle changes.

Such exercise and wellness health strategy framework needs a supportive software infrastructure that besides others promotes collection of health related data and metadata followed by their further annotation, processing and visualization. This paper introduces the first version of such software infrastructure,a software prototype that focuses on definition and automation of the data collection process in order to capture a huge amount of heterogeneous health related data from many users in various environment in a short time. It is assumed that the procedure of data collection has to be as short and user friendly as possible to significantly promote the initial participants' motivation to cope with the most important step, a desired change of participants' behavior leading to their better physical, emotional and mental health.

The paper is organized in the following way. The next section discusses some questions related to chronic illnesses, personal physical, emotional and mental well-being and the necessity of exercise and wellness strategy framework to cope with these matters. The third section introduces a supportive software infrastructure for such a framework. The subsections then present the first version of this software infrastructure from architectural, implementation, deployment and testing points of view. The last section brings concluding remarks and introduces future steps and improvements.

## 2 EXERCISE AND WELLNESS HEALTH STRATEGY FRAMEWORK

Chronic illness has profound effects on a persons physical, emotional, and mental well-being, which often make it difficult to carry on with daily routines and relationships. Over the past decades, considerable knowledge has accumulated concerning the significance of exercise in the treatment of a number of diseases, including diseases that do not primarily manifest as disorders of the locomotive apparatus. Today, exercise is indicated in the treatment of a large number of additional medical disorders. In the medical world, it is traditional to prescribe the evidence-based treatment known to be the most effective and entailing the fewest side effects or risks. The evidence suggests that an exercise therapy is just as effective as medical treatment in selected cases and even more effective or improving the final effect in special situations. In this context, exercise therapy does not represent a paradigm change, it is rather that the accumulated knowledge is now so extensive that it has to be implemented.

There is a growing interest in the use of exercise for clients with chronic diseases and disabilities. It is thus suitable to provide a framework for determining functional capacity and developing exercise strategy in persons with chronic diseases and/or disabilities. The basic principles for exercise testing and training stated provides the foundation for program design. However, some special situations created by a disease pathology, disability, or treatment alter these basic principles. For example, exercise testing is an important aspect of the approach used, but some people will not have completed an exercise test before starting an exercise program. Participation in regular physical activity can enhance functional capacity, and a primary goal is to get more individuals physically active. Thus, for many people, exercise testing may not be absolutely necessary before starting a low-level physical activity program.

Many people who have chronic disease or disability enter a downward spiral toward exercise intolerance, so exercise intervention programs should be designed to resist this spiral and optimize functional capacity. Within any given population, there is a wide range of abilities determined by several factors; progression of the disease, response to treatment, and presence of other concomitant illnesses. Expected outcomes of exercise training are not always known. Realistically, optimal exercise and medical programming may yield improvements or merely prevent further deterioration. There may be recommended tests or programs that have not been validated, but that experience has shown to be successful. It is hoped that optimal management will bring the individual greater independence and improved quality of life.

In general, our society has a bias toward curative rather than palliative medicine, toward making the

disease go away rather than finding ways to cope with disease. An unfortunate consequence of this perspective is that for persons with chronic disease or disability, we devalue the palliative benefits of preserving functionality and well-being. Since the 1960s, exercise has been promoted as a method of extending life, largely through prevention and moderation of cardiovascular disease. In recent years we discovered, however, that perhaps the greatest potential benefit of exercise is its ability to strengthen musculoskeletal system structure and function, enhance functioning of cardiovascular, digestive and endocrine system and to augment mental capacity through changes in body chemistry. Its side effects unlike other pharmacological interventions are positive (improvement and preservation in functional capacity, freedom, and independence).

The frameworks and applications that cope with the questions of health and wellness and that are available to the broad public audience are described e.g. in (Banos et al., 2015; Joo et al., 2012; Laakko et al., 2008).

# 3 SUPPORTIVE SOFTWARE INFRASTRUCTURE

## 3.1 Software Requirements Specification

This section introduces a supportive software infrastructure for the exercise and wellness health strategy framework. This infrastructure will provide means for definition of the data collection procedure and data collection itself, repository for storing and long term management of health related data and metadata in a standardized way, data processing and interpretation methods and/or defined interfaces for the application of data processing and interpretation methods. Last but not least visualization tools providing views on collected/analyzed data and their interpretations will be included.

An important aspect of this supportive software infrastructure is its wide range that includes collecting data from classic measurements such as blood pressure to relatively rare kinds of measurements such as acquisition of brain frequencies or brain event related potentials. It means that also experience of experimenters from the neuroinformatics lab (reference omitted for a review process) in measuring brain waves highly contributed to the definition of the first set of collected health related human body parameters.

Since big health related data (from at least thousands of people) are planned to be stored in the repository of this software infrastructure, the future analysis of interdependence between brain parameter values, other physiological and body proportion values and personal lifestyle records would bring valuable results for the targeted application of exercise and wellness health framework strategies.

Thus the whole infrastructure will enable to find out complex information related to health conditions of each participant and derive statistical results when defining various cohorts. With long-term and repeated measurements appropriate software modules will be able to detect and depict personal trends and changes in measured values.

However, the preliminary version of the software infrastructure presented in this article (hereafter called a software prototype) is required to help experimenters to efficiently collect all health related data and metadata outside the laboratory environment in one session according to the procedure that could be defined even on site. The schema of such experimental procedure (that contains a limited set of measurements) is depicted in Figure 1.

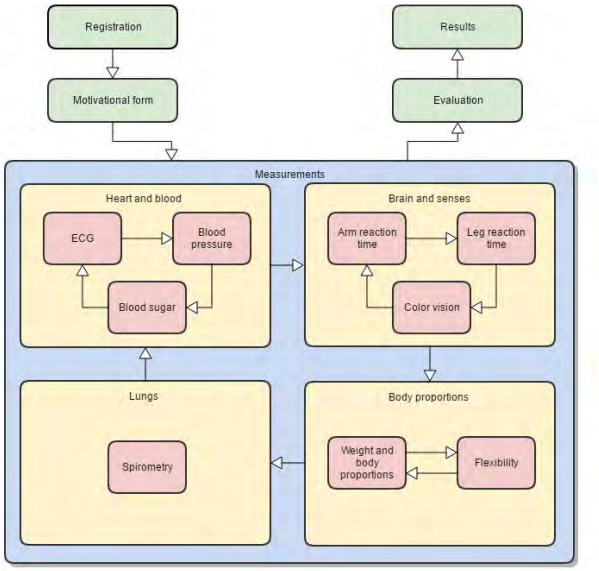



Figure 1: Experimental procedure.

The measured person starts with his/her registration, where he/she agrees with the collection and processing of his/her personal and health related data (informed consent) and provides basic information about himself/herself (name, sex, age, etc.). After the registration he/she proceeds to fill in a motivational questionnaire containing a set of 10-15 single choice questions related to his/her lifestyle. Then the measurement cycle starts (as shown in Figure 1). Individual

measurements are grouped into sections depending on measured parameters of human body. Every section includes one or more physical sites where the experimenter takes the measurement. When a single measurement is completed, the acquired data are inserted via a user interface into the system and sent to the server. When the whole set of measurements is completed, the data are evaluated, sent back to the client application and the participant can see his/her results extended with additional information. The whole procedure shown in Figure 1 takes approximately 10–15 minutes.

The following list contains the current set of defined physical sites and health related data acquired at these sites:

1. Brain
   - P300 latency [ms]
   - Fundamental frequency during concentration [Hz]
   - Fundamental frequency during meditation [Hz]
2. Body proportions
   - Height [cm]
   - Weight [kg]
   - Muscle mass [kg] or [%]
   - Body fat [kg] or [%]
   - Body water [kg] or [%]
   - Body Mass Index (BMI)
3. Electrocardiography (ECG)
   - Heart rate [BPM]
   - ST segment [ms]
   - QRS complex [ms]
4. Blood pressure
   - Heart rate [BPM]
   - Systolic pressure [mm Hg]
   - Diastolic pressure [mm Hg]
5. Blood sugar
   - Concentration of glucose [mmol/l]
6. Spirometry
   - FVC (Forced Vital Capacity) [l]
   - FEV1 (Forced Expiratory Volume in 1st second) [l]
   - PEF (Peak Expiratory Flow) [l/s]
7. Hand reaction time
   - Average reaction time [ms]
   - Number of falsely pressed buttons
   - Number of missed buttons
8. Leg reaction time
   - Average reaction time [ms]
   - Best reaction time [ms]
   - Worst reaction time [ms]
   - Standard deviation [ms]
9. Flexibility
   - Difference between fingers and foot during deep forward bend [cm]
10. Color vision
    - Number of wrongly recognized pseudoisochromatic tables

The whole software infrastructure as well as the current software prototype is designed in a way that allows easy enrollment of any future health related data category, it means with any new collection of health related data and metadata that could be defined even on site (just before the measurement). This put further requirements on the flexibility of the data model itself and the technologies used for its representation.

The software prototype can work in the online (default) or offline mode. When it is connected to the Internet, it sends all input data directly to the main server and receives back results and additional data. When the Internet connection is not available, the prototype switches to the offline mode. In this mode all data are stored to a local repository and sent to the server immediately when the Internet connection is again available.

The parametric requirements on the prototype application include its performance (input of health related data does not produce any visible delay), mobility (application can be used outside laboratory conditions when several hundreds people are measured during a day), simplicity and user friendliness (experimenters are able to work with the application after five minutes of training in the worst case).

## 3.2 Architecture and Design

The architecture of the software prototype is shown in Figure 2. It follows the MVC architectural pattern and client server architectural style.

The functionalities are encapsulated in modules that can be added as plugins.

- The General module covers functionalities affordable also for non-logged users.
- The Admin module serves for the administration of users and application setting.
- The Auth module is responsible for user registration and login.

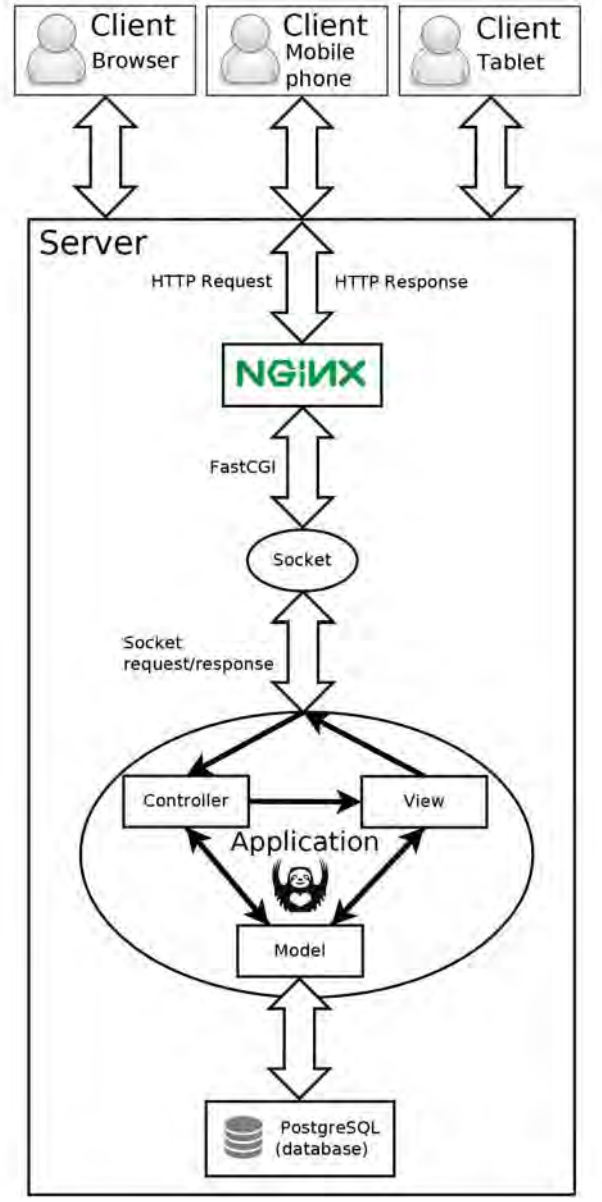



Figure 2: Software prototype architecture.

- The Measurement module includes definition of measurement procedure and overall data management.
- The Experiment module provides features for adding, editing and deleting experiments.
- Every experiment needs own equipment for measuring devices. The Equipment module stores them in the database and provides tools for their management.
- The QR generator module generates QR codes into a PDF document given to the participant. Each person has his/her number identifier included in his/her QR code.
- The Statistics module currently includes a function for simple statistics of reaction time.

Rest API is defined for collecting data from client devices.

## 3.3 Implementation and Deployment

The software prototype has been written in Python 3 language and uses the Flask micro framework that is based on the Werkzeug toolkit and Jinja2 template engine. The used languages and technologies include Python, HTML5, CSS3, Flask, Sqlalchemy, and Postgre SQL.

The Postgre SQL database is used for storing data. As a representative of open source object-relational database system it offers significant support for storing and retrieving JSON documents. Object relational mapping is provided by the SqlAlchemy framework. The current ERA model of the software prototype is shown in Figure 3.

The application is hosted by a virtualized server having the technical parameters provided in Table 1. The web server NGINX listens on two ports (the first port is used for the release version and the second one for the development version). NGNIX extends HTTP requests for non-standard headers CORS (cross-origin resource sharing) technology which is important for communication with mobile devices. NGINX passes requests to the python application with FastCGI protocol by a socket. Both development and release versions run five instances. The process describing how the requests are handled by the server is shown in Figure 2.

Table 1: Server hardware and software specifications.

| | |
|---|---|
| Processor | Intel(R) Xeon(R) CPU E5-4620 v2 @ 2.60GHz |
| Memory | 8036MB |
| Hard disk capacity | 100GB |
| Operating system | Debian 8.4 with kernel 2.6 |
| Web server | NGINX 1.6.2 |
| Database | Postgres 9.4 |
| Language and technologies | Python 3 using framework Flask |

## 3.4 Testing

The software prototype has been tested on 346 people in real environment (e.g. Days of science and technologies that were held in large tents on the square)

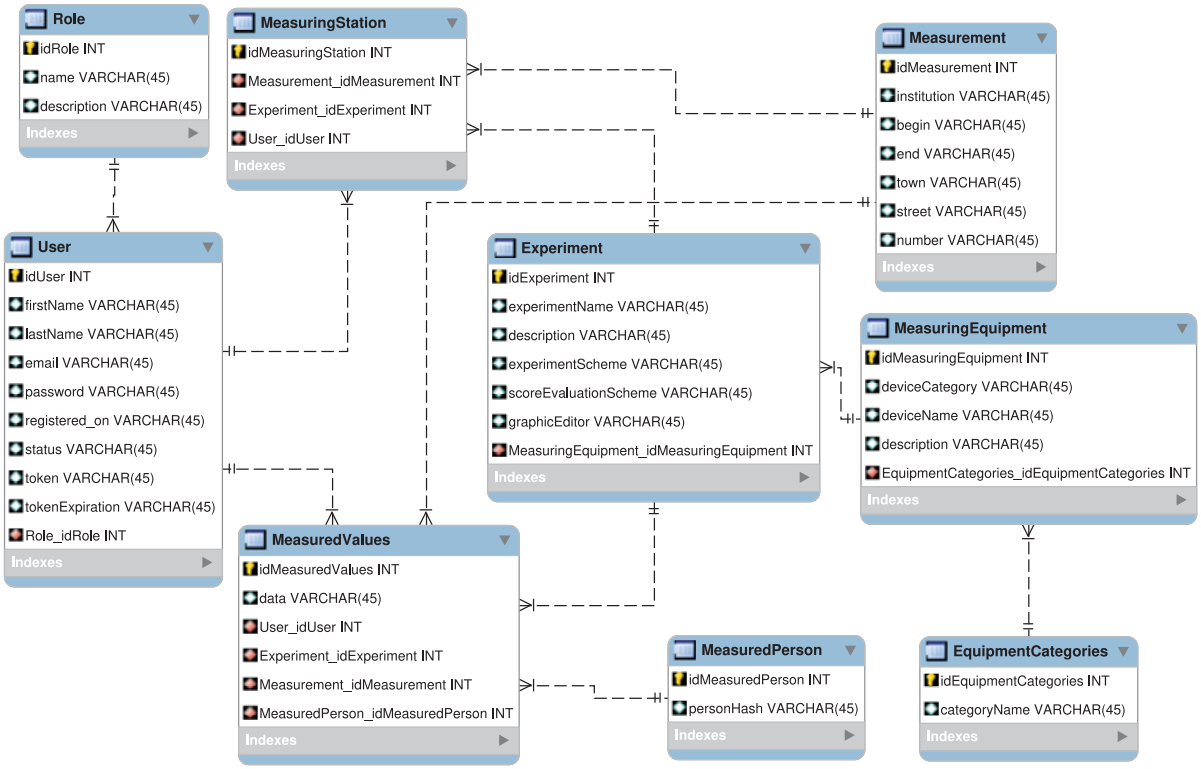



Figure 3: ERA model.

and has been continuously improved according to operation difficulties.

The list of devices which were used during the testing is shown in Table 2. The control software for impact dance is described in (omitted for the review process).

Table 2: Used devices.

| Site | Device name |
|---|---|
| Body proportions | Medisana BS 440 Connect |
| Electrocardiography | ReadMyHeart Handheld ECG |
| Blood pressure | Omron M6 Comfort IT |
| Blood sugar | FORA Diamond Mini |
| Spirometry | SP10W |
| Hand reaction time | Device for cognitive research |
| Leg reaction time | Impact Dance Pad |
| Flexibility | Podium and ruler |
| Color vision | Pseudoisochromatic tables |

# 4 CONCLUSIONS

In this paper we presented an idea of the exercise and wellness health strategy framework and the first steps that were done to support this framework by an appropriate software architecture. The proposed and developed software prototype covering a part of the whole infrastructure, namely rapid collection, storage and visualization of heterogeneous health related data, was successfully tested on more than three hundreds participants outside laboratory conditions.

The presented software prototype enables its users to quickly set the health related data to be collected and the whole data collection procedure. Then the data are collected, stored and visualized in an efficient way. Currently the prototype enables not only collection of classical data, but also non-traditional data (such as brain data) gathering. The software prototype design follows proven architectural patterns and styles, its modular structure facilitates its further extension.

In the future work we will focus on continuous extension of the software prototype to cover other parts of the intended software infrastructure for exercise and wellness health strategy framework. This include e.g. the modules for food and exercises recommenda-

tions, overall data evaluation or improved data visualization.

## ACKNOWLEDGMENTS

This publication was supported by the UWB grant SGS-2016-018 Data and Software Engineering for Advanced Applications, the project LO1506 of the Czech Ministry of Education, Youth and Sports under the program NPU I and the 2nd Internal grant scheme of UWB School of Computing, 2016.

## REFERENCES

Banos, O., Amin, M. B., Khan, W. A., Ali, T., Afzal, M., Kang, B. H., and Lee, S. (2015). Mining minds: An innovative framework for personalized health and wellness support. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2015 9th International Conference on*, pages 1–8. IEEE.

Blair, S. N., Cheng, Y., and Holder, J. S. (2001). Is physical activity or physical fitness more important in defining health benefits? *Medicine and science in sports and exercise*, 33(6; SUPP):S379–S399.

Busse, R. (2010). *Tackling chronic disease in Europe: strategies, interventions and challenges*. Number 20. WHO Regional Office Europe.

Joo, M.-I., Chung, G.-S., and Lee, T.-G. (2012). Wellness information system for smart healthcare. *Journal of Advanced Information Technology and Convergence*, 2(2):25–34.

Laakko, T., Leppanen, J., Lahteenmaki, J., and Nummiaho, A. (2008). Mobile health and wellness application framework. *Methods of Information in Medicine*, 47(3):217–222.

Lee, I.-M. and Skerrett, P. J. (2001). Physical activity and all-cause mortality: what is the dose-response relation? *Medicine and science in sports and exercise*, 33(6; SUPP):S459–S471.

Organization, W. H. et al. (2010). Global recommendations on physical activity for health.

Shephard, R. J. (2001). Absolute versus relative intensity of physical activity in a dose-response context. *Medicine and science in sports and exercise*, 33(6 Suppl):S400–18.

Taylor, R. S., Brown, A., Ebrahim, S., Jolliffe, J., Noorani, H., Rees, K., Skidmore, B., Stone, J. A., Thompson, D. R., and Oldridge, N. (2004). Exercise-based rehabilitation for patients with coronary heart disease: systematic review and meta-analysis of randomized controlled trials. *The American journal of medicine*, 116(10):682–692.

Tunstall-Pedoe, H. (2006). Preventing chronic diseases. a vital investment: Who global report. geneva: World health organization, 2005. pp 200. chf 30.00. isbn 92 4 1563001. also published on http://www. who. int/chp/chronic_disease_report/en. *International Journal of Epidemiology*, 35(4):1107–1107.

Webber, L., Divajeva, D., Marsh, T., McPherson, K., Brown, M., Galea, G., and Breda, J. (2014). The future burden of obesity-related diseases in the 53 who european-region countries and the impact of effective interventions: a modelling study. *BMJ open*, 4(7):e004787.

# An Adaptive Scrum Model for Developing Disease Registries

Hatem Bellaaj[1,3], Afef Mdhaffar[2,3], Mohamed Jmaiel[3] and Bernd Freisleben[4]
[1]*Department of Mathematics and Computer Science, IPEIS, University of Sfax, Sfax, Tunisia*
[2]*Department of Mathematics and Computer Science, ISSAT, University of Sousse, Sousse, Tunisia*
[3]*Digital Research Center of Sfax, Sfax, Tunisia*
[4]*Department of Mathematics and Computer Science, University of Marburg, Marburg, Germany*
{*hatem.bellaaj, afef.mdhaffar, mohamed.jmaiel*}*@redcad.org, freisleb@informatik.uni-marburg.de*

Keywords: Disease Registry, Agile, Scrum, Sprint, Transparency, Inspection, Adaptation.

Abstract: This paper presents an adaptive model for developing disease registries. The proposed model is based on the Scrum methodology. It can be used to draw a road map to identify priorities, inputs, outputs, team members and exchanges for all tasks required to develop disease registries. Our model has been applied to real cases from several Tunisian hospitals where it has improved the efficiency of the team members. The developed disease registries are currently used in Tunisia. They allow medical researchers to identify origins of diseases, establish new protocols, perform surveys and compute morbidity.

## 1 INTRODUCTION

Disease registries are used to collect and analyze epidemiological information related to the frequency and distribution (i.e., incidence and prevalence) of a specific disease. In addition to the detection of known symptoms and diagnosis parameters of diseases, statistics obtained from disease registries can help doctors to discover new risk factors. These are important to assess the medical situation and to provide framework conditions, preventive measures and management plans. A clinical information system refers to the organization of clinical data from medical records of patients to coordinate the delivery of interventions and self-management support activities for the entire population. Building a disease registry is crucial for managing a population of patients suffering from chronic diseases (McCulloch et al., 1998).

A disease registry can be considered as a medical information system (MIS) and as an instance of a general information system (IS). The development of this kind of software system is a resource-intensive process that involves different groups of people in an organization. Several software development models have emerged. Among them are: the systems-development life cycle (SDLC), rapid application development (RAD), and agile methodologies. Agile methodoloies are widely used for software project management, and Scrum is the most common agile method (Schwaber and Sutherland, 2001).

In this paper, we present a new model for developing disease registries based on Scrum. The model is adaptive in the sense that it provides support for incorporating recommendations that satisfy specific requirements of this domain. The proposed model has been applied to real cases from several hospitals in Tunisia, such as the Tunisian Fanconi Anemia Registry, the Tunisian Gaucher Disease Registry and the Tunisian Non-Hodgkin's Lymphoma Registry. In these projects, the use of our model has (a) accelerated the establishment of paper-based disease forms as functional specifications, (b) established a strong relation with the doctors, (c) minimized losses due to changes in fields' data types by fixing the sprint duration between 7 to 15 days, (d) found the right balance between documentation and discussion, (e) increased chances that the final product is as originally specified.

The paper is organized as follows. Section 2 introduces disease registries. Section 3 discusses related work. In Section 4, we summarize Scrum. Section 5 presents the proposed methodology. Applications are described in Section 6. Section 7 concludes the paper.

## 2 DISEASE REGISTRIES

A disease registry stores medical information related to the same disease. This allows medical researchers



Bellaaj, H., Mdhaffar, A., Jmaiel, M. and Freisleben, B.
An Adaptive Scrum Model for Developing Disease Registries.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 484-491
ISBN: 978-989-758-213-4

to (1) review and study health records of many individuals, (2) answer questions about diagnostic disease parameters and (3) build a knowledge database for each disease.

A disease registry is a tool for tracking the clinical care and outcomes of a defined patient population. Most disease registries are used to support care management for groups of patients suffering from one or many chronic diseases, such as diabetes, coronary artery disease and asthma. In contrast to paper-based registries that have been used to track patients suffering from chronic diseases, digital registries provide users with an automated way to (1) store data, (2) create, sort, and display lists of patients and (3) extract data used for planning, quality improvement, reporting, and direct care delivery (Hummel, 2000).

Digital disease registry functionality is included (i.e., available as an add-on) in electronic health record (EHR) products. In addition, stand-alone options can be implemented and are usually simpler to set up than EHRs. Based on their priorities, some organizations may choose to implement digital disease registries as an interim step prior to implementing a more comprehensive EHR system. Disease registries can also provide more flexibility for reporting and aggregating data from multiple data sources. To implement a digital disease registry, an organization needs to analyze and adjust practice workflows to support new data collection requirements and to integrate the new information from their registry into decision making and planning (Hummel, 2000).

# 3 RELATED WORK

## 3.1 The Process

A software development methodology is used to structure, plan, and control the process of developing an information system. Several methodologies to develop software have been proposed, such as Agile Software Development, Crystal Methods, Dynamic Systems Development Model (DSDM), Extreme Programming (XP), Feature Driven Development (FDD), Joint Application Development (JAD), Lean Development (LD), Rapid Application Development (RAD), Rational Unified Process (RUP), Scrum, Spiral, and Systems Development Life Cycle (SDLC). Three often used methods are discussed below.

### 3.1.1 Systems Development Life Cycle (SDLC)

SDLC is considered to be the oldest project execution framework. It can be used to manage large software projects, associated with corporate systems running on mainframes. It is a structured methodology, designed to manage complex projects that implicate many programmers and systems, having a high impact on the organization (Bourgeois, 2016).

### 3.1.2 Rapid Application Development (RAD)

RAD is a software development methodology that favors rapid prototyping on complex planning. In RAD, modules are developed in parallel as prototypes. Later, they are merged to the final product for faster delivery (Vickoff, 2000). The RAD method presents a secured and short development cycle following different phases: framing, design, building with fixed duration. Each phase takes about 90 to 120 days. RAD includes methods, techniques and tools to achieve four potentially contradictory objectives: budget, deadlines, technical quality, functional quality and visibility (Vickoff, 2000). The most important aspect for this model to be successful is to make sure that developed prototypes are reusable.

### 3.1.3 Agile Methods

Agile methods aim to reduce the life cycle of software development by creating a minimal version and then integrating functionality by an iterative process based on a customer listening and tests throughout the development cycle. The origin of agile methods is linked to the instability of the technological environment and the fact that the client is often unable to define his or her needs exhaustively from the beginning of the project. The term ”agile” thus refers to the ability to adapt to changes in the context and changes in specifications occurring during the development process. It was first coined in 2001 in the Manifesto for Agile Software Development (Agile Manifesto) (Beck et al., 2001). Now, agile refers to any process that follows concepts of the Agile Manifesto. There are four main points in the Agile Manifesto:

1. Individuals and interactions over processes and tools
2. Working software over comprehensive documentation
3. Customer collaboration over contract negotiation
4. Responding to change over following a plan

The Agile Manifesto lists 12 principles to guide teams on how to execute with agility. The principles are described below.

1. *Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.*

2. *Welcome changing requirements, even late in development. Agile processes harness change for the customers competitive advantage.*
3. *Deliver working software frequently, from a couple of weeks to a couple of months, with preference to the shorter timescale.*
4. *Business people and developers must work together daily throughout the project.*
5. *Build projects around motivated individuals. Give them the environment and support their need, and trust them to get the job done.*
6. *The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.*
7. *Working software is the primary measure of progress.*
8. *Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.*
9. *Continuous attention to technical excellence and good design enhances agility.*
10. *Simplicity: the art of maximizing the amount of work not done – is essential.*
11. *The best architectures, requirements, and designs emerge from self-organizing teams.*
12. *At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly (Beck et al., 2001).*

## 3.2 The Product

The basic architecture of a digital disease registry consists of four layers. They are described below.

### 3.2.1 Data Layer

Data is stored in a specific electronic health record (EHR), with data corresponding to parameters related to the diagnosis of the disease, such as responses to a treatment. The included data respects specific properties, such as persistence, accuracy (Staroselsky et al., 2008) and validity (Carroll et al., 2007). There are three main ways to populate the registry with data:

1. directly through the fields in the disease form.
2. importing data using standard interoperability mechanisms, such as HL7 or CCD. Data is automatically inserted into the database.
3. combining these two approaches.

The data layer must respect the international standards. Among them are:

- ISO 13119:2012 describes metadata that relates to resources including medical knowledge. This standard applies mainly to digital documents, such as WWW resources, accessible from databases or file transfers. It can also be applied to paper documents, such as articles in the medical literature (ISO 13119, 2012).
- ISO 13606-1:2008 describes the communication of some or all of the electronic health record (EHR). The record of a patient is identified between the DSEs, or between this latest and a centralized repository. It can be used to communicate an EHR system or repository with clinical applications or middleware components that need to access or provide EHR data (ISO 13606-1, 2008).

### 3.2.2 Security Layer

We distinguish between different kinds of users: system administrator, group administrator (who establishes the disease registry form, statistics parameters, therapy protocol, etc.), reference doctors, participant doctors, analysts, developers, patients, simple users, etc. A set of rules that defines the behavior of each kind of users must be specified at the start of the project. Different security norms are defined for such systems. Two of them are described below:

- ISO / TR 11633-1, 2009 concerns remote maintenance services (RMS) for information systems in healthcare facilities. It presents an example of a risk analysis that protects the information in the medical information system (ISO/TR 11633-1, 2009).
- ISO / TS 14441, 2013 examines the electronic patient records systems at the point of clinical care that are also interoperable with EHRs. It treats their protections in terms of security and privacy through a set of security and privacy rules. It includes guidelines and practices for conformity assessment (ISO/TS 14441, 2013).

### 3.2.3 Dashboard Layer

The dashboard layer is the main goal of disease registries. It shows morbidity, mortality, survey curves, treatment responses and illness origins. The plurality of dashboard components is associated with a plurality of types of health-care content and are based on parameters received from a user.

### 3.2.4 Interoperability Layer

Interoperability consists of communication between healthcare organizations. The medical information

system is specific for each organization. Building a global medical information system for many organisations is a complex task. Indeed, there is an array of healthcare-related applications that supports multiple needs but remains isolated or incompatible. Thus, interoperability is a challenge (Hammami et al., 2014). Different interoperability norms are defined for such systems. Some of them are presented below.

- ISO / TR 16056-1, 2004 introduces the interoperability of telehealth systems and networks. It also defines telehealth and related terms.
- ISO EN 13606: Medical record communication is based on a two-model approach using paradigms. This ensures that the data collected from heterogeneous systems are correctly interpreted. The dual model provides the basis for a stable architecture. It separates information from knowledge.

# 4 SCRUM

Scrum originates from the sporting term rugby meaning: melee. Like this technical aspect of the game, the methodology asks its actors to be united in the accomplishment of a project, in achieving a goal. A melee is not a unique process. This is a part of the game that is often found to move the team forward. In the same concept Scrum uses a procedure that we name sprint. Each iteration or sprint provides a functional part of product. Three pillars uphold every implementation of empirical process control: transparency, inspection, and adaptation (Schwaber and Sutherland, 2001). Scrum includes definitions, descriptions, concepts and methods for better running projects. It defines details for: the Scrum team, the product owner (PO), the development team, the Scrum master, the Scrum events, the sprint, the daily Scrum, the sprint review, the sprint retrospective, the artifacts, the product backlog, the sprint backlog and the artifact transparency.

# 5 AN ADAPTIVE SCRUM MODEL FOR DISEASE REGISTRIES

This section presents all adapted pillars, mechanisms, and concepts for developing disease registries using the Scrum model.

## 5.1 Three Pillars

### 5.1.1 Transparency

Transparency insurance is difficult between people who do not share the same discourse. Thus, we propose to build:

- a medical dictionary, including term definitions comprehensible by all players
- a crossing table between the variables, specifying: mathematical equations, if they exist and logical relations between values, e.g., it is not logical to find the symptom S1 true while the variable V1 is normal
- an effective way to present a variable: selection between alphanumeric value, checkbox, text field to be specified

### 5.1.2 Inspection

There are three levels of inspections that can be done sequentially and at the same time according to the iteration in question:

- functional validation by computer scientists
- medical validation of the distribution of the fields to be entered, the logical and mathematical links of the different variables
- validation of the possibility of statistical interpretation of the different variables

The second type of validation often leads to the addition, deletion or modification of the type of presentation of some fields. Development constraints cause developers to modify a type of representation or an operating logic that can essentially lead to poor statistical quality. Inspection becomes foolish and can greatly slow down the development process.

### 5.1.3 Adaptation

Scrum prescribes four formal events for inspection and adaptation: sprint planning, daily scrum, sprint review and sprint retrospective. The fact that the project is carried out by people of different specialties, a misunderstanding of need can delay and burden the concept of adaptation. Reports of daily meetings at a frequency of maximally three days can be sent to the head of the medical study group.

## 5.2 The Scrum Team

The Scrum team consists of a product owner (PO), the development team, and a Scrum master. Scrum

teams are self-organizing and cross-functional. In the following, the needed skills, the specific mission, relations and input / output of each of them are detailed.

For a disease registry, three teams should be established, which will work together and simultaneously: the doctors who are members of the study group, statisticians and developers. The PO leads these three teams and takes decisions after meetings including all the teams or representative members of each of them. Therefore, (s)he is the first person responsible for the product backlog.

The team model in Scrum is designed to optimize flexibility, creativity, and productivity. For disease registries, flexibility is guaranteed by trust, designing, and good communication. Trust means the capability of achieving a good job, possibly by young employees. Designing consists of giving priority to dispatching the members of the team and not the tasks. Productivity is increased by communication (e.g., regular meetings), tools (e.g., management version applications) and documentations.

Scrum teams deliver products iteratively and incrementally, maximizing opportunities for feedback. An average incremental delivery period can be defined as about 10% of the number of fields for form and dashboard pages. For the user management module, 25% of the total number of user groups is appreciated. For other modules, these values must be defined at the start of the project but cannot exceed two weeks as a period to increase profitability.

#### 5.2.1 The Product Owner

The product owner is in charge of maximizing the value of the product and the work of the development team. The PO is the sole person responsible for managing the product backlog (Sverrisdottir et al., 2014).

For disease registries, it is recommended that the product owner communicates periodically the updated backlog to the leader of the disease study group, discusses and may change some priorities.

The PO insures that the product backlog is visible, transparent, and clear to all, and shows what the Scrum team will work on in the next step. The draft version of backlog should be validated at the last meeting with doctors before the kick-off of the project.

#### 5.2.2 The Development Team

The development team consists of professionals who are in charge of delivering a feasible increment of the done product at the end of each sprint. Only members of the development team create the increment.

In Scrum, it is recommended that the development team includes 3 to 9 members to insure efficiency and global efficacy. For a disease registry, we have adopted this structure: 1 designer, 1 tester, 1 architect, 1 to 2 developers for security management and 3 to 9 Java developers.

#### 5.2.3 The Scrum Master

The Scrum master is responsible for ensuring that Scrum is understood and implemented. It will play exactly the same roles as defined in the Scrum guide. Thus, it will ensure the availability of tools necessary for the good understanding and adherence to Scrum for the product owner and the development team.

Ideally, the Scrum master is a member of the development team. He or she is supposed to master notions of statistics. In his or her profile, conducting research in medical informatics or participating in a similar project is appreciated.

### 5.3 The Scrum Events

In Scrum, we attempt to fix regular events and to minimize the need for unplanned meetings. All events are timed, so that each event has a maximum duration. Once a sprint starts, it can not be shortened or lengthened. Events end each time the associated goal is reached. Appropriate time must be allocated for each sprint to avoid having waste in the process.

#### 5.3.1 The Sprint

The heart of Scrum is a sprint. It is a time block resulting in an increment of the potentially deliverable product. It lasts for one month or less, and a deliverable product increment is created. A new sprint starts immediately after the conclusion of the previous sprint (Schwaber and Sutherland, 2001).

Sprints contain and consist of the sprint planning, daily Scrums, the development work, the sprint review, and the sprint retrospective (Schwaber and Sutherland, 2001). For a medical disease registry, the average duration of a sprint is ideally between 7 and 15 working days.

#### 5.3.2 Sprint Planning

One of the first methods spread around the world is the V-Cycle method. It is a project organization logic that limits the problem of reactivity in the event of an anomaly, limiting the return to the preceding steps. It is represented by a V whose descending branch contains all the phases of the design of the project, and the rising branch all the stages of tests of the project.

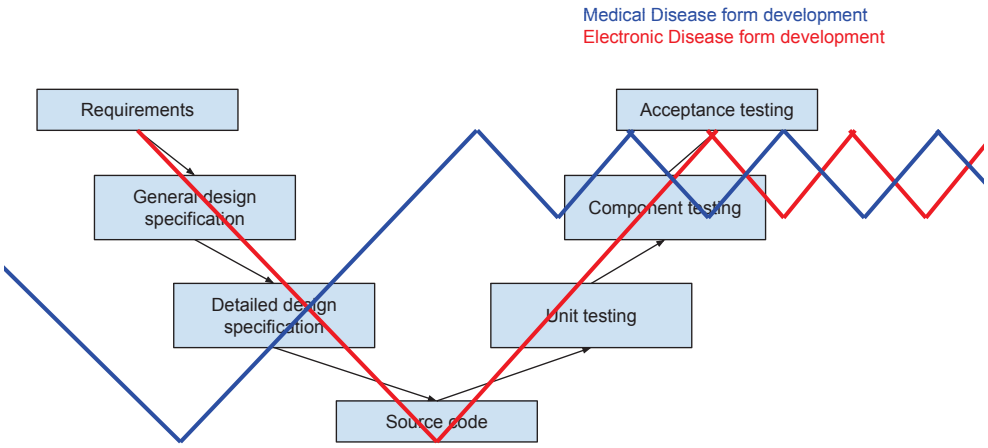



Figure 1: The W Model for disease registry development.

The tip of the V represents the stage of realization of the project. Each phase of the descending branch is related to a phase of the rising branch.

In disease registry development, we have adopted, for the first time, this kind of project organization method. The same thing is done with the medical team to establish the disease form and the list of included elements in a dashboard (statistics). For this purpose, two V cycles (i.e., a W cycle) are established with intersection points (see Figure 1). These points represent meetings between members of medical and development staff. In addition to the complicated problems of return in the cycle for both teams, the meeting management presented by the intersection points in Figure 1 must be handled.

In disease registry development, we recommend that a sprint duration is between 1 and 2 weeks. The reduction of this value increases the profitability of the team, but may make the management of the different tasks complicated. In this stage, it is important to take into account the source version management process. Two methods can be used: distribution according to functionalities or according to teams. The first is strongly recommended for disease registries.

#### 5.3.3 Daily Scrum

The daily scrum is a 15-minute time-boxed event for the development team to synchronize activities and create a plan for the next 24 hours (Schwaber and Sutherland, 2001). It is recommended that a member of the study group participates in the daily scrum. Since doctors are usually solicited by their patients, it is recommended to establish these meetings at the end of the day. Thus, a meeting should answer the following questions:

- What did I do today that helped the development team meet the sprint goal?
- What will I do tomorrow to help the development team meet the sprint goal?
- Do I see any impediment that prevents me or the development team from meeting the sprint goal?

#### 5.3.4 Sprint Review

A sprint review is held at the end of the sprint to inspect the increment and adapt the product backlog if needed. During the sprint review, the Scrum team and doctors (two or three doctors who have different specialities), and members of the study group collaborate about what was done in the sprint.

For the first six months, the sprint review should be done twice a month. After that, the frequency can be decreased to once a month. The meeting can take between 15 minutes and 1 hour.

In addition to the elements defined in the Scrum guide, the sprint review includes (1) the team problems to understand field types and relations; (2) doctors' comments and validation about cognitive workload; (3) the steps to do by participating doctors and (4) discussion of technical issues, system interoperability, privacy, confidentiality and lack of health information data standards.

#### 5.3.5 Sprint Retrospective

The purpose of the meeting is to improve the process for the next sprint. The entire Scrum team participates in the meeting. The retrospective takes place just after the sprint review, and the speakers who have attended can remain for the retrospective as observers (Schwaber and Sutherland, 2001). The retrospective sprint clarifies the points of interference with the doc-

tors; a sharp intervention by them to verify the understanding may be planned for the next sprint.

## 5.4 Scrum Artifacts

Scrum artifacts represent information that the Scrum team needs to insure inspection and adaptation. They help the team to understand the product under development, the activities done, and the activities being planned in the project. Scrum defines the following artifacts: product backlog, sprint backlog, increment and burn-down chart. For a disease registry, the modification proposal is limited to the three first artifacts (Schwaber and Sutherland, 2001).

### 5.4.1 Product Backlog

The Scrum product backlog is a prioritized feature list, containing short descriptions of the functionality desired for the product. In a disease registry project, the product backlog includes:

- Disease form management operations: insert, modify, list and delete. The deletion must be logic. A module for logging the various actions carried out by the user must be set up. Data verification algorithms may be required.
- User management operations: insert, modify, list and delete. A group right must be clearly defined to access data and to consult statistics. In the general case, the doctor should consult and modify only forms that (s)he has introduced.
- Data mining includes histogram presentations, pie chart, or curves for:
  - enumerating parameters (example: consanguinity)
  - numerical parameters (example: weight), with a possibility of zooming on particular zones
- Security requirements: who can do what?

### 5.4.2 Sprint Backlog

The sprint backlog is created during the sprint planning event, which is the first event in a sprint. A critical point in a disease registry project is the establishment of the detailed plan for delivery of the items and realization of the sprint goal during the sprint. Some distributions may occur due to insufficient explanations of requirements by the doctor for mathematical or logical relations between fields. This detailed plan will continue to be updated during the sprint.

### 5.4.3 Increment

An increment represents items made during the current sprint and those that preceded it. At the end of a sprint, the tasks included in the new increment must be performed. This means it must be finished (i.e., developed and tested) and meet the Scrum teams definition of done. In a disease registry, a task is done when it is validated by at least one doctor.

# 6 APPLICATIONS

Three disease registries were developed based on the proposed approach: The Tunisian Fanconi Anemia Registry (TFAR), The Tunisian Non-Hodgkin Lymphoma Registry (NHLR) and the Tunisian Gaucher Disease Registry (TGDR).

The TFAR was developed within one year. The disease form has more than 200 fields. Doctors from 5 medical areas participated: hematology, oncology, biology, pediatrics, cytogenetics. They belong to 10 hospitals and institutes. The project was developed by 8 members: 5 developers, 1 statistician, 1 human machine interface designer and 1 product owner. Scrum daily meetings included one doctor and usually took about 1 hour. The first backlog sprints contain explanations of data fields and relations between them.

The NHLR was developed, in cooperation with MDSoft [1], during three years. The statistics module is still in the works. The disease form includes more than 1000 fields. Doctors from 2 medical areas participated: hematology and oncology. They belong to 6 hospitals and institutes. The project was developed by 7 members: 4 developers, 1 statistician, 1 human machine interface designer and 1 product owner. The duration of Scrum daily meetings is about a half hour, with the participation of one doctor. A particular difficulty was experienced during Scrum planning. The disease form is divided into sections. Each section contains several input fields. During the merging operation of the source code of the different sections, we have encountered difficulties especially in the case where there are relationships between the fields that they contain. The sprint retrospective was complicated due to the new requests raised by the doctors after each trial of the product.

The second edition of the TGDR was established in November 2016 in collaboration with the MDSOFT team. The disease form includes more than 250 fields. Doctors from 5 medical areas participate: hematology, pediatrics, internal medicine, rheumatology and gastroenterology. They belong to 6 hospitals

[1] http://www.mdsoft-int.com

and institutes. The project was developed by 6 members. The product backlog has been updated several times due to the instability of both medical and developer teams.

During these projects, our new methodology has improved the following indicators: dedication (additional sponteneous working hours reached up to 80%), focus (90% of the tasks were completed within deadlines), openness (the rate of misunderstanding between team members was less than 5%), audacity (the team did not hesitate to name things, to ask questions and to propose new solutions), and continuity (despite major changes in team composition, the projects did not have any delays in delivery). Moreover, the use of our methodology has reduced the risk of failure by 95% in the case of TFAR.

# 7 CONCLUSION

In this paper, we have presented a new methodology based on Scrum for disease registry development. Several actions have been proposed to improve team performance: (a) minimize the time of different iterations, (b) facilitate code retrieval in the majority of iterations, (c) clarify the descriptions and interactions between different fields, (d) maximize collaboration between the different teams and specialists involved, namely doctors, computer scientists and statisticians. Our methodology has been applied to the Tunisian Fanconi Anemia Registry, the Tunisian Non-Hodgkin Lymphoma Registry and the Tunisian Gaucher Disease Registry. The developed registries are currently used in several hospitals in Tunisia.

In the future, we plan to define an effective methodology for managing source code and deliverables to improve team profitability.

# ACKNOWLEDGEMENTS

This work is supported by the German Academic Exchange Service (DAAD) (Transformation Partnership: Theralytics Project).

# REFERENCES

Beck, K., Beedle, M., Bennekum, A., Alistair, C., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R., Mellor, S., Schwaber, K., Sutherland, J., and Thomas, D. (2001). The agile manifesto. *Agile Alliance*, 2001.

Bourgeois, D. (2016). *Information Systems for Business and Beyond*. Textbook Equity Edition.

Carroll, N., Ellis, J., Luckett, C., and Raebel, M. (2007). Improving the validity of determining medication adherence from electronic health record medications orders. *J. Am. Med. Inform. Assoc.*, 18(5):717–720.

Hammami, R., Bellaaj, H., and Kacem, A. (2014). Interoperability for medical information systems: an overview. *Health and Technology*, 4(3):261–272.

Hummel, J. (2000). Building a computerized disease registry for chronic illness management of diabetes. *Clinical Diabetes*, 18(5).

ISO 13119 (2012). Health informatics – clinical knowledge resources – metadata. Standard, International Organization for Standardization.

ISO 13606-1 (2008). Health informatics – electronic health record communication – part 1: Reference model. Standard, International Organization for Standardization.

ISO/TR 11633-1 (2009). Health informatics – information security management for remote maintenance of medical devices and medical information systems – part 1: Requirements and risk analysis. Standard, International Organization for Standardization.

ISO/TS 14441 (2013). Health informatics – security and privacy requirements of EHR systems for use in conformity assessment. Standard, International Organization for Standardization.

McCulloch, D., Price, M., Hindmarsh, M., and Wagner, E. (1998). A population-based approach to diabetes management in a primary care setting: early results and lessons learned. *Effect. Clin. Pract.*, 1(1):12–22.

Schwaber, K. and Sutherland, J. (2001). The Scrum guide.

Staroselsky, M., Volk, L., Tsurikova, R., Newmark, L., Lippincott, M., Litvak, I., Kittler, A., Wang, T., Wald, J., and Bates, D. (2008). An effort to improve electronic health record medication list accuracy between visits: patients' and physicians' response. *Int. J. Med. Inform.*, 77(3):153–160.

Sverrisdottir, H., Ingason, H., and Jonasson, H. (2014). The role of the product owner in Scrum - comparison between theory & practice. *Procedia - Social & Behavioral Sciences*, 19(3):257–267.

Vickoff, J. (2000). RAD, le developpement rapide d'applications. *RAD.fr*.

# Special Session on Smart Medical Devices - From Lab to Clinical Practice

# Full Papers

# A Methodology to Reduce the Complexity of Validation Model Creation from Medical Specification Document

Francesco Gargiulo, Stefano Silvestri, Mariarosaria Fontanella and Mario Ciampi
*Institute for High Performance Computing and Networking, ICAR-CNR, Via Pietro Castellino 111 - 80131, Naples, Italy*
{*francesco.gargiulo, stefano.silvestri, mariarosaria.fontanella, mario.ciampi*}*@icar.cnr.it*

Keywords: Clustering, Medical Specification Document, Validation, Natural Language Processing (NLP), Schematron.

Abstract: In this paper we propose a novel approach to reduce the complexity of the definition and implementation of a medical document validation model. Usually the conformance requirements for specifications are contained in documents written in natural language format and it is necessary to manually translate them in a software model for validation purposes. It should be very useful to extract and group the conformance rules that have a similar pattern to reduce the manual effort needed to accomplish this task. We will show an innovative cluster approach that automatically evaluates the optimal number of groups using an iterative method based on internal cluster measures evaluation. We will show the application of this method on two case studies: i) Patient Summary (*Profilo Sanitario Sintetico*) and ii) Hospital Discharge Letter (*Lettera di Dimissione Ospedaliera*) for the Italian specification of the conformance rules.

## 1 INTRODUCTION

The availability of medical information processing systems and the digitalization of almost all information in hospital and clinical processes provide an important support for the tasks of healthcare professionals. Dealing with digital documents and using custom processing systems can improve their work, offering a lot of innovative tools and instruments, ranging from improved information retrieval systems to intelligent image and text processing.

Focusing especially on text documents, we know that an important part of the work of healthcare professionals is the editing of many different clinical documents such as Patient Summaries, Laboratory Tests Reports and Medical Prescriptions. All of them are structured or semi-structured text documents and, furthermore, they even require the presence of certain information, like, for example, a doctor name, a date or a disease code. In addition, their structure and content must respect official guidelines, often established by law. These specifications propose to standardize the structure of these digital documents, ensuring the correctness and the completeness of the content and of the text format.

A standard like the ones promoted by HL7 not only can ensure the semantic and formal correctness of the digital version of these documents, but it supports an effective and reliable automatic processing and the interoperability between different systems too (Ciampi et al., 2016). In other words, it is crucial that the exchanging of these documents between different hospitals or physicians is error-free, without loss of information.

Due to the importance of these tasks, the definition of the conformance rules is a long and critical process, that involves many specialists from medical, clinical, legal and computer science fields and, of course, the governments and health-care agencies. The results of their work are usually documents written in natural language, containing a set of conformance requirements rules for specifications that define the format and the content each of them.

The need of conformance rule documents arises from requirements of standards. Official medical natural language text documents must not only be automatically processed easily, but even respect a format and contain specific information. In Italy, Agencies and government representatives, at this aim, have produced the conformance specifications documents for the digital version of the Patient Summary, the Laboratory Medicine Report, the Hospital Discharge Letter and the Medical Prescription, that are actually part of HL7 International standards in the Italian context.

As explained before, the specifications are documents written in natural language format, describing the whole conformance rules for specifications and the details of the implementation guide for each of



Gargiulo, F., Silvestri, S., Fontanella, M. and Ciampi, M.
A Methodology to Reduce the Complexity of Validation Model Creation from Medical Specification Document.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 497-507
ISBN: 978-989-758-213-4

digital medical certificates listed above.

Among all the possible uses of conformance rules, one of them could be the development of a validation model, that ensures and tests the complete conformance of the digital certificate to the standard statements.

To implement this kind of functions, computer scientists and engineers must perform a long and tedious task, analysing the natural language text in the conformance specifications document to realize a complete and reliable validation schema for each rule listed in the standard (Gargiulo et al., 2016). This task can be performed only by an hand-made translation of each natural language rule in a software model for validation purposes, using, for example, Schematron (ISO/IEC 19757-3:2016) (Jelliffe, 2001), or other rule-based validation languages. Nowadays, it is a critical task to extract automatically a validation schema from a set of rules described in natural language.

A great boost in the realization of the validation schema can be obtained simply reducing the complexity of the problem, decreasing the number of assertions that has to be manually built. This task can be accomplished grouping the rules following the same pattern: in this way, the same assertion function could be applied to more rules, speeding up the development of the validation model.

In this paper we propose an innovative methodology based on unsupervised machine learning techniques, namely clustering, that extracts automatically the text of the rules from the specification documents and groups them together. Each group contains all the rules that belong to the same assertion schema. The experiments have been performed on Italian language specification rule documents of medical topic, but the proposed techniques are language independent and they can be applied on documents in different languages, or to any kind of specification document.

The paper is structured as follow: in Section 2 it will be given a critical review of the state of the art related to automatically validation and clustering optimization fields; in Section 3 it will be shown the methodology and in Section 4 it will be detailed the designed architecture; in Section 5 the methodology correctness will be demonstrated for two use cases: i) *Patient Summary* and ii) *Hospital Discharge Letter*. Finally, in Section 6 it will be given the conclusion and it will be draw up some key issues for future works.

## 2 RELATED WORKS

Nowadays there is a big interest of scientists about creation and automatic validation of conformance rules in natual language, especially for medical domain. In (Boscá et al., 2015) the authors proposed and described the archetypes to generate a rules in Natural Language text and Schematron rules for the validation of data instances. The goal was creating a formal document with a formal value of archetype, but at same time understandable by non-technical users.

In (Boufahja et al., 2015) the authors demonstrated the conformance of their samples with HL7 CDA requirements and evaluated the capability of the tools to check those requirements. They looked at the conformance of the provided samples with the basic HL7 CDA requirements as specified within the *Clinical Document Architecture, R2 Normative Edition*, and analysed the capability of the tools provided to check the requirements. At the first time, the authors revisited the CDA specifications and extract the requirements not covered by the CDA Schema, then they checked the coverage of the requirements with another validation tools.

In (Hamilton et al., 2015) the authors described a method in which users realize the benefits of a standards-based method for capturing and evaluating verification and validation (V&V) rules within and across metadata instance documents. The rule-based validation and verification approach presented has the primary benefit that it uses a natural language based syntax for rule set, in order to abstract the computer science-heavy rule languages to a domain-specific syntax. As a result, the domain expert can easily specify, validate and manage the specification and validation of the rules themselves.

In (Jafarpour et al., 2016) is evaluated the technical performance and medical correctness of their execution engines using a range of Clinical Practice Guidelines (CPG). They demonstrated the efficiency of CPG execution engines in terms of CPU time and validity of the generated recommendation in comparison to existing CPG execution engines.

Clustering is an unsupervised machine learning technique, that can well group together objects that show similarity between each others. One of the main problem in clustering, being unsupervised, is the cluster validation, that, in fact, has long been recognized as one of the crucial issues in clustering applications. Validation is a technique to find a set of clusters that best fits natural partitions without any class information, finding the optimal number of clusters (Halkidi and Vazirgiannis, 2001).

The measures used for cluster validation purposes

can be categorized into two classes: external and internal. The first case can be used when a gold case is available, verifying the correctness of results through measures like F-measure, Entropy, Purity, Completeness, Homogeneity, Jaccard coefficient, Fowlkes and Mallows index, Minkowski Score and others (Rendón et al., 2011), (Wu et al., 2009), (Handl et al., 2005), (Rosenberg and Hirschberg, 2007). These papers analysed and compared all the aspects of each measure to understand how well it fits specific cluster algorithm, application or topic, revealing the goodness of the clustering. A common ground of external measures is that they can often be computed by the contingency matrix (Wu et al., 2009).

When a gold case is not available, the second class of cluster validation measures, namely the internal ones, must be used. In this case, the goodness of clustering results is based only on spatial characteristics of cluster members, like their compactness or separation. One of the first internal cluster measure proposed in literature is the silhouette (Rousseeuw, 1987). The silhouette is a numeric parameter that takes in account the tightness and separation of each cluster, showing which objects lie well within their cluster and which ones are merely somewhere in between clusters. Many other internal measures have been defined in literature, like Dunns indices, SD and SD_bw validity indexes and others (Liu et al., 2010), taking into account different aspects of the clustering results in addition to the separation and compactness, like monotonicity, noise, density, sub-clusters and skewed distributions, that can better show different aspects of the results.

Internal cluster measures have been often used to set the correct cluster number, not only optimizing their global value (Kaufman and Rousseeuw, 2009), but even obtaining some specific new measures from the classical ones, to identify cluster characteristics of a specific domain, as, for example, they did in (Pollard and Van Der Laan, 2002). In (Dhillon et al., 2002) an iterative clustering method is proposed to improve spherical *K*-means algorithm results, that, when applied to small cluster sizes, can tend to get stuck at a local maximum far away from the optimal solution. They presented an iterative local search procedure, which refines a given clustering by incrementally moving data points between clusters, thus achieving a higher objective function value.

# 3 METHODOLOGY

In this Section we explain the details of the methodology applied in our experiments. We developed an iterative cluster strategy, that aims to obtain the best clustering solution. This is achieved through an internal measure cluster selection, described in 3.1 and in 3.2. Then, to assess the whole methodology, we manually built a gold case, validated using a custom cluster external validation measure described in 3.3. Gold case construction and validation assessment are described in Section 5

## 3.1 Clustering Algorithm and Internal Measures

Following the literature (Alicante et al., 2016a), we decided to use the spherical *K*-means cluster algorithm, a slight variation of the *K*-means algorithm, and the *cosine* similarity. It has been shown that the optimal distance for *K*-means based cluster applications for Italian natural language text of medical topic is the *cosine distance* (Alicante et al., 2016b), that is equals to inverse cosine similarity (eq. 1).

$$1 - \sum_{i=1}^{M} \frac{x_i \cdot y_i}{\mid x_i \mid\mid y_i \mid} \tag{1}$$

The cosine similarity measure allows to use the spherical *K*-means (Zhong, 2005) algorithm, that uses a slight variation of the *K*-means algorithm exploiting the cosine similarity measure: the classical *K*-means minimizes the mean squared error from the cluster centroid (eq. 2)

$$\frac{1}{N} \sum_{\mathbf{x}} \|\mathbf{x} - \mu_{k(\mathbf{x})}\| \tag{2}$$

where N is the total number of feature vectors and $\mu_{k(\mathbf{x})}$ is the most similar centroid; instead, in spherical *K*-means the objective function is defined as (eq. 3)

$$\sum_{\mathbf{x}} \mathbf{x} \cdot \mu_{k(\mathbf{x})} \tag{3}$$

that is strictly related to the cosine similarity. Our experiments confirm the goodness of these choices (see Section 5).

The determination of optimal partition is performed through an iterative loop, based on cluster internal measure, described in details in Section 3.2.

As assessment of clustering results we can only use internal measures, having no labelled data. For validation purposes we have chosen the *silhouette* (Rousseeuw, 1987), a classic cluster internal validation measure, that takes into account two important aspects of a clustering solution: i) the similarity among elements of the same cluster and ii) the dissimilarity among elements belong to different clusters.

Let call $i$ a generic point of the data set and $a(i)$ the average dissimilarity of the point with the elements

of the same cluster. Dissimilarity is calculated with inverse cosine similarity. A small $a(i)$ means that the point is quite close to all the other points in the cluster. We define $b(i)$ as the smallest average dissimilarity between $i$ and the elements of any cluster different from the one $i$ belongs, estimating how far the current point is from the closest point not in the same cluster.

Then the silhouette $s(i)$ of each point of a cluster is defined as:

$$s(i) = \frac{b(i) - a(i)}{max(a(i), b(i))} \tag{4}$$

where the opposite of $a(i)$ is considered so that its effect is in favour of compactness. The silhouette value is in the range $[-1, 1]$, and a larger silhouette indicates a better assignment of that point to its cluster. The silhouette is negative whenever the other points in the cluster are, on average, farther from the point $i$ than the closest point outside of the cluster. Silhouette can then be averaged on all points of a cluster to assess the compactness of that cluster with respect to the others. In this case, a negative number of silhouette means that the diameter of the cluster is larger than the distance from the closest point out of the cluster.

The average silhouette over all elements $i$ could be used to compare clustering results and to select the optimal number of clusters $k$ by maximizing it over a range of possible values for $k$ (Kaufman and Rousseeuw, 2009). The method of maximizing average silhouette can be used with any clustering algorithm and any distance metric, but it has the disadvantage that measures only the global structure of the solution. To take in account finer behaviour we have proposed an alternative parameter. Let consider the average silhouette of the $j^{th}$ cluster as $S_j$. We then call *MAS* the median of average silhouettes $S = \{S_1, S_2, \ldots, S_k\}$, a value equals to:

$$MAS = median(S) \tag{5}$$

The *MAS* can give a synthetic clue about the goodness of the entire cluster solution, but, differently from the simple average of all silhouettes $s(i)$, it can take into account each cluster validity.

## 3.2 Iterative Cluster Optimization

To obtain a more precise clustering we proposed an iterative clustering optimization algorithm, based on *MAS* optimization. In Figure 1 is depicted flow chart diagram, representing the proposed methodology.

After constructing a Vector Space Model (VSM) of the Conformance Rules, we have defined an iterative cycle. The first task is a de-noising of the input data, using a Principal Component Analysis

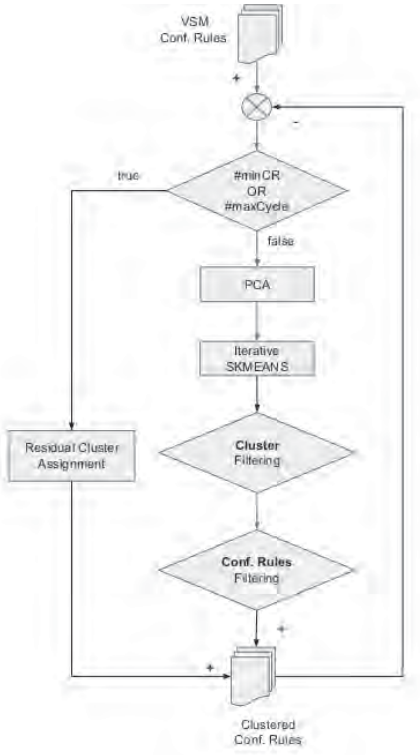



Figure 1: Flow chart of the methodology used to cluster the conformance rules.

(PCA) methodology for feature reduction (Cao et al., 2003). We set the selection of information content of PCA at 96%: this value has been obtained observing the higher mean silhouette value of clustering experiments. The feature reduction is performed at each step of the iterative cluster algorithm, reducing each iteration the number of extracted features.

Then, an Iterative Spherical K-means algorithm, depicted in Figure 2, is applied to evaluate the optimal cluster solution. We perform Spherical K-means with cluster number ranging from 2 to total Conformance Rules number. The optimal cluster solution is the one with highest *MAS* value in the range of all solutions obtained during the iteration. From the whole solution with highest *MAS* we select only clusters whose mean silhouette is bigger than *MAS* (*inter cluster selection*); then, in these selected clusters, we filter out the elements whose silhouette is smaller *MAS* (*intra cluster selection*). The clusters obtained with this filtering operations are selected as part of the final solution and the remaining elements are iteratively re-processed in the same way, until the number of remaining documents is smaller than a given threshold *#minCR* or the number of iterations is bigger than a threshold *#maxCycle* (see Figure 1).

When the termination condition is reached, the

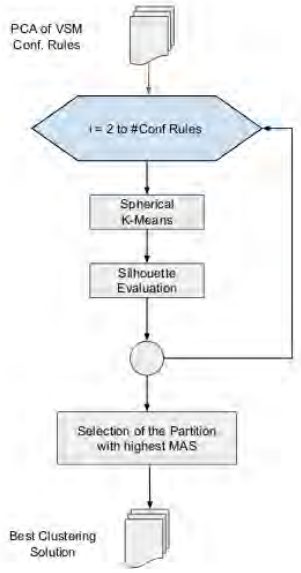



Figure 2: Flow chart of the implemented Iterative Spherical K-Means.

residual conformance rules are assigned to their own cluster (one element cluster), considering that each rules have a low silhouette value.

### 3.3 Cluster Validation Measure

The evaluation of the clustering goodness considering an handmade gold case is obtained using external measures, that are often computed by the contingency matrix (Wu et al., 2009).

The contingency matrix (see Tab.1) is defined as follow: given a data set $D$ with $n$ objects, assume that we have a partition $C = \{C_1, \ldots, C_{K'}\}$ of $D$, where $\bigcup_{i=1}^{K'} C_i = D$ and $C_i \bigcap C_j = \emptyset$ for $1 \leq i \neq j \leq K'$ , and $K'$ is the number of clusters. If we have a *Gold Case*, we can have another partition on $D : P = \{P_1, \ldots, P_K\}$, where $\bigcup_{i=1}^{K} P_i = D$, $P_i \bigcap P_j = \emptyset$ and $K$ is the number of classes. Each element $n_{ij}$ of the matrix denotes the number of objects in cluster $C_i$ from class $P_j$.

Table 1: The Contingency Matrix.

| | $C_1$ | $C_2$ | ... | $C_{K'}$ | $\sum$ |
|---|---|---|---|---|---|
| $P_1$ | $n_{11}$ | $n_{12}$ | ... | $n_{1K'}$ | $n_{1\cdot}$ |
| $P_2$ | $n_{21}$ | $n_{22}$ | ... | $n_{2K'}$ | $n_{2\cdot}$ |
| . | . | . | ... | . | . |
| $P_K$ | $n_{K1}$ | $n_{K2}$ | ... | $n_{KK'}$ | $n_{K\cdot}$ |
| $\sum$ | $n_{\cdot 1}$ | $n_{\cdot 2}$ | ... | $n_{\cdot K'}$ | $n$ |

From the contingency matrix it is possible to define for each obtained cluster $C_j$ and for each gold case cluster $P_i$ the following two measures (Rosenberg and Hirschberg, 2007):

- **Homogeneity** $Hom(C_j)$**:** a clustering must assign only those data-points that are members of a single class to a single cluster. It can be calculated as:

$$Hom(C_j) = \frac{1}{n_{\cdot j}} \max_j (n_{ij}) \tag{6}$$

- **Completeness** $Com(P_i)$**:** a clustering must assign all of those data-points that are members of a single class to a single cluster. Completeness is symmetrical to Homogeneity.

$$Com(P_i) = \frac{1}{n_{i\cdot}} \max_i (n_{ij}) \tag{7}$$

These two measures are both needed to characterize the goodness of the clustering partition, taking into account two complementary aspects. Using them, we defined a new measure for the whole dataset partition named as *Clustering Goodness* (*CG*) defined as the weighted mean of the *Hom* and *Com* (see eq. 8). The weighting is necessary because the goodness of cluster solution is related even to the correct choice of cluster number and not only to *Hom* and *Com*. In other words, if clusters number is close to documents number, the $mean(Hom)$ value tends towards one; on the other hand, if the clustering solution is made by only one cluster, the $mean(Com)$ tends towards one. This extreme cases demonstrate that an arithmetic mean of these measures does not capture clustering goodness in every case.

$$CG(C) = \frac{1}{K+K'} \sum_{i=1}^{K} \sum_{j=1}^{K'} \alpha \cdot Com(P_i) + (1-\alpha) \cdot Hom(C_j) \tag{8}$$

The $\alpha$ value must balance the negative effects previously described, taking into account the cluster number in function of the gold-case cluster number. So we defined $\alpha$ as:

$$\alpha = \begin{cases} \frac{1}{2 \cdot K'}, & \text{if } K \leq K' \\ \frac{1}{2 \cdot (n-K')}, & \text{otherwise} \end{cases} \tag{9}$$

In this way, the value of equation 8 varies in range $(0, 1]$ and a perfect clustering solution has a *CG* value equal to 1, meaning that the clustering is identical to gold case partition, but the $\alpha$ value as defined in 9 can weight the importance of *Hom* and *Com* in function of optimal cluster number too. We used the *CG* in the experimental assessment in Section 5, showing the effectiveness of the proposed methodology.

## 4 SYSTEM ARCHITECTURE

The system architecture is divided into six different blocks, as shown in Figure 3.

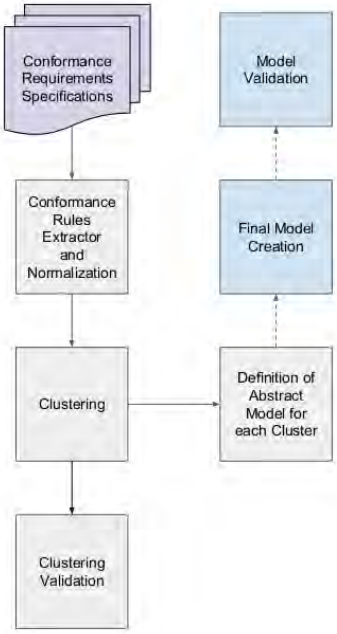



Figure 3: Main System Architecture. In purple, the input data; in grey, the blocks that are evaluated in this paper and in light-blue the blocks that will be considered as future works.

The first block consists in a pre-processing stage where the input specification document is converted into a more structured file, extracting and normalizing the conformance rules from the text. In the second block, a vector space model is created, extracting the features from the text of conformance rules and the iterative clustering technique, previously described, is applied, to obtain the group of rules that respect the same pattern. In the third block, a clustering evaluation is made considering hand-made gold cases. In the fourth block, an implementation of an abstract rule for each cluster is defined. In the fifth block we plan to create a module that implements each conformance rule according with its own abstract rule and, finally, in the last block is planned to evaluate the correctness of the *final* model obtained using hand-made gold cases.

The whole pipeline has been implemented in *Knime* environment (Berthold et al., 2007), an open platform for machine learning that natively supports external scripts in the most common language (C, Python, Java, R, Matlab, Weka and other). Using Knime is possible to integrate many tools in a single environment and design an optimized pipeline. In Figure 4 is shown the workflow implemented for the experiments.

The following subsections 4.1, 4.3, 4.4 describe the details of each block and the tools used to realize the system.

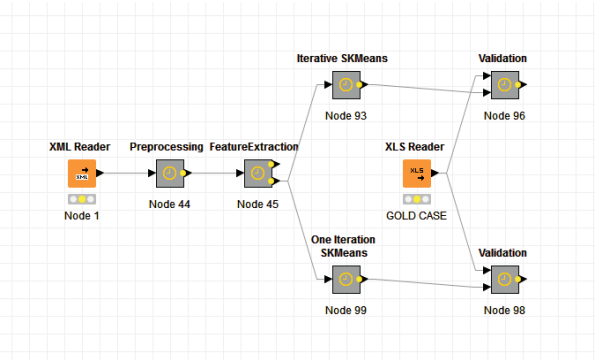



Figure 4: Example of Knime workflow. The grey blocks represent a *metanode* that is a group of nodes which perform complex operations.

## 4.1 Conformance Rule Extraction and Normalization

The first task of the system is the extraction of rules text and its normalization, obtained after an analysis of documents and conformance text structure. A specification document is often in pdf file format and the first operation needed to perform any kind of processing is the conversion in a plain-text *UTF-8* file.

We converted the conformance rule documents used in our experiment using *pdftotxt*[1], an open source command-line utility, included by default within many Linux distributions. The text file outputted from *pdftotxt* preserves most of the indentation and the structure of the original pdf file and this is really important for the subsequent task of the system. In fact, we need to extract only the natural language text where each rule is defined and stated. This task has been accomplished writing some Python scripts that, using the regular patterns of the text, extract the index of the document, the paragraph names and the rule texts. In our case, for example, all the conformance rules have a well-defined pattern: their definitions start with a tag (CONF-xx-yy) and end with another rule tag (CONF-xx-zz) or a dot sign. The next Figure 5 shows an example of the regular pattern from the original conformance rules file.

Different conformance documents can be processed only analysing their specific patterns. Python scripts have in input the start and end rule delimiter, but they can be easily modified to be applied to different and more complex patterns. The scripts perform a normalization phase too, deleting punctuation, the not-standard characters, symbols and stop words (a list of Italian language stop words is provided by Lucene[2]). Then, using regular expressions,

[1] http://www.foolabs.com/xpdf/home.html

[2] https://lucene.apache.org/core/4_4_0/analyzers-common/org/apache/lucene/analysis/it/ItalianAnalyzer.html

DEVE essere presente un elemento *patient/name* contenente nome e cognome del paziente (vedi § 2.3 - Persone ed Organizzazioni). Non può essere utilizzato il nullFlavor per indicare l'indisponibilità del dato.

**CONF-14:** L'elemento *patientRole/patient* **DEVE** contenere l'elemento *patient/administrativeGenderCode* (sesso).

**CONF-15:** L'elemento *patientRole/patient* **PUÒ** contenere l'elemento *patient/birthTime* (data di nascita).

**CONF-16:** L'elemento *patientRole/patient* **PUÒ** contenere l'elemento *patient/birthPlace/place/addr/censusTract* che riporta il codice ISTAT del luogo di nascita dell'assistito.

Per i dettagli relativi agli elementi *patient/name patient/administrativeGenderCode* e *patient/birthTime* si rimanda a Rif 10.

Figure 5: An example of input document used for experiment assessment. It is possible to observe the regular pattern of conformance rules to be extracted. In addition, each rules lies on a grey background.

we replace *Logical Observation Identifiers Names and Codes* (LOINC[3]), TemplateId codes, Paragraphs and Key Names with a generic identifier (i.e. LOINC 33882 – 2 is substituted with the word *LOINC*). We did this further normalization to reduce the noise induced by different terminology associated to the same concept, obtaining a better clustering results.

The output of this module is an xml file, whose structure is depicted in Figure 6. As shown, the body of CONF tags contains only the normalized text of each conformance rule. The used tags are the following:

- *documento*: the xml root, its body contains the document title and all the paragraphs will be its children;
- *paragrafo*: it contains the paragraph name in its body and the paragraph number as *id* attribute. All the CONF associated to it are its own children;
- *CONF*: it contains in its body the normalized text of the original conformance rule. Its attributes are: i) *num* in which is indicated the rule number and ii) *par* that indicates the paragraph number.

```
<paragrafo id="2.4.2.1.3">
    2.4.2.1.3 Paziente (Human Patient)
    <CONF num="12" par="2.4.2.1.3">-12: L'elemento paragrafo DEVE contenere almeno
    un elemento id con §chiave§ valorizzato a §template§
    ed in cui nell'attributo §chiave§ è riportato il Codice Fiscale del
    soggetto, oppure con §chiave§ valorizzato a "[OID ROOT STP REGIONALI]" ed in
    cui nell'attributo §chiave§ è riportato il Codice STP, oppure con
    §chiave§ valorizzato a §template§ ed in cui
    nell'attributo §chiave§ è riportato il Numero di Identificazione Personale TEAM.
    </CONF>
    <CONF num="13" par="2.4.2.1.3">-13: Il documento DEVE contenere l'elemento
    paragrafo DEVE essere presente un elemento §context§
    contenente nome e cognome del paziente (vedi paragrafo 2.
    3 Persone ed Organizzazioni). Non può essere utilizzato il nullFlavor per
    indicare l'indisponibilità del dato.</CONF>
    <CONF num="14" par="2.4.2.1.3">-14: L'elemento
    paragrafo DEVE contenere l'elemento
    §context§ (sesso).</CONF>
    <CONF num="15" par="2.4.2.1.3">-15: L'elemento paragrafo PUÒ
    contenere l'elemento paragrafo (data di nascita).</CONF>
    <CONF num="16" par="2.4.2.1.3">-16: L'elemento
    paragrafo PUÒ contenere l'elemento
    §context§ che riporta il codice ISTAT del luogo
    di nascita dell'assistito.</CONF>
</paragrafo>
```

Figure 6: Part of the output xml file obtained from text extraction and normalization module.

[3]http://loinc.org/

## 4.2 Feature Selection

The input to machine learning applications is represented through a *Vector Space Model* (VSM). In VSM a vector is associated to each sample (in this case the Conformance Rule) in which the elements of the vector correspond to the feature values.

Vectors of size $M$ correspond to points in an $M$-dimensional space; the main hypothesis underlying the VSM is that similar objects are represented by points which are closed in the $M$-dimensional space. Achieving optimal results with a machine learning technique based on VSM is strictly related to the correct choice of feature space (Amato et al., 2013).

In our case, the entity to be clustered are the conformance rules in natural language text, identified by their name. The rules in a VSM are mapped as $n$-grams of words. The correct selection of the $n$-gram size, namely the length of $n$, is both language and topic dependant (Cavnar and Trenkle, 1994) and so there is not an absolute rule (Eder, 2011). In our case we selected all $n$-grams with $n$ ranging from 2 to 6, observing the highest *MAS* (see equation 5) obtained in different clustering experiments, varying both $n$ and the number of $n$-grams together. The high value of $n$ obtained (often only uni-grams, bi-grams and tri-grams are used in literature) can be explained by the repetitive structure of the patterns in the description of a rule. We extract the features using internal KNIME modules.

The VSM obtained can be represented by a high dimensional sparse matrix. To reduce the noise caused by not discriminant features and consequently the space dimension, improving clustering performance and providing a faster computation, we applied Principal Component Analysis (PCA) as feature reduction method. The PCA has been implemented through the Cran R built in function *prcomp*, a really fast and accurate PCA algorithm. We set the selection of information content of PCA at 96%: this value has been obtained observing the higher mean silhouette value within all clustering experiments. The feature reduction is performed at each step of the iterative cluster module, described in the next Section 4.3, reducing each time the number of extracted features.

The use of n-grams directly extracted from the dataset makes the whole process totally language independent; the same methodology can be applied on conformance rules in any language and even to mixed languages, or medical slang documents. Changing the input dataset affects only the scripts for the normalization and rule extraction, that must be slightly modified as described in previous Section 4.1, but none of the other modules, included the feature extraction one.

### 4.3 Iterative Clustering

As described in Section 3, to group the conformance rules we applied iteratively a spherical *K*-means algorithm, selecting at each step the best solution according to *MAS* (equation 5), a cluster internal measure based on silhouette. After applying PCA feature reduction, we used the Cran R *skmeans* package (Hornik et al., 2012) with CLUTO algorithm (Karypis, 2002) to iteratively calculate spherical *K*-means with a cluster number range between 2 and the total number of rules, as described in Section 3.2. To speed up the iterative clustering process we used the *doParallel* Cran R package (Weston and Analytics, 2014), running more cluster processes in parallel.

### 4.4 Abstract Model Definition

The last implemented module performs the abstract model definition. At this aim we use a functionality of the standard Schematron that allows to define abstract patterns. In this way, it is possible to implement for each obtained cluster only one abstract model, obtaining a reduction of the complexity evaluable as:

$$\Delta(\text{Complexity}) = \left(1 - \frac{\text{Cluster Number}}{\text{Conf. Rules Number}}\right) \cdot 100 \quad (10)$$

The Figure 7 represents the conceptual schema for the creation of a *Final Implemented Rule* starting from a *Clustered Conformance Rule* and an *Abstract Pattern Template*.

An abstract pattern template is a way to generalize a class of possible instances of conformance rules and, like the concept of *Abstract Class* in the Object Oriented paradigm, it is possible to instantiate a specific *Final Implemented Rule* starting from it. The

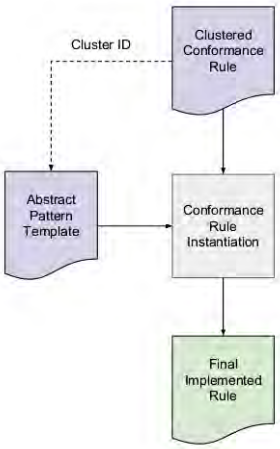



Figure 7: Main Schema of the Conformance Rule Implementation starting from an Abstract Pattern Template and a Clustered Conformance Rule.

```
<sch:pattern abstract="true" id="Cluster_ID">
    <sch:rule id="$Context">
        <sch:assert test= "assertion($par1, ..., $parN)> <!— es: count($par1) = 1
            and count($parN) lt6; 2 —>
        Error/Warning: $body"/>
        </sch:assert>
    </sch:rule>
</sch:pattern>
```

Figure 8: Main Schema of the Conformance Rule Implementation starting from an Abstract Cluster Template.

Figure 8 shows a generic example written according to the standard Schematron where, considering a cluster partition identified by *Cluster_ID*, it is created an abstract pattern with all the parameters defined as generic variables (ex. $Context, $par1, $par2, etc.). In the example we also defined a generic function $assertion(\cdot)$ to obtain complex tests using the defined variables.

In Figure 9, the abstract pattern is used to instantiate a specific conformance rule that belongs to that cluster. In this case the instantiation consists to declare the abstract pattern to use and to specify each parameter involved.

```
<sch:pattern is-a="Cluster_ID" id="CONF_XXX">
    <sch:param name="$Context" value="value of the context""/> <!—es:cda:patient—>
    <sch:param name= "$par1" value= "value of par1"/> <!— es: cda:code —>
    <sch:param name= "$parN" value= "value of parN"/> <!— es: cda:codeName —>
    <sch:param name="$body" value="body content "/>
</sch:pattern>
```

Figure 9: Main Schema of the Conformance Rule Implementation starting from an Abstract Cluster Template.

At the moment the abstract pattern template implementation is manual and involves a human processing. As future work, we are planning to automatize this task, using NLP tools. In details, the use of a Part of Speech (PoS) tagger and of a dependency parser will automatically identify the subject, the main verb and its objects of each cluster member. In addition, a dedicated entity extraction can help to classify the object types. Then, a rule based system can build a pattern for each cluster.

## 5 EXPERIMENTAL RESULTS

To verify the effectiveness of described approach we will show the application of the proposed methodology on two case studies, namely the Italian localization of specification of the conformance rules of: i) Patient Summary[4](in Italian *Profilo Sanitario Sintetico*, PSS) and ii) Hospital Discharge Letter[5](in Italian *Lettera di Dimissione Ospedaliera*, LDO). The conformance requirements and specifications are part

[4]Patient Summary: http://www.hl7italia.it/sites/default/files/Hl7/docs/public/HL7Italia-IG_CDA2_PSS-v1.2-S.pdf

[5]Hospital Discharge Letter: http://www.hl7italia.it/webfm_send/1709

Table 2: Gold case cluster number of each dataset.

| Specification | Rules number | Gold case Cluster number |
|---|---|---|
| **LDO** | 104 | 42 |
| **PSS** | 259 | 129 |
| **PSS+LDO** | 363 | 159 |

of HL7-Italia (see Section 1 for more details). PSS and LDO are both conformance requirements and specification documents written in semi-structured natural language text. PSS contains 259 conformance rules, while LDO a total number of 104. To extend the experimental assessment, we have applied our methodology even to the sum of the rules from both documents, clustering a new data set with a total of 363 rules, named PSS+LDO. It could be useful in real application group together similar conformance rules documents, identifying the rules with the same patterns from different documents.

The assessment is based on gold cases, formed by the ideal grouping of the conformance rules of each dataset belonging to the same pattern. The goodness of the cluster results have been measured through the *CG* (equation 8) applied on those gold cases. Each gold case has been manually built by the software developers who previously implemented the whole conformance rule validation schema: they well know the rules text and their patterns and so they produced a reliable gold cases for each dataset used. The number of conformance rules grouped in each gold case is shown in next Table 2.

We have compared the results obtained with our approach, namely Iterative Spherical K-Means (IT-SKM), with the ones obtained using a One Iteration Spherical K-Means (1-SKM) method. In this case, only one step of iteration process is performed, choosing the cluster number of the partition with the MAS function (equation 5), without selecting the elements to be clustered in the following steps.

In Table 3 is shown the effectiveness of using IT-SKM for evaluating the optimal number of clusters through the synthetic external measure *CG* (eq. 8), previously defined in Section 3.3. We compared the IT-SKM results with the 1-SKM results through the *CG* measure for PSS, LDO and PSS+LDO cases. In all experiments the best results have been obtained with iterative approach IT-SKM. It is even worth noting that cluster number obtained with IT-SKM is really close to the gold case.

To better understand and explain the results of our experiments, we show in Figure 10 the $Hom$ (in red) and *Com* (in blue) percentage value distribution for 1-SKM and IT-SKM for all data sets. In details, the figures depict the cluster distribution whose $Hom$ and *Com* have a certain value. All 1-SKM experiments have an high number of clusters whose $Hom$ is high, due to the fact that the number of clusters obtained is close to the total conformance rules number and many clusters have only one element. So the high value of $Hom$ is caused simply by cluster formed by only one element, not by a good cluster solution. On the other side, the number of clusters with an high *Com* value is only a little fraction of the whole partition, suggesting a bad clustering.

Instead, IT-SKM experiments show in all cases a very high fraction of clusters with both $Hom$ and *Com* equal to 100%. A perfect solution (identical to gold case) has $Hom$ and *Com* equal to 100% for each cluster. The results in Figure 10 for IT-SKM show that this condition is verified for an high number of clusters, demonstrating the effectiveness of the proposed methodology. In addition, the figure confirms that *CG* measure follows the correct behaviour and it is an useful external measure.

# 6 CONCLUSION AND FUTURE WORK

In this paper we proposed a novel approach to reduce the complexity of the definition and implementation of a medical document validation model.

We defined an architecture to automatically produce a software specification starting from a set of conformance rules in semi-structured natural language format. At this aim, we presented an innovative cluster approach that automatically evaluates the optimal number of groups using an iterative method based on internal cluster measures evaluation.

The effectiveness of the proposed approach is evaluated on two case studies: i) Patient Summary (*Profilo Sanitario Sintetico*) and ii) Hospital Discharge Letter (*Lettera di Dimissione Ospedaliera*) for the Italian localization specification of the conformance rules.

As future works we are planning to realize the remaining blocks of the architecture depicted in the Figure 3 and, in particular, the *Final Model Creation* and *Model Validation* (the blocks have light-blue background in the Figure). Furthermore, we are considering to automatize the creation of the abstract pattern template starting from a cluster, with the support of natural language tools. At least, we are also investigating more deeply on other unsupervised methods to automatically grouping the conformance rule and in particular on deep-learning approaches.

Table 3: Results. The best results are highlighted in bold.

| Specification | Method | Mean(CG) | Mean(COM) | Mean(HOM) | #Cluster | #Gold | #Conf | Δ(Complexity) |
|---|---|---|---|---|---|---|---|---|
| LDO | Iterative | **74.21**% | 70.00% | 76.77% | 46 | 42 | 104 | 55.77% |
| | One Iteration | 70.85% | 53.85% | 98.72% | 77 | | | 25.96% |
| PSS | Iterative | **67.17**% | 75.88% | 63.53% | 108 | 129 | 259 | 58.30% |
| | One Iteration | 65.36% | 58.37% | 95.48% | 211 | | | 18.53% |
| PSS+LDO | Iterative | **66.88**% | 64.90% | 68.16% | 167 | 159 | 363 | 53.99% |
| | One Iteration | 60.34% | 50.00% | 98.43% | 313 | | | 13.77% |

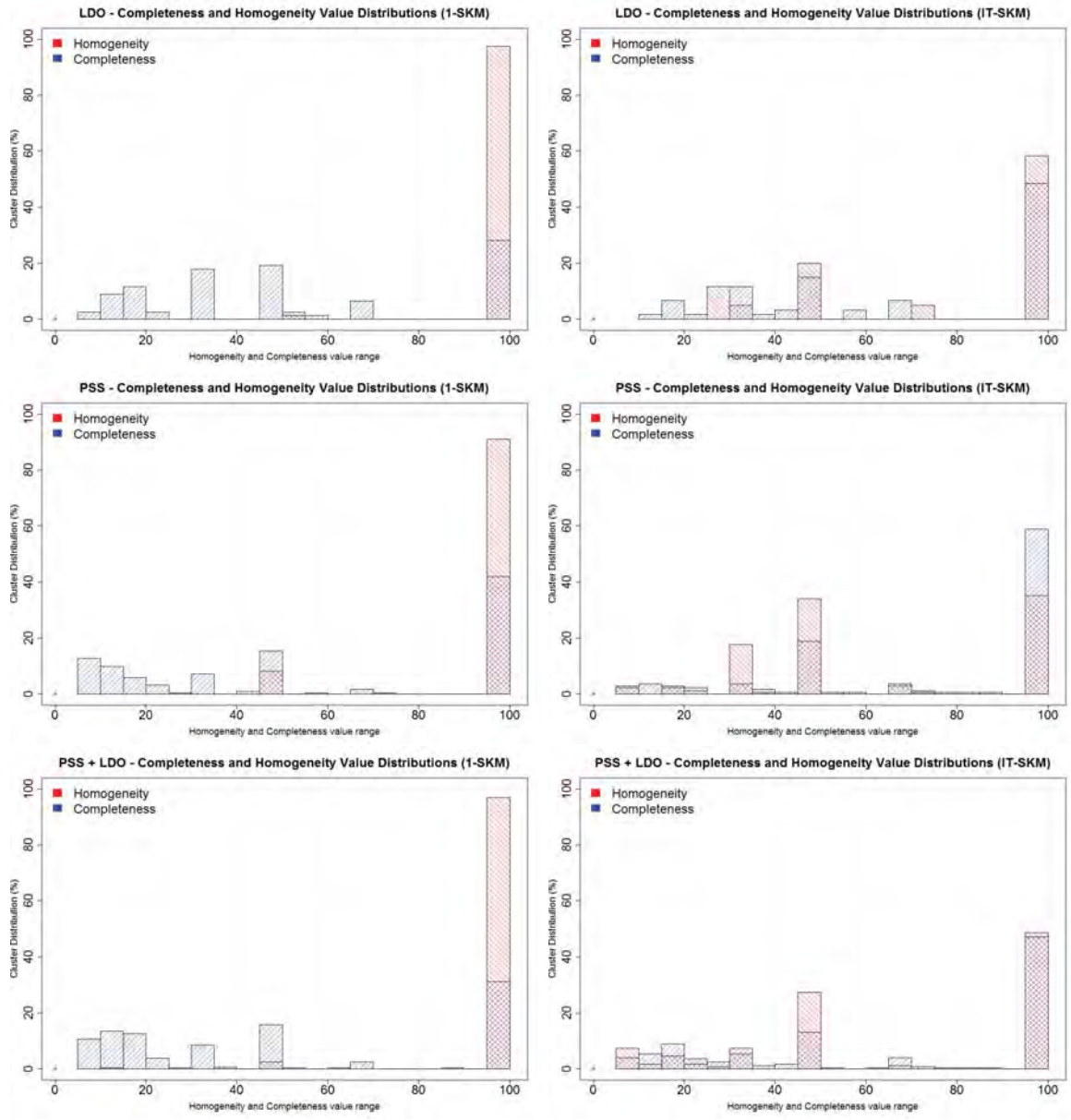



Figure 10: *Hom* and *Com* value distributions for all experimental assessment.

## ACKNOWLEDGEMENTS

This work has been partially supported by the Italian project *Realization of services of the national infrastructure for interoperability for Electronic Health Records*, a Convention between the Agency for Digital Italy and the Italian National Research Council.

## REFERENCES

Alicante, A., Corazza, A., Isgrò, F., and Silvestri, S. (2016a). Semantic cluster labeling for medical relations. *Innovation in Medicine and Healthcare 2016*, 60:183–193.

Alicante, A., Corazza, A., Isgrò, F., and Silvestri, S. (2016b). Unsupervised entity and relation extraction from clinical records in Italian. *Computers in Biology and Medicine*, 72:263–275.

Amato, F., Gargiulo, F., Mazzeo, A., Romano, S., and Sansone, C. (2013). Combining syntactic and semantic vector space models in the health domain by using a clustering ensemble. In *HEALTHINF 2013 - Proceedings of the International Conference on Health Informatics*, pages 382–385.

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., and Wiswedel, B. (2007). KNIME: The Konstanz Information Miner. In *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. Springer.

Boscá, D., Maldonado, J. A., Moner, D., and Robles, M. (2015). Automatic generation of computable implementation guides from clinical information models. *Journal of Biomedical Informatics*, 55:143–152.

Boufahja, A., Poiseau, E., Thomazon, G., and Bergé, A.-G. (2015). Model-based analysis of hl7 cda r2 conformance and requirements coverage. *EJBI*, 11(2).

Cao, L., Chua, K. S., Chong, W., Lee, H., and Gu, Q. (2003). A comparison of pca, kpca and ica for dimensionality reduction in support vector machine. *Neurocomputing*, 55(1):321–336.

Cavnar, W. B. and Trenkle, J. M. (1994). N-gram-based text categorization. *Ann Arbor MI*, 48113(2):161–175.

Ciampi, M., Esposito, A., Guarasci, R., and Pietro, G. D. (2016). Towards interoperability of ehr systems: The case of italy. In *Proceedings of the International Conference on Information and Communication Technologies for Ageing Well and e-Health - Volume 1: ICT4AWE,*, pages 133–138.

Dhillon, I. S., Guan, Y., and Kogan, J. (2002). Iterative clustering of high dimensional text data augmented by local search. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pages 131–138. IEEE.

Eder, M. (2011). Style-markers in authorship attribution a cross-language study of the authorial fingerprint. *Studies in Polish Linguistics*, 6(1):99–114.

Gargiulo, F., Fontanella, M., and Ciampi, M. (2016). Validazione di documenti sanitari strutturati in hl7 cda rel. 2.0 con schemi schematron. Technical report, Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR) del Consiglio Nazionale delle Ricerche (CNR).

Halkidi, M. and Vazirgiannis, M. (2001). Clustering validity assessment: Finding the optimal partitioning of a data set. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 187–194. IEEE.

Hamilton, J., Darr, T., Fernandes, R., Jones, D., and Morgan, J. (2015). Rule-based constraints for metadata validation and verification in a multi-vendor environment. In *International Telemetering Conference Proceedings*. International Foundation for Telemetering.

Handl, J., Knowles, J., and Kell, D. B. (2005). Computational cluster validation in post-genomic data analysis. *Bioinformatics*, 21(15):3201–3212.

Hornik, K., Feinerer, I., Kober, M., and Buchta, C. (2012). Spherical *k*-means clustering. *Journal of Statistical Software*, 50(10):1–22.

Jafarpour, B., Abidi, S. R., and Abidi, S. S. R. (2016). Exploiting semantic web technologies to develop owl-based clinical practice guideline execution engines. *IEEE Journal of Biomedical and Health Informatics*, 20(1):388–398.

Jelliffe, R. (2001). The schematron assertion language 1.5. *Academia Sinica Computing Center*.

Karypis, G. (2002). Cluto-a clustering toolkit. Technical report, DTIC Document.

Kaufman, L. and Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons.

Liu, Y., Li, Z., Xiong, H., Gao, X., and Wu, J. (2010). Understanding of internal clustering validation measures. In *2010 IEEE International Conference on Data Mining*, pages 911–916. IEEE.

Pollard, K. S. and Van Der Laan, M. J. (2002). A method to identify significant clusters in gene expression data. In *Proceedings of SCI World Multiconference on Systemics, Cybernetics and Informatics*, pages 318–325.

Rendón, E., Abundez, I., Arizmendi, A., and Quiroz, E. (2011). Internal versus external cluster validation indexes. *International Journal of computers and communications*, 5(1):27–34.

Rosenberg, A. and Hirschberg, J. (2007). V-measure: A conditional entropy-based external cluster evaluation measure. In *EMNLP-CoNLL*, volume 7, pages 410–420.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Weston, S. and Analytics, R. (2014). *doParallel: Foreach parallel adaptor for the parallel package*. R package version 1.0.8.

Wu, J., Xiong, H., and Chen, J. (2009). Adapting the right measures for k-means clustering. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 877–886. ACM.

Zhong, S. (2005). Efficient online spherical K-means clustering. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, volume 5, pages 3180–3185.

# A Statistical Analysis for the Evaluation of the Use of Wearable and Wireless Sensors for Fall Risk Reduction

Giovanna Sannino, Ivanoe De Falco and Giuseppe De Pietro
*Institute on High-Performance Computing and Networking (ICAR), National Research Council of Italy (CNR), Naples, Italy*
{*giovanna.sannino, ivanoe.defalco, giuseppe.depietro*}*@icar.cnr.it*

Keywords: Falling Risk, Physical Activity, Body Mass Index, Statistical Analysis, Correlation, Wearable Sensors, Mobile Devices.

Abstract: The aim of this study is to investigate the correlation between, on the one hand, personal and life-style indicators and, on the other hand, the risk of falling. As indicators we consider here for each subject age, body mass index, and information about physical activity habits, while a subject's risk of falling is estimated by the Mini-BES test score. Three different groups of subjects are taken into account, namely healthy, suffering from metabolic diseases and suffering from cardiovascular diseases. Firstly, we aim at finding explicit linear correlations for any pair of parameters. Secondly, we wish to pay attention to whether or not these correlations change as the health state of the subjects does. The final goal is to move the first steps towards the design of a system composed by wearable sensors, a mobile device, and an app that would be able to help people in improving their life-style so as to decrease their falling risk.

## 1 INTRODUCTION

Falls have been shown to result in increased morbidity and are considered the cause of the yearly loss of more than 17 million years worldwide that are spent in disability (World Health Organization . Ageing and Life Course Unit, 2008). Many studies have been dedicated to fall detection, as e.g. (Sannino et al., 2015). Identifying individuals with a high fall risk is often a significant part of prevention programs. The assessment of the risk of falling is a major and effective prevention tool that allows identifying intrinsic and extrinsic risk factors. These latter help determine the most suitable interventions, thus reducing, or in some cases even eliminating, falls.

The goal of this study is threefold.

Firstly, we aim to carry out a statistical analysis to inquire into the existence of clear correlations between, on the one hand, some of the most widely considered body parameters, as age and Body Mass Index, and physical activity tests, and, on the other hand, the risk of falling, represented through the score of the Mini-Balance Evaluation Systems (Mini-BES) test (Franchignoni et al., 2010).

Secondly, we wish to diversify our statistical analysis, so as investigate whether or not these correlations change when healthy or unhealthy subjects are considered. We wish to take into account here two different wide classes of diseases. The first class contains metabolic problems such as hypo- and hyper-thyroidism, hypo- and hyper-glycemia, and so on. Approximately 34% of the worlds adult population has the cluster of risk factors that is metabolic syndrome (Mozumdar and Liguori, 2011). The second class, instead, makes reference to diabetes, hypo- and hyper-tension, vascular and heart-related problems. Cardiovascular diseases (CVD) are responsible for 30% of all deaths (17.5 million) (World Health Organization and others, 2005).

Thirdly, we wish to move the first steps towards the opening of a path to the use of wearable sensors and mobile devices for the on-line monitoring and the real-time evaluation of a subject's falling risk through the consideration of the above found relationships.

This latter goal would make fall risk assessment much easier, because subjects would not need to undergo the classical Mini-BES test, rather they could estimate it at home in their everyday life by simply using a small set of wearable sensors. Namely, a sensor could estimate the Body Mass Index (BMI), whereas a second could keep track of the subject's physical activity. Based on the measured data, an app on the subject's mobile device could act as a kind of an advisor, by providing them with a view of their general health state, and with useful suggestions as well. Moreover, subjects with a potentially moderate-to-high falling



Sannino, G., Falco, I. and Pietro, G.
A Statistical Analysis for the Evaluation of the Use of Wearable and Wireless Sensors for Fall Risk Reduction.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 508-516
ISBN: 978-989-758-213-4

risk assessment would be adviced to meet a doctor so as to possibly undergo a real test. This approach could lead to easily performing a kind of a 'mass screening' with reference to the risk of falling. To fulfill the two first above statistical goals, in this paper we will make use of a set of personal and life-style information contained in a real-world database making reference to to the risk of falling. Then, we will describe the body system we propose, based on some wearable senors, a mobile device, and an app.

This paper is organized as follows. Section 2 reports on the related work on finding correlations between personal parameters and falling risk. Information about the database is provided in Section 3. The statistical analysis is shown and discussed in Section 4. In Section 5 some considerations are given on the use of the results of the statistical analysis for the design of a monitoring system based on wearable sensors. Finally, our conclusions and future works follow in Section 6.

## 2 RELATED WORKS

One of the first papers trying to find correlations between personal parameters and falling risk was (Gardner et al., 2000). In it, the objective was to assess the effectiveness of exercise programs in preventing falls (and/or lowering the risk of falls and fall related injuries) in older people. Their conclusion was that exercise is effective in lowering falls risk in selected groups and should form part of falls prevention programs.

In (Hue et al., 2007) the aim of the study was to determine the contribution of body weight to predict balance stability. Their experiments suggest that body weight may be an important risk factor for falling.

In (Faulkner et al., 2009) the authors examined potential independent effects of lifestyle on fall risk. Not smoking and going outdoors frequently or infrequently were independently associated with more falls, indicating lifestyle-related behavioral and environmental risk factors are important causes of falls in older women.

Although not directly tied to fall risk, the paper (Shekharappa et al., 2011) dealt with similar ideas, in fact the aim was to find a correlation between body mass index and cardiovascular parameters in obese and non-obese in different age groups. The results showed a statistically significant increase in heart rate, systolic blood pressure and diastolic blood pressure in obese subjects when compared to non-obese in all age group. Moreover, there was a positive correlation between body mass index and heart rate, systolic blood pressure, diastolic blood pressure, mean blood pressure and pulse blood pressure.

The relationship between Body Mass Index and stability has been investigated in (Ku et al., 2012). Namely, the aims of that study was to examine the impact of BMI and gender on static postural control. Their conclusion was that BMI do have an impact on postural control during both bipedic stance and unipedic stance.

The effect of the type, level and amount of physical activity in falls and fall-related injuries was examined in (Pereira et al., 2014). Their conclusions were that being active, especially sufficiently active, reduces fall-related injuries by decreasing falls and by safeguarding against severe injuries when falls occur.

A study was conducted in (Shahudin et al., 2016) to investigate the effects of age on physical activity level, strength and balance towards fall risk index (FRI) among women, as well as identifying the main contributing factors towards FRI test performance. That study suggested that women aged 20–73 years were found to associate their FRIs mostly with age, followed by strength, balance, and lastly, physical activity.

## 3 THE DATABASE

To carry out our investigation, we have taken advantage of the Human Balance Evaluation database, collected at the Biomechanics and Motor Control Laboratory (BMCLab) of the Federal University of ABC, Sao Paulo, Brazil (http://demotu.org/datasets/balance/), and freely available in PhysioNet (Goldberger et al., 2000).

This database was collected while performing stabilography tests over a set of subjects. Each of those subjects had to perform standing tasks under four different conditions: by keeping their eyes opened or closed, and while standing on a rigid surface or on an unstable one. Each condition was tested three times, with the order of the conditions being randomized among subjects. A total of 1930 trials performed by 164 different subjects are given in this database. Each 1 minute recording is sampled at 100Hz and low pass filtered at 10Hz.

Moreover, and most importantly to us, the following qualitative tests were employed on each subject, and the replies/outcomes recorded in the database: Short Falls Efficacy Scale International (FES-I) (Kempen et al., 2008) (seven questions plus the score), the Short version of the International Physical Activity Questionnaire (IPAQ) (Craig et al., 2003) (eight questions plus the score), Trail Making Test

(four pieces of information), Mini Balance Evaluation Systems (Mini-BES) Tests (Franchignoni et al., 2010) (fourteen values plus the score). Furthermore, the subjects were also interviewed about some of their socio-cultural, demographic, and health information, including their age, medications, and illnesses.

Consequently, each database item contains 63 attributes. The database, apart from the raw data recordings, also includes a BDSinfo file that contains metadata describing the conditions of the stabilography trials, the information from the anamnesis, and the results of the qualitative evaluations. Because, as stated above, a subject has 12 files for the force platform data, there are 12 rows for each subject in this file. In these 12 rows, the only column that has rows with different values is the column identifying the trial (the file name). The content of all the other columns are simply repeated over the 12 rows. As result, the BDSdata file has the header plus 1930 rows and 64 columns. The complete list of the attributes can be found in (Santos and Duarte, 2016).

Starting from this database, we have conducted an analysis phase by creating a new database composed by 6 items for each of the 164 subjects. The parameters taken into account in our study are:

- $x_1$: age group
- $x_2$: Body Mass Index (BMI)
- $x_3$: IPAQ_1: minutes per week of vigorous physical activity according to the short IPAQ questionnaire
- $x_4$: IPAQ_2: minutes per week of moderate physical activity according to the short IPAQ questionnaire
- $x_5$: IPAQ_3: minutes per week of low physical activity according to the short IPAQ questionnaire
- $x_6$: the total score of the Mini-BES test

In the short IPAQ questionnaire used to create the Human Balance Evaluation database, the vigorous physical activities are defined as: heavy lifting, digging, aerobics, or fast bicycling. The moderate ones, instead are considered as: carrying light loads, bicycling at a regular pace, or doubles tennis. Finally, the low physical activities include: walking at work and at home, walking to travel from place to place, and any other walking that is done solely for recreation, sport, exercise or leisure.

As concerns the value for IPAQ_1 for a subject in our database, this is computed starting from the subject's answers to short IPAQ questions 1a (days per week of high-level physical activity) and 1b (hours per day of high-level physical activity through: $IPAQ\_1 = IPAQ\_1a \cdot IPAQ\_1b$. The same mechanism holds true for the computation of IPAQ_2 and IPAQ_3.

The value of the score for the Mini-BES test is computed through the answers of the subject to 14 questions, each of which can be assigned a value equal to 0, 1, or 2, the higher the better. Therefore, the value of the Mini-BES test score can range within 0 and 28, where a higher value means that the subject has a lower falling risk.

Moreover, we have divided the subjects in the database into three groups:

- *healthy*: they are the subjects with no disease at all. This has resulted in a number of 56 individuals;
- *metabolic diseases*: this group contains all the individuals who declared problems related to hyper - or hypo-thyroidism, hyper- or hypo-glicemia, and so on. This group contains 32 subjects;
- *cardiovascular diseases*: this group is composed by all the individuals with hyper- or hypo-tension, cardiovascular problems, or diabetes. There are 41 people in this group.

It should be pointed out that we excluded from the groups 48 subjects who were not healthy, yet they suffered from diseases other than those reported in the above two groups. As examples, some of them suffered from melanoma, breast cancer, hepatitis, Parkinson, arthrosis, asthma, dermatitis, rhinitis, gastritis, kidney stones, sickle cell anemia,tendinitis, and so on. Moreover, there are 13 people in the database who suffer from both endocrinological and cardiovascular diseases. These have been assigned to both groups.

# 4 STATISTICAL ANALYSIS

For each of the three groups of subjects described in the previous section we have performed a correlation analysis among the chosen database parameters. By doing so, we have been able to obtain the correlation value for each pair of parameters. Let's recall here that a correlation value between two parameters is in the range [-1.0, 1.0], where positive values represent direct correlations and negative values inverse correlations, and the higher the absolute value the stronger the correlation.

Moreover, for each of these pairs we have created a figure, in which we report the raw data, and have computed and drawn the best line for the linear regression that best fits the data, and have also reported the related R-squared ($R^2$) value. R-squared is a statistical measure of how close the data are to the fit-

Table 1: Correlation values between pairs of parameters for healthy subjects.

| | age | BMI | IPAQ_1 | IPAQ_2 | IPAQ_3 | Mini-BES |
|---|---|---|---|---|---|---|
| age | 1.00 | | | | | |
| BMI | 0.45 | 1.00 | | | | |
| IPAQ_1 | 0.56 | 0.29 | 1.00 | | | |
| IPAQ_2 | 0.30 | -0.08 | 0.20 | 1.00 | | |
| IPAQ_3 | -0.05 | -0.02 | -0.06 | 0.21 | 1.00 | |
| Mini-BES | -0.30 | -0.24 | -0.14 | -0.06 | 0.05 | 1.00 |

ted regression line. A value of 0 for $R^2$ indicates that the model explains none of the variability of the response data around its mean, whereas an $R^2$ of 1 indicates that the regression line perfectly fits the data. These regression lines and their $R^2$ values are very useful because from them fruitful information can be obtained. In the next three subsections all these findings are shown for the three groups, respectively.

## 4.1 Healthy Subjects

Table 1 reports the correlation values for all of the considered parameters.

In it the generic cell $(i, j)$ contains the correlation value between the pair of parameters $i$ and $j$. Very high values ($\geq 0.50$) and very low ones ($\leq -0.50$) are shown in dark grey. Moderate values, lying in [-0.49, 0.30] and [0.30, 0.49], are highlighted in light grey. All the pairs in which the Mini-BES test score appears have been considered for further analysis. For each of them the corresponding figures are shown, which contain information about the linear regression too.

A first remark that can be made concerns the pairs of parameters for which the correlation values are high or moderate, i.e. those for which the values in the tables are shown in dark grey or light grey, respectively.

As concerns the healthy subjects, the only strong correlation is between the age and the IPAQ_1, i.e. the vigorous activity, and it a positive correlation. Basically, this says that as long as healthy people get old, they go on exercising vigorously. Also quite high is the moderate direct correlation between age and BMI, meaning that the older a healthy subject, the more obese she/he is. Moreover, a moderate direct correlation is shown between age and IPAQ_2, similar to that between age and IPAQ_1, but with reference to moderate physical activities. Furthermore, a moderate inverse correlation exists also between age and Mini-BES. This suggests that for healthy subjects the higher the age the lower the value of the Mini-BES test score, hence the more probable the subject will be prone to falls.

Fig. 1 shows that this group is mainly composed by young adults. In fact the average age is 31.32 years

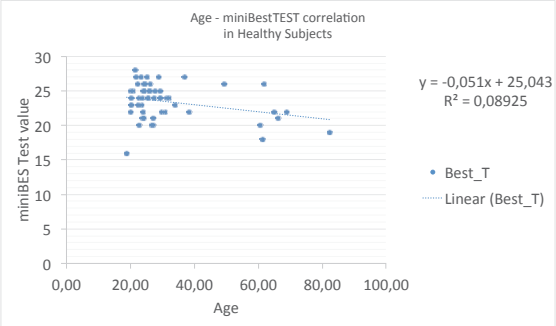



Figure 1: Analysis of parameters age and Mini-BES Test score for healthy subjects.

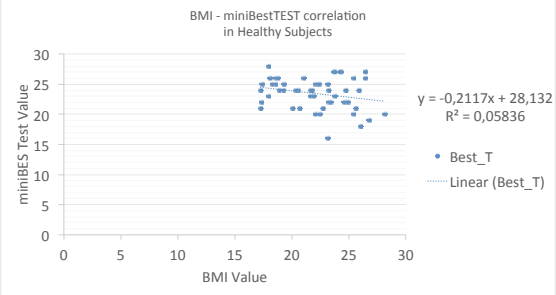



Figure 2: Analysis of parameters BMI and Mini-BES Test score for healthy subjects.

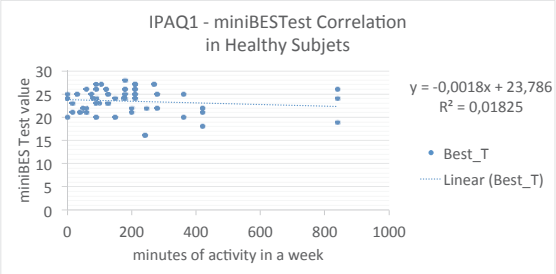



Figure 3: Analysis of parameters IPAQ_1 and Mini-BES Test score for healthy subjects.

$\pm 14.90$. Moreover, the average value for the Mini-BES test score is $23.44 \pm 2.54$, which is quite a high value suggesting that healthy people have a scarce fall risk.

A closer examination of Figures 1, 2, 3, 4, and 5 provides more precise information about the relation-

Table 2: Correlation values between pairs of parameters for metabolic subjects.

| | age | BMI | IPAQ_1 | IPAQ_2 | IPAQ_3 | Mini-BES |
|---|---|---|---|---|---|---|
| age | 1.00 | | | | | |
| BMI | 0.52 | 1.00 | | | | |
| IPAQ_1 | 0.04 | -0.17 | 1.00 | | | |
| IPAQ_2 | 0.13 | 0.08 | 0.09 | 1.00 | | |
| IPAQ_3 | 0.08 | 0.26 | -0.01 | 0.03 | 1.00 | |
| Mini-BES | -0.78 | -0.54 | -0.06 | -0.03 | -0.25 | 1.00 |

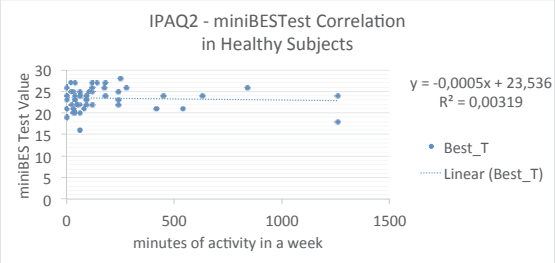



Figure 4: Analysis of parameters IPAQ_2 and Mini-BES Test score for healthy subjects.

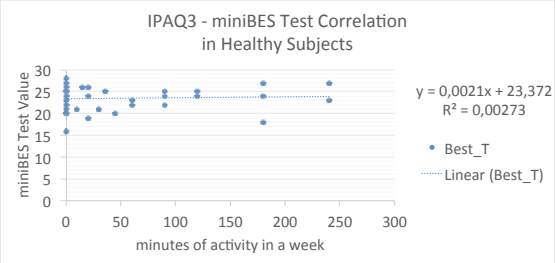



Figure 5: Analysis of parameters IPAQ_3 and Mini-BES Test score for healthy subjects.

ship between the Mini-BES test score and each of the other parameters considered in this study. Namely, the precise form of their relationship, under a linear hypothesis, is shown.

The slope of the line drawn in each figure provides intuitive visual understanding of the relationship: a down-bound line means an inverse linear relationship, an up-bound one a direct dependence, and the more inclined the line the higher the amount of this relationship. Consequently, lines that are almost horizontal imply a substantial independence between the two parameters. As an example, this is the case shown in Fig. 5.

## 4.2 Subjects with Metabolic Diseases

Table 2 reports the correlation values for all of the considered parameters.

As it has been for the healthy group, here too the generic cell $(i, j)$ contains the correlation value between the pair of parameters $i$ and $j$. The same convention used for that group applies also in this case to highlight some specific cells in the table. Similarly to the previous case, here too all the pairs in which the Mini-BES test score appears have been considered for further analysis. For each of them the corresponding figures are shown, which contain information about the linear regression too.

As far as the metabolic patients are taken into account, three parameter pairs have high correlation values, namely age-BMI, age-Mini-BES, and BMI-MiniBES. The first is a direct correlation, meaning that as the age increases so does BMI, as it is quite frequent in humans, be they healthy or suffering from some disease. Of higher interest for our purposes are the other two correlations. Age and Mini-BES test score are strongly and inversely correlated, which means that as the age of these diseased subjects increases the Mini-BES test score decreases, so older subjects suffering from metabolic diseases are more prone to falls. Moreover, also BMI and Mini-BES test score are strongly and inversely correlated, meaning that the more obese a metabolic subject, the higher probability she/he has of falling.

In this case, as Fig. 6 reveals, the age of this group is quite higher than that for healthy subjects. In fact, the average is 62.80 years $\pm$17.91. The average value for the Mini-BES test score for these subjects, instead, is $19.67 \pm 4.10$, i.e. about four points worse that that for healthy people.

By looking at Figures 6, 7, 8, 9, and 10, it can be visually understood that for people suffering from metabolic diseases changes in IPAQ_2 and IPAQ_3 almost do not affect the Mini-BES test score, while the opposite is true for the age, BMI, and IPAQ_1. In particular, the $R^2$ value for the correlation between age and Mini-BES test score is equal to 0.60908, so we are confident that the regression line well fits the data.

## 4.3 Subjects with Cardiovascular Diseases

Table 3 reports the correlation values for all of the considered parameters.

Also for this group, the generic cell $(i, j)$ in the table reports the correlation value between the pair of parameters $i$ and $j$. The cells in this table have

Table 3: Correlation values between pairs of parameters for cardiovascular subjects.

| | age | BMI | IPAQ_1 | IPAQ_2 | IPAQ_3 | Mini-BES |
|---|---|---|---|---|---|---|
| age | 1.00 | | | | | |
| BMI | -0.08 | 1.00 | | | | |
| IPAQ_1 | -0.10 | -0.29 | 1.00 | | | |
| IPAQ_2 | -0.34 | -0.18 | 0.15 | 1.00 | | |
| IPAQ_3 | -0.08 | 0.17 | -0.03 | 0.03 | 1.00 | |
| Mini-BES | -0.39 | 0.01 | -0.07 | 0.17 | 0.22 | 1.00 |

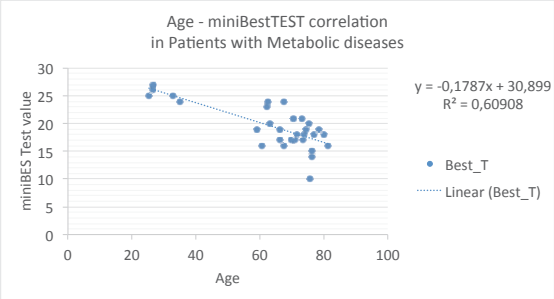



Figure 6: Analysis of parameters age and Mini-BES Test score for metabolic subjects.

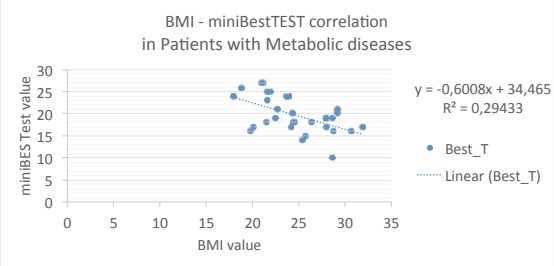



Figure 7: Analysis of parameters BMI and Mini-BES Test score for metabolic subjects.

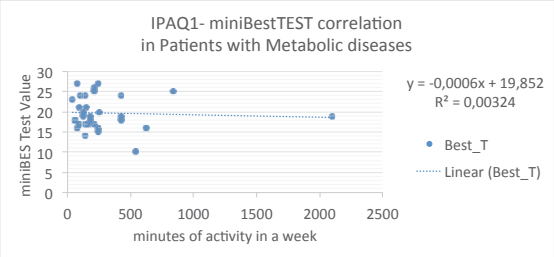



Figure 8: Analysis of parameters IPAQ_1 and Mini-BES Test score for metabolic subjects.

been highlighted by using the same convention as done for the two previous groups. Similarly to the two above described cases, also for this group all the pairs in which the Mini-BES test score appears have been considered for further analysis. For each of them the corresponding figures are shown, which contain information about the linear regression too.

Finally, when the cardiovascular subjects are considered, no correlation can be defined as strong, the

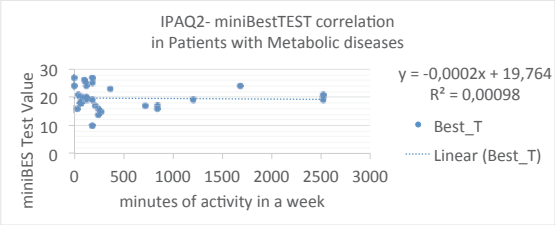



Figure 9: Analysis of parameters IPAQ_2 and Mini-BES Test score for metabolic subjects.

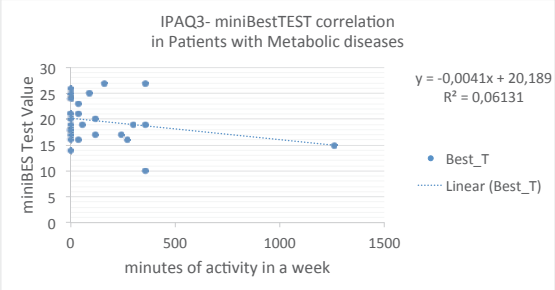



Figure 10: Analysis of parameters IPAQ_3 and Mini-BES Test score for metabolic subjects.

highest one being a moderate inverse correlation between age and Mini-BES test score. This is quite similar to that already seen for the metabolic subjects, although with a lower tie between the two parameters. Also, age and IPAQ_2 show a moderate inverse correlation, that is the opposite as that for healthy subjects: healthy people tend to exercise when they get older, whereas cardiovascular ones tend to not work out.

Also for the cardiovascular subjects, as it was for the metabolic ones, the average age is quite higher than that for the healthy subjects. In fact, as shown in Fig. 11, it is equal to $72.27 \pm 6.61$, which is higher than that of metabolic people too. As for the average score of the Mini-BES test for this group, it results to be equal to $18.02 \pm 3.81$, i.e. even worse that that of the metabolic subjects.

For this group the Figures 11, 12, 13, 14, and 15 do not show any particularly strong correlation, nor do they report any sufficiently high value for $R^2$, apart from, possibly, the case of age and Mini-Best test score.

In conclusion, the main result from this statisti-

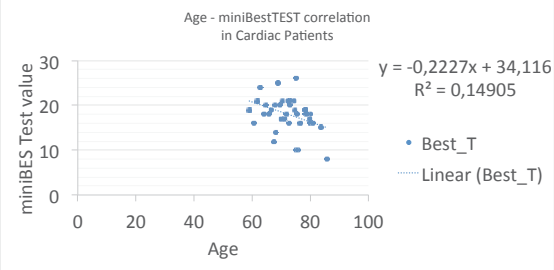



Figure 11: Analysis of parameters age and Mini-BES Test score for cardiovascular subjects.

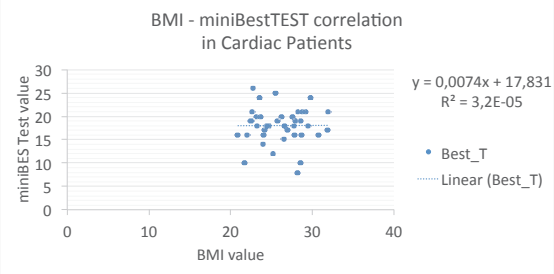



Figure 12: Analysis of parameters BMI and Mini-BES Test score for cardiovascular subjects.

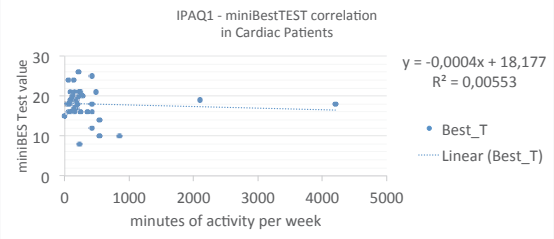



Figure 13: Analysis of parameters IPAQ_1 and Mini-BES Test score for cardiovascular subjects.

cal analysis is that, when a subject suffers from a metabolic disease, she/he has a probability of falling that is higher than that of an equally aged cardiovascular subject, and much higher than that of a healthy peer.

# 5 USE OF THE STATISTICAL ANALYSIS RESULTS IN A MONITORING SENSOR-BASED SYSTEM

The statistical analysis made in this preliminary study has shown that, even though moderate, a correlation exists between the risk of falling (the mini-BES test score) and the personal and lifestyle indicators.

These results mean that it is imaginable to realize a monitoring system in order to give specific rec-

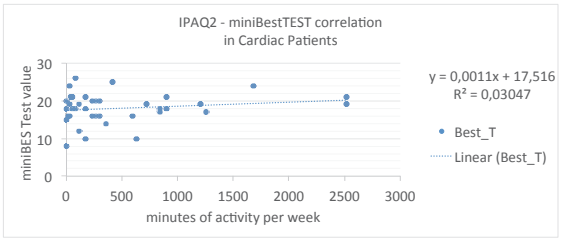



Figure 14: Analysis of parameters IPAQ_2 and Mini-BES Test score for cardiovascular subjects.

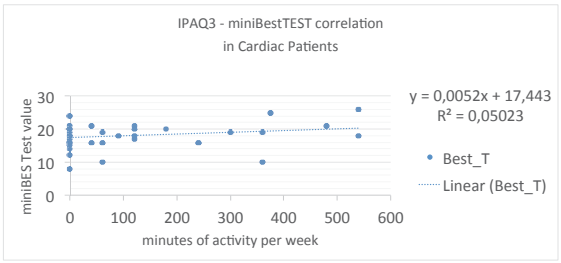



Figure 15: Analysis of parameters IPAQ_3 and Mini-BES Test score for cardiovascular subjects.

ommendations about the diet and the amount and the type of physical activity so as to improve the subject well-being with respect to the risk of falling.

Nowadays, mobile devices, such as smartphones or tablets, wearable devices, such as smartwatches or bands, and wireless healthcare devices, such as smart digital scales, are widely used and it is demonstrated that they are valid tools to monitor body and lifestyles parameters. For these reasons, thanks to these devices, it is possible to collect in real time data about, for example, the weight of a subject or the type and the amount of activity performed during a day or a week.

This collected information could be analyzed in real time on a mobile device in order to give a prompt feedback to the subject about her/his risk of falling and in order to guide the subject to develop new habits to reduce the estimated risk of falling.

As an example, for the subjects with metabolic disease the statistical analysis has shown that there is a strong and inverse correlation between the BMI and the Mini-BES test score, so a wearable sensor-based monitoring system could give specific recommendations in order to not only reduce the obesity, but also to reduce the risk of falling knowing that more obese a metabolic subject, the higher probability she/he has of falling.

Within our laboratory, several mobile applications have been implemented aimed to monitor different kinds of healthcare parameters, as for example (Forastiere et al., 2016; Sannino and De Pietro, 2014). The correlation results found in this study for healthy subjects and cardiac subjects will be easily added into

them respectively.

Unfortunately, there is no unique app with the possibility to have a specific knowledge base for each subject in order to suggest different recommendations for the three different groups of subjects.

For this reason, a new mobile health application is under development in order to take into account the different results obtained for the different classes of people examined, e.g. healthy subjects, subjects with metabolic diseases or subjects with cardiac diseases. The app will be able to monitor body indicators, physiological data, and physical activity information by using wearable sensors, be they or not compliant to the Continua Alliance guidelines (Carroll et al., 2007). Of course, some of these sensors will be used for long periods, as e.g. those for physical activity monitoring, therefore they are affected by the typical problems related to battery charge. Other types of sensors, instead, will be employed more rarely, as for example smart digital scales that are typically used just once in a day.

# 6 CONCLUSIONS AND FUTURE WORK

In this paper the correlation between personal and life-style indicators and the risk of falling has been investigated.

As indicators we have considered here for each subject age, body mass index, and information about physical activity habits, while a subject's risk of falling has been estimated by the Mini-BES test score. Three different groups of subjects have been taken into account, namely healthy, suffering from metabolic diseases and suffering from cardiovascular diseases.

Firstly, explicit linear correlations have been found for any pair of parameters. Secondly, attention has been paid to whether or not these correlations change as the health state of the subjects does.

Finally, some first steps have been moved towards a system, composed by wearable sensors, a mobile device, and an app, that would be able to help people in improving their life-style so as to decrease falling risk.

In the near future we aim at implementing the sensor-based system.

Moreover, due to the fact that the data set from Physionet used in this paper looks highly clustered with little outliers, e.g. most healthy patients are around 20 years old, we plan to start a cooperation phase with the University of Naples "Federico II" in which they will provide us with some volunteers with different ages in order to better balance the database.

Within this cooperation we will supply the system to the volunteers, so as to test its effectiveness and usefulness.

# ACKNOWLEDGEMENTS

This work has been supported by the project "eHealthNet: Ecosistema software per la Sanità Elettronica" (PON03PE_00128_1) financed within the P.O.N. "Research and Competitiveness" call of the Italian Ministry for University and Research.

# REFERENCES

Carroll, R., Cnossen, R., Schnell, M., and Simons, D. (2007). Continua: An interoperable personal healthcare ecosystem. *IEEE Pervasive Computing*, 6(4):90–94.

Craig, C. L., Marshall, A. L., Sjstrm, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J. F., and Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine and science in sports and exercise*, 35(8):13811395.

Faulkner, K. A., Cauley, J. A., Studenski, S. A., Landsittel, D. P., Cummings, S. R., Ensrud, K. E., Donaldson, M., Nevitt, M., of Osteoporotic Fractures Research Group, S., et al. (2009). Lifestyle predicts falls independent of physical risk factors. *Osteoporosis international*, 20(12):2025–2034.

Forastiere, M., De Pietro, G., and Sannino, G. (2016). An mhealth application for a personalized monitoring of ones own wellness: Design and development. In *Innovation in Medicine and Healthcare 2016*, pages 269–278. Springer.

Franchignoni, F., Horak, F., Godi, M., Nardone, A., and Giordano, A. (2010). Using psychometric techniques to improve the balance evaluation systems test: the mini-bestest. *Journal of Rehabilitation Medicine*, 42(4):323–331.

Gardner, M. M., Robertson, M. C., and Campbell, A. J. (2000). Exercise in preventing falls and fall related injuries in older people: a review of randomised controlled trials. *British journal of sports medicine*, 34(1):7–17.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220.

Hue, O., Simoneau, M., Marcotte, J., Berrigan, F., Doré, J., Marceau, P., Marceau, S., Tremblay, A., and Teasdale, N. (2007). Body weight is a strong predictor of postural stability. *Gait & posture*, 26(1):32–38.

Kempen, G. I., Yardley, L., Van Haastregt, J. C., Zijlstra, G. R., Beyer, N., Hauer, K., and Todd, C. (2008). The short fes-i: a shortened version of the falls efficacy scale-international to assess fear of falling. *Age and ageing*, 37(1):45–50.

Ku, P., Osman, N. A., Yusof, A., and Abas, W. W. (2012). Biomechanical evaluation of the relationship between postural control and body mass index. *Journal of biomechanics*, 45(9):1638–1642.

Mozumdar, A. and Liguori, G. (2011). Persistent increase of prevalence of metabolic syndrome among us adults: Nhanes iii to nhanes 1999–2006. *Diabetes care*, 34(1):216–219.

Pereira, C. L., Baptista, F., and Infante, P. (2014). Role of physical activity in the occurrence of falls and fall-related injuries in community-dwelling adults over 50 years old. *Disability and rehabilitation*, 36(2):117–124.

Sannino, G., De Falco, I., and De Pietro, G. (2015). A supervised approach to automatically extract a set of rules to support fall detection in an mhealth system. *Applied Soft Computing*, 34:205–216.

Sannino, G. and De Pietro, G. (2014). A mobile system for real-time context-aware monitoring of patients health and fainting. *International journal of data mining and bioinformatics*, 10(4):407–423.

Santos, D. A. and Duarte, M. (2016). A public data set of human balance evaluations. *PeerJ*, 4:e2648.

Shahudin, N. N., Yusof, S. M., Razak, F. A., Sariman, M. H., Azam, M. Z. M., and Norman, W. M. N. W. (2016). Effects of age on physical activity level, strength and balance towards fall risk index among women aged 20–73 years. In *Proceedings of the 2nd International Colloquium on Sports Science, Exercise, Engineering and Technology 2015 (ICoSSEET 2015)*, pages 25–34. Springer.

Shekharappa, K., Smilee, J. S., Mallikarjuna, P. T., Vedavathi, K. J., and Jayarajan, M. P. (2011). Correlation between body mass index and cardiovascular parameters in obese and non obese in different age groups. *International Journal of Biological & Medical Research*, 2(2):551–555.

World Health Organization . Ageing and Life Course Unit (2008). *WHO global report on falls prevention in older age*. World Health Organization.

World Health Organization and others (2005). Preventing chronic diseases: a vital investment: Who global report.

# Modular Health Kiosk based on Web Technologies

João Silva, Pedro Brandão and Rui Prior
*Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Portugal*
*{jsilva, pbrandao, rprior}@dcc.fc.up.pt*

Keywords: Health Kiosk, Personal Health Devices, Data Collection, Autonomous Measurements, Health Monitoring.

Abstract: We describe a modular and easily reconfigurable Health Kiosk based on common, off-the-shelf Personal Health Devices and a computer with a touchscreen interface. The Kiosk is implemented in standard web technologies (JavaScript, HTML and CSS) on top of the Electron platform. It is intended to be used autonomously by the patients. It is highly modular, can easily be adapted and reconfigured by health professionals with little to no computer expertise, using a graphical interface, to adapt to different groups of patients and use cases. We document our findings, identifying problems faced throughout the development and solutions to those problems.

## 1 INTRODUCTION

With the increase of the population at a worldwide level, the expenses associated with health care are proportionally increasing. Despite that fact there are still some gaps that need to be fulfilled since the resources are mainly allocated in central areas that most of the time are not easily accessible to all the population. Moreover, due to that centralization of resources, the ratio of patients for each health professional is expected to increase, making it more difficult for them to proper assess all patients that use a medical facility.

There is a clear demand for alternatives to visiting a medical facility, due to the expenses associated with it, not only financial but also in terms of time. Due to the high demand and a low offer in terms of medical professionals available, there is a chance that when visiting a medical facility a patient may not be evaluated by a professional. This situation can lead to a delay in finding abnormal biometric values, causing a later treatment start.

Health has always been a field with a high-level of acceptance of innovation, and using technology with the intent of improving patients health is something that has been evolving over the time. This evolution allowed for the development of medical devices accessible not only to hospitals but also for personal use at home. (Suggs, 2006) There was an evolution from different types of devices that facilitated the access to health information. The phone acted as a way to solve simple questions for the patients, which, when the Internet access was globalized, was made easier through forums or e-mails.

A shortage of these types of devices or technologies in rural areas is visible and due to that fact, it should be a main goal of our society to try to reduce the discrepancy between these areas and others more developed. In rural areas where the offer in terms of medical services is low, the impact of a system that can make evaluation of the patients without taking time from the medical professionals can be extremely relevant. (Das and Padhy, 2014)

Creating a system such as the health kiosk would allow the patients, without the requirement of allocating human resources in these areas, hospitals or other possible locations, to obtain information on their health, ultimately saving time and resources to the user, and in the case of the hospital both to the patient and to the health professional that has access to the information without having to collect it.

Possible applications of this system are hospitals or health centers, but it is not limited to those as the system is being developed in order to adapt to the needs of the situation. This is accomplished by having a configuration that allows for modifications. With the right configuration it can be deployed in rural areas, based on the needs of those areas, or elderly communities in which its inhabitants require constant medical attention.

In this work, the possibilities of improving an existing health kiosk are analyzed, what modifications were made, what benefits those modifications bring to the table, as well as the difficulties faced on the



Silva, J., Brandão, P. and Prior, R.
Modular Health Kiosk based on Web Technologies.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 517-525
ISBN: 978-989-758-213-4

development and milestones that are to be reached in order to have a fully functional system, capable of being deployed in different areas.

## 2 RELATED WORK

Using computers or other internet capable devices has always been of great help in health related developments, not only with respect to machines used, the type of devices available in hospitals but also in the way the general population has access to health information.

The evolution of technology made possible the adaptation of several of those developments to health areas, such as the phone, video, clarification of doubts via e-mail, medical websites, and the creation and usage of electronic health records. (Verma et al., 2008)

With the increase in the variety and offer of medical devices, it rose the idea of grouping a set of these devices associated with an application in order to create a system capable of taking measurements of its users to provide some feedback on their vital signs, which otherwise would not be possible if its users lacked the resources to acquire these devices or to make visits to health centers. Currently the health kiosk is a system capable of collecting data from a blood pressure monitor, a weighing scale and a pulse oximeter. Nonetheless, it is not limited solely to these devices as new ones can easily be added to the system.

The idea of a system capable of collecting biometric data from its users and show them the results is not new, as it can be seen in public locations, where most of the time static, offline systems with a small set of instructions provide the user valuable information. In parallel some other types of kiosks can be seen in some locations such as the information kiosks that serve the purpose of disseminating information throughout the population (Nicholas et al., 2003; Leeman-Castillo et al., 2010), they can be internet connected with the intent of remotely changing the available information.

A more approximated system to our proposal is the Multi-User Health Kiosk developed in a joint effort of the University of Pittsburgh and Carnegie Mellon University. (Courtney et al., 2013) Their findings provided helpful information on how to tackle the development of a health kiosk. Creating a modular architecture, to change in accordance to what is needed or providing helpful, step by step, instructions on how to use the devices, is information that, despite being simple, has a major impact on the usability of the system.

Since our aim with this system is to deploy it at communities, public locations or medical facilities, a multi-user approach must be made instead of creating a simple collection tool for a single user to use at home. This has to take into account that the system will be used by different users, with different characteristics and different health needs.

A specific application case of this type of system are elderly communities, as the need for attention on their vital signs is higher due to the increased fragility of human health over time. This type of deployment has proven to be very positive with a high-level of acceptance from its users. (Demiris et al., 2013)

The system is being developed making use of commercially available Personal Health Devices (PHDs). With the increased availability of these types of devices, the price is more accessible, reducing the total cost of assembling a system with these requirements. Some of the devices that are already working with this system, and possible new devices, communicate under a standardized form, which allows adding new devices in a simpler way since the communication module handles all the devices.

The ISO/IEEE 11073 Personal Health Data (PHD) Standards allow for the intended interoperability between some of the different devices that make up the system. This standard appeared as a merge of different standards such as ISO TC215, CEN TC251, IEEE 1073. (Nam et al., 2011) The Continua Health Alliance has a major impact both in the standard as well as in the health care industry, trying to standardize health devices in an orderly process, and also to make it easier for developers to work around with these devices.

Not having this standardization in some devices is a problem since it means that an individual approach must be made in order to interact with the devices. A way to establish the communication has to be created, as well as an interpretation of the messages sent by the devices, and if needed to the devices.

The objective of reducing consultation times, and keeping a more detailed patient's history is achieved by having access to these records. This implies that the health kiosk in the future must be able to add data to the patient's Electronic Health Record (EHR) this would improve the view the health professional has on the patient's history since the number of measurements taken outside a medical facility should be higher than inside one. Despite its benefits, it is not an easy task to converge the user's medical data of a user and make it available at all medical facilities that the user could visit. (Kalra, 2006) For now, a local database serves the purpose of giving the patients the possibility of saving and evaluating their history.

It is now possible, when using the health kiosk to

make use of a smart card reader, this option makes it possible to extract data from the Portuguese Citizenship card. This function helps not only to avoid human error when inserting personal data, but also as a way of reducing usage time of the application. Further testing needs to be made in order to establish the true value of this feature. The data available for collection is public data that is visible in the physical card, no private data is accessible without a Personal Identification Number (PIN) that only the user has knowledge of, besides text data it is also possible to extract the picture of the user.

With the objective of being used in an autonomous way by the patients, the usability of the health kiosk is a concern. The current version of this system is an improvement to an existing version, which was assessed in respect of the usability of the application. The differences between the two versions will be address in the next section. (Soares et al., 2016) Little changes were made relative to the execution flow of the application in order to take advantage of the information that was collected regarding the usability. This evaluation of the application was based on observer-filled questionnaires alongside the option of keeping track of user clicks on the screen in order to evaluate the number and position of clicks per screen to see the changes to be made in the presented content.

## 3 OVERVIEW OF THE SYSTEM

An example of a health kiosk physical deployment is visible in Figure 1, which has three devices associated with it, a blood pressure monitor, a pulse oximeter and a weighing scale that is not visible in the figure. The interaction from the user is made through the use of the touchscreen, if the user chooses to, it can insert the citizenship card onto the smart card reader. After the exams are finished, the results are printed in the printer that is also visible in the figure.

Figure 1: Physical health kiosk.

Relative to the system architecture a representation is visible in Figure 2. The system has two possible types of devices, Continua Alliance Certified and Non Certified Devices, at the moment all the devices communicate via Bluetooth. Nonetheless, the development of a way to communicate with the devices takes one of two possible approaches, either the device is Continua Alliance certified and communicates with an Antidote IEEE 11073 PHD interface, or it is not certified and an individual approach must be made in order to communicate with that specific device. The process of creating a way of communicating with a non-certified device can be applied to other devices, creating a skeleton that is the base for the development. Consequently, reducing the time of development.

In the future we intend to have a connection to a central EHR system, in order to make the patient's data available in different health kiosks, as well as to health professionals. By sharing data along all different locations that the users visit the impact the collected information has is higher.

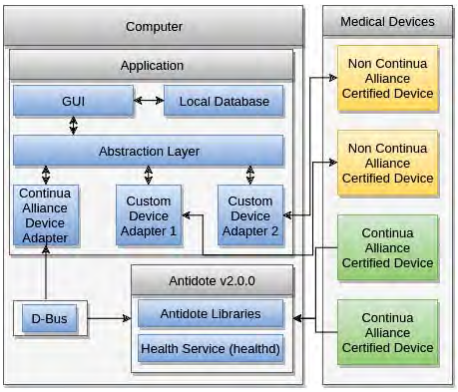



Figure 2: Overview of the System.

## 4 IMPLEMENTATION

The current development of the health kiosk has been based on a previous existing version that, despite already having a functional flow, did not have the tools for a proper continuous development. With that came the idea of developing a version of the health kiosk that used web technologies. The previous version was developed using JavaFX, where some challenges were found before the transition into the use of web technologies. This transition came as an alternative that could replace the existing version and at the same time take advantage of features that are available with the usage of this type of technologies, such as Web Real Time Communication (WebRTC).

The application was developed on top of Electron, a framework for building cross-platform applications

based on JavaScript, HTML and CSS. The Electron[1] framework was selected because it makes it possible to communicate with the operating system directly using Node.js, while providing all the advantages of using web technologies (including the vast amount of JavaScript libraries for building web applications).

The previous version was already assessed in terms of usability (Soares et al., 2016). With that in mind, the development of the new interface was made trying to keep the same flow of interaction.

Some changes had to be made, either because of the advantages of using web technologies, or because the creation of new modules such as the smart card reader implied new screens that the user has to interact with. In the following subsections the adaptations made to the system will be addressed.

Regarding the usage of the application, there are several possible cases of use due to the modularity of the application. Figure 3 represents a full usage of the application, this includes language selection, authentication selection, performing exams and an end summary. Figure 4 represents a smaller application case, where no user data is collected, just the exams to be performed and a summary with respect to the collected data. These variations allow for different field applications since what is intended is to adapt the application to the users and/or scenario and not the other way around.

The flows visible in the figures goes through several steps, these steps being optional in some cases. By configuring the system to follow the flow visible in Figure 3 it gives the users the possibility of taking full advantage of all the features. Starting with the language selection, the user is presented flags representing all available languages. After selecting the language, all texts and voice instructions to be presented to the user are in the selected language. The user is presented now with the authentication method selection, if the user chooses to use the citizenship card all the necessary personal data is collected automatically. If not the user is presented with several screens in order to insert the data by himself. After having all the data, the user is presented with a screen showing all exams that are to be performed. Subsequently the user iterates over each exam, following a set of instructions, either by image or video, and evaluating the collected information on an intermediary results page. In the end a table is shown presenting the final results, and it is also visible in the screen a QR Code containing all the information of that table that is inserted as a calendar entry, all this is printed out to give out to the user. After that, the system restarts and is ready for a new user.

[1]http://electron.atom.io/

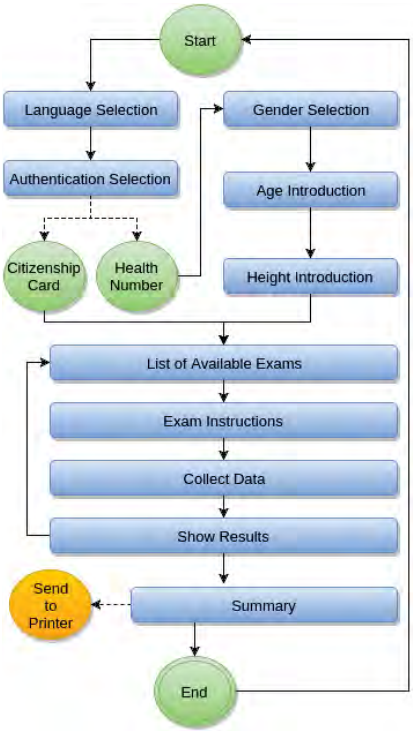



Figure 3: Example of the flow of execution.

What is intended with this system is not only to have the ability to collect data, process it and present the results, but to give the option to the responsible person to adapt the health kiosk in order to fulfill the needs of its users. There are different use cases to the health kiosk, some differences are related to the type of data that is intended to be collected, other to the necessity or not to collect user personal data, for that the health kiosk must be modular and easily adaptable to the circumstances.

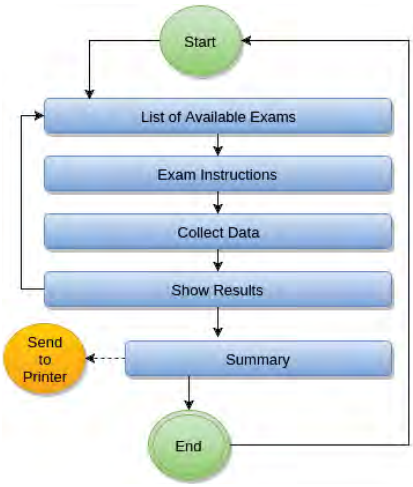



Figure 4: Example of a simpler flow of execution.

## 4.1 User Interface

The usage of this system can be divided upon three different stages, collecting personal data from the user, collecting data from the devices, presenting results to the user.

When the patient starts the process of using the health kiosk, it has the choice of either inserting a citizenship card, and the application itself collects relevant data from the card, or the user can himself insert its personal data, such as the national health identification number, gender, height and age.

After having all the user data collected, the next step is to go through all the defined exams, after each exam the user is presented with the results and a chronological chart with all the measurements taken with that device associated with the user. The collected measurements are displayed in a bar, colored green on what are considered default values for that measurement with a gradient to red as it increases the distance to those values. The intermediary results screen are visible in Figure 5, which represents what it looked like in the older version of the health kiosk, and Figure 6 represents the newer version of that screen.

The proposed changes included an usage of chronological graph, with the $x$ axis being a time series, which means that the points take into consideration the distance between the presented dates; centering the next button and using a green color instead of blue.

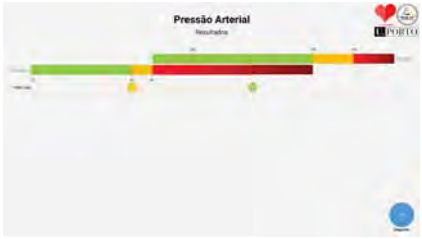



Figure 5: Blood Pressure Results on the previous version.

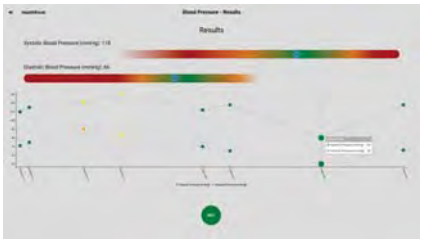


Figure 6: Blood Pressure Results on the new version.

In the final screen a table is presented with all collected data, as well as a QR Code that upon reading by a smartphone inserts in the default calendar app an entry with all the collected data.

A page is also printed containing the user personal data, the collected data and the QR Code. It also contains a logo on the top that can be configured to fit the needs of the situation.

The idea of using a QR Code to share data serves not only the purpose of saving the data in the users smartphones, but can also be used in the future as a way of transmitting data to a health kiosk smartphone application that if developed can act as a place where the user can evaluate at any moment their vital signs registry and even add other measurements that were not made in the health kiosk.

The printing of the results, as well as the generation of a QR Code are optional functions that can be disabled by the person responsible for the configuration. This goes along with the idea of creating a system that is highly configurable and capable of dealing with the needs of its users.

## 4.2 Application Modularity

From the beginning one of principles of the development of this health kiosk was the idea of modularity. By creating a system that is based on this idea, adding or removing elements, changing their order or using the system with different configurations is a goal easier to achieve.

With that in mind, one of the main focus when developing this application is the possible configurations that the application should have. The use cases can increase by creating a modular application. As such, a configuration file was created, in which all the possible options of the system are inserted. The configuration file allows for the definition of which screens to show, which exams are to be performed and in which order to show them.

When having the possibility of choosing which screens to be used, or which exams to perform, the system is easily adapted to different use cases, from a situation where no user data is needed such as a specific event where only a specific parameter is being evaluated in a population to detect anomalies, to cases where the system is deployed in a local place where all the population can access it, and a record specific to each user has more meaning.

A approach was made, when the system was being developed, to try and simplify the process of adding new exams or devices. A skeleton, based on web components, is created for adding new exams to the application, and if the device is certified, it can easily be added to the system, needing only specific codes for the parameters being read from the device.

Figure 7 represents an exam component, this component can be reused in different cases, the compo-

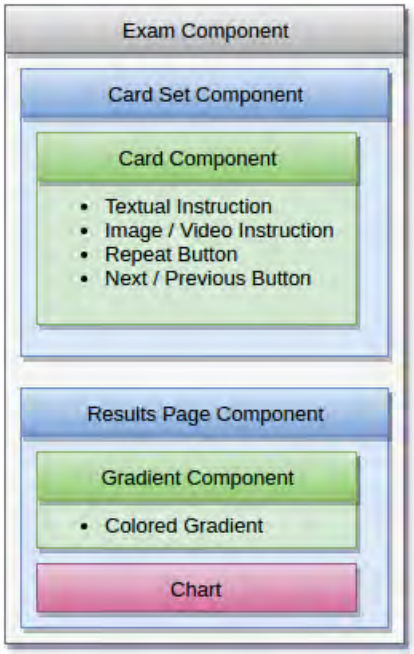



Figure 7: Representation of an Exam Component.

nent itself contains two different components, the card set that is shown to the user with the instructions, and the results component that is shown after the user has passed through all the instruction cards. This approach was made throughout all the application in order to facilitate adding new screens or elements to existing ones.

## 4.3 Internationalization

Although at the moment the smart card component collects only data from Portuguese Citizenship cards, and the national health identification number asked is also the Portuguese one, this will not always be the case, and the wide range of possible users of the health kiosk must be taken into consideration. With that in mind, we implemented in the system the possibility of using different languages, making use of an internationalization framework[2].

Having each text component associated with an id, and for each language a set of ids with the respective text values, the process of making the application available in new different languages was made easy. As for adding a new language the development process passes only through creating a file with all the id and values pairs and the system is capable of dealing with that information.

Although the previous version of the health kiosk already had voice instructions, these were recorded by a person. This has some limitations in terms of development. This initial approach made possible a first development and evaluation on how sound instructions can help the user take the available exams. However, it has several problems associated with it, for perfect instructions no noise in the sound would be preferable and this is not possible to achieve with low cost recording devices. Another fact to take into account is that even small changes to the audio instructions imply that a new recording session has to be made.

To tackle these issues, we added digital voice instructions to the application. These voice instructions are generated making use of the text-to-speech technology and saved into files to ensure that the produced speech is the same across all systems. The process of creating these instructions is the same as for the text instructions, a set of values associated with an Id have to be created, and then all these values are read and a file for each available language is generated. Since the voice's used are the available ones on the operative system, and new ones can be added, it is easily possible to add voice instructions in different languages.

## 4.4 Hardware

Currently the prototype of the health kiosk is deployed in an all-in-one PC with a 22" touchscreen. This prototype is capable of collecting from any possible combination of a blood pressure monitor, a weighing scale, and a pulse oximeter. This system has connected to it a printer in order to handout the results to its users and a smart card reader to extract data from the citizenship card.

The current supported and tested devices in the system are two Continua Alliance Certified devices, a blood pressure monitor (AND A&D Medical UA-767 Plus BT-Ci) and a weighing scale (AND A&D Medical UC-351PBT-Ci), and a non-certified device, a pulse oximeter (Nonin 3230), this device communicates via Bluetooth Low Energy (BLE). Although Nonin has a Continua Alliance compliant device we used this one to develop and test the integration of non Continua devices. Moreover, we also experimented with the differences for a BLE device.

### 4.4.1 Device Communication

During the development of the new version, we created modules to communicate with the devices. By using Electron to develop the application, Node.js can be used in order to access the operating system directly, which would not be possible if a browser application were to be developed.

[2]See http://i18next.com/.

Using a Node.js module (`node-dbus`), made it possible to develop a new module capable of communicating with the existing Continua Alliance certified devices, and to easily add new devices as long as information about the devices is previously given to the developers in order to properly extract relevant information produced by the device.

Continua Alliance certified devices, generate data in eXtensible Markup Language (XML) format, having the relevant medical data associated with a specific Id for the parameter, it also provides information relative to the date of the measurement. It also contains parameters that can hold information about the physical device.

Since the objective of the application is to adapt to the users and not the other way around a decision had to be made in order to support non certified devices since there is a possibility of these types of devices being cheaper for some possible use cases. This decision goes along with the idea of creating an adaptive system that is not closed in terms of compatible devices.

For the case of the Nonin Oximeter 3230, as it is not a Continua Alliance certified device a different approach had to be made. Since this device communicates via BLE it was decided that taking advantage of a Node.js modules (`Noble`) was an appropriate way to develop the means to communicate with these types of devices. For that a module was created that takes the MAC Address of a device, starts scanning until the device is found, after connecting to the device it has to activate the characteristic of the device responsible for sending data via Bluetooth, when that is done a stream of continuous data is received by the application, being decided that upon a number of repetitions that value was to be considered and the stream stopped.

The development of these new modules will allow a simpler development in the future since the base of the communication is already implemented and what is required is to adapt it to the new devices is an evaluation of the device in order to proper assess what messages it sends, how it sends them and what is the best way to extract information from them.

## 4.5 Usability

Being an application that evolved from an existing one, and since the previous version had an usability evaluation, what was made when developing this application was to follow the flow of interaction provided in that application. The usability of the previous version was studied in different scenarios, at the university open days (by 195 users), at a health day in one of the university schools (46 users), during a week at a local city hall (127 users) and for thirteen days in thirteen different villages in Brazil (465 users). This covered different age groups, and usage information was retrieved not only from the application but also from the evaluation on how people used the system. The evaluation used observer-filled questionnaires and user click tracking. The evaluation did not use the standard user questionnaires as we perceived them as either too long or less appropriate than a researcher observing the usage. More details can be seen in (Soares et al., 2016).

At the moment the usability of the current system has not been tested by a large population. The tools for assessing time spent on each screen and where the user has clicked throughout the usage are developed and will soon be tested. This will allow us to evaluate if the interface is easily usable, if the user has the ability to use the application from start to finish without any assistance, or if so which screens are taking more time from the flow of interaction.

The collection of time spent on each screen can help us determine if the instructions on the screen are easy to understand, and for instance, if having the ability to use a citizenship card is quicker and more adopted than introducing the data manually.

By also collecting the coordinates of the clicks and group all clicks made in a single screen presenting them on top of a screen representation it is possible to determinate if the screen is usable, and if it is being used in the most correct way. If a large group of people click somewhere in the screen that is not intended to be clicked then something is not right about that screen and must be evaluated.

More consideration has to be made in order to proper evaluate the usability by using these methods, such as the elimination of outliers that could affect the visualization of the problems on the screen. A small amount of users could make excessive clicks in wrong places or spent too much time in certain screens and with that alert for a non existing problem.

The usability of a previous version of the health kiosk has been tested (Soares et al., 2016) with the same idea of representing clicks on a screen representation, which is visible on Figure 8.

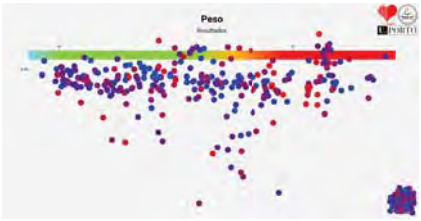



Figure 8: Weight measurement result screen usage pattern.

In Figure 8 it is possible to evaluate that users try to interact with the color gradient before clicking on the button to show the new screen. This type of information allow us to consider what steps to take to make the application more fluid, with quicker and better responses from the users.

### 4.6 Configuration Tool

Both on the previous version as well as in the new version there was a unique point in the system that was responsible for the configuration of the system.

The configuration file is currently responsible to allow for the definition of which screens to be shown or wich personal data to collect, if features such as QR Code, printing, voice instructions should be used or not in the application among other different options.

For easier configuration, all possible configurable values must be in this file. At this moment it is under development a tool that will allow for non-technical users to configure the health kiosk. A preview of the application interface is visible in Figure 9, this application will be divided in several blocks of configuration grouped by what is being configured, such as available languages, what type of authentication is to be used, what exams are to be performed. The development of this tool is not only important to make the health kiosk configurable without technical help but also to evaluate what truly is configurable in this system.


Figure 9: Current status of the Configuration Tool.

## 5 CONCLUSIONS

A system such as the health kiosk could have a major impact in different possible scenarios, such as rural areas lacking medical resources or health centers with higher attendance than the one they can process. Deploying this system in these locations, makes it possible for the population to, by themselves, measure their vital signs.

The current prototype makes use of a simple weighing scale, a portable blood pressure monitor and pulse oximeter, an evaluation of possible changes to this setup is undergoing, with the idea in mind of having devices simpler to use and that possibly collect more data. One example is going from a normal blood pressure monitor that the user has to set up and adjust, to one where the user simple introduces the arm in the fixed device and waits for the results. Also, weighing scales capable of evaluating body fat are also an alternative to the current existing ones. One point that has to be considered is relative to the powering of the devices, presently all the devices are battery powered, which is not feasible in large scale.

At it was already referred, the configuration of the health kiosk is made by editing an existing configuration file with the desired values, but a Graphical User Interface (GUI) is being integrated into the health kiosk system soon that will allow for anyone responsible for the health kiosk to edit the configuration without knowledge of the technical details.

It is also being developed under a master's thesis project a module that will allow the establishment of a video conference, using WebRTC, in which the patient can get help on how to use the device or ask for a medical opinion on the collected data. Due to the possibilities of WebRTC, the health professional is not only able to establish a video conference, but it is also able to have access to data from the application, and to send data to the application. This creates the possibility of having remote instructions that can change the status of the application.

## ACKNOWLEDGEMENTS

This article is a result of the project NanoSTIMA Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics, NORTE-01-0145-FEDER-000016, supported by Norte Portugal Regional Operational Programme (NORTE 2020), through Portugal 2020 and the European Regional Development Fund.

## REFERENCES

Courtney, K. et al. (2013). Designing the community multi-user health kiosk. *Enabling Health and Healthcare Through ICT: Available, Tailored and Closer*, 183:79.

Das, R. and Padhy, H. M. (2014). Health monitoring kiosk: An effective system for rural health management. *International Journal of Innovations in Engineering Research and Technology, IJIERT*, 1(2).

Demiris, G., Thompson, H., Boquet, J., Le, T., Chaudhuri, S., and Chung, J. (2013). Older adults' acceptance of a community-based telehealth wellness system. *Informatics for Health and Social Care*, 38(1):27–36.

Kalra, D. (2006). Electronic health record standards. *IMIA Yearbook of Medical Informatic*, pages 136–144.

Leeman-Castillo, B., Beaty, B., Raghunath, S., Steiner, J., and Bull, S. (2010). LUCHAR: Using computer technology to battle heart disease among latinos. *American Journal of Public Health*, 100(2):272–275.

Nam, J.-C., Seo, W.-K., Bae, J.-S., and Cho, Y.-Z. (2011). Design and development of a u-Health system based on the ISO/IEEE 11073 PHD standards. In *The 17th Asia Pacific Conference on Communications*, pages 789–793. IEEE.

Nicholas, D., Huntington, P., and Williams, P. (2003). Three years of digital consumer health information: A longitudinal study of the touch screen health kiosk. *Inf. Process. Manage.*, 39(3):479–502.

Soares, E., Oliveira, C., Maia, J., Almeida, R., Coimbra, M., Brandão, P., and Prior, R. (2016). Modular health kiosk for health self-assessment. In *Computers and Communication (ISCC), 2016 IEEE Symposium on*, pages 278–280. IEEE.

Suggs, L. S. (2006). A 10-year retrospective of research in new technologies for health communication. *Journal of health communication*, 11(1):61–74.

Verma, A., Dhand, H., and Shaha, A. (2008). Healthcare kiosk next generation accessible healthcare solution. In *e-health Networking, Applications and Services, 2008. HealthCom 2008. 10th International Conference on*, pages 194–199. IEEE.

# Numerical and Implementation Issues in Food Quality Modeling for Human Diseases Prevention

A. Galletti[1], R. Montella[1,2], L. Marcellino[1], A. Riccio[1], D. Di Luccio[1], A. Brizius[2] and I. Foster[2]
[1]*Department of Science and Technology, University of Naples "Parthenope", Centro Direzionale, Isola C4, 80143, Napoli, Italy*
[2]*Computation Institute, University of Chicago, 5735 South Ellis Avenue, 60637, Chicago, IL, U.S.A.*
*{ardelio.galletti, raffaele.montella, livia.marcellino, angelo.riccio, diana.diluccio}@uniparthenope.it, {abrizius, foster}@uchicago.edu*

Keywords: Lagrangian Methods, Numerical Interpolation, Food Quality, Human Diseases, Cloud Computing, High Performance Computing, Scientific Workflow, Smart Devices, Internet of Things, Marine Data Crowdsourcing.

Abstract: Monitoring nearshore sea water pollution using connected smart devices could be nowadays impracticable due to the aggressive saline environment, the network availability and the maintain and calibration costs. Accurate forecast of marine pollution is most needed to evaluate the adverse effects on coastal inhabitants' health when fishes and mussels farming economically characterizes the local social background. In an operational context, numerical simulations are performed routinely on a dedicated computational infrastructure producing space and temporal high-resolution predictions of weather and marine conditions of the Bay of Naples. In this paper we present our results in developing a community open source Lagrangian pollutant transport and dispersion model, leveraging on hierarchical parallelism implying distributed memory, shared memory and GPGPUs. Some numerical details are also discussed. This system has been used to develop an alarm system to help local authorities in making decisions regarding the collection of mussels. The model setup and the simulation results will be improved using FairWind, an under development system dedicated to coastal marine crowdsourced data gathering and sharing, based on smart devices and Internet of Things afloat.

## 1 INTRODUCTION

Human health can be adversely affected by pollutants emission into sea water, specially from seafood contamination caused by inshore discharges or offshore spills in areas close to aquaculture farms.

Fish and mussel farms are critically sensitive to coastal water quality, and thus require continuous monitoring to enforce food security and quality and to prevent any possible disease affecting human health.

The potentially toxic substances emitted from point sources can, in more or less short time, reach the mussel farms and promote, in relation to the mussel-pollutant contact time, the bioaccumulation in filter feeders organisms. Several studies have shown that pathogens, such as bacteria and enteric viruses, can be transmitted by mussels and the widespread habit of consuming raw or slightly cooked shellfish contributes to maintaining the incidence of hepatitis A cases in the southern Italy at high level (Croci et al., 2003).

On the other hand, nevertheless the availability of technologies for remote water quality monitoring system using wireless sensors (Haron et al., 2009), the livestock sampling and the microbiological spottily analysis fails if the goal is a consistent data time series needed by any process aimed to make inference with human health. The use of connected smart devices is fully feasible in a context where the farms are in a limited environment, while the challenges rise for fish and mussel farms in marine nearshore, but open waters: the extreme weather events, the aggressive saline environment, the network and energy availability and, last but not least, the need for continuous maintenance and sensors calibration could have a negative impact on the use of a technical solution fully based on the Internet of Things afloat approach.

To face the above depicted scenario, we designed and implemented WaComM (Water Community Model), a three dimensional Lagrangian model enforcing the decision support system and enabling the simulation and prediction of pollutant spills, transport and dispersion in both inshore and offshore environments (Giunta et al., 2005).

Here, we provide some details about the way in which the input data of our Lagrangian model are ob-



Galletti, A., Montella, R., Marcellino, L., Riccio, A., Luccio, D., Brizius, A. and Foster, I.
Numerical and Implementation Issues in Food Quality Modeling for Human Diseases Prevention.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 526-534
ISBN: 978-989-758-213-4

tained. Moreover, we give some insights about the related parallel implementation.

In order to support the numerical model solution, we leverage on crowdsourced data acquired by the system called "FairWind" a smart, cloud-enabled, multifunctional navigation system for leisure and professional vessels. This system has been developed as an open technologies developing platform based on the Android operating system for smart devices.

In this context, the system is used to collect data from on board marine instruments such as, but not limited to, GPS, heading, pitch, roll, yaw and speed sensors, water temperature, depth and weather sensors. Collected data are sent on the land using grid file transfer technologies, and then processed in order to improve data quality and consistency.

Data, made available to the community as open data, are used to improve the model setup (i.e. high resolution coastal digital depth model) and will be used for model evaluation and improvement (surface temperature, surface current).

The rest of the paper is organized as follows: section 2 is about related work; section 3 focuses on the numerical issues driving the design of our Lagrangian model; section 4 is on details about the WaComM implementation with focus on parallel cores; section 5 is about the novel contribution of the crowdsourcing approach in this application research field; section 6 describes some preliminary computational evaluations; finally section 7 presents conclusions and future developments.

## 2 RELATED WORK

The Regional Ocean Modeling System (ROMS) is a free-surface, terrain-following, primitive equations ocean model widely used by the scientific community to characterize and simulate the meso- and sub-mesoscale ocean and coastal water dynamics (Haidvogel et al., 2000; Wilkin et al., 2005; Shchepetkin and McWilliams, 2003; Shchepetkin and McWilliams, 2005). In this application ROMS is forced by a high-resolution atmospheric forecast and provides the flow velocity on a curvilinear boundary-fitted grid.

ROMS was used to deploy a real-time forecast for the transport and deposition of water pollutants using the particle-tracking WaComM model, which implements a Lagrangian technique consistent with the advection-diffusion equation (Rodean, 1996).

In this kind of model, the dispersion phenomenon is reproduced by imaginary numerical particles; at each of these particles different characteristics are assigned, e.g. pollutant concentration and settling velocity.

At each time step, the particle position is calculated on the basis of the flow velocity, computed by the hydrodynamic model ROMS, and a random jump representing the turbulence diffusion.

GVirtuS (Montella et al., 2011) is a general-purpose virtualization service for high performance computing applications on cloud environments, focusing on NVidia CUDA GPGPU virtualization and MPI based virtual clusters.

The RAPID GVirtuS incarnation (Montella et al., 2016b) was used as GPGPU remoting provider for hierarchical parallelism, sending the instruction set kernel to the accelerating hardware, processing data on the device and then sending back results to the general purpose CPU.

The FACE-IT (Framework to Advance Climate, Economic, and Impact Investigations with IT) project (Pham et al., 2012) has been developed, and continues to be developed, to provide a cloud-based science gateway for the web-based access to a range of data projects, simulation models and analysis tools (Montella et al., 2015). FACE-IT builds on the Globus Galaxies platform, which has been developed over the past several years at the University of Chicago, initially in support of the Globus Genomics project (Madduri et al., 2015). FACE-IT is used as main computational playground for the implementation of the WaComM Lagrangian model running as an on-demand and routinely workflow (Montella et al., 2016a).

## 3 DESIGN

WaComM is the evolution of the LAMP3D model. We optimized the algorithms in order to improve its performance on high performance computing environment adding features as restarting and shared memory parallelization. The description of the underlying mathematical model is as follows.

Pollutants are considered as inert Lagrangian particles, tracing the marine circulation without feedback interactions with sea current fields and other particles. Each particle is assumed to have:

- position $r(t) = (x(t), y(t), z(t))$ at time $t$;
- initial position $r_0 = r(0) = (x_0, y_0, z_0)$ at the initial time $t = t_0 = 0$;
- velocity $v(r(t),t) = U(r(t),t) + \eta(r(t),t)$ at time $t$, where $U(r(t),t)$ denotes the deterministic velocity, and $\eta(r(t),t)$ is the stochastic fluctuation

arising from the Langevin equation model in order to describe the Brownian motion of particles (Rodean, 1996).

Given $U(r(t),t)$, or an estimate of it, for each position $r(t)$ and at each time $t$, the final particle position $r(t_{k+1}) = r(t_k + \Delta t)$, at time $t_{k+1} = t_k + \Delta t$, can be computed by means of the equation:

$$r(t_k + \Delta t) = r(t_k) + \int_{t_k}^{t_k+\Delta t} v(r(t),t)\,dt, \quad (k = 0,1,\ldots), \tag{1}$$

where $r(t_k)$ is a starting position, at the starting time $t_k$, and $\Delta t > 0$ denotes a time interval length (in WaComM we set $\Delta t = 1\,h$).

Numerical integration of (1) could be made in several ways; in our approach we use the Eulero method that considers a discretization of the time interval $[t_1, t_1 + \Delta t]$ in the grid

$$\tau_{j,k} = t_k + j \cdot d\tau, \qquad j = 0,\ldots,N$$

where $d\tau = \Delta t/N$ denotes the discretization step. To do this, at each time $\tau_j$ the evaluations of $U(r(\tau_{j,k}),\tau_{j,k})$ and $\eta(r(\tau_{j,k}),\tau_{j,k})$ are required. However, these values are provided by ROMS only at some discrete time instants, and on a discrete irregular three-dimensional grid. Such a grid can be thought of as the set of vertexes of a finite number of polyhedrons $V_i$ (cells). These cells are all topologically homeomorphic to a cube and their union is the space domain of $U$.

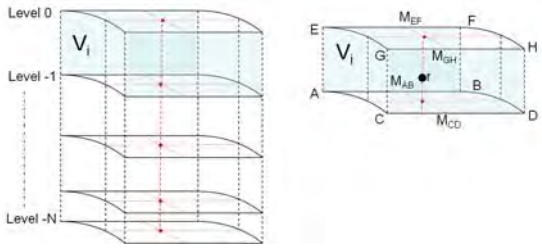



Figure 1: Example of cell. The cells form the irregular grid where the values of $U$ are known.

An example of the form of such cells is in Figure 1. Notice that the irregular polyhedron is defined by assigning its eight vertexes.

Possible choices of interpolants, which do not need any information about the mesh (i.e. the so-called mesh-free methods) are the radial basis functions methods (Cuomo et al., 2013; Fasshauer, 2007). However, when some structure of the grid is assigned, i.e. when a smart subdivision of the domain in geometrical polyhedral structures is known, one can take advantage of this fact and so several kinds of interpolants, exploiting the geometry of the cells that form the mesh, can be defined (Galletti and Maratea, 2016; Cuomo et al., 2014a; Cuomo et al., 2014b; Cuomo et al., 2015). In this case we choose a simple trilinear interpolation approach using barycentric coordinates. In particular, by referring to the notations introduced in Figure 1, we compute velocities at any spatial location, e.g. at particle position $r$ and desired time, by the linear interpolation of velocities made available by ROMS at the vertexes of any grid cell and regular time intervals.

In order to assign the stochastic fluctuations, WaComM relies on the standard 'K-theory', based on a diffusion coefficient which is estimated by preprocessed ROMS data. An exponential decay which uses the T90 parameter (the time required to degrade 90% of the biodegradable matter in a given environment) is applied to takes into account decaying processes. A sedimentation velocity, $w_{\text{sed}} = (0,0,-w_{\text{sed}})$, is added to the deterministic component of velocity to simulate settling particles. At the end of each suitably chosen time interval, a scaled concentration field $C_{i,j,k}$ is computed by simply counting the number of particles found within each grid cell.

# 4 IMPLEMENTATION

The modeling system can be used in an ex-ante fashion, as a decision support tool to aid in the selection of the best suitable areas for farming activity deployment, or in an ex-post fashion, in order to achieve a better management of offshore activities. We tested the system on several case studies where pollutants are spilled out from well known punctual sources located along the coasts of Campania region.

As arguing from the numerical approach, the model is computing intensive and parallelization is needed for its usage in real-world applications. The problem size increases with the number of emission sources and the number of emitted particles. Moreover, the computing time is influenced by the integration time step which should be short enough to correctly represent the turbulent diffusion processes.

Although consistent results can be guaranteed using the sequential implementation of the WaComM model, the wall-clock performance actually makes the production unfeasible.

Hence, the growing need of on-demand results, which involves a large computational effort for WaComM, suggests to use general purpose GPUs in order to efficiently perform computing intensive tasks.

In particular, a GPU implementation is considered for the WaComM main cycles involved for the interpolation and evaluation steps of the 3D momentum and dispersion parameters. This concerns with a parallel design schema hierarchical and heterogeneous.

The implementation of the GPGPU enabled code is realized with the NVIDIA CUDA programming model, and using heterogeneously both the CPU and GPU (Ortega et al., 2016) supported by an MPI based distributed memory approach. Such an implementation considers a dynamical load balancing on the particle number.

More in detail, the distributed memory parallelization has been introduced in the hourly inner cycle of WaComM in order to enhance the performance. Such a cycle computes the path of each particle and, since no interaction between particles is assumed, each particle path can be virtually tracked independently of the others. As explained in the design section, the interpolation stage is time consuming. The problem size scales with the input data (the resolution of the momentum, the sea current 3D vector components - U, V, W - and the vertical T-diffusion - AKT - grid). Then, previous discussion justifies even more the shared memory parallel approach for the evaluation of the interpolation model.

# 5 DATA CROWDSOURCING

In the field of ocean modeling, the need for computational and storage resources is currently satisfied by the availability of cloud based services that reduce the total cost of ownership and accelerate the democratization of science (Foster, 2011). Nevertheless, to have more robust and accurate models, there is a need for detailed, high resolution, spatio-temporal data for initialization and validation. While data can be hard to obtain from traditional sources, due the lack of available public data in some coastal areas, the challenges of surveying large areas and other technical constraints, this data can be easily obtained using internet of things based crowdsourcing tools like FairWind

(Montella et al., 2016c).

FairWind is an integrated, multifunctional, navigation software based on open technologies designed and developed by a very interdisciplinary team in order to maximize the benefits and the advantages. It is a marine single board computer device leveraging a custom-built infrastructure on top of stable and well documented mobile and cloud technologies.

From the marine electronics point of view, the most remarkable innovation introduced by FairWind are the Boat Apps that extend the FairWind basic features, integrating with already present onboard instruments and straightforwardly interacting with industrial or self - made internet of things based instruments. The board dataset, collected by FairWind, is a scientifically intriguing source of huge amounts of geolocated data (big-data) about marine coastal environment (weather and sea conditions, surface sea currents, water temperature, water depth, etc.), boat engine status, boat performances (speed, heading, pitch, roll), presence of board water and waste management, fuel consumptions and, above all, safety at sea and search and rescue systems (Figure 2).

Data is collected on board and, when possible, sent to cloud storage and computing facilities using reliable, affordable, and safe technologies such as the Globus data transport services. Users can choose what data to share in a named or anonymous way. Operationally, once data is collected, it will be analyzed and processed in order to extract sensor calibration using big-data algorithms, evaluated with a quality model comparing it with data acquired by trustful equipment and, finally, made available as open data for ocean model initialization and/or validation.

# 6 EVALUATION

This section describes the WaCoMM model use case for the Campania Region (Italy), applied to prevent the consumption of contaminated food harvest in mussel farms. This farms have, generally, a long lines organization in which the mussels are attached to submerged ropes hung from a back-bone supported by large plastic floats.

In 2006, the Campania Region published the "*Guidelines for mussel farms: classification and monitoring system of mussels production and relaying areas*" to delineate the guideline to mussel farms monitoring.

This document identified the skills to perform the analysis provided by the *Experimental Zooprophylactic Institute of Southern Italy* (IZSM) for mussels samples and by the *Campania Regional Environmental Protection Agency* (ARPAC) for water samples. The *in situ* analysis also included the compulsory microbiological parameters: *Escherichia Coli* and *Salmonella spp*. In accordance with the current European legislation (2073/2005EC), the concentration of *E. Coli* and *Salmonella spp* in mussels must be less than 230 MPN/100g (Most Probable Number per 100 grams) and zero respectively.

The considered mussels rows ($MF_{PT}$ in Figure 3) are located about 500 m distance to the coast in Punta Terone-Capo Miseno and cover an area of about 257 $m^2$, with a depth of about 20 m. In this case, the reared mussels are mostly *Mytilus Galloprovincialis*. The $MF_{PT}$ mussels are allowed to be placed on the market for human consumption directly, without fur-

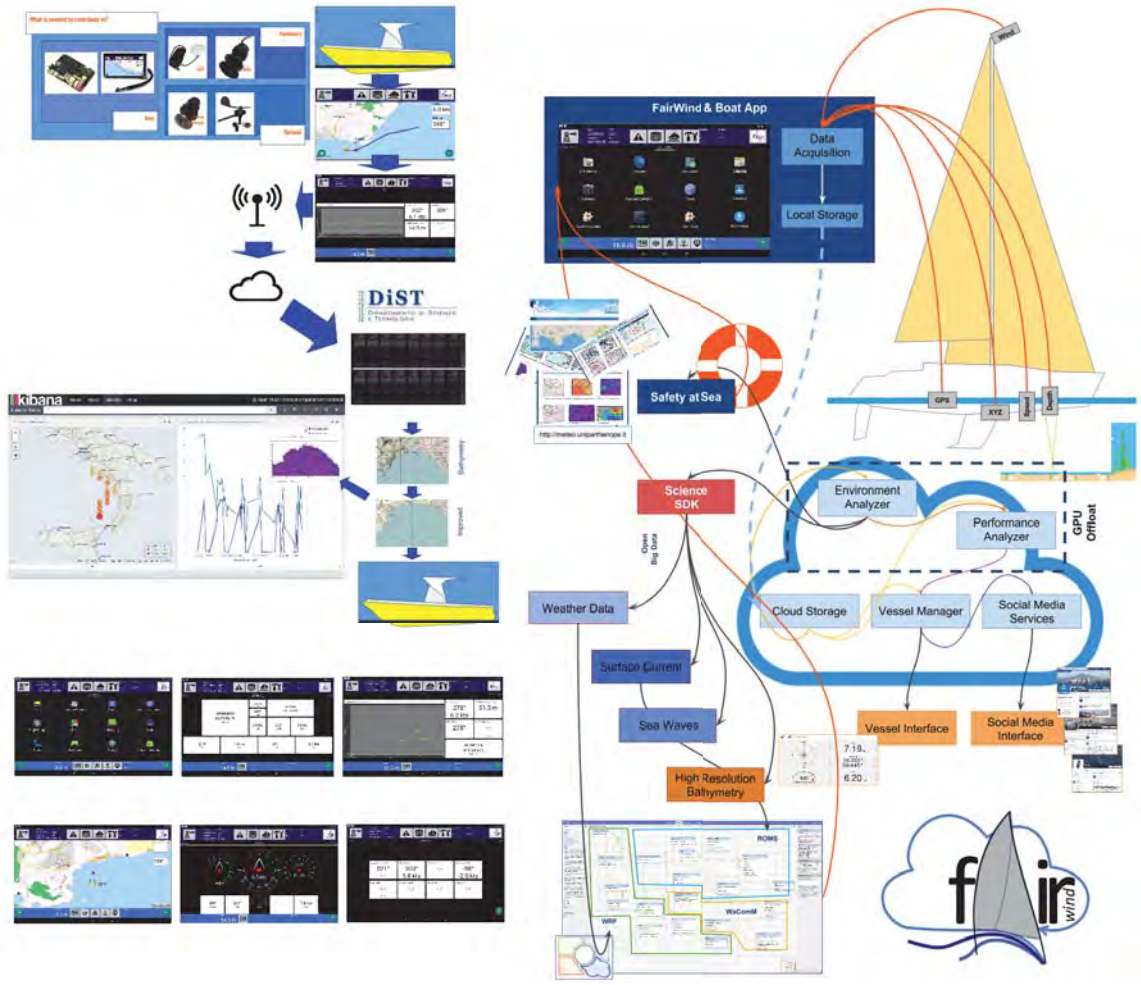



Figure 2: Smart devices in a context of internet of things afloat make data crowdsourcing affordable. FairWind equipped professional and leisure boats can contribute to model setup improvement and data assimilation.

ther treatment in a purification centre or after relaying (class type A). This makes the mussels quality very important to human diseases prevention.

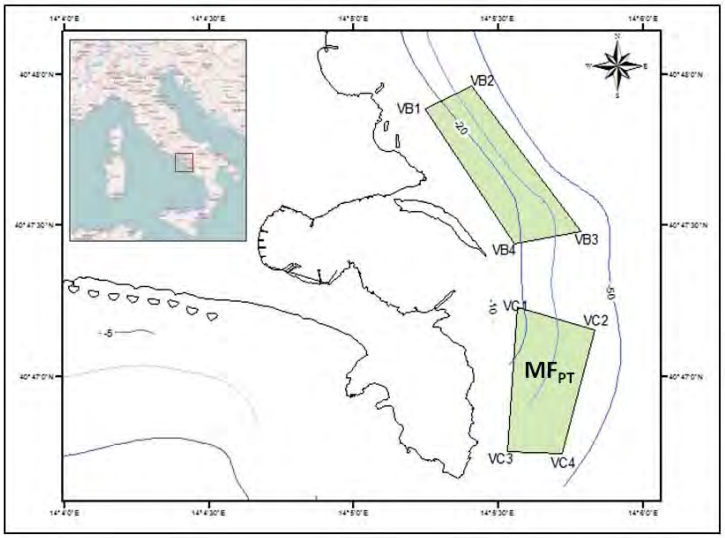
In some cases flooding events can bring too much pollutants to shellfish farms, banning them from harvesting.

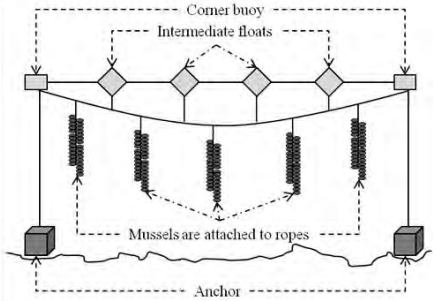
In this context, a forecasting system of the meteorological and hydrodynamic circulation, coupled with the WaComM model, can be a valid support to the mussel farm management.

WaComM is included in a scientific workflow to ingest the forecast input data needed to initialize the simulation and track particles trajectories.

The *Weather Research and Forecasting Model* (WRF) (Skamarock et al., 2001) simulates the weather conditions for driving the ROMS model. The WRF model has already been used to simulate weather conditions on the Campania region (Barone et al., 2000; Ascione et al., 2006). WaComM model domain is $715 \times 712 \times 11$ grid points ($Lat_{min}$=40.48N, $Lat_{max}$=40.95N; $Lon_{min}$=13.91E, $Lon_{max}$=14.52E). The pollutant sources in the Gulf of Naples are considered as points all along the coast, spilling out 100 particles for each simulation hour. Actually we used 50 particle sources. Our system has been tested comparing the numerical forecast and mussel microbiological analysis. The simulation spanned the time interval 07/12/2015 Z00 – 21/12/2015 Z23 and the output was stored at a hourly interval. On days 09/12/2015 and 21/12/2015 the local authorities carried out the microbiological analysis on *Mytilus Galloprovincialis* in Punta Terone mussel farm. Results showed a concentration of *E. Coli* much greater than the legal limits in the first day (5400 MPN/100g) and lower in the second day (45 MPN/100g); *Salmonella spp* was absent in both samples.

The mean wind direction started blowing from NW from 7 to 9 December 2015; in this period parti-




| Mussel farm MF$_{PT}$ | | |
|---|---|---|
| **Vertex** | **Latitude** | **Longitude** |
| VC1 | 40°47'10''.14 | 14°05'30''.66 |
| VC2 | 40°47'05''.64 | 14°05'46''.80 |
| VC3 | 40°46'41''.10 | 14°05'42''.49 |
| VC4 | 40°46'41''.64 | 14°05'28''.62 |

Figure 3: Mussel farms locations in the Bay of Pozzuoli. The studied mussel farm is in Punta Terone (MF$_{PT}$) area.

cles spilled out by sources in the eastern part moved towards the center, while particles emitted by sources in the western part remained close to the coast (see Figure 4). After, the mean wind shifted to NE and all particles moved towards the center of the gulf with a progressive increase in the concentration of the tracer in the area surrounding the two mussel farms with a maximum value on day 15/12/2015. Subsequently, the rotating ocean currents contributed to the dispersion of the tracer away from the mussel farms. This picture is also confirmed by microbiological analysis carried out on *Mytilus Galloprovincialis* mussels (Figure 5). The comparison between numerical forecast and microbiological analysis showed a remarkable similarity in trends, although this kind of analysis have to be performed in a more extensively fashion. That confirmed the possibility to use the system as a decision maker tool for applications correlated with sea quality and as a support system for experimental observations and controls.

# 7 CONCLUSION

The quality of coastal marine waters depends strictly on the impact of human activities. Urban settlements, industries, agriculture, livestock farming and weather conditions produce effects which, individually or together, can heavily compromise or even disrupt the equilibrium of aquatic ecosystems.

In this paper we presented our research efforts in designing and developing WaComM, a community water quality model, with the main aim, but not limited to, to develop a forecast system and perform operational numerical predictions in the context of mussel farms management, in order to prevent *E. Coli* and *Salmonella* human diseases with a strong effort in data dissemination for local management decision support (Montella et al., 2007).

WaComM is under continuous active development.

From the implementation point of view a deep refactoring is needed in order to better exploit the hierarchical parallelism. The current implementation leverages on a naive application designed load balancing (Laccetti et al., 2013). A Portable, Extensible Toolkit for Scientific Computation (PETSc) approach could enhance the overall perforance and, above all, the problem scalability (Carracciuolo et al., 2015).

A short-term goal of our project is to extend the studied area to the whole coasts of the Campania Region in order to promote its use as an effective tool dedicated to improve the management of coastal farms. In order to achieve this target, we need to improve the robustness of the WaComM model and the scalability of the offline coupling system.

From the scientific point of view, we will enhance the simulation quality with data collected using the FairWind technology as depicted above improving the data acquisition from boat sensor as interconnected smart devices using the Internet of Things afloat technologies. This issue is a source of novelty even because the whole FairWind ecosystem is based on the SignalK marine data interchange open protocol (http://signalk.org). The proposed system is extensible in order to collect data from other sensors as, but not limited to, surface ph and salinity sensors that could improve the simulation quality and the model validation process and, finally, if supported by

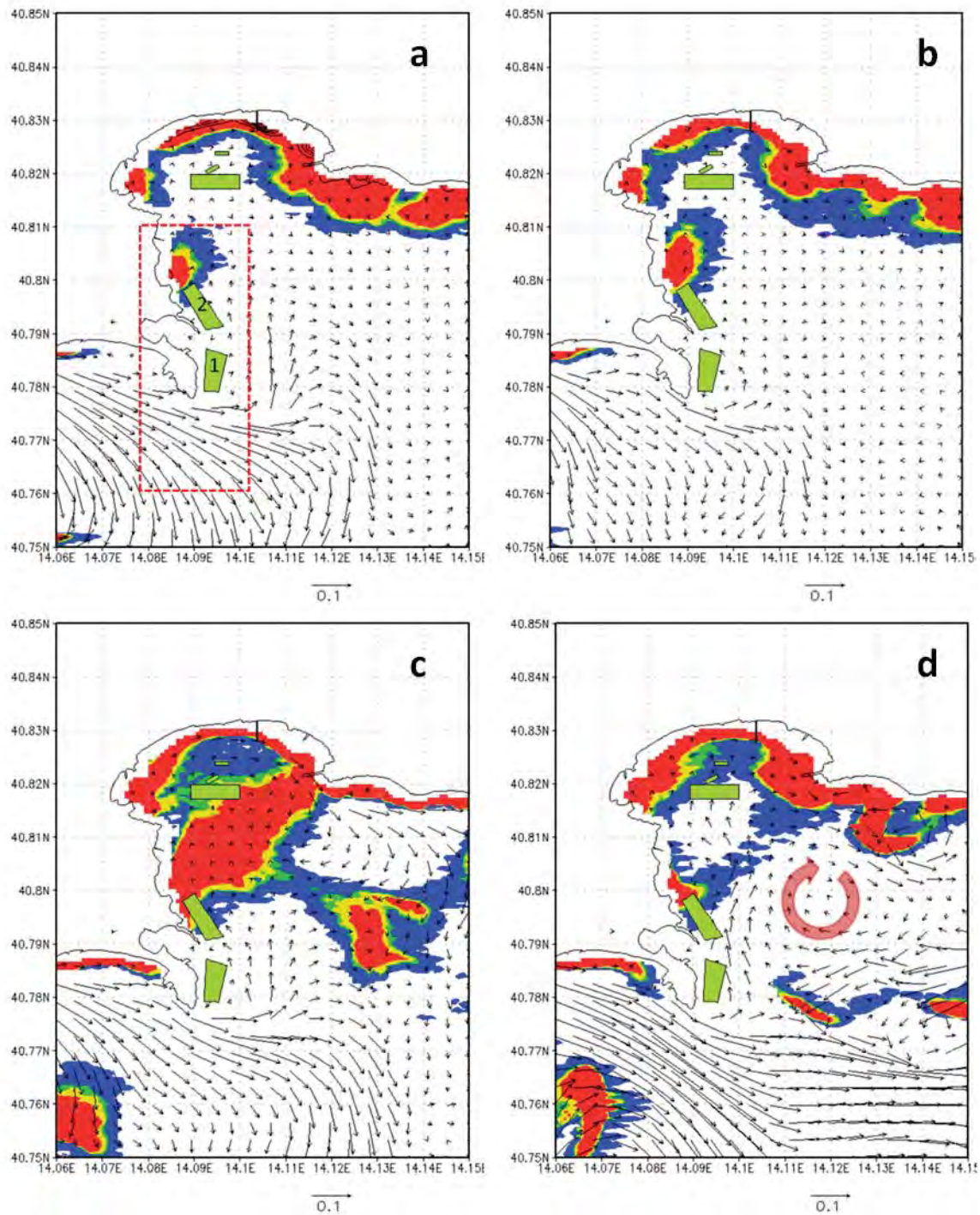



Figure 4: Sea surface currents (vectors) and pollutants concentration (red=high; yellow=medium; green=low; blue=very low; white=absent) in in Gulf of Pozzuoli (Campania Region, Italy) in days 08/12/2015 Z12 (figure A), 09/12/2015 Z12 (figure B), 15/12/2015 Z12 (figure C) and 20/12/2015 Z23 (figure D). The red dotted line is the area of Study with mussel farms in Punta Terone (number 1 in figure A) area.

a ground true based on better microbiological analysis and consistent epidemiological studies on mussels originated enterogastric diseases (Suffredini et al., 2014), enhance the overall system trustability.

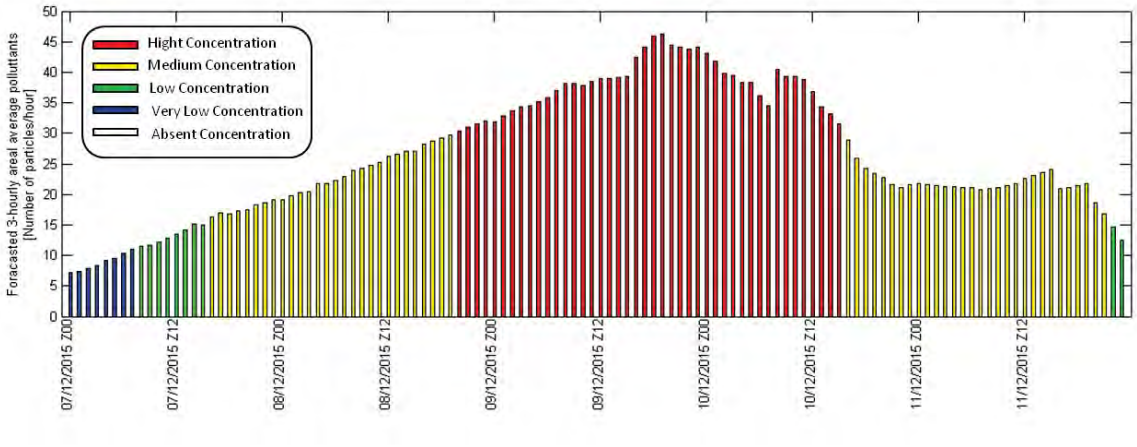



Figure 5: Forecasted averaged particles concentration timeseries in the study area.

## ACKNOWLEDGEMENTS

This work is supported in part by the European Union Grant Agreement number 644312-RAPID–H2020-ICT-2014/H2020-ICT-2014-1 "Heterogeneous Secure Multi-level Remote Acceleration Service for Low-Power Integrated Systems and Devices", using GVirtuS open source software components; project "Development of a software for modeling pollution events based on high-performance and distributed computing technologies with advanced web interface"; project "**MytiluSE** - **M**odelling m**ytilu**s farming **S**ystem with **E**nhanced web technologies"'; NSF cyberSEES program award ACI-1331782; NSF Decision Making Under Uncertainty program award 0951576; DOE contract DE-AC02-06CH11357. IBM-P8 resources were generously provided through an agreement between IBM and the University of Naples "Parthenope" - Department of Science and Technology - High Performance Scientific Computing Smart-Lab.

## REFERENCES

Ascione, I., Giunta, G., Mariani, P., Montella, R., and Riccio, A. (2006). A grid computing based virtual laboratory for environmental simulations. In *European Conference on Parallel Processing*, pages 1085–1094. Springer.

Barone, G., D'Ambra, P., di Serafino, D., Giunta, G., Murli, A., and Riccio, A. (2000). Application of a parallel photochemical air quality model to the campania region (southern italy). *Environmental Modelling & Software*, 15(6):503–511.

Carracciuolo, L., D'Amore, L., and Mele, V. (2015). Toward a fully parallel multigrid in time algorithm in petsc environment: A case study in ocean models. In *High Performance Computing & Simulation (HPCS), 2015 International Conference on*, pages 595–598. IEEE.

Croci, L., De Medici, D., Ciccozzi, M., Di Pasquale, S., Suffredini, E., and Toti, L. (2003). Contamination of mussels by hepatitis a virus: a public-health problem in southern italy. *Food control*, 14(8):559–563.

Cuomo, S., Galletti, A., Giunta, G., and Marcellino, L. (2014a). A class of piecewise interpolating functions based on barycentric coordinates. *Ricerche di Matematica*, 63(11):87–102.

Cuomo, S., Galletti, A., Giunta, G., and Marcellino, L. (2014b). A novel triangle-based method for scattered data interpolation. *Applied Mathematical Sciences*, 8(134):6717–6724.

Cuomo, S., Galletti, A., Giunta, G., and Marcellino, L. (2015). Piecewise hermite interpolation via barycentric coordinates. *Ricerche di Matematica*, 64(2):303–319.

Cuomo, S., Galletti, A., Giunta, G., and Starace, A. (2013). Surface reconstruction from scattered point via rbf interpolation on gpu. In *Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on*, pages 433–440. IEEE.

Fasshauer, G. E. (2007). *Meshfree approximation methods with MATLAB*, volume 6. World Scientific.

Foster, I. (2011). Globus Online: Accelerating and democratizing science through cloud-based services. *IEEE Internet Computing*, 15(3):70.

Galletti, A. and Maratea, A. (2016). A bound for the accuracy of sensors acquiring compositional data. *Procedia Computer Science*, 98:485 – 490. The 7th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2016)/The 6th International Conference on Current and Future Trends

of Information and Communication Technologies in Healthcare (ICTH-2016)/Affiliated Workshops.

Giunta, G., Montella, R., Mariani, P., and Riccio, A. (2005). Modeling and computational issues for air/water quality problems: A grid computing approach. *Nuovo Cimento C Geophysics Space Physics C*, 28:215.

Haidvogel, D. B., Arango, H. G., Hedstrom, K., Beckmann, A., Malanotte-Rizzoli, P., and Shchepetkin, A. F. (2000). Model evaluation experiments in the north atlantic basin: simulations in nonlinear terrain-following coordinates. *Dynamics of Atmospheres and Oceans*, 32(3):239–281.

Haron, N. S., Mahamad, M. K., Aziz, I. A., and Mehat, M. (2009). Remote water quality monitoring system using wireless sensors. In *Proceedings of the 8th WSEAS International Conference on Electronics, Hardware, Wireless and Optical Communications (EHAC'09), Cambridge, UK*, pages 148–154.

Laccetti, G., Lapegna, M., Mele, V., and Romano, D. (2013). A study on adaptive algorithms for numerical quadrature on heterogeneous gpu and multicore based systems. In *International Conference on Parallel Processing and Applied Mathematics*, pages 704–713. Springer.

Madduri, R., Chard, K., Chard, R., Lacinski, L., Rodriguez, A., Sulakhe, D., Kelly, D., Dave, U., and Foster, I. (2015). The Globus Galaxies platform: Delivering science gateways as a service. *Concurrency and Computation: Practice and Experience*, 27(16):4344–4360. CPE-15-0040.

Montella, R., Coviello, G., Giunta, G., Laccetti, G., Isaila, F., and Blas, J. G. (2011). A general-purpose virtualization service for HPC on cloud computing: an application to GPUs. In *International Conference on Parallel Processing and Applied Mathematics*, pages 740–749. Springer.

Montella, R., Di Luccio, D., Troiano, P., Brizius, A., and Foster, I. (2016a). WaComM: A parallel Water quality Community Model for pollutant transport and dispersion operational predictions. In *International Workshop on Numerical Algorithms and Methods for Data Analysis and Classification, 2016. NAMDAC'16. IEEE*. IEEE.

Montella, R., Ferraro, C., Kosta, S., Pelliccia, V., and Giunta, G. (2016b). Enabling android-based devices to high-end gpgpus. In *Algorithms and Architectures for Parallel Processing*, pages 118–125. Springer.

Montella, R., Giunta, G., and Riccio, A. (2007). Using grid computing based components in on demand environmental data delivery. In *Proceedings of the second workshop on Use of P2P, GRID and agents for the development of content networks*, pages 81–86. ACM.

Montella, R., Kelly, D., Xiong, W., Brizius, A., Elliott, J., Madduri, R., Maheshwari, K., Porter, C., Vilter, P., Wilde, M., et al. (2015). FACE-IT: A science gateway for food security research. *Concurrency and Computation: Practice and Experience*, 27(16):4423–4436.

Montella, R., Luccio, D. D., Ferraro, C., Izzo, F., Troiano, P., and Giunta, G. (2016c). Fairwind: a marine data crowdsourcing platform based on internet of things and mobile/cloud computing technologies. In *8th International Workshop on Modeling the Ocean (IWMO)*, Bologna, Italy.

Ortega, G., Puertas, A., de Las Nieves, F. J., and Martin-Garzón, E. (2016). Gpu computing to speed-up the resolution of microrheology models. In *Algorithms and Architectures for Parallel Processing*, pages 457–466. Springer.

Pham, Q., Malik, T., Foster, I. T., Di Lauro, R., and Montella, R. (2012). SOLE: Linking Research Papers with Science Objects. In *IPAW*, pages 203–208. Springer.

Rodean, H. C. (1996). *Stochastic Lagrangian models of turbulent diffusion*, volume 45. Springer.

Shchepetkin, A. F. and McWilliams, J. C. (2003). A method for computing horizontal pressure-gradient force in an oceanic model with a nonaligned vertical coordinate. *Journal of Geophysical Research: Oceans*, 108(C3).

Shchepetkin, A. F. and McWilliams, J. C. (2005). "the regional oceanic modeling system (roms): a split-explicit, free-surface, topography-following-coordinate oceanic model". *Ocean Modelling*, 9(4):347 – 404.

Skamarock, W. C., Klemp, J. B., and Dudhia, J. (2001). Prototypes for the wrf (weather research and forecasting) model. In *Preprints, Ninth Conf. Mesoscale Processes, J11–J15, Amer. Meteorol. Soc., Fort Lauderdale, FL*.

Suffredini, E., Lanni, L., Arcangeli, G., Pepe, T., Mazzette, R., Ciccaglioni, G., and Croci, L. (2014). Qualitative and quantitative assessment of viral contamination in bivalve molluscs harvested in italy. *International journal of food microbiology*, 184:21–26.

Wilkin, J. L., Arango, H. G., Haidvogel, D. B., Lichtenwalner, C., Glenn, S. M., and Hedström, K. S. (2005). A regional ocean modeling system for the long-term ecosystem observatory. *Journal of Geophysical Research: Oceans*, 110(C6).

# Short Papers

# Discussions of a Preliminary Hand Hygiene Compliance Monitoring Application-as-a-service

Peng Zhang[1], Marcelino Rodriguez-Cancio[2], Douglas Schmidt[1], Jules White[1] and Tom Dennis[3]
[1]*Institute for Software Integrated Systems, Vanderbilt University, Nashville, TN, U.S.A.*
[2]*Research Institute of Computer Science and Random Systems, University of Rennes 1, Rennes, France*
[3]*ZH Solutions, Nashville, TN, U.S.A.*
{*peng.zhang, jules.white*}*@vanderbilt.edu, marcelino.rodriguez-cancio@irisa.fr, schmidt@dre.vanderbilt.edu, tdennis@zhs.io*

Keywords: Machine Learning, Deep Learning, Data Mining, Hospital and Healthcare, Hand Hygiene Compliance, Re-admissions.

Abstract: Hospital Acquired Infections (HAIs) are a global concern as they impose significant economic consequences on the healthcare systems. In the U.S. alone, HAIs have cost hospitals an estimated $9.8 billion a year. An effective measure to reduce the spread of HAIs is for Health Care Workers (HCWs) to comply with recommended hand hygiene (HH) guidelines. Unfortunately, HH guideline compliance is currently poor, forcing hospitals to implement controls. The current standard for monitoring compliance is overt direct observation of hand sanitation of HCWs by trained observers, which can be time-consuming, costly, biased, and sporadic. Our research describes a hand hygiene compliance monitoring app, Hygiene Police (HyPo), that can be deployed as a service to alleviate the manual effort, reduce errors, and improve existing compliance monitoring practice. HyPo exploits machine learning analyses of handwashing compliance data from a 30-bed intensive care unit to predict future compliance characteristics. Based on the results, HyPo then provides HWCs with timely feedback and augments the current monitoring approach to improve compliance.

## 1 INTRODUCTION

**Emerging Concerns in Healthcare.** Hospital Acquired Infections (HAIs) are occupational infections acquired by Healthcare Workers (HCWs) or by patients in healthcare facilities that appear after patient discharge (WHO, 2009). HAIs represent significant health problems, with a considerable economic impact on patients and hospitals worldwide.

Persistent exposures to diseases and lack of appropriate hand hygiene (HH) practice can cause HCWs' hands to become carriers of infections transmitted to patients or other staff through physical contact. To reduce re-admission rates, therefore, HCWs are expected to comply with HH guidelines to prevent the spread of HAIs in healthcare facilities. The current standard practice of compliance monitoring is for covert human auditors to unobtrusively observe and record HH compliance of medical workers. Unfortunately, this approach is costly and subject to bias (Boyce et al., 2009) due to the Hawthorne effect (Eckmanns et al., 2006), which occurs when subjects of a study change their behavior due to awareness of being observed.

**Contribution.** Based on our preliminary study on HH compliance characteristics using machine learning (Zhang et al., 2016), this work proposes a hand hygiene compliance monitoring app, Hygiene Police (HyPo), that can be deployed as a service. The goal of this app is to mitigate the laborious effort and reduce errors of direct observation.

**App Workflow.** HyPo is implemented as a Java-based desktop app that communicates to and from Bluetooth Low-Energy (BLE) devices equipped at the facility from our previous study (Zhang et al., 2016). The schematic in Figure 1 depicts the overall app workflow, which is divided into the following three stages (the last two are the core components of HyPo):

1. **Data Acquisition**, where raw data is acquired from the BLE devices.
2. **Data Mining**, where the raw data undergoes a data mining process provisioned by HyPo to produce a set of features that is fed to Feature Selection algorithms to obtain a Sanitized Dataset. The Feature Selection is done to improve the execu-



Zhang, P., Rodriguez-Cancio, M., Schmidt, D., White, J. and Dennis, T.
Discussions of a Preliminary Hand Hygiene Compliance Monitoring Application-as-a-service.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 537-544
ISBN: 978-989-758-213-4

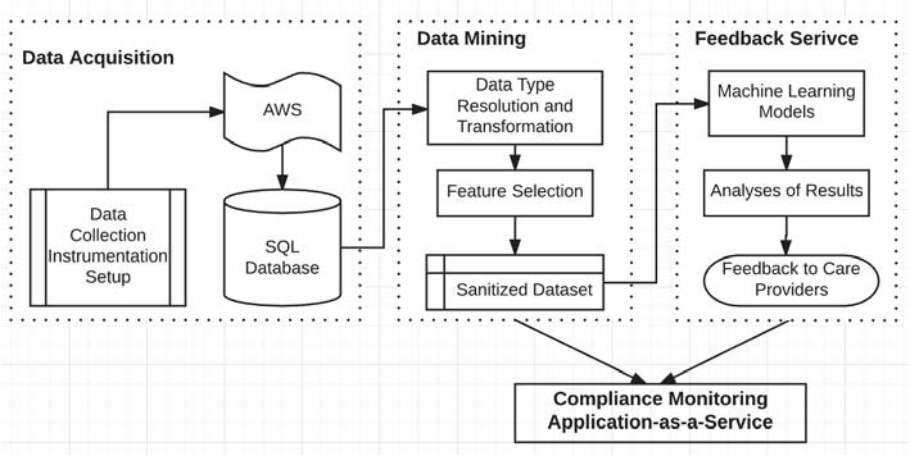



Figure 1: The Overall System Flow: from Data Collection to Post-prediction Analysis and Feedback Service.

tion performance of the Machine Learning (ML) methods that will follow by determining the most relevant features and removing the others from the Sanitized Dataset.

3. **Feedback Service**, where the ML models are run over the Sanitized Dataset to produce feature set that can be used to provide timely feedback to healthcare providers.

**Paper Organization.** The remainder of this paper is organized as follows: Section 2 defines key terms frequently referenced throughout the paper; Section 3 describes the data collection setup; Section 4 details the data preparation and mining process; Section 5 describes Hypo's feedback service and how it complements the direct observation approach; Section 6 surveys and compares existing research in the area of hand hygiene compliance improvement and monitoring applications; and Section 7 presents concluding remarks and outlines future extensions of this work.

## 2 HAND HYGIENE COMPLIANCE OVERVIEW

This section defines the following terms that are used frequently in the paper:

1. **Hand hygiene opportunity**: an opportunity for hand cleaning is presented before each care provider's entry/exit of a patient room.
2. **Hand hygiene/handwashing compliance**: each hand hygiene opportunity requires one hand hygiene action, which should be a corresponding positive (compliance) or negative (non-compliance) action (EFORE, 2009).
3. **Entry compliance**: hand hygiene compliance observed at staff's entry to a patient room, determined by *wash on entry*.
4. **Exit compliance**: hand hygiene compliance observed at staff's exit from a patient room, determined by *wash on exit*.
5. **Wash on entry**: hand hygiene action at patient room entry that determines entry compliance, true if performed and false otherwise.
6. **Wash on exit**: hand hygiene action at patient room exit that determines exit compliance, true if performed and false otherwise.

Our previous study collected 60 days of care providers' real-time location and handwashing data, from an intensive care unit (ICU) equipped with 30 beds, and observed two major correlating factors to compliance: (1) entry compliance has an 89% accuracy on predicting exit compliance and (2) exit compliance can predict entry compliance at the next visit (for the same staff) at an accuracy as high as 77%. Likewise, location data was observed to have a minor impact on predicting exit compliance (Zhang et al., 2016).

Based on this previous study, in the HyPo app we compiled the following rules of hand hygiene compliance that ICU staff should abide by:

1. All on-duty staff at the ICU were required to wear a BLE badge.
2. All staff were required to sanitize their hands within a short interval of 2 minutes upon entering a patient room and before exiting the same room.
3. Each compliant action should be associated with an activation of a specific soap dispenser with disinfectant solution against Clostridium difficile,

a common HAI spread through physical contact (Shrestha et al., 2016). These dispensers are located both inside and outside each patient room.

These rules only apply to this ICU but can be configured to work with other caregiving settings. The rest of this paper describes the application of HyPo using the same device-equipped 30-bed ICU from our previous study (Zhang et al., 2016) as an example.

# 3 DATA ACQUISITION

This section describes the data acquisition process, where real-time location data and handwashing station activation data is recorded, and then provides an overview of the essential data fields extracted from the collection. The process described in this section is one approach of obtaining the hand hygiene compliance data to provide input for our app, but it is by no means the only option to acquire this type of data.

## 3.1 Instrumentation Configurations

The ICU with HyPo deployment was equipped with a Bluetooth Low-Energy (BLE) indoor positioning system that provides room-level accuracy for reporting staff locations in real-time. The system produced the location data for all staff with BLE badges.

The ICU also deployed Gojo Active Monitoring handwashing stations, which record each dispenser activation. These activation events were then combined with real-time location data to track individual staff handwashing compliance. The system expected to receive at least one handwashing event from either a sanitation station inside of the room or a station immediately outside the room within two minutes prior to entry, abiding the facility rules described in Section 2. Similarly, two minutes before or after room exit, the system expected one handwashing event from either sanitation stations.

Overall, the dataset collected at the studied ICU contains 8 weeks of events recording activities from 180+ soap dispensers activated by 60 badged nurses 24 hours a day. All raw event data was streamed to a data storage on Amazon Web Services (AWS), which was post processed and output to a SQL database. We then extrapolated the data fields of interest for compliance predictions and analyses.

## 3.2 Dataset Limitations

Although real-time location data was acquired and handwashing station activations recorded at the ICU, the dataset was still an estimate rather than a ground truth of hand hygiene compliance. The dataset collected has a number of key threats to validity, including: (1) not all staff wore their BLE badges at all times, (2) the system could not differentiate activations from badged vs. non-badged visitors/staff, and (3) subsets of the monitoring equipment went offline at some intervals and prevented data capture in certain rooms.

However, we did not consider these limitations as fatal to our study results because we could either easily eliminate the data entries associated with these threats or discard the marginal impact that the threats had on our findings.

## 3.3 Dataset Schema

From the SQL database we obtained an initial dataset by omitting certain data fields with extraneous information, such as device IDs of the wearable badges, internally-used identifiers of the patient rooms, etc. The data fields associated to each patient room visit event that we deemed essential thus extracted from the database include:

1. *Staff ID* - ID of badge worn by the staff who has been associated with a patient room visit
2. *Location* - patient room number visited by the badged staff
3. *Entry time* - timestamp (in CDT) at which the badged nurse entered the patient room
4. *Exit time* - timestamp (in CDT) at which the badged nurse exited the patient room
5. *Wash on entry* - a boolean value indicating whether the staff properly performed hand hygiene on patient room entry
6. *Wash on exit* - a boolean value indicating if the staff properly performed handwashing on patient room exit
7. *Duration* - for how long (in milliseconds) the staff was in the patient room

# 4 DATA PREPARATION

This section discusses how we prepared the collected data to maximize the utilization of our machine learning classifiers, which is an important capability offered by HyPo. This process is the same as that employed in (Zhang et al., 2016) to assist the analyses and characterization of hand hygiene compliance. Other influencing factors of hand hygiene compliance may be discovered as more relevant data becomes

available, such as patient admittance details, medical records of admitted patients, facility regulations of compliance, etc.

Despite the specificity of the dataset used throughout this paper, the data mining process provided by HyPo as described below can be generalized to support transformations of different forms of data collected in other facilities.

## 4.1 Data Transformation

Most machine learning (ML) classifiers yield better results when the input dataset is structured in certain ways. For example, suppose we want to know if the day of week (Monday to Sunday) influences compliance, some ML classifiers will yield better results if we express date as a set of integers ranging from 1 to 7, as opposed to a real continuous stream of timestamps expressed in milliseconds.

As another example, our location data consists of room numbers, which provides little information regarding spatial distribution of the rooms. If we want to know whether compliance decreases in nearby locations, we must first transform the room numbers into coordinates on the facility's floor plan, for instance.

To obtain a transformed schema that can be better handled by our classifiers, we took the collected dataset and performed the following transformations over it:

1. We converted all event data from the original timestamp format into an integer field with range 1 to 7 to represent day of week, an integer field with range 1 to 4 to represent time of day in morning, afternoon, evening, bedtime, and another integer data field of 0-23 to represent hour of day. The numeric representations of the original nominal time stamp data will allow our classifiers to achieve higher classification accuracy.
2. We mapped each patient room on the ICU floor plan to a set of *x* and *y* coordinates to identify the spacial location. We then extended each entry in the dataset to include these corresponding coordinates of the patient room.
3. For each data point we added new fields to include the previous record of the corresponding badged staff's handwashing data, *i.e.*, duration, location, washed on entry, and washed on exit. To ensure data integrity, we removed all entries that did not have previous handwashing records.

As a result of these transformations, we obtained a new schema consisting of a minimal set of features that our application expects to receive for best accuracy:

1. staff ID - integer
2. location (room number) - integer
3. washed on entry - TRUE/FALSE
4. washed on exit - TRUE/FALSE
5. duration (s) - length of patient room visit in seconds, integer
6. entry hour - hour of day on room entry, 0-23
7. exit hour - hour of day on room exit, 0-23
8. entry time - time of day on recorded room entry in Morning (1), Afternoon (2), Evening (3), and Bedtime (4)
9. exit time - time of day on recorded patient room exit, 1-4
10. entry day of week - day of week on recorded patient room entry, 1-7
11. exit day of week - day of week on room exit, 1-7
12. location X coordinate - x coordinate of patient room on the ICU floor plan
13. location Y coordinate - y coordinate of patient room on the ICU floor plan
14. previous duration (s) - duration of the same staff's previous patient room visit in seconds
15. previous washed on entry - dispenser activation on previous room entry TRUE/FALSE
16. previous washed on exit - dispenser activation on previous room exit TRUE/FALSE
17. previous location - previously visited patient room number

## 4.2 Feature Selection

After we transformed the dataset into a features set, we executed a feature selection process to automatically select feature subsets in our transformed data that best (1) reduced overfitting of data, (2) improved classification accuracy, and (3) decreased model training time (Guyon and Elisseeff, 2003). Although we do not have a significantly large feature list produced for this ICU, it is still useful to apply this technique to select the most relevant subsets of features to help produce the most accurate feedback in the next step.

To automatically select features from the transformed dataset, HyPo applies a supervised attribute selection filter from the open source Weka Java library (Hall et al., 2009). The filter is composed of two pieces: (1) a feature *Evaluator* to determine how

features are evaluated and (2) a *Search Method* to navigate the feature's search space. Our app runs feature selection using the following pairs of *Evaluators* and *Search Methods*, as shown in Table 1:

1. *Evaluator*: CfsSubsetEval that evaluates a subset of features by considering each feature's predictive ability and the degree of redundancy between them.

   *Search Method*: GreedyStepwise with a backward search through the space of attribute subsets.

2. *Evaluator*: InfoGainAttributeEval that evaluates an attribute's worth by measuring the information gain with respect to the class variable to classify.

   *Search Method*: Ranker that ranks features by their individual evaluations with an optional parameter of 6 features in the output subset

3. *Evaluator*: WrapperSubsetEval (Kohavi and John, 1997) with NaiveBayes (John and Langley, 1995) as the basic learning scheme and a 10-fold cross validation to use for estimating accuracy.

   *Search Method*: GeneticSearch that performs a search using the simple genetic algorithm (Goldberg, 1989)

Table 1: Evaluator and Search Method Pairs Used in Feature Selection

| Evaluator | Search Method |
|---|---|
| CfsSubsetEval | GreedyStepwise |
| InfoGainAttributeEval | Ranker |
| WrapperSubsetEval | GeneticSearch |

Our previous study (Zhang et al., 2016) observed two highly correlating factors of compliance using the data collected in the same 30-bed ICU. We could configure HyPo to select only these two features from the dataset to use for determining feedback provision. However, automatic feature selection is an integral piece because as the dataset increases in size and dimension, our enhanced app can continuously combine features or intelligently adjust the correlating features to maximize classification accuracy.

# 5 FEEDBACK SERVICE

This section first describes the machine learning models employed by HyPo and then presents the feedback service that uses these models to provide timely feedback and to complement the direct observation approach to hand hygiene compliance monitoring.

## 5.1 Machine Learning Models

After preparing the dataset, we split the data to 65% for training, 10% for cross validation, and the remaining 25% for testing the ML models. Based on the compliance prediction observations from the previous study in (Zhang et al., 2016), we employed the top three classifiers, one from Weka (Hall et al., 2009) and two deep nets from DeepLearning4J (DL4J) (Team, 2016) to serve as our models for classifying *washed on entry* and *washed on exit*. HyPo then uses the results with highest accuracy.

- The **Sequential Minimal Optimization** (SMO) implementation of the Support Vector Machine (SVM), which uses heuristics to partition the training problem into smaller sub-problems and uses pairwise linear regression to classify. This method is usually resilient to data overfitting and by default normalizes the input data (Platt et al., 1998).
- The **Feed-Forward Neural Network** (FFNN), which is a one direction (from input to output) artificial neural network that performs classifications based on weight calculations of the network nodes (Glorot and Bengio, 2010). Using the DL4J Java library, we developed a 3-layer FFNN with a random seed of 6, 1000 iterations, a learning rate of 0.1, and the Stochastic gradient descent optimization algorithm (Gardner, 1984).
- The **Recurrent Neural Network** (RNN), which has a feedback loop whereby the immediately previous step's output is fed back to the net to affect the outcome of the current step. We used a 3-layer RNN with two Graves' Long Short-Term Memory (LSTM) layers (Graves, 2012) (input and hidden) and an output layer along with the same parameters as the FFNN.

## 5.2 Just-in-Time Alerting

With our previously characterized predictability of compliance (Zhang et al., 2016), as described in Section 2, and necessary pre-configurations to the data collection instruments, HyPo can provide just-in-time alerting to remind HCWs to perform hand hygiene when they are predicted not to comply, using either a *singular* prediction or a *chain-prediction* scheme, depending on if there is adequate time to provide such notifications between each hand hygiene opportunity.

Suppose that HyPo has just observed a staff nurse's compliance on a patient room entry, then the ML classifiers will predict the same staff's exit compliance. For instance, if the staff is predicted to be

non-compliant, an alert of red flashing light can be sent to either the wearable badge or the device at the appropriate dispenser activation station as a reminder to the staff; otherwise, no alert is necessary.

If duration of the visit is too short of an interval to send the notification signal to the devices, then we can use the probability chain rule (Schum, 1994) to provide a backup alert to the same staff if necessary. In this case, the ML models will use the predicted entry compliance for the current visit (from the staff's exit compliance of the previous visit) to determine exit compliance of the current visit at a probability of 89% * 77% = 69%. It is less ideal, but the likelihood of the visit interval being too short is minimal because the grace period for compliance is set at two minutes, and if a room visit is within two minutes, hand hygiene compliance is not required.

### 5.3 Recommend Training Material

If a staff member is frequently predicted as non-compliant over a long observation period, HyPo (with integrated email capabilities) can recommend hand hygiene guidelines or appropriate training materials to the staff via email. The goal is to improve compliance on an individual basis.

### 5.4 Assist Direct Observation

The compliance prediction results can also be used to assist the current standard practice of direct observation. With predicted non-compliance reoccurring at a certain location (*i.e.*, a patient room), HyPo can deploy a human auditor (*e.g.*, by sending a notification) to observe compliance at the location that should be given most attention.

## 6 RELATED WORK

Due to worldwide high demands of HAI prevention, a number of other researchers have studied approaches to improve hand hygiene compliance. Although the gold standard monitoring method is human-centric (WHO, 2009), (Gould et al., 2011), a wide rage of studies propose electronic or electronically assisted hand hygiene compliance monitoring and intervention systems (Ellingson et al., 2014), (Ward et al., 2014). This section compares our work on the HyPo app with common electronic intervention systems including (1) technology-assisted direct human observation, (2) counting systems, and (3) automated monitoring systems.

**Technology-assisted Human Observation.** Direct observation is the most precise way of controlling compliance. Several studies use technologies such as handheld devices and cameras to aid human observation, aiming at reducing input errors, costs, and time consumption. Handheld devices are used for data entry, and video cameras provide opportunities to reduce the Hawthorne effect and observe locations that are remote or hard to access.

Chen et al (Chen et al., 2013), used wireless data entry devices and a website to allow human observers to audit compliance. University of North Carolina Hospitals implemented a "clean-in, clean-out" system that allowed covert observers and designated nurses to track compliance using a mobile app and a web portal (Sickbert-Bennett et al., 2016).

Cameras have been used by Armellino (Armellino et al., 2013) to increase compliance in an ICU. The study connected motion sensors near the sinks that would activate cameras being monitored by remote auditors. The study by Davis (Davis, 2010) placed a discreet camera at the entrance of a ward and assessed compliance before and after a sink was placed pointing to the dispenser.

Unfortunately, these methods still require human interaction and can bias the results, as the medical workers know they are being directly observed. Moreover, audits require trained personnel who are regularly monitored to ensure quality control.

**Counting Systems.** Installing counting devices to measure the remaining sanitation product volume or the number of dispenser activation times is a quiet method that is not subject to the Hawthorne effect. A counter may detect usage patterns and frequency changes.

Marras (Marra et al., 2010) used dispenser counters along with direct observation to assess whether positive deviance in hand hygiene behaviors could have an impact on reducing HAIs. A downside to this approach, however, is that counter systems cannot tell who used the dispensers and therefore are unable to evaluate compliance by itself. Morgan et al (Morgan et al., 2012) provided evidence to support the claim that dispenser usage data could be more reliable than direct human observation to estimate hand hygiene compliance.

**Automated Monitoring Systems using Wearables.** Many automated monitoring systems are capable of producing feedback or reminders in real or near real time without human intervention, similar to our approach.

Fakhry used a motion-triggered system with audible hand washing reminders at each medical department entrance (Fakhry et al., 2012). Sahud and Bhanot developed an electronic hand hygiene feedback device that reports real-time compliance rate on a liquid-crystal display visible to all staff in the intervention unit (Sahud and Bhanot, 2009). Edmond et al installed a sensor network using a credit-card sized sensor badge on each alcohol dispenser, which when not activated on room entry or exit beeped with a red indicating light (Edmond et al., 2010). Similarly, Marra et al employed a wireless network with sensors on the alcohol dispensers that provide real-time flashing light feedback to HCWs for hygiene activity detection (Marra et al., 2014). Most recently, Ellison et al proposed a prospective electronic hand hygiene room entry/exit audible reminder system (Ellison et al., 2015) that provides a combination of 24-hour-a-day recording of hand hygiene activities and real-time computer monitor performance feedback.

**Differentiating Factors of Our Approach.** All the prior research we reviewed collected data to propose strategies that increased hand hygiene performance or gather conclusions regarding the efficacy of a specific approach. Our HyPo app is unique since it uses the gathered data to predict future compliance behavior instead of notifying appropriate caregivers after non-compliance has been detected. Other approaches *react* to non-compliance, while ours *predicts* compliance ahead of time.

We presented a novel methodology using ML algorithms, which is also unique to our work. Hence, the aim of our work is also a differentiating factor. In particular, HyPo evaluates the prediction capabilities of different ML algorithms to predict compliance ahead of time.

# 7 CONCLUDING REMARKS

This paper presented a hand hygiene monitoring app called Hygiene Police (HyPo) that can be deployed as a service to complement the current monitoring approach and improve compliance. We showed an example data collection process taken place at a 30-bed ICU where we acquired the handwashing compliance data. We also described the data transformation process HyPo employs to maximize the utilization of the selected machine learning (ML) classifiers.

Combining the results of real-time compliance predictions using the correlations identified in (Zhang et al., 2016), HyPo can provide three types of services: (1) just-in-time alerting to remind predicted non-compliant staff to perform hand hygiene, (2) recommending training materials to habitually non-compliant staff via email, and (3) assisting the direct observation approach by deploying human auditors at the opportune time and place when and where non-compliance is frequently predicted to occur. We also compared our app to related research work and found that our approach *predicted* future compliance behavior instead of *reacted* to non-compliance as in other approaches. Our methodology using ML algorithms is unique and is the only work that evaluates ML prediction capabilities in this domain.

In future work, we plan on collecting more compliance data, ideally using the same process as discussed in the paper. We will use this data to fine tune the parameters in our ML classifiers to increase the prediction accuracy. We will also run simulations that test whether our HyPo app can improve compliance in general and if the improvement can be sustained over time in a range of caregiving settings.

# REFERENCES

Armellino, D., Trivedi, M., Law, I., Singh, N., Schilling, M. E., Hussain, E., and Farber, B. (2013). Replicating changes in hand hygiene in a surgical intensive care unit with remote video auditing and feedback. *American Journal of Infection Control*, 41(10):925–927.

Boyce, J. M., Cooper, T., and Dolan, M. J. (2009). Evaluation of an Electronic Device for Real-Time Measurement of Alcohol-Based Hand Rub Use. *Infection Control & Hospital Epidemiology*, 30(11):1090–1095.

Chen, L. F., Carriker, C., Staheli, R., Isaacs, P., Elliott, B., Miller, B. A., Anderson, D. J., Moehring, R. W., Vereen, S., Bringhurst, J., Rhodes, L., Stritholt, N., and Sexton, D. J. (2013). Observing and improving hand hygiene compliance: Implementation and refinement of an electronic-assisted direct-observer hand hygiene audit program. *Infection control and hospital epidemiology: the official journal of the Society of Hospital Epidemiologists of America*, 34(2):207–210.

Davis, C. R. (2010). Infection-free surgery: how to improve hand-hygiene compliance and eradicate methicillin-resistant Staphylococcus aureus from surgical wards. *Annals of the Royal College of Surgeons of England*, 92(4):316–319.

Eckmanns, T., Bessert, J., Behnke, M., Gastmeier, P., and Rüden, H. (2006). Compliance with antiseptic hand rub use in intensive care units the hawthorne effect. *Infection Control*, 27(09):931–934.

Edmond, M. B., Goodell, A., Zuelzer, W., Sanogo, K., Elam, K., and Bearman, G. (2010). Successful use of alcohol sensor technology to monitor and report hand hygiene compliance. *The Journal of Hospital Infection*, 76(4):364–365.

EFORE, B. (2009). Hand hygiene technical reference manual.

Ellingson, K., Haas, J. P., Aiello, A. E., Kusek, L., Maragakis, L. L., Olmsted, R. N., Perencevich, E., Polgreen, P. M., Schweizer, M. L., Trexler, P., VanAmringe, M., and Yokoe, D. S. (2014). Strategies to Prevent Healthcare-Associated Infections through Hand Hygiene. *Infection Control and Hospital Epidemiology*, 35(8):937–960.

Ellison, R. T., Barysauskas, C. M., Rundensteiner, E. A., Wang, D., and Barton, B. (2015). A Prospective Controlled Trial of an Electronic Hand Hygiene Reminder System. *Open Forum Infectious Diseases*, page ofv121.

Fakhry, M., Hanna, G. B., Anderson, O., Holmes, A., and Nathwani, D. (2012). Effectiveness of an audible reminder on hand hygiene adherence. *American Journal of Infection Control*, 40(4):320–323.

Gardner, W. A. (1984). Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique. *Signal Processing*, 6(2):113–133.

Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Aistats*, volume 9, pages 249–256.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley.

Gould, D. J., Drey, N. S., and Creedon, S. (2011). Routine hand hygiene audit by direct observation: has nemesis arrived? *The Journal of Hospital Infection*, 77(4):290–293.

Graves, A. (2012). Supervised sequence labelling. In *Supervised Sequence Labelling with Recurrent Neural Networks*, pages 5–13. Springer.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.

John, G. H. and Langley, P. (1995). Estimating continuous distributions in bayesian classifiers. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 338–345, San Mateo. Morgan Kaufmann.

Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324. Special issue on relevance.

Marra, A. R., Guastelli, L. R., Arajo, C. M. P. d., Santos, J. L. S. d., Lamblet, L. C. R., Silva, M., Lima, G. d., Cal, R. G. R., Paes, n. T., Neto, M. C., Barbosa, L., Edmond, M. B., and Santos, O. F. P. d. (2010). Positive Deviance A New Strategy for Improving Hand Hygiene Compliance. *Infection Control & Hospital Epidemiology*, 31(1):12–20.

Marra, A. R., Sampaio Camargo, T. Z., Magnus, T. P., Blaya, R. P., Dos Santos, G. B., Guastelli, L. R., Rodrigues, R. D., Prado, M., Victor, E. d. S., Bogossian, H., Monte, J. C. M., dos Santos, O. F. P., Oyama, C. K., and Edmond, M. B. (2014). The use of real-time feedback via wireless technology to improve hand hygiene compliance. *American Journal of Infection Control*, 42(6):608–611.

Morgan, D. J., Pineles, L., Shardell, M., Young, A., Ellingson, K., Jernigan, J. A., Day, H. R., Thom, K. A., Harris, A. D., and Perencevich, E. N. (2012). Automated hand hygiene count devices may better measure compliance than human observation. *American Journal of Infection Control*, 40(10):955–959.

Platt, J. et al. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines.

Sahud, A. G. and Bhanot, N. (2009). Measuring hand hygiene compliance: a new frontier for improving hand hygiene. *Infection Control and Hospital Epidemiology*, 30(11):1132.

Schum, D. A. (1994). *The evidential foundations of probabilistic reasoning*. Northwestern University Press.

Shrestha, S. K., Sunkesula, V. C., Kundrapu, S., Tomas, M. E., Nerandzic, M. M., and Donskey, C. J. (2016). Acquisition of clostridium difficile on hands of healthcare personnel caring for patients with resolved c. difficile infection. *Infection Control & Hospital Epidemiology*, 37(04):475–477.

Sickbert-Bennett, E. E., DiBiase, L. M., Schade Willis, T. M., Wolak, E. S., Weber, D. J., and Rutala, W. A. (2016). Reducing health careassociated infections by implementing a novel all hands on deck approach for hand hygiene compliance. *American Journal of Infection Control*, 44(5, Supplement):e13–e16.

Team, D. D. (2016). Deeplearning4j: Open-source distributed deep learning for the jvm. *Apache Software Foundation License*, 2.

Ward, M. A., Schweizer, M. L., Polgreen, P. M., Gupta, K., Reisinger, H. S., and Perencevich, E. N. (2014). Automated and electronically assisted hand hygiene monitoring systems: A systematic review. *American Journal of Infection Control*, 42(5):472–478.

WHO (2009). A guide to the implementation of the WHO multimodal hand hygiene improvement strategy.

Zhang, P., White, J., Schmidt, D., and Dennis, T. (2016). A preliminary study of hand hygiene compliance characteristics with machine learning methods. (ISIS-16-101).

# A Real-time m-Health Monitoring System: An Integrated Solution Combining the Use of Several Wearable Sensors and Mobile Devices

Salvatore Naddeo[1], Laura Verde[2], Manolo Forastiere[3], Giuseppe De Pietro[1] and Giovanna Sannino[1]
[1]*Institute of High Performance Computing and Networking (ICAR), National Research Council of Italy (CNR), via Pietro Castellino 111, 80131 Naples, Italy*
[2]*Department of Technology, University of Naples Parthenope, 80143 Naples, Italy*
[3]*Neatec S.p.a. R&D Support, Via Campi Flegrei 34, 80078 Pozzuoli, Italy*
{*salvatore.naddeo, giuseppe.depietro, giovanna.sannino*}*@icar.cnr.it, laura.verde@uniparthenope.it, m.forastiere@neatec.it*

Keywords: Physiological Monitoring, Wearable Device, m-Health System, Continua Alliance Guidelines, Chronic Diseases.

Abstract: Nowadays the upsurge in the prevalence of chronic diseases represents an increasing burden on individuals and society. Chronic diseases can require repeated and frequent hospital treatment to control vital parameters of interest. The use of automatic instruments for a real-time monitoring of biological parameters constitutes a valid instrument to improve the patient's quality of life. The integration of mobile communications with wearable devices has facilitated the shift of healthcare assistance from clinic-centric to patient-centric monitoring. In this paper, a real-time monitoring system is proposed. The system is conceptualized to provide an instrument for patients, by means of which they can easily monitor, analyse and save their own vital signs using wearable sensors and an Android device such as a smartphone or tablet, offering an efficient solution in terms of a decrease in time, human error and cost.

## 1 INTRODUCTION

During the recent decade the demographic changes in developed countries resulting in a more elderly population and the increasing prevalence of chronic diseases have contributed to the need for a constant monitoring of the state of patients' health.

According to the World Health Organization (World Health Organization, 2016), chronic diseases such as coronary heart disease, cancer, chronic obstructive pulmonary disease and diabetes mellitus type 2 constitute the leading cause of mortality in the world, representing about 60% of all deaths. Chronic diseases are primarily attributable to heart failure, currently the main cause of death in most western countries. The 2016 Heart Disease and Stroke Statistics update of the American Heart Association (AHA) reported that 15.5 million people in the USA suffer from cardiovascular disease, this prevalence increasing with age for both women and men (Sanchis-Gomar et al., 2016). In Europe, in the last 12 months the prevalence of people reporting heart problems was 9.2% for both sexes (Townsend et al., 2016). In particular, in Italy cardiovascular diseases are the main cause of death, responsible for 44% of all deaths (Centro nazionale di epidemiologia, sorveglianza e promozione della salute dell'Istituto superiore di sanità, 2016).

Chronic diseases also have a negative impact on the quality of people's life. Patients suffering from these pathologies must, often, carry out a monitoring of physiological parameters such as heart rate and blood pressure as well as take control of the main risk factors that can aggravate their state of health. In less dangerous cases, it is convenient to monitor patients outside the hospital. On the one hand,in fact, such patients can face their illness in a family context that helps to speed up their recovery time. On the other, this strategy implies a considerable saving of resources, allowing social health facilities and personnel to be assigned to patients with more severe diseases.

Therefore, an efficient solution for the monitoring of a patients' state of health is required, which is able to collect, record and analyse vital signs, and so support preventive care, diagnosis and rehabilitation planning. Moreover, the automation of physiological data capture and its visualisation on a device



Naddeo, S., Verde, L., Forastiere, M., Pietro, G. and Sannino, G.
A Real-time m-Health Monitoring System: An Integrated Solution Combining the Use of Several Wearable Sensors and Mobile Devices.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 545-552
ISBN: 978-989-758-213-4

reduces the risk of errors in manual harvesting. This lower risk of error results in increased efficiency, reduced costs and improved vital qualities in an area such as healthcare one where human error can really make the difference between life and death. The most innovative aspect of the adoption of a tele-monitoring system is represented by the means by which patients and healthcare professionals communicate and interact. Patients are directly involved in managing their health and wellness.

Mobile devices, such as smartphones or tablets, constitute the perfect instrument to monitor the vital parameters of a patient. Thanks to the use of appropriate wearable sensors, it is possible to collect and analyse the data coming from these devices to monitor the patient's state of health.

In this work we present a real-time m-health monitoring system for people suffering from chronic diseases that enables the collection, sharing and exchange of physiological data, such as blood pressure, heart and respiration rate and ECG signals. The system is not limited only to the acquisition of such data but also enables the analysis of vital signs. The physiological data are captured using wearable systems compliant and not to Continua Health Alliance guidelines (Carroll et al., 2007), one of the most respected communication standardization protocol.

## 2 RELATED WORK

Mobile healthcare applications constitute an instrument for individuals to keep track of their health condition, to take more responsibility for their lifestyle and to improve the efficiency of care by providing high quality data to health professionals. Such medical data are often acquired by means of the use of wearable devices. Interest in the research and development of smart wearable systems is increasing in both the academic world and industry, to promote the realization of devices that comply with the required standards of effectiveness, interoperability, cost, privacy and ethics (Lmberis and Dittmar, 2007; Konstantas, 2007; Chan et al., 2012). These systems are able to measure vital signs, such as body and skin temperature, heart rate, blood pressure, blood oxygen saturation (SpO2), electrocardiograms (ECG), electroncephalograms (EEG) and respiration rate.

To implement an m-health solution a level of standardization is necessary to ensure an easy and quick integration between the wearable device (e.g. a blood pressure monitor or pulse-oximeter) and the gateway device (e.g. a smartphone). The interoperability among all devices that compose a telehealth system is regulated by the guidelines of the Continua Health Alliance.is necessary which describe a set of internationally established standards and frameworks to ensure interoperability of devices (Carroll et al., 2007).

Several examples of monitoring systems are described in literature useful for the monitoring of physiological data taken from wearable devices. Unfortunately, not all the presented systems use devices that follow the Continua Health Alliance guidelines. Kakria et al. (Kakria et al., 2015), for example, propose a real-time monitoring system to collect data about heart rate, blood pressure and body temperature using wearable devices non-compliant with the Continua Health Alliance guidelines. The same limitation applies to the device used by PhysioDroid (Banos et al., 2014), a mobile system to monitor heart rate, body temperature and respiration rate, and to the system realized by Forastiere et al.(Forastiere et al., 2016) for a monitoring of a patient's own wellness. The proposed system, instead, supports devices compatible with the Continua Health Alliance design Guidelines. In addition, the open architecture of our framework allows an integration with wearable devices that use other communication and data access protocols to provide a system that can monitor patients integrating data coming from several medical devices, unlike the system proposed by (Park et al., 2016). They develop an m-health application limiting the integration to only certified devices. While, Szydlo et al (Szydło and Konieczny, 2016) present a data acquisition module, implemented as a mobile application to integrate several medical devices, whether or not compliant or not with the Continua Health Alliance guidelines, but without analysing the acquired data.

Many mobile healthcare systems are aimed at patients with cardiac diseases, diabetes, hypertension, or hyperthermia, limiting the acquisition and monitoring to only one set of physiological data, and sometimes providing an analysis such as (Gay and Leijdekkers, 2007; Lv et al., 2010; Rodrigues et al., 2014; Guo et al., 2013; Lee et al., 2016), SenseView (http://www.senseview.mobi/), and SimpleEye (https://simpleeye.com/platforms/android/bluetooth-pulse-oximeter/). Our proposed system, however, is not limited to capturing and monitoring in real time vital signs but also to allows data processing and analysis locally, to better evaluate whether health activities have been performed planned and to assess whether the desired results are being achieved.

Moreover, in literature there are platforms that collect the physiological data and send it to a care center for processing and analysis. These solutions do not process the medical data locally on mobile de-

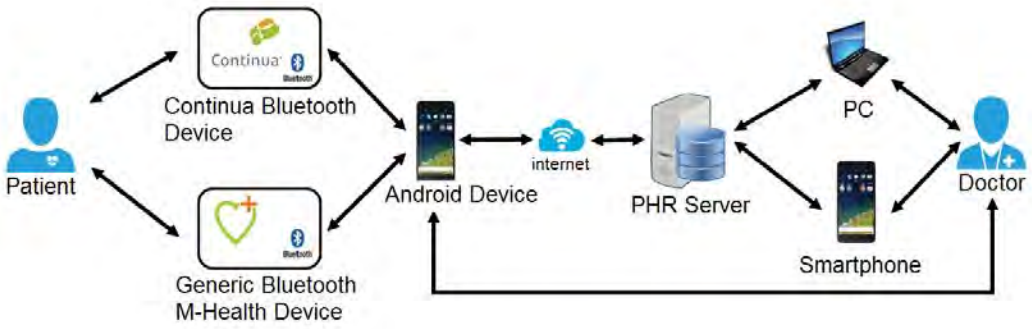



Figure 1: System Structure.

vice, but the signal needs to be continuously transferred to a health center. For example, the European Union-funded research project AMON (Anliker et al., 2004) has an objective realizing a wrist-worn medical monitoring and alert system. The data are transmitted via GSM to a medical center. Unfortunately, for heart monitoring the system is technically limited by the noisy ECG signal that is acquired on the wrist through the realized device and is not appropriate for the diagnosis of cardiac abnormalities. The Epi-medics project (Rubel et al., 2005) is another project for ECG monitoring that records and analyses ECG signals and generates alarms. In this case, also, the monitoring system uses its own sensor device, not interfaced with other sensor devices. Bisio et al. (Bisio et al., 2015), instead, proposed a platform for patients suffering from a heart failure, able to monitor data coming from the pulse oximeter to measure the saturation of peripheral oxygen and a weighting scale to measure the body weight, using devices compliant with the Continua Health Alliance standards. Moreover, Al-Taee et al.(Al-Taee et al., 2015) present a platform to support the self-management of diabetes using several medical sensors which are not compliant the guidelines.

# 3 MONITORING SYSTEM

The proposed system is able to monitor several health parameters using multiple sensors. The acquired physiological data are processed and analysed locally on a mobile device. A report of the analysis and the original files are saved, can be visualized whenever the user wants and are sent to the medical specialist.

The scheme in the figure 1 shows the structure of the system, a set of interconnected blocks in which the data flows start from the patient and arrive at the medical specialist. A patient can monitor his/her own vital signs using an Android device, which is easy to use and able to capture and store the data from the wearable sensors. Once the acquisition has been completed, the Android device sends the data to a remote Personal Health Record (PHR) server, which the doctor can access to monitor the patient's health state. In the following section the system architecture is explained in detail.

## 3.1 System Architecture

The system has a modular structure, in which every module performs a specific set of operations, as shown in the figure 2.

The *Health Device Profile (HDP) module* is responsible for the transmission of data from the wearable devices to the mobile devices using the Bluetooth protocol. Moreover, it is possible to connect the proposed system with an external PHR, as shown in the figure 1, and/or Decision Support Systems. In the interaction with these systems, the adoption of the Continua Alliance standard, using an appropriate wrapper, allows a fast and accurate data translation in accordance with the two Health Level 7 standards (http://www.hl7.org/implement/standards): Fast Healthcare Interoperability Resources (FHIR) and Personal Healthcare Monitoring Report (PHMR). The FHIR Standard (Bender and Sartipi, 2013) is useful to represent and exchange information between several types of monitoring devices, for a correct syntactic and semantic interoperability and appropriate efficiency between information systems and biomedical devices. Moreover, in order to improve the data dissertation in a tele-monitoring application, the PHMR (Wolf et al., 2009) was introduced, which is able to translate personal healthcare monitoring information into electronic medical record systems including a representation of the measurements captured.

The data, that are received by the HDP module, will be forwarded to the other modules connected with it, the *Storage module* and the *Signal Analysis* module. The first one is used to save physiological data and a patient's analysis report in the device in-

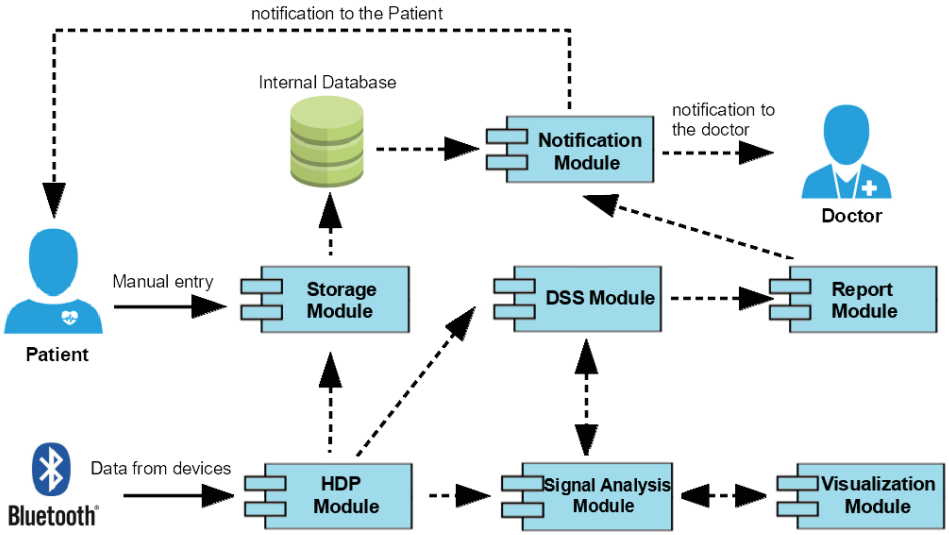



Figure 2: System Architecture.

ternal memory. SQLite database is used to store measurement data acquired from the mobile health device and/or data from manual measurement that the patient performs with devices not equipped with a Bluetooth connection. In the first case, the Storage Module receives data from the HDP Module, while in the second the data are directly inserted by the user. Collected and analysed data, during the monitoring, constitute the Personal Health Record (PHR) of the patient (Tang et al., 2006). To access the data contained in the PHR the doctor, through appropriate devices, must interact with a remote server to visualize any desired data. This operation is fundamental when there are anomalies that generate an alert. In this situation the specialist visualizes all the patient's measurements and can decide to change the therapy, perform in depth examinations or modify the monitoring plan.

The monitored data, besides being saved, can be analyzed, using the *Signal Analysis Module*. In detail,such analysis can include:

- Filtering data to remove the additive noise that can distort the signal and the resulting interpretation. It is possible to choose between four types of filter: high-pass and low-pass (Chen and Chen, 2003), recursive (Cuomo et al., 2016) and Hanning filters (Verde et al., 2015). The implemented filters are characterized by appropriate parameters such as the cut-off frequency that the user can set indicating the desirable value;
- Data processing, whereby the signals can be analysed to extract characteristic parameters, useful to evaluate the patient's health state.

Data obtained from the HDP and Signal Analysis modules are sent to the *DSS Module*. This module constitutes the Decision Support System, apable of supporting and improving the real-time monitoring and analysis using "the medical specialist's knowledge". The data are, in fact, compared with selected threshold values to check the patient's state of health. If the module discovers a warning situation, it asks the Report Module to generate a message to be sent to the doctor via the Notification Module.

All acquired data and analyses are saved in the appropriate reports, realized by the *Report Module*. The system saves a report with all measurement data once a day, which is then sent to the medical specialist at the time scheduled by the user. Moreover, the system can create a warning report message, attaching the report files, if an analysis of the data using the DSS Module reveals an abnormal situation in the patient's state of health.

The *Notification Module* is the part of the software that is responsible for notifying messages generated by the system to the doctor and the patient. The messages can be of two types:

- Alerts, that instructs the user to perform the desired measurements at scheduled time;
- E-mail messages, sent to the medical specialist indicated by the user. Such messages include any report of the patient's measurements.

Finally, the *Visualization Module* is the interface between the system and the users. It shows the real time measurements acquired from the wearable devices, the analysis data and the report files, whenever the user chooses.

# 4 USE-CASE

The realized monitoring system can be used with patients suffering from cardiovascular diseases. These patients have to constantly monitor characteristic phisiological data, such as blood pressure and heart rate. Therefore, they must report these data, very often collected manually over several days, to their medical specialists during rountine follow-up visits.

The manual collection of physiological data can increase the risk of errors, and the recording of erroneous values can change the evaluation of patient's state of health. To avoid this problem it is possible to automate the process for capturing and monitoring vital signs, as performed by the proposed system. Moreover, this system allows the patients to monitor their own physiological data at home using appropriate wearable devices without the necessity of going to a hospital center, providing a complete clinical picture to the medical specialist.

The fundamental vital signs, in order to perform a monitoring of a patient suffering from cardiovascular disease, are the heart rate, respiration rate, ECG signal, blood pressure and oxygen saturation. Therefore, the patient, using an appropriate wereable device, such as Zephyr Bioharness BH3, A&D Medical UA-767PBT-C and Nonin 9560 Onyx 2 (described in the following subsection), can capture and monitor the relevant parameters, by means of a mobile device, such as a smartphone or tablet, on which the tele-monitoring application has been installed.

## 4.1 Medical Devices

The measuring accuracy of the sensor has a direct impact on the accuracy of the heart parameter measurement in real-time monitoring systems. Therefore, the selection of an accurate heart parameter monitoring device plays an important role in the early detection of any possible heart disease.

*Zephyr Bioharness BH3* is a multisensory belt used to monitor heart rate, breathing rate and posture in real-time. The selection of the Zephyr Bioharness BH3 device was made on the basis of the accuracy of the acquired physiological data, low cost and comfort for patient comfort. This instrument is able to capture the vital signs in three different ways: with an elastic belt in which the sensors are integrated, with a shirt with the sensors integrated inside or with a and holder connected with common cardiac electrodes, as shown in the figure 3.

The second measurement parameter in our monitoring system is the evaluation of blood pressure. Blood pressure is an index of the force of the blood

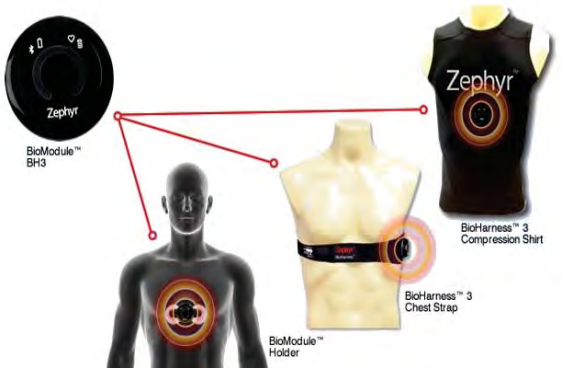



Figure 3: Zephyr Bioharness BH3.

pushing against the walls of the arteries as the heart pumps blood. It is measured as both the pressure when the heart beats pump blood (the systolic pressure) and the pressure when the heart is at rest between beats (the diastolic pressure). To monitor the patient's blood pressure the realized system uses the *A&D Medical UA-767PBT-C* blood pressure monitor.

An acute rate of these parameters is considered as an early indicator which supports medical specialists in the diagnosis of serious diseases. Additionally, the blood oxygen saturation and pulse rate can be indices of possible alterations, monitored in our system by using the *Nonin 9560 Onyx 2*. The blood pressure monitor and the pulse oximeter are illustred in the figure 4.


Figure 4: A&D Medical UA-767PBT-C and Nonin 9560 Onyx 2.

These three devices uses the Bluetooth protocol to communicate with the monitoring software installed on the Android device. The communication between the realized system and the Nonin 9560 Onyx 2 and the UA-767PBT-C follows the Continua Health Alliance guidelines. Although the Zephyr Bioharness BH3 is not certified for this standard, a dedicated library is available enabling its connection with the mobile devices.

## 4.2 Mobile Application

The system was implemented for mobile devices, such as a smartphone or tablet, developed by using the Java Programming Language through the use of

Eclipse IDE and the Android Software Developer Kit (SDK). The mobile system offers several functionalities. At the first access the user must complete a Registration form, in which he/she inserts personal information such as name, surname, date of birth, gender, email address. In detail, the e-mail address is both that of the user and that of the medical specialist who will receive the user's report containing the measurement of the physiological data and the analysis of the estimated values. Such an e-mail will be sent at the time indicated in the registration phase.

After the registration, the user can choose any of the operations in the main menu shown in the screenshot in the figure 5.

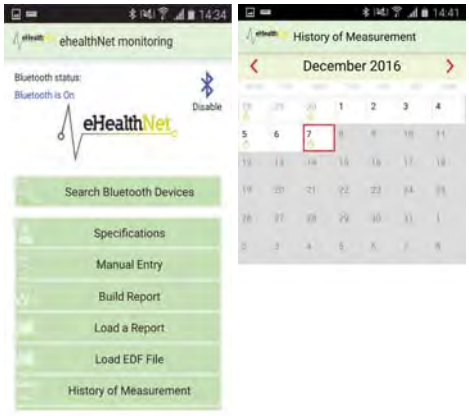



Figure 5: Screenshots of the Homepage and History of Measurement.

In detail, the user can perform the following functionalities:

- Search Bluethooth Devices: the system searches for Bluethooth devices and shows them to the user. The user can choose to collect his/her preferred wearable devices able to execute different the measurements and analyses. If the user connects to the Zephyr Bioharness BH3, for example, he/she can monitor in real-time physiological data such as the heart and breath rate or the ECG signal, as shown in the figure 6.

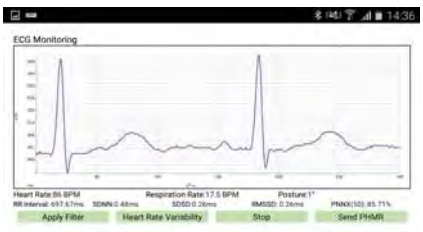



Figure 6: Screenshot of the ECG monitoring.

During the signal capture, noise can add to the useful signal distorting its interpretation and the resulting diagnosis. To avoid this problem the user can filter the captured signal choosing a filter offered by our system, described in section 3.1. In addition, it is possible to extract the characteristic parameters of the Heart Variability Index (HRV). The HRV describes the variations of both the instantaneous heart rate and RR intervals, useful to evaluate the heart state and any possible anomalies due to the presence of arrhythmias (Townsend et al., 2016);

- Specifications: information about cardiac diseases and healthy lifestyles is provided, useful for the prevention of these disorders;
- Manual Data Entry: the user can directly insert measurement data such as blood pressure, blood oxygen saturation, pulse rate and weight, indicating the date and time when he/she made this measurement;
- Loading Report and EDF file: any report with the recorded measurements can be saved and exported in pdf and European Data Format (EDF) format. EDF is the most used standard to exchange and store biological and physical signals (Kemp and Olivan, 2003). The user can choose this functionality to upload these reports and files, thanks to which the specialist can monitor the user's progress;
- History of Measurements: a calendar with a history of all the measurements performed is displayed.

## 4.3 Evaluation of Performance

The performance of the proposed system has been evaluated in terms of the allocation of memory and CPU used by the mobile application during the real-time monitoring. The first test consisted of tracking memory and CPU usage during a monitoring period of 15 minutes so as to assess over time the trend of these quantities. The real-time monitoring feature was tested on two different devices with two different Android versions: a Samsung GT-i9505 (S4) with the Android 5.0.1 version and a HUAWEI GRA-L09 (P8) with the Android 6.1 version. The performance of these analyses showed good results, as reported in the figures 7 and 8.

The obtained results indicate that the application is efficient in the use of system resources, which is very important considering that it is designed to run on a mobile device. The memory used by the application in the real-time monitoring feature is constantly less than 32 MB, increasing slightly applying when the filters are applied, and the use of the CPU is low.

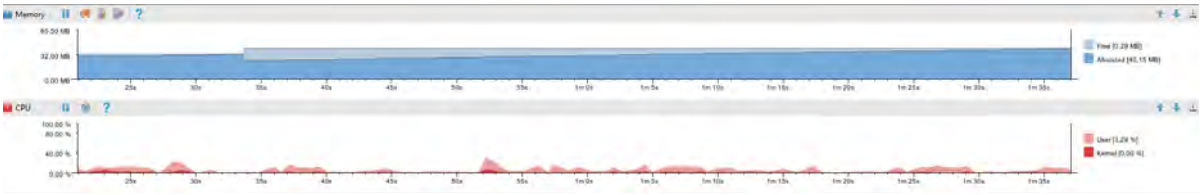


Figure 7: Performance on Samsung GT-i9505

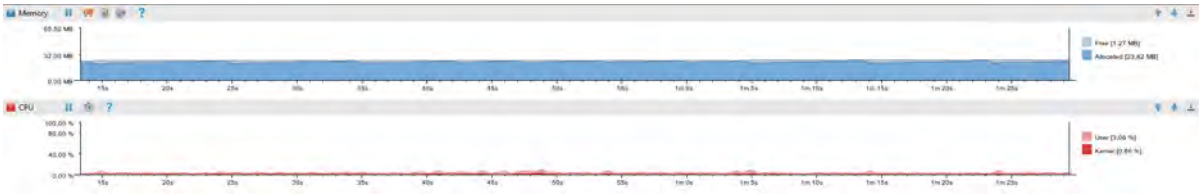


Figure 8: Performance on HUAWEI GRA-L09

To evaluate the CPU use, we compared the performance of our system with that of the SenseView App (http://www.senseview.mobi/), which is one of the most used apps on the market to connect with the Zephyr Bioharness BH3. The results show that our proposed system occupies, during the instantaneous monitoring, about 0,4 % of CPU, while the CPU used by the SenseView App is about 0,2 %. The RAM used, instead, is less for our proposed app (21 MB) than that used by the SenseView App (52 MB). This means that the realized system can be run on mid-range mobile devices without degrading the device performance.

## 5 CONCLUSIONS

In recent years the prevalence of chronic diseases has increased due to the rise in life expectancy and changes in lifestyle, and thus people suffering from these pathologies often need a continuos monitoring of their vital signs. The mobile health field can provide new access opportunities for treatment and medical services, constituting a valuable support instrument for both patients and doctors.

In this paper we have presented a physiological monitoring application designed for Android mobile devices. The user can visualize his/her own vital signs information, collected by means of an instrument, easy and fast to use, such as a smartphone or tablet and appropriate wearable devices. The system does not only monitor these parameters but also analyzes them, allowing the medical specialist to make a more accurate analysis. The system provides a good performance in terms of memory and CPU used, in comparison with other systems on the market.

## ACKNOWLEDGEMENTS

This work has received funding from the European Unions Horizon 2020 Framework Programme for Research and Innovation under grant agreement no 727528 (KONFIDO) and the project ”eHealthNet” PON03PE_00128_1. The authors would like to acknowledge support from Neatec S.p.A.

## REFERENCES

Al-Taee, M. A., Al-Nuaimy, W., Al-Ataby, A., Muhsin, Z. J., and Abood, S. N. (2015). Mobile health platform for diabetes management based on the internet-of-things. In *Applied Electrical Engineering and Computing Technologies (AEECT), 2015 IEEE Jordan Conference on*, pages 1–5. IEEE.

Anliker, U., Ward, J. A., Lukowicz, P., Troster, G., Dolveck, F., Baer, M., Keita, F., Schenker, E. B., Catarsi, F., Coluccini, L., et al. (2004). Amon: a wearable multiparameter medical monitoring and alert system. *IEEE Transactions on information technology in Biomedicine*, 8(4):415–427.

Banos, O., Villalonga, C., Damas, M., Gloesekoetter, P., Pomares, H., and Rojas, I. (2014). Physiodroid: Combining wearable health sensors and mobile devices for a ubiquitous, continuous, and personal monitoring. *The Scientific World Journal*, 2014.

Bender, D. and Sartipi, K. (2013). Hl7 fhir: An agile and restful approach to healthcare information exchange. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 326–331. IEEE.

Bisio, I., Lavagetto, F., Marchese, M., and Sciarrone, A. (2015). A smartphone-centric platform for remote health monitoring of heart failure. *International Journal of Communication Systems*, 28(11):1753–1771.

Carroll, R., Cnossen, R., Schnell, M., and Simons, D. (2007). Continua: An interoperable personal health-

care ecosystem. *IEEE Pervasive Computing*, 6(4):90–94.

Centro nazionale di epidemiologia, sorveglianza e promozione della salute dell'Istituto superiore di sanità (2016). Il portale dell'epidemiologia per la sanità pubblica. http://www.epicentro.iss.it/focus/cardiovascolare/cardiovascolari.asp/.

Chan, M., Estève, D., Fourniols, J.-Y., Escriba, C., and Campo, E. (2012). Smart wearable systems: Current status and future challenges. *Artificial intelligence in medicine*, 56(3):137–156.

Chen, H. and Chen, S. (2003). A moving average based filtering system with its application to real-time qrs detection. In *Computers in Cardiology, 2003*, pages 585–588. IEEE.

Cuomo, S., De Pietro, G., Farina, R., Galletti, A., and Sannino, G. (2016). A revised scheme for real time ecg signal denoising based on recursive filtering. *Biomedical Signal Processing and Control*, 27:134–144.

Forastiere, M., De Pietro, G., and Sannino, G. (2016). An mhealth application for a personalized monitoring of ones own wellness: Design and development. In *Innovation in Medicine and Healthcare 2016*, pages 269–278. Springer.

Gay, V. and Leijdekkers, P. (2007). A health monitoring system using smart phones and wearable sensors. *International journal of ARM*, 8(2):29–35.

Guo, X., Duan, X., Gao, H., Huang, A., and Jiao, B. (2013). An ecg monitoring and alarming system based on android smart phone. *Communications and Network*, 5(03):584.

Kakria, P., Tripathi, N., and Kitipawang, P. (2015). A real-time health monitoring system for remote cardiac patients using smartphone and wearable sensors. *International journal of telemedicine and applications*, 2015.

Kemp, B. and Olivan, J. (2003). European data format plus(edf+), an edf alike standard format for the exchange of physiological data. *Clinical Neurophysiology*, 114(9):1755–1761.

Konstantas, D. (2007). An overview of wearable and implantable medical sensors. *Yearbook of medical informatics*, 7(1):66–69.

Lee, J. M., Hirschfeld, E., and Wedding, J. (2016). A patient-designed do-it-yourself mobile technology system for diabetes: Promise and challenges for a new era in medicine. *JAMA*, 315(14):1447–1448.

Lmberis, A. and Dittmar, A. (2007). Advanced wearable health systems and applications-research and development efforts in the european union. *IEEE Engineering in Medicine and Biology Magazine*, 26(3):29–33.

Lv, Z., Xia, F., Wu, G., Yao, L., and Chen, Z. (2010). icare: a mobile health monitoring system for the elderly. In *Proceedings of the 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing*, pages 699–705. IEEE Computer Society.

Park, H. S., Cho, H., and Kim, H. S. (2016). Development of a multi-agent m-health application based on various protocols for chronic disease self-management. *Journal of medical systems*, 40(1):1–14.

Rodrigues, D. F., Horta, E. T., Silva, B. M., Guedes, F. D., and Rodrigues, J. J. (2014). A mobile healthcare solution for ambient assisted living environments. In *e-Health Networking, Applications and Services (Healthcom), 2014 IEEE 16th International Conference on*, pages 170–175. IEEE.

Rubel, P., Fayn, J., Nollo, G., Assanelli, D., Li, B., Restier, L., Adami, S., Arod, S., Atoui, H., Ohlsson, M., et al. (2005). Toward personal ehealth in cardiology. results from the epi-medics telemedicine project. *Journal of electrocardiology*, 38(4):100–106.

Sanchis-Gomar, F., Perez-Quilis, C., Leischik, R., and Lucia, A. (2016). Epidemiology of coronary heart disease and acute coronary syndrome. *Annals of Translational Medicine*, 4(13).

Szydło, T. and Konieczny, M. (2016). Mobile and wearable devices in an open and universal system for remote patient monitoring. *Microprocessors and Microsystems*, 46:44–54.

Tang, P. C., Ash, J. S., Bates, D. W., Overhage, J. M., and Sands, D. Z. (2006). Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. *Journal of the American Medical Informatics Association*, 13(2):121–126.

Townsend, N., Wilson, L., Bhatnagar, P., Wickramasinghe, K., Rayner, M., and Nichols, M. (2016). Cardiovascular disease in europe: epidemiological update 2016. *European heart journal*, page ehw334.

Verde, L., De Pietro, G., Veltri, P., and Sannino, G. (2015). A noise-aware methodology for a mobile voice screening application. In *Advances in Sensors and Interfaces (IWASI), 2015 6th IEEE International Workshop on*, pages 193–198. IEEE.

Wolf, K.-H., Schirmer, S., Marschollek, M., and Haux, R. (2009). Representing sensor data using the hl7 cda personal healthcare monitoring report draft. In *MIE*, pages 480–484. Citeseer.

World Health Organization (2016). Chronic diseases and health promotion. http://www.who.int/chp/en/.

# PerDMCS: Weighted Fusion of PPG Signal Features for Robust and Efficient Diabetes Mellitus Classification

V. Ramu Reddy[1], Anirban Dutta Choudhury[2], Srinivasan Jayaraman[1], Naveen Kumar Thokala[1], Parijat Deshpande[3] and Venkatesh Kaliaperumal[4]

[1]*TCS Research & Innovation Lab, Bangalore, India*

[2]*TCS Research & Innovation Lab, Kolkata, India*

[3]*TRDDC, Pune, India*

[4]*Institute of Ayurveda and Integrative Medicine, School of Health Sciences, Transdisciplinary University, Bangalore, India*
{*ramu.vempada, anirban.duttachoudhury, srinivasa.j, naveen.thokala, parijat.deshpande*}*@tcs.com,*
*pondyvenkatesh@gmail.com*

Keywords: Pervasive Healthcare, Type 2, Diabetes Mellitus, Classification, Heart Rate Variability, Fusion, Photoplethysmogram, Pulse Oximeter, Morphological.

Abstract: Non-invasive detection of Diabetes Mellitus (DM) has attracted a lot of interest in the recent years in pervasive health care. In this paper, we explore features related to heart rate variability (HRV) and signal pattern of the waveform from photoplethysmogram (PPG) signal for classifying DM (Type 2). HRV features includes time-domain ($F_1$), frequency domain ($F_2$), non-linear features ($F_3$) where as waveform features ($F_4$) are one set of features such as height, width, slope and durations of pulse. The study was carried out on 50 healthy subjects and 50 DM patients. Support Vector Machines (SVM) are used to capture the discriminative information between the above mentioned healthy and DM categories, from the proposed features. The SVM models are developed separately using different sets of features $F_1$, $F_2$, $F_3$,and $F_4$, respectively. The classification performance of the developed SVM models using time-domain, frequency domain, non-linear and waveform features is observed to be 73%, 78%, 80% and 77%. The performance of the system using combination of all features is 82%. In this work, the performance of the DM classification system by combining the above mentioned feature sets with different percentage of discriminate features from each set is also examined. Furthermore weight based fusion is performed using confidence values obtained from each model to find the optimal set of features from each set with optimal weights for each set. The best performance accuracy of 89% is obtained by scores fusion where combinations of mixture of 90% features from the feature sets $F_1$ and $F_2$ and mixture of 100% features from the feature sets $F_3$ and $F_4$, with fusion optimal weights of 0.3 and 0.7, respectively.

## 1 INTRODUCTION

Diabetes is a malfunction of glucose-insulin regulatory system that leads to onset of various complications. It has been recognized as fourth leading cause of death in developed countries (Tabish, 2007). From the recorded data in health centres worldwide it is predicted that it is reaching epidemic proportions in many developing and newly industrialized nations. In terms of diabetic population, the top three countries in the world are China, India and USA (Collaboration et al., 2016). In India it has shot up from 11.9 million in 1980 to 64.5 million in 2014. International Diabetes Federation (IDF), has raised a serious alarm for India by saying that nearly 52% of Indians are not aware that they are suffering from high blood sugar and it is expected to cross 100 million mark by 2030[1]. Risk of cardiovascular disease (CVD) is two or four times greater for diabetic individuals than normal ones and there is a trend in increased risk of cardiac mortality[2] However, till date there is limited medical equipment and awareness of the severity of this disease largely aggravated due to prevalence of bad diet, no physical exercise, abnormal body weight, and use of tobacco. Furthermore, the symptoms of cardiac patients and diabetes patients are similar due

[1]http://ccebdm.org/news.php

[2]www.world-heart-feaderation.org/cardiovascular-disease-risk-factors/diabetes/



Reddy, V., Choudhury, A., Jayaraman, S., Thokala, N., Deshpande, P. and Kaliaperumal, V.
PerDMCS: Weighted Fusion of PPG Signal Features for Robust and Efficient Diabetes Mellitus Classification.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 553-560
ISBN: 978-989-758-213-4

to change in the arterial stiffness and hence likely to be mis-classified. The symptoms of this disease are high blood sugar include frequent urination, increased thirst, and increased hunger. If left untreated it results in long-term complications include heart disease, stroke, chronic kidney failure, foot ulcers, and damage to the eyes.

These problems are addressed by few existing solutions; such as C-peptide test, fasting plasma glucose test, GAD antibodies test, Hba1c test, oral glucose tolerance test, type-2 diabetes indication test (Association et al., 2015). It should be noted that most of the above-mentioned technique are either invasive or minimal invasive (finger prick) in nature. This study aims to identify the individual's diabetic status by assessing the vascular pulse function and other vital features by using the non-invasive PPG signal. In addendum, continuous monitoring of diabetes patients can aid in assisting the short and long-term complication risks as well. Hence, there is an inherent demand to explore the feasibility for the continuous, non-invasive monitoring and estimation of the type 2 diabetes. In (Schroeder et al., 2005), (Seyd et al., 2012) researchers had explored diabetes detection using HRV features from the time domain, in (Elgendi, 2012) Mohamed has explored PPG features and its applications where as Rohan *et al* has identified HRV and pulse waveform features for identifying coronary artery disease (Banerjee et al., 2016).

Non-invasive, quick, easy, low cost and on time recognizing diabetes with simple method and portable technology for the primary care and community-based clinical settings is the main goal of researchers in this area. The pulse plethysmogram technology has been used in a wide range of commercially available medical devices for measuring oxygen saturation, blood pressure and cardiac output (Allen, 2007). Due to change in glucose level, the amount of blood volume in the finger changes, this variation can be measured by PPG. When a fixed source of infrared radiation is used, the variation of blood volume act as a phototransistor and the receive signal is changed. This is why we use the PPG signal for identifying the diabetic subjects. In this work a low-cost FDA approved pulse oximeter is used to collect vital physiological signal such as PPG signal from finger. Heart rate variability (HRV) features from the time domain along with some useful features related to shape of the pulse (morphological information) are extracted from PPG to discriminate healthy and diabetic subjects. In addition to the above features we have also explored other HRV features extracted from the frequency domain, non-linear and poincare features. Our data-driven approach enables visualization of PPG signals and captures specific features such as heart rate variability (HRV) and features related to shape of the pulse from PPG signal related to change in blood flow which in turn caused due to change arterial stiffness due to diabetes. We have developed pervasive diabetes mellitus classification system (PerDMCS) and has achieved an accuracy of 82%, sesitivity 84% and specificity 90% using above mentioned features. Further weight based fusion technique is proposed for more robust detection of diabetes. Field data shows this method works properly and achieved an improved accuracy of 89% with sensitivity and specificity of 90% and 88%, respectively.

## 2 DIABETES MELLITUS DATASET

In this study, we have collected data from 50 confirmed diabetic patients and 50 healthy subjects. Diabetic subjects were aged between 34 $\pm$ 10 years where as healthy subjects were aged between 41 $\pm$ 13. The subjects are selected from IAIMS Research Center located in Bangalore, India. Experimental protocol has been approved by the hospital ethical committee. PPG data were collected from right hand index finger of each subject for 5 minutes using a USB based commercial pulse oximeter (Contec CMS 50D+ [3]) at 60 Hz.

## 3 PRE-PROCESSING

The collected PPG data is fed as an input to the pre-processing block to obtain accurate RR intervals as an output. This is achieved by following the sequence of steps like baseline removal, peak detection and removal of outliers obtained due to motion artifacts. The signals obtained across each stage are also depicted in Fig. 1, where Fig. 1(a) represents raw PPG signal obtained during data collection and Fig. 1(b) is the corresponding PPG signal obtained after baseline drift removal followed by peak detection. The region where erroneous peak is obtained due to motion artifact is marked with an elliptical region in Fig. 1(b). Subsequently, peaks are calculated to obtain the RR intervals and the outliers of RR intervals are then removed. The steps involved in pre-processing stage are briefly discussed as follows:

- *Baseline Removal:* The baseline removal on the PPG signal is carried out using beads technique (Ning et al., 2014).

[3]http://www.contecmed.com/

- *Peak Detection:* The peak detection is carried out on the baseline removal of PPG signal using peak detection algorithm.
- *Outlier Removal:* After peak detection, RR intervals are computed from peak to peak intervals. The outlier RR intervals resulted due to motion artifacts are removed using the percentage filter. As mentioned in [4] if the percentage change of succeeding RR interval from the preceding RR interval is greater than 20% can be treated as an outlier.

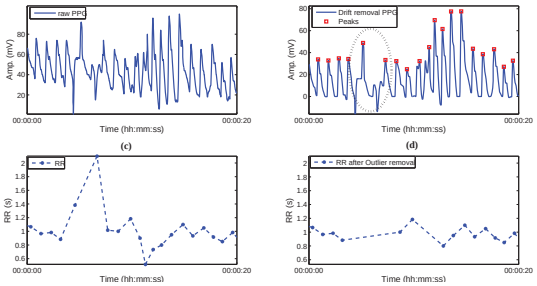



Figure 1: (a) Raw PPG signal, (b) Baseline drift PPG signal (c) RR intervals and (d) RR intervals after Outlier removal.

After preprocessing different features were extracted as shown in Table 1.

## 4 NUMERICAL RESULTS

In this work, Support Vector Machines (SVM) are explored to discriminate the diabetic and healthy subjects. SVM classification is an example of supervised learning. SVMs are useful due to their wide applicability for classification tasks in many signal processing applications such as emotion recognition (Koolagudi et al., 2010), crowd noise and activity classifications (Reddy et al., 2013) (Reddy and Chattopadhyay, 2014), and physiological signal based CAD detection (Banerjee et al., 2016). A classification task usually involves training and testing data which consist of some data instances. In the training set, each instance contains one target class label and many attributes. The main goal of SVM for classification problem is to produce a model which predicts target class label of data instances in the testing set, given only the attributes. The SVM model was developed as-one against-rest principle by using feature vectors derived from the intended class as positive examples and the feature vectors derived from the other class as negative examples.

Radial Basis Function (RBF) is used in this work. This is because RBF, unlike linear kernel can handle the case where the relationship between the class labels and attributes is non-linear. Another advantage of RBF kernel is its universal approximation properties. Also, it offers good generalization as well as good performance in solving practical problems (Reddy et al., 2014). The basic architecture of diabetes mellitus classification system using SVMs with above mentioned features is shown in Fig. 2.

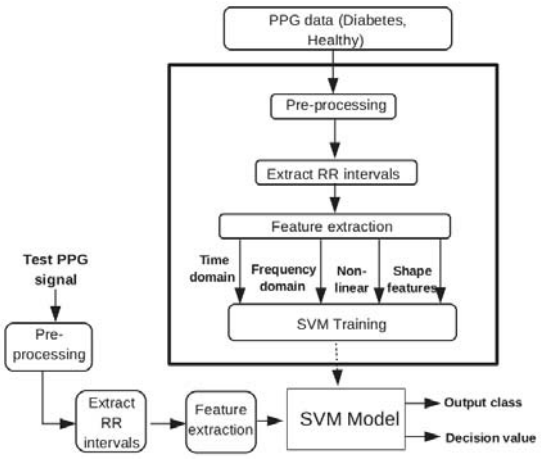



Figure 2: Architecture of Pervasive Diabetes Mellitus Classification System (PerDMCS).

In this study, 100 subjects of data is used out of which 50 subjects are healthy and 50 are diabetic. We have used 5-fold validation technique where 4 folds are used for training and remaining fold is used for testing.

In this work, we first analyzed the capability of individual feature sets $F_1$, $F_2$, $F_3$ and $F_4$ for discriminating the diabetes and healthy. Five DMCS systems are developed, which are summarized as follows:

1. PerDMCS-1: Pervasive diabetes mellitus classification system using $F_1$ features.
2. PerDMCS-2: Pervasive diabetes mellitus classification system using $F_2$ features.
3. PerDMCS-3: Pervasive diabetes mellitus classification system using $F_3$ features.
4. PerDMCS-4: Pervasive diabetes mellitus classification system using $F_4$ features.
5. PerDMCS-5: Pervasive diabetes mellitus classification system using combination of all features.

Performance of the pervasive DMCSs using the features discussed earlier, is represented in the form of a consolidated confusion matrix as shown in Table 2. The diagonal elements of the confusion matrix represent the correct classification performance of class. Non-diagonal elements indicate the performance of

[4]http://circ.ahajournals.org/content/93/5/1043

Table 1: List of Features.

| **Name** | **Description** | **DM Range** $mean \pm std$ | **Non DM Range** $mean \pm std$ |
|---|---|---|---|
| ***Time Domain Features ($F_1$)*** | | | |
| meanNN | mean values of NN intervals (ms) | $787 \pm 133$ | $735 \pm 111$ |
| medianNN | median values of NN intervals (ms) | $787 \pm 134$ | $735 \pm 114$ |
| SDNN | standard deviation of NN intervals (ms) | $26.58 \pm 14.13$ | $41.85 \pm 18.45$ |
| RMSSD | root mean square of successive NN differences (ms) | $27.89 \pm 15.27$ | $34.81 \pm 16.75$ |
| NN50 | total # of of successive NN intervals differing by $\geq 50$ ms | $4.52 \pm 6.9$ | $9.14 \pm 9.62$ |
| pNN50 | percentage of successive NN intervals differing by $\geq 50$ ms | $0.07 \pm 0.1$ | $0.12 \pm 0.13$ |
| HRVti | Ratio of number of all NN intervals to maximum number | $3.39 \pm 1.32$ | $4.89 \pm 1.71$ |
| ***Frequency domain Features ($F_2$) (Welch, 1967)(Billman, 2007)*** | | | |
| aVLF | raw area of VLF (0-0.04 Hz) band ($ms^2$) | $100 \pm 42$ | $294 \pm 97$ |
| aLF | raw area of LF (0.04-0.15 Hz) band ($ms^2$) | $134 \pm 25$ | $351 \pm 39$ |
| aHF | raw area of HF (0.15-0.5 Hz)band ($ms^2$) | $143 \pm 29$ | $284 \pm 92$ |
| aTotal | total raw area of VLF, LF and HF bands | $892 \pm 116$ | $2062 \pm 177$ |
| LFHF | ratio of LF and HF areas | $1.49 \pm 1.79$ | $2.1 \pm 1.83$ |
| nLF | normalized LF area w.r.t to LF+HF | $0.15 \pm 0.08$ | $0.2 \pm 0.08$ |
| nHF | normalized HF area w.r.t to LF+HF | $0.15 \pm 0.08$ | $0.14 \pm 0.07$ |
| %VLF | relative VLF area w.r.t to total area | $12.04 \pm 7.34$ | $12.59 \pm 7.89$ |
| % LF | relative LF area w.r.t to total area | $13.15 \pm 7.08$ | $18.03 \pm 7.6$ |
| % HF | relative HF area w.r.t to total area | $13.79 \pm 7.99$ | $12.66 \pm 7.01$ |
| peakVLF | freq. of highest power in VLF band | $0.02 \pm 0.02$ | $0.02 \pm 0.02$ |
| peakLF | freq. of highest power in LF band | $0.02 \pm 0.03$ | $0.03 \pm 0.03$ |
| peakHF | freq. of highest power in HF band | $0.14 \pm 0.08$ | $0.09 \pm 0.08$ |
| ***Non-linear Features ($F_3$)*** | | | |
| pSD1 | Poincaré SD i.e., standard deviation of points perpendicular to the axis of line-of-identity | $19.86 \pm 10.89$ | $24.77 \pm 11.93$ |
| pSD2 | Poincaré SD i.e., standard deviation of points along the axis of line-of-identity | $31.43 \pm 17.66$ | $53.31 \pm 24.19$ |
| sampEN | sample entropy estimates | $1.57 \pm 0.45$<br>$1.4 \pm 0.42$<br>$1.33 \pm 0.42$ | $2.04 \pm 0.48$<br>$1.73 \pm 0.46$<br>$1.66 \pm 0.5$ |
| alpha | detrended fluctuation analysis i.e., slope of log-log plot of integrated RR vs window size | $0.4 \pm 0.31$<br>$0.84 \pm 0.26$<br>$0.81 \pm 0.28$<br>$0.43 \pm 0.29$<br>$0.89 \pm 0.7$<br>$0.34 \pm 0.88$ | $0.5 \pm 0.31$<br>$0.98 \pm 0.23$<br>$1.04 \pm 0.27$<br>$0.44 \pm 0.3$<br>$0.85 \pm 0.68$<br>$0.65 \pm 0.86$ |
| ***Waveform Features ($F_3$)*** | | | |
| meanFS | mean value of falling slopes | $0.03 \pm 0.01$ | $0.03 \pm 0.01$ |
| meanRS | mean value of rising slopes | $0.07 \pm 0.02$ | $0.09 \pm 0.02$ |
| meanPWp75 | mean value of pulse widths at 75% | $0.18 \pm 0.02$ | $0.15 \pm 0.03$ |
| meanPWp50 | mean value of pulse widths at 50% | $0.29 \pm 0.04$ | $0.27 \pm 0.06$ |
| meanPWp25 | mean value of pulse widths at 25% | $0.46 \pm 0.06$ | $0.45 \pm 0.07$ |
| meanCT | mean value of crest times | $0.21 \pm 0.03$ | $0.17 \pm 0.03$ |
| meanDT | mean values of diastolic times | $0.58 \pm 0.11$ | $0.56 \pm 0.1$ |
| meanPH | mean values of pulse heights | $50.46 \pm 9.5$ | $52.91 \pm 7.65$ |
| meanPI | mean values of pulse intervals | $0.79 \pm 0.13$ | $0.73 \pm 0.11$ |

misclassification. Columns 3-4 indicate the performance of the PerDMCS systems. Other performance measurements like true positive, false positive, true negative, false negative, precision, recall, sensitivity , specificity and overall model accuracy are presented in Table 3.

*Analysis:* From Tables 2 and 3, it is observed that feature sets $F_1$, $F_2$, $F_3$, and $F_4$ have discriminatory information related to diabetes. It is also observed that the diabetes is well discriminated compared to healthier subjects using shape related features i.e., $F_4$ , whereas using non-linear information i.e. $F_3$, classification of healthier shows better performance compared to diabetes. From this we can hypothesize that both $F_3$ and $F_4$ are complementary in nature, and if integrated can lead into better classification. Though combination of all features (i.e. PerDMCS-5) yields the best model accuracy, one can observe that PerDMCS-3 outperforms PerDMCS-5 in some of the performance measurements (True Positive, True Negative, Precision and Specificity). This is due to the inclusion of unimportant features from $F_1$ and $F_2$. Results indicate that there is a scope of minute features selection from individual feature sets. Hence for improving performance of the entire system, different fusion technologies are explored.

Table 2: Performance of pervasive diabetes mellitus classification systems developed using different features. The entries in the table indicate the subjects of classification. Act: Actual, Pred: Predicted.

| PerDMCS | Act. / Pred. | Diabetic | Healthy |
|---|---|---|---|
| PerDMCS 1 | Diabetic | **34** | 11 |
| | Healthy | 16 | **39** |
| PerDMCS 2 | Diabetic | **37** | 9 |
| | Healthy | 13 | **41** |
| PerDMCS 3 | Diabetic | **38** | 8 |
| | Healthy | 12 | **42** |
| PerDMCS 4 | Diabetic | **40** | 13 |
| | Healthy | 10 | **37** |
| PerDMCS 5 | Diabetic | **42** | 10 |
| | Healthy | 8 | **40** |

# 5 WEIGHTED FUSION: PROPOSED METHOD

## 5.1 Features Fusion

In this study, the fusion at feature level is performed by concatenation of the different percentage of discriminative features from each set i.e., $F_1$ to $F_4$. The concatenation process of features is carried out as follows.

1. The features are ranked using correlation of features and labels (Hall, 2000).
2. Different percentage of features are selected from ranked features. In this work, we have explored top 50% features to 100% with increments of 10% i.e., 6 variations such as 50%, 60% 70%, 80%, 90% and 100% most discriminative features.
3. Finally we have concatenated different percentage of features from each set to build the PerDMCS model.

Different technologies of features level fusion are employed e.g. different percentages from each domain separately (One vs One vs One vs One), different percentage from the combinations of two feature sets and different percentages from each of the rest (Two vs One vs One) etc. The comparisons of different feature level fusions are presented in 3.

As shown in 3(a), 24 (out of 1296) feature combinations result in accuracy of 83%. In Fig. 3(b), it can be observed that the model accuracy is improved slightly i.e., 2% for some different feature combinations compared to combination of all individual features as shown in Table 3. Here, 8 feature combinations results in high accuracy of 84% and which is 1% improvement compared to the earlier combination. The best model accuracy remains similar 84% (for the combination of 50% features from $F_3$ and 90% features from ($F_2$ + $F_4$) as shown in Fig. 3(c)). In Fig. 3(d), the best model accuracy is 83% for the combination of 50% features from ($F_1$ + $F_4$) and 80% features from ($F_2$ + $F_3$). However, the average accuracy is slightly less than the best combinations of *Two vs One vs One* and *Two vs One* fusions. Fig. 3(e) shows that the maximum accuracy achieved in this fusion approach is 81% and it marks a clear degradation in performance compared to the combinations mentioned above. Here, among all 144 combinations, 8 feature combinations results in 81% accuracy.

## 5.2 Scores Fusion

Score level fusion is performed by summing the weighted confidence scores (evidences) derived from the different PerDMCSs. The weighing rule for combining the confidence scores of individual modalities is as follows:

$$C = \frac{1}{m}\sum_{i=1}^{m} w_i c_i \quad (1)$$

Table 3: Objective parameters of different PerDMCSs.

| Performance Measurements | PerDMCS-1 | PerDMCS-2 | PerDMCS-3 | PerDMCS-4 | PerDMCS-5 |
|---|---|---|---|---|---|
| True Positive | 34 | 37 | 38 | 40 | **42** |
| False Positive | 11 | 9 | **8** | 13 | 10 |
| False Negative | 16 | 13 | 12 | 10 | **8** |
| True Negative | 39 | 41 | **42** | 37 | 40 |
| Precision | 0.76 | 0.80 | **0.83** | 0.75 | 0.81 |
| Sensitivity | 0.68 | 0.74 | 0.76 | 0.80 | **0.84** |
| Specificity | 0.78 | 0.82 | **0.84** | 0.74 | 0.80 |
| Model Accuracy | 0.73 | 0.78 | 0.80 | 0.77 | **0.82** |

where C is the weighted confidence score, $w_i$ and $c_i$ are weight and confidence score of the $i^{th}$ modality, and $m$ indicates number of modalities used for combining the scores. In this work, we have combined different modalities as described in section VI-A, such as two modalities (Two vs One, Two vs Two and Three vs One), three modalities (Two vs One vs One) and four modalities (One vs One vs One vs One).

In our study, for two modality systems one of the weights ($w_i$) is varied in steps of 0.1 from 0 to 1, and the other weight is determined using the formula: $w_j$ = 1- $w_i$ such that total weight $w_i + w_j$ = 1. Hence, we get 11 combinations of weighing factors. Similarly, for three modality systems two of the weights ($w_i$ and $w_j$) are varied in steps of 0.1 from 0 to 1 and the other weight is determined using the formula: $w_k$ = 1- $w_i$-$w_j$ such that total weight $w_i + w_j + w_k$ = 1 and $w_k \geq 0$. Hence, we get 60 combinations of weighting factors. For four modality systems three of the weights ($w_i$, $w_j$ and $w_k$) are varied in steps of 0.1 from 0 to 1 and the other weight is determined using the formula: $w_l$ = 1- $w_i$-$w_j$ -$w_k$ such that total weight $w_i + w_j + w_k + w_l$ = 1 and $w_l \geq 0$. The classification performance of the combined system for various combinations of the weighting factors are as follows.

#### 5.2.1 One vs One vs One vs One

It is observed that out of 313632 fusion of score combinations only 12 instances found to produce highest accuracy of 85%. The accuracy of fusion based models at score level is performed slightly better than the models obtained by simple feature level combinations.

#### 5.2.2 Two vs One vs One

It is observed that out of 77760 fusion of score combinations only 3 instances found to produce highest accuracy of 85%. The accuracy of fusion based models at score level is performed slightly better than the models obtained by simple feature level combinations.

#### 5.2.3 Two vs One

It is observed that out of 4752 fusion of score combinations only 20 instances found to produce highest accuracy of 84%. The accuracy of fusion based models at score level is not improved compared to the accuracy obtained by simple feature level combinations.

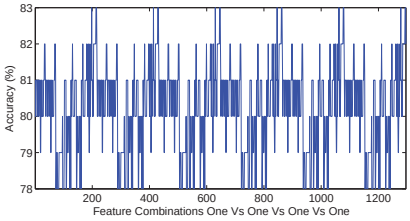
#### 5.2.4 Two vs Two

It is observed that out of 1188 fusion of score combinations only 2 instances found to produce highest accuracy of 89%. The accuracy of fusion based models at score level is performed better than the models obtained by simple feature level combinations.

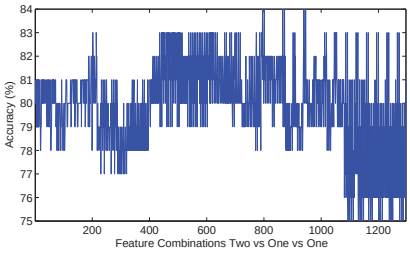
#### 5.2.5 Three vs One

It is observed that out of 1584 fusion of score combinations only 14 instances found to produce highest accuracy of 82%. The accuracy of fusion based models at score level is performed slightly better i.e.1% than the models obtained by simple feature level combinations. However, still the accuracy is very low compared to other feature level and score level fusion models.

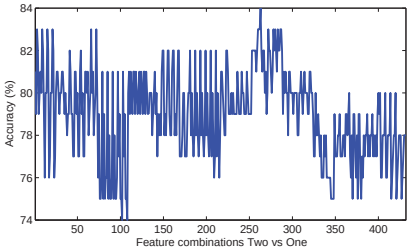
From the results it is found that the fusion of scores performed better than the models simple feature level fusion. Among all the models the highest accuracy is obtained for score level fusion of Two vs Two combinations. The best accuracy is observed to be 89% for the feature combinations of mixture of 90% features from the feature sets $F_1$ and $F_2$ and mixture of 100% features from the feature sets $F_3$ and $F_4$, with optimal weights of 0.3 and 0.7, respectively. Similarly, another best combination is observed to be for the feature combinations of mixture of 100% features from the feature sets $F_1$ and $F_2$ and mixture of



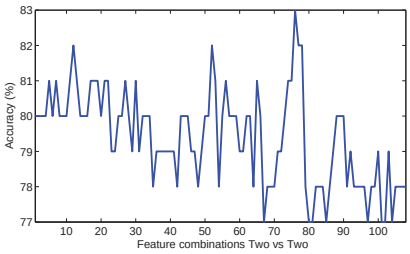

(a) One vs One vs One vs One



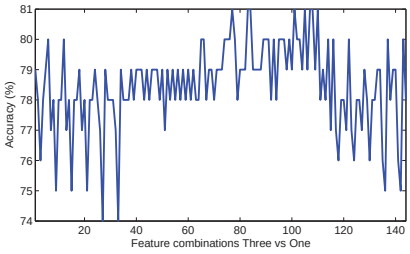

(b) Two vs One vs One




(c) Two vs One

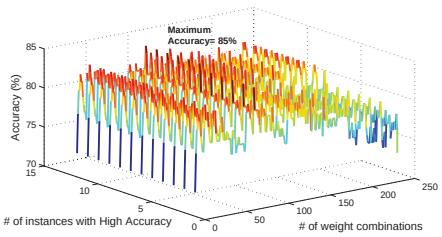



(d) Two vs Two

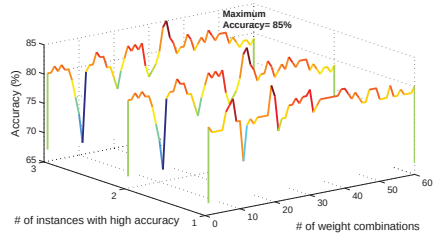



(e) Three vs One

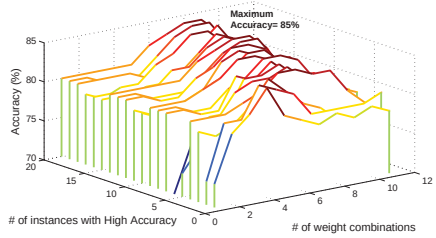
Figure 3: Accuracy of Feature Fusions.



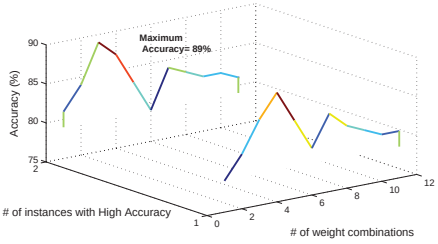

(a) One vs One vs One vs One



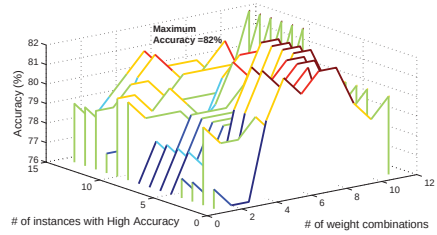

(b) Two vs One vs One




(c) Two vs One




(d) Two vs Two




(e) Three vs One

Figure 4: Accuracy of Score Fusion.

100% features from the feature sets $F_3$ and $F_4$, with optimal weights of 0.2 and 0.8, respectively.

It is noted that the best combination in feature level fusion at Two vs Two level showed the best accuracy of 83% for the combination of mixture of 50% features from the feature sets $F_1$ and $F_4$ and mixture of 80% features from the feature sets $F_2$ and $F_3$. However, in score level fusion it is not improved. In score level fusion it picked the different set of fea-

ture combinations and the performance of the system improved from 83% at feature level to 89% at score level. It can be seen that the best feature combination observed in score level fusion exhibits 81% accuracy in feature level fusion from the Fig. 3(d) with equal weights i.e. 0.5 and 0.5 (6th combination across # of weight combination). This indicates that fusion is able to combine the complementary nature of evidence obtained from different sets of features. The performance measures for the best combination is observed to be same and given in Table 4. From the results, it is observed that the score based fusion based PerDMCS system is outperformed compared to individual system performances (Table 2).

Table 4: Performance of best pervasive diabetes mellitus classification systems developed using fusion technique. The entries in the table indicate the subjects of classification.

| Predicted \ Actual | Diabetic | Healthy |
|---|---|---|
| Diabetic | **45** | 6 |
| Healthy | 5 | **44** |

## 6 SUMMARY

In this work, HRV features related to time domain, frequency domain and non-linear and shape (morphological) related features extracted from PPG signal are used to discriminate between diabetic and healthy. SVMs are used as classification models for developing different PerDMCS systems. The performance of the PerDMCS systems developed by individual features are improved by exploring fusion techniques, by combining different percentage of discriminate features from different combinations of feature sets and scores of the individual systems and different combination systems as well. An improvement in classification performance of the system is observed at score level fusion with average classification performance of 89%. This is attributed to the complementary nature of evidence present in the features.

## ACKNOWLEDGEMENTS

We are immensely grateful to the subjects for spending their valuable time and providing the data for this study. We are thankful to IAIMS Research Center. Further, we express our gratitude towards Dr. Arpan Pal and Dr. Balamurali P. for their guidance and support.

## REFERENCES

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3):R1.

Association, A. D. et al. (2015). 2. classification and diagnosis of diabetes. *Diabetes care*, 38(Supplement 1):S8–S16.

Banerjee, R., Vempada, R., Mandana, K., Choudhury, A. D., and Pal, A. (2016). Identifying coronary artery disease from photoplethysmogram. In *Int. Joint Conf. on Pervasive and Ubiquitous Computing: Adjunct*, pages 1084–1088. ACM.

Billman, G. E. (2007). The lf/hf ratio does not accurately measure cardiac sympatho-vagal balance. *Heart Rate Variability: Clinical Applications and Interaction between HRV and Heart Rate*, page 54.

Collaboration, N. R. F. et al. (2016). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4· 4 million participants. *The Lancet*, 387(10027):1513–1530.

Elgendi, M. (2012). On the analysis of fingertip photoplethysmogram signals. *Current cardiology reviews*, 8(1):14–25.

Hall, M. A. (2000). Correlation-based feature selection of discrete and numeric class machine learning.

Koolagudi, S. G., Reddy, V. R., and Rao, K. S. (2010). Emotion recognition from speech signal using epoch parameters. In *Proc. of SPCOM*, pages 1–5, India.

Ning, X., Selesnick, I. W., and Duval, L. (2014). Chromatogram baseline estimation and denoising using sparsity (beads). *Chemometrics and Intelligent Laboratory Systems*, 139:156–167.

Reddy, V. R., Chakravarty, K., and Aniruddha, S. (2014). Person identification in natural static postures using kinect. In *ECCV*, pages 793–808. Springer.

Reddy, V. R. and Chattopadhyay, T. (2014). Human activity recognition from kinect captured data using stick model. In *HCI. Adv. Interaction Modalities and Techniques*, pages 305–315. Springer.

Reddy, V. R., Sinha, A., and Seshadri, G. (2013). Fusion of spectral and time domain features for crowd noise classification system. In *Intelligent Systems Design and Applications*, pages 1–6. IEEE.

Schroeder, E. B., Chambless, L. E., Liao, D., Prineas, R. J., Evans, G. W., Rosamond, W. D., and Heiss, G. (2005). Diabetes, glucose, insulin, and heart rate variability the atherosclerosis risk in communities (aric) study. *Diabetes care*, 28(3):668–674.

Seyd, A., Joseph, P. K., and Jacob, J. (2012). Automated diagnosis of diabetes using heart rate variability signals. *Journal of medical systems*, 36(3):1935–1941.

Tabish, S. A. (2007). Is diabetes becoming the biggest epidemic of the twenty-first century? *International Journal of health sciences*, 1(2):V.

Welch, P. D. (1967). The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73.

# Patient-centric Handling of Diverse Signals in the mHealth Environment

Jan Sliwa
*Bern University of Applied Sciences, Quellgasse 21, CH-2501 Biel, Switzerland*
*sliwa@swissonline.ch*

Keywords: Mobile Health, Signals, Cyber-Physical Systems, Real Time.

Abstract: In the context of the Mobile Health (or Telemedicine) many signals (data items) are exchanged between the medical devices, data storage units and involved persons. They are often treated as a uniform mass of "medical data", especially in the Big Data community. At a closer look, they unveil various characteristics, like type (single / continuous), required quality and tolerance for missing values. As in medical care time is crucial, real-time characteristics are important, like the sampling rate and the overall reaction time of the emergency system. The handling of data depends also on the severity of the medical case. Data are transferred and processed by external systems, therefore the overall function depends on the environment and the persons involved: from the user/patient to a dedicated medical emergency team. The collected data can be anonymously reused to gain or verify knowledge, what calls for a fair trade-off between the interests of the individual patient and the community. This paper discusses the semantics of mHealth data, like medical requirements, physical constraints and human aspects. This analysis leads to a more precise mathematical definition of the required data handling that helps to construct mHealth systems that better fulfill the health support function.

## 1 INTRODUCTION

A mobile health system, serving to treat a serious medical case has to meet many demands. For a designer, it is easy to be biased by his/her own experience. A developer of smartphone applications, a Big Data analyzer, a sensorics specialist - each of them has a different viewpoint and puts stress on different aspects of the problem. As the technical issues are truly challenging and crucial for the success, they tend to play the dominant role, especially in small and young enterprises. We will try to connect various technical viewpoints with the medical perspective in order to build a unified picture.

Mobile health systems generate and process large amounts of data. It is important to see their medical significance which is not equal for all of them. Some of them are required in a predefined frequency, some are helpful but optional. Some need a certain precision, some serve only as orientation values. Some have to ensure guaranteed reaction times, for some timing constraints are irrelevant. Some, when analyzed statistically, need a well balanced, unbiased sample, for some sheer quantity makes them valuable.

This paper takes a closer look at the above mentioned aspects of mHealth data and brings to a technically-oriented reader a better view on their *meaning*. This semantic view allows in the next step to formalize the requirements regarding time constraints and data extraction, conversion, transmission and storage. The final goal is to help the reader to construct better *patient-centric* medical systems.

Remark: We use here interchangeably the terms signals or data (items). The word *signal* stresses the capture from sensors, whereas *data* emphasizes the information content. We also use the term *patient*, although in the case of fitness tracking *user* would be more correct. As the borders between the application areas are not sharp, we stay with the first term.

## 2 RELATED WORK

There are several good overviews of today's mobile health technology and applications (Baig et al., 2015), (Islam et al., 2015), (Silva et al., 2015), (Soh et al., 2015). A detailed analysis of remote blood glucose monitoring is given in (Lanzola et al., 2016). Cardiovascular health and disease is discussed in (Eapen et al., 2016). The authors also propose a roadmap for mobile health in this area. An Android-based system for ECG arrhythmia classification and heart rate variability analysis is presented in (Xue et al., 2015).

It is useful to look at the mHealth from the medical perspective. (Itrat et al., 2016) discuss the



Sliwa, J.
Patient-centric Handling of Diverse Signals in the mHealth Environment.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 561-568
ISBN: 978-989-758-213-4

telemedicine in prehospital stroke evaluation. (Agboola et al., 2016) presents a bitter reality check for smartphone apps. A top-selling app for managing and diagnosing skin cancer was only able to accurately classify 10 of 93 biopsy-proven melanomas. As for insulin dose calculation, 67% of the apps gave inappropriate output dose recommendations that put users at risk (Huckvale et al., 2015).

If we analyze the function of the systems not in the lab, but in real life, practical experience is of use. (Espinoza et al., 2016) present the design of a telemedicine management system in Ecuador where specific local challenges of a rural environment in a less developed country had to be addressed.

# 3 SIGNAL TYPES

## 3.1 Basic Signals

We can divide the basic signals in following categories:

- Single measurements: *(t, value)*
- Point events: *(t, type)*
- Continuous waveforms: *(y = f(t))*

If we take a closer look, the divisions between them are blurred. In the case of a chronic disease, single measurements form a sequence of values. Point events may be entered as such, like an epileptic seizure or tumbling. If they are detected by sensors, they are derived from other measurements, as shown below. Continuous signals are created by sampling a property, so formally they are sequences of values, at a high sampling rate. When we speak of such signals like ECG (electrocardiogram) or EEG (electroencephalogram), they may come from a single sensor (in a very basic version) or from a set of sensors placed on the chest or on the head in well defined locations. In the latter case we obtain a set of synchronized waveforms.

## 3.2 Derived Signals

From the basic signals we can derive more condensed information. It is especially useful in the case of continuous signals the volume of which is too large to store or transfer. Equally, in the raw form it is not yet very useful. For the electrocardiogram, a well known condition is the ST Segment Elevation Myocardial Infarction (STEMI). The cardiac cycle has characteristic points P-Q-R-S-T, and the elevation of the S-T segment is a signal warning about the risk of a myocardial infarction, i.e. heart attack. Similarly, irregularities of the cycle frequency can be identified as arrhythmia, in several variants, like tachycardia or bradycardia (heartbeat too fast or too slow). For arrhythmia to be detected, the basic signal has to be analyzed over many cycles. The severity of the case depends on the intensity and duration of the abnormal condition.

As mentioned above, a point event can be detected by sensors. For example, tumbling of the patient can be detected by accelerometers, e.g. in a smartphone. In this case, an algorithm extracts a characteristic waveform from a continuous signal.

When we monitor vital signals, we have to treat adequately missing values. The fact that the measurement is missing may be an information itself. For example, missing a required value for a prolonged time may indicate that the device is not working (e.g. battery empty), that something has happened to the patient or that the patient is not using the device because it is obtrusive, he/she went on travel and: has forgotten it at home / has left it because of its weight / has no plug adapter for the charger. In a similar way, systematic outlier values may mean a health problem or a wrong placement or poor body contact of the device. We list those cases in order to stress that the same observed situation may have very different causes. Some of them may require intervention of the system operator (hospital).

## 3.3 Complex Signals

Until now we have discussed signals coming from single sources. From the medical point of view, it is often useful to combine information from many sources. One of the main reasons is making sharper distinctions between the cases and eliminating false alarms. There are many papers presenting methods to detect patients' falls with the use of the accelerometers. Typically the accelerometers built in the smartphones are considered as described in (Sannino et al., 2015). They are reasonably ubiquitous, however the assumption that they are worn all the time seems not to hold. In any case, if such a device detects the patient tumbling on the floor, it is useful to verify it with more data. Especially if the detected condition is severe, requires an action and this action is costly - like sending an ambulance - it is crucial to detect only real cases. False alarms, even not frequent, will erode the confidence in the service.

If we want to design a novel architecture that connects various devices from unrelated producers, we face the problem of the interoperability. Such devices - microphone, ECG sensor, EEG headset - typically

are delivered with a connection to the smartphone or the cloud and visualization and/or analysis application. If we want to connect them into a combined device, we have to go on the level of the internal interfaces (rarely disclosed) and to write our own application. The system described in (Sannino and De Pietro, 2014) detects fainting and falling of patients by connecting the heart rate variability in the ECG with the information from accelerometers and other body and environmental sensors. In this way the decisions generated by the system take the context into account what increases their reliability.

Not only data formats may pose a problem, also communication protocols may be different (periodic sending / on demand), time and data granularity or measurement precision. When detecting complex events, we need synchronized data. If they come marked by internal timestamps, we have to compensate possible differences or to force the synchronization of the clocks.

# 4 CASE SEVERITY

The treatment of the signal depends on the severity of the medical case. We can list here following classes:

- fitness, general health, behavior modification
- chronic disease
  - mild
  - severe
- life saving

The influenced factors are:

- required quality
- necessity, sampling rate
- reaction time

If the devices are used for ***fitness improvement***, the measurements are performed or checked according to the interest of the user. The value is rather used for general information, and there are no timing requirements. The user often loses interest for the measurements after a certain time. If in meantime he/she changed to a healthier lifestyle, the basic goal has been achieved.

There is however a risk that the user gets obsessed with the fitness goals. Some people measure their weight on bathroom scales many times per day, not taking account of normal daily variation and measurement errors. In the same way, trying to constantly increase the daily step count, especially when comparing to the group and obtaining (verbal) rewards from the device, may pose a health risk. There is a certain optimum and not always more is better. On the other hand, the device typically just counts steps, so climbing a mountain is like walking, or even of smaller value, as the covered distance is smaller. Equally swimming, when the device is left in a locker, does not count at all. Therefore the user has to be aware that his/her virtual health is only an imperfect model of the real one.

This becomes a problem, if the user has an agreement with an insurance company to have a healthy lifestyle in exchange for lower primes. A simple device, like an accelerometer in a smartphone, registers only certain types of activities. On the other hand, the user may be tempted to cheat the device, by simulating realistic oscillations.

In a case of a ***chronic disease***, the device is typically used to monitor the state of the patient, detect the abnormalities and inform the doctor or hospital that handles his/her case. For a mild condition, the measurement can be done occasionally, more or less periodically or if the patient does not feel well. Therefore missing values are not problematic, possibly the patient feels no necessity to act. He/she should not be burdened too much by reminders. He/she can be called for a periodic check, as it is for apnea patients using a ventilator. The measurement should have a reasonable precision, determined on the basis of the medical science. It is necessary to eliminate the variability caused by imprecise placement of the sensor, too high (or too low) humidity of skin, wrong operating mode of the device, or similar. If - as we assume - the measurement is performed by the patient at home, the handling should be entirely simple and clear. The device has to be approved by the doctor what eliminates most cheap sensors and easily downloadable smartphone applications.

For a severe chronic disease, the requirements for quality and regularity of the measurements are more stringent. Regular measurements may show the increased risk before the actual event occurs and the patient may come for an extensive check or a preventive hospital stay. As the deterioration of the health state may be rapid and serious, time plays an important role. Ergonomics has to be carefully designed, as in an emergency the patient's capabilities are impaired. For example, during an insulin shock the patient is dizzy and nervous and his/her vision is blurred. If the patient is not able to act on his/her own and has no assistance, an automatic communication with the hospital is necessary.

For a ***life saving*** condition, the precision and reaction times are even more important. Let us fix our attention on a patient with a cardiac disease, having one or more wearable / implantable devices connected

via a wireless Body Area Network (BAN) to a smartphone that can alarm the hospital in case of emergency. The devices are active, i.e. can induce a life saving action locally. As the patient is at risk, he/she has to adapt his/her life habits accordingly. As the emergency call system depends on functioning communication, the loss of the phone signal or of data roaming is a problem. This may occur in a rural area, in a location not properly covered by the patient's provider or in a restaurant restroom in the basement under a thick concrete floor. Therefore the patient should avoid such locations, limit the stay and take care. The current risk level may be evaluated by the devices and communicated to the patient. In a better phase less precautions have to be taken. The user interaction design is delicate, as the patient has to know about the increased risk but on the other hand should not be overwhelmed by messages. Too many warnings will themselves increase his/her anxiety or with time will be ignored. An intervention is costly, possibly includes sending an ambulance, therefore making a clear distinction between real and false alarms is extremely important. When in doubt, the emergency team may try to contact the patient. However, if the risk condition is generated by a complex algorithm combining many factors, the hospital emergency team may know about the upcoming event earlier than the patient him-/herself. This evidently is true if the measurement system is reliable and the algorithm is correct. We see that taking quickly correct, resolute decisions is not easy.

## 5 REAL TIME ANALYSIS

In the systems that handle serious medical conditions where emergencies may occur, reaction time is essential. At the lowest level we treat basic communications issues, like network architecture, data quantity, channel capacity and similar. Several papers discuss these problems and propose various feasible architectures (Thelen et al., 2015), (Castellano et al., 2015), (Hossain and Muhammad, 2016), (Kang et al., 2015).

Let us analyze the sequence of events (Figure 1) that if not handled on time, can lead to a catastrophic outcome, like death or a durable health damage.

We monitor the health state of the patient with periodic measurements. The time from the occurrence of the emergency state to a catastrophic outcome is a medical fact, as well as the possible outcome itself (risk level). The time from noticing the emergency to an intervention depends on technology. The period of the measurements has to be adapted adequately. For fast events, like a heart attack, where the time limit

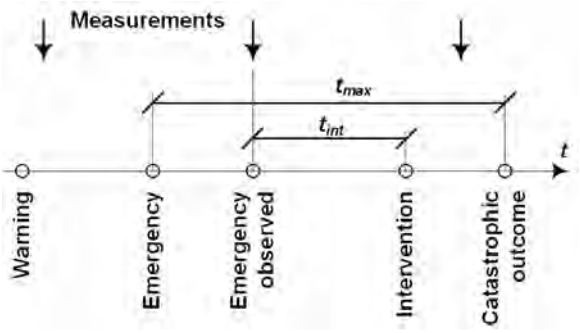



Figure 1: Timing of a medical intervention.

for an intervention is around one hour, the monitoring should be continuous (measurement period reduced to almost zero) and the intervention delay reduced to a minimum. For slow events, like breast cancer, where the disease develops in months, the measurement period is decisive. Not observing this limits makes the monitoring system virtually useless.

Under *warning*, as indicated in the figure 1, we understand some early signals that suggest an increased risk (*yellow alarm*). Early warning permits to extend our time reserve for action. As the indication in this stage is less decisive, full scale intervention is not appropriate. The patient may however reduce the risk by taking a rest, performing additional tests or visiting a doctor.

Breast cancer screening with mammography is a well studied example of risk analysis (Wegwarth and Gigerenzer, 2013). The authors show we should be wary of overdiagnosis and overtreatment.

When deciding what and how to measure and what actions to take, following factors have to be considered:

- disease
  - development time
  - medical risk when not treated on time
  - cost of intervention
- measurements (possible overdiagnosis)
  - cost (financial and organizational)
  - medical side effects
- false positives (overtreatment)
  - probability
  - cost of unnecessary treatment
  - medical side effects

The figure 2 depicts the basic trade-offs in this process. Mobile health technology mainly permits to reduce the cost of the measurement and in this way to make more frequent measurements without visiting a doctor - at home and in travel. It also permits to send

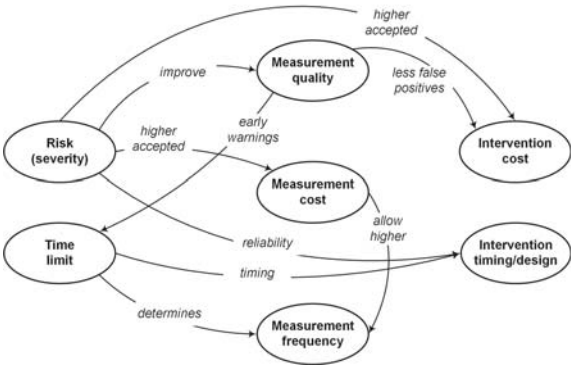



Figure 2: Trade-offs between various factors influencing a successful intervention.

quickly the data and emergency alarms, allowing the patient to maintain the same risk level even having an active life.

In the case of a disease with high risk and short required reaction times (like heart disease), it is important to design carefully the emergency service, ensuring high reliability and respecting the time limits. As this problem goes beyond the desine of the device itself, it will be treated in the following section, regarding the Cyber-Physical-Social Systems.

## 6 CYBER-PHYSICAL-SOCIAL SYSTEMS

It is important to see that mHealth systems consist of more than the purely technical elements. They interact intensely with the physical world. The sensors themselves are material - they need electrical power, their probes can break, sensing surfaces can have poor contact to the skin, output nozzles can be clogged. If we speak of Body Area Network, we have to remember that tissue and clothing damp the wireless signal.

If the patient moves in the environment, the wireless signal connecting him/her with the server may be blocked by obstacles (e.g. when visiting the restroom in a restaurant's basement). He/she can also move into the area where his provider has a poor signal or data roaming is impossible.

The emergency systems (Figure 3) are operated

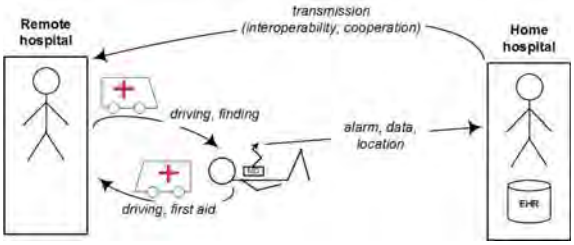



Figure 3: Emergency service.

by humans (*Human in the Loop*) and depend on their cognitive skills and information state. For example, if a heart monitoring system issues an alarm and the ambulance scheduler has no experience with automated systems, his/her thinking time adds directly to the delay of the rescue. The hospital that sends the ambulance is located near to the current position of the patient is not necessarily the same that handles his/her case and owns his/her health record. This requires to arrange the cooperation in advance, including a smooth interoperable data transfer and financial agreements. In the case of a heart attack the whole process has to be concluded during *the golden hour* and leaves no reserve for doubts and clarifications.

This shows that we should ensure that our assumptions about the system reliability and availability are not simplistic. We have to remember that in real life even a trivial cause (empty battery of a sensor, dry skin under the sensor patch, loosened contact of a cable) may have dramatic consequences for the patient's health. Therefore when analyzing the system with formal methods we should be aware that the model and the real object are different and imagination and common sense will be helpful.

## 7 VERIFICATION OF HYPOTHESES

Data generated in an mHealth supported therapy can be reused to verify our assumptions and obtain new knowledge. Let us consider a heart monitoring system issuing alarms and warnings in emergency cases. If the system is working continuously, it can detect significant events also if the patient is not aware of anything. They may consist of changed ECG waveforms, slower / faster / less periodic frequency or else. The intensity and duration of such events may play a role, as well as their sequence. Combining them with other vital signals will make the detection more specific.

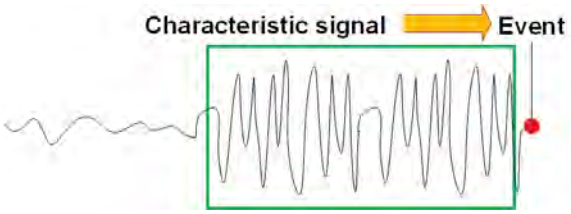



Figure 4: Using knowledge / hypothesis.

Normally, as shown in the figure 4, we assume that we *know* the rules and just *detect* the characteristic signals that indicate that an emergency event is imminent.

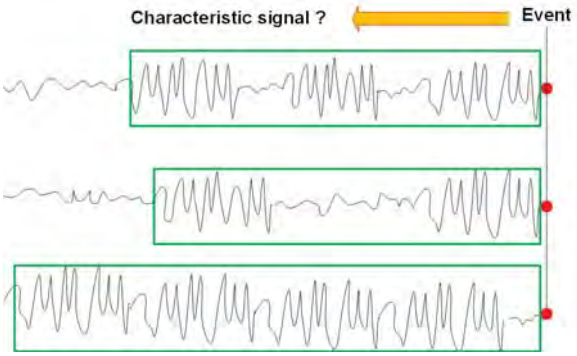



Figure 5: Searching for a new hypothesis.

The area of continuous ECG analysis, especially combined with other signals, is however fairly new. Therefore such event detection rules have to be verified, possibly on many patients, in many situations. Mobile health systems provide us such opportunity. We can analyze the actually occurring events and *look back* at the signals (Figure 5) searching for telltale data events, determine their characteristics and verify if the correlation is significant. In this way, we can create new hypotheses based on data. Evidently, quantifying variability of continuous signals is not easy, especially if we do not know beforehand what are we looking for. This shows that storing raw data (or extensive excerpts of them) for later analysis is useful, even if the benefit is not immediately evident.

# 8 COMMUNICATION AND STORAGE - SECURITY

Data collected from a mHealth device are sent to a receiver. Locally, it may be a smartphone or another data aggregator. A device can also have a direct connection to a server in the cloud. It is interesting what happens to those data later: where are they processed and stored, who owns them, for which purposes can they be used.

## 8.1 Architectures

Following factors are important in the analysis of the transmission and processing of data:

- data quantity
- network independence
- local use of data
  - user feedback
  - local action
- data reuse at a server
- cost of local computation
- cost of transmission

Especially in handling continuous signals there is a trade-off between storage, computation and communications. For taking decisions, we are interested in global parameters (pulse frequency, oxygen saturation) or special states or events (arrhythmia). If the rules to extract such condensed information are well defined, it is better to do the processing locally on the strongest device that can handle this task. This could mean that the implanted sensor device sends raw data to the smartphone, and the smartphone executes the complex calculation. This reasoning is only an example, in a specific case a precise analysis of computational power, transmission channel capacity, energy costs, etc. has to be performed.

Measurements can be used for local action with strong requirements for precision and reaction time, like in glucose management system. It is essential to ensure this function locally, also in the absence of the connection with the remote system. In this case we often speak of Fog Computing - like in Cloud Computing we have storage and processing nodes but not *high in the sky*, but rather locally, *near to the ground* (Gia et al., 2015), (Chakraborty, 2016).

## 8.2 Security

As mobile health systems handle very personal data, they have to be properly secured. This concerns both data transmission and storage. It is easier said than done. Sensor and actuator nodes are low power devices and handling strong security is costly. Basically all exposed networks should be properly managed - this includes applying security patches if necessary. However distributing software updates via network is itself an attack vector.

We have a set of heterogeneous devices coming from different sources, with different lifetimes. Establishing secure communications with the Public Key Infrastructure (PKI) is difficult. We have also to be aware of the trade-off between security and function. On one hand, the devices should be sure they talk to trusted partners. On the other hand, if the security certificate expires on a node and the communication is blocked, this would stop the proper function of the device that may be critical for the health or life of the patient.

Various architectural options, also in the context of the Internet of Things (IoT) and cloud computing, with the stress on the security aspects are presented in (AlTawy and Youssef, 2016), (Gejibo et al., 2015), (Samie et al., 2016), (Suciu et al., 2015), (Larburu et al., 2016) and (Sukor et al., 2015).

# 9 INDIVIDUAL PATIENT AND POPULATION - PRIVACY

Data regarding individual patients can be collected and reused for many purposes. Fitness tracking applications typically permit to send own data to the pool and to compare personal results with the community. This community has some basic stratification, e.g. with respect to gender and age. The quality of input data is unproven, no sources of bias are considered. The power of the solution lies in quantity of data. Such comparison is used mostly for personal satisfaction and motivation for further effort. It has to be mentioned that more effort is not always good, especially for elderly with osteoporosis and worn-out knees. The competition may cause an addiction where gaining points and virtual rewards count more than the actual health.

In order to participate, the patient has to agree, i.e. to express consent to share data. However, the exact conditions of this consent, if published, are never consulted. All detailed data reside on the servers of the provider and the patient somehow assumes that exact times and locations of his/her walks will not be disclosed to third persons or sold.

In the treatment of more serious medical cases, data can be aggregated and analyzed in order to obtain and verify knowledge. This can help to identify risk factors or to issue recommendations for healthy behavior.

If statistics based on collected data are used for generating decisions (actionable knowledge), we have to ensure adequate quality and statistical validity. Various aspect of medical data reuse are discussed in depth in (Sliwa, 2016b). If the decisions apply to the entire population, we should eliminate bias. If the population shows strong variations, it should be properly stratified, and the statistics for each category have to satisfy the quality criteria. This is not easy, as it is more practical to observe the population than to execute a formal clinical trial. To put it simply: if mostly young, technically oriented patients are willing to share their data, those data should not be used as a benchmark for the entire population, including the elderly.

Evidently, the system provider has access to all data and can use them at least to improve the service. Due to the fast pace of the technical development, much faster than the legislation, from formal legal point of view the area of reusing mobile health data is a gray zone. The question of data ownership in a multi-party Internet of Things (IOT) system, with smart medical devices as one of the examples, is discussed in (Sliwa, 2016a).

# 10 CONCLUSIONS AND FUTURE WORK

The basic goal of this analysis is to raise awareness for the semantic aspects of data processing in mHealth systems. It is important to understand the properties of the signals flowing from the sensors and their relevance to the overall health support function of the system. Their properties are determined by the medical factors, like the severity of the case, the possible outcome. the necessary intervention and its time constraints.

After the semantic analysis the elementary properties of the signals can be extracted, which permit to build a formal model:

- data type
- conversion algorithm
- timing requirements
- data quantity
- duration and location of storage
- ownership and protection

Nevertheless, it has to be stressed that for complex Cyber-Physical-Social-Systems, as in Mobile Health, the formal model is only an approximation and has to be constantly verified. The environmental conditions are diverse, the technology changes rapidly and human behavior is difficult to predict, therefore the usage of such systems has to observed and the design assumptions have to be periodically reviewed.

# REFERENCES

Agboola, S. O., Bates, D. W., and Kvedar, J. C. (2016). Digital health and patient safety. *JAMA*, 315(16):1697–1698.

AlTawy, R. and Youssef, A. M. (2016). Security tradeoffs in cyber physical systems: A case study survey on implantable medical devices. *IEEE Access*, 4:959–979.

Baig, M. M., GholamHosseini, H., and Connolly, M. J. (2015). Mobile healthcare applications: system design review, critical issues and challenges. *Australasian Physical & Engineering Sciences in Medicine*, 38(1):23–38.

Castellano, N. N., Gazquez, J. A., Salvador, R. M. G., Gracia-Escudero, A., Fernandez-Ros, M., and Manzano-Agugliaro, F. (2015). Design of a real-time emergency telemedicine system for remote medical diagnosis. *Biosystems Engineering*, 138:23–32.

Chakraborty, S. (2016). *Data Aggregation in Healthcare Applications and BIGDATA set in a FOG based Cloud System*. PhD thesis, University of Cincinnati.

Eapen, Z. J., Turakhia, M. P., McConnell, M. V., Graham, G., Dunn, P., Tiner, C., Rich, C., Harrington, R. A., Peterson, E. D., and Wayte, P. (2016). Defining a mobile health roadmap for cardiovascular health and disease. *Journal of the American Heart Association*, 5(7).

Espinoza, J., Chandy, D., Ochoa, S., Jiménez, J., Huerta, M., Soto, A., Sagbay, G., and Avila, R. (2016). Design of telemedicine management system in ecuador. In *Ecuador Technical Chapters Meeting (ETCM), IEEE*, volume 1, pages 1–6. IEEE.

Gejibo, S., Mancini, F., and Mughal, K. A. (2015). Mobile data collection: A security perspective. In *Mobile Health*, pages 1015–1042. Springer.

Gia, T. N., Jiang, M., Rahmani, A.-M., Westerlund, T., Liljeberg, P., and Tenhunen, H. (2015). Fog computing in healthcare internet of things: A case study on ecg feature extraction. In *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on*, pages 356–363. IEEE.

Hossain, M. S. and Muhammad, G. (2016). Cloud-assisted industrial internet of things (iiot)–enabled framework for health monitoring. *Computer Networks*, 101:192–202.

Huckvale, K., Adomaviciute, S., Prieto, J. T., Leow, M. K.-S., and Car, J. (2015). Smartphone apps for calculating insulin dose: a systematic assessment. *BMC Medicine*, 13:106.

Islam, S. R., Kwak, D., Kabir, M. H., Hossain, M., and Kwak, K.-S. (2015). The internet of things for health care: a comprehensive survey. *IEEE Access*, 3:678–708.

Itrat, A., Taqui, A., Cerejo, R., Briggs, F., Cho, S.-M., Organek, N., Reimer, A. P., Winners, S., Rasmussen, P., Hussain, M. S., et al. (2016). Telemedicine in prehospital stroke evaluation and thrombolysis: taking stroke treatment to the doorstep. *JAMA neurology*, 73(2):162–168.

Kang, K., Wang, Q., Hur, J., Park, K.-J., and Sha, L. (2015). Medical-grade quality of service for real-time mobile healthcare. *IEEE Computer*, 48(2):41–49.

Lanzola, G., Losiouk, E., Del Favero, S., Facchinetti, A., Galderisi, A., Quaglini, S., Magni, L., and Cobelli, C. (2016). Remote blood glucose monitoring in mhealth scenarios: A review. *Sensors*, 16(12):1983.

Larburu, N., Bults, R., van Sinderen, M., Widya, I., and Hermens, H. (2016). Quality-of-data broker for quality-of-data-aware telemedicine systems. *IRBM*, 37(4):210–218.

Samie, F., Bauer, L., and Henkel, J. (2016). Iot technologies for embedded computing: A survey. In *Hardware/Software Codesign and System Synthesis (CODES+ ISSS), 2016 International Conference on*, pages 1–10. IEEE.

Sannino, G., De Falco, I., and De Pietro, G. (2015). A supervised approach to automatically extract a set of rules to support fall detection in an mhealth system. *Applied Soft Computing*, 34:205–216.

Sannino, G. and De Pietro, G. (2014). A mobile system for real-time context-aware monitoring of patients health and fainting. *International journal of data mining and bioinformatics*, 10(4):407–423.

Silva, B. M., Rodrigues, J. J., de la Torre Díez, I., López-Coronado, M., and Saleem, K. (2015). Mobile-health: a review of current state in 2015. *Journal of biomedical informatics*, 56:265–272.

Sliwa, J. (2016a). A generalized framework for multi-party data exchange for iot systems. In *2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pages 193–198. IEEE.

Sliwa, J. (2016b). Reuse of data from smart medical devices for quality control and evidence based medicine. In Dobre, C. and Xhafa, F., editors, *Pervasive Computing: Next Generation Platforms for Intelligent Data Collection*, chapter 9, pages 255–297. Morgan Kaufmann.

Soh, P. J., Vandenbosch, G. A., Mercuri, M., and Schreurs, D. M.-P. (2015). Wearable wireless health monitoring: Current developments, challenges, and future trends. *IEEE Microwave Magazine*, 16(4):55–70.

Suciu, G., Suciu, V., Martian, A., Craciunescu, R., Vulpe, A., Marcu, I., Halunga, S., and Fratu, O. (2015). Big data, internet of things and cloud convergence–an architecture for secure e-health applications. *Journal of medical systems*, 39(11):1–8.

Sukor, J. A., Mohktar, M. S., Redmond, S. J., and Lovell, N. H. (2015). Signal quality measures on pulse oximetry and blood pressure signals acquired from self-measurement in a home environment. *IEEE journal of biomedical and health informatics*, 19(1):102–108.

Thelen, S., Czaplik, M., Meisen, P., Schilberg, D., and Jeschke, S. (2015). Using off-the-shelf medical devices for biomedical signal monitoring in a telemedicine system for emergency medical services. *IEEE journal of biomedical and health informatics*, 19(1):117–123.

Wegwarth, O. and Gigerenzer, G. (2013). Overdiagnosis and overtreatment: evaluation of what physicians tell their patients about screening harms. *JAMA internal medicine*, 173(22):2086–2087.

Xue, S., Chen, X., Fang, Z., and Xia, S. (2015). An ecg arrhythmia classification and heart rate variability analysis system based on android platform. In *Future Information and Communication Technologies for Ubiquitous HealthCare (Ubi-HealthTech), 2015 2nd International Symposium on*, pages 1–5. IEEE.

# Special Session on Analysis of Clinical Processes

# Full Papers

# Enabling Interactive Process Analysis with Process Mining and Visual Analytics

P. M. Dixit[1,2], H. S. Garcia Caballero[1,2], A. Corvò[1,2], B. F. A. Hompes[1,2], J. C. A. M. Buijs[1] and W. M. P. van der Aalst[1]

[1]*Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands*

[2]*Philips Research, Eindhoven, The Netherlands*

{*p.m.dixit, h.s.garcia.caballero, a.corvo, b.f.a.hompes, j.c.a.m.buijs, w.m.p.v.d.aalst*}*@tue.nl*

Keywords: Process Mining, Visual Analytics, Compliance Analysis, Performance Analysis, Classification.

Abstract: In a typical healthcare setting, specific clinical care pathways can be defined by the hospitals. Process mining provides a way of analyzing the care pathways by analyzing the event data extracted from the hospital information systems. Process mining can be used to optimize the overall care pathway, and gain interesting insights into the actual execution of the process, as well as to compare the expectations versus the reality. In this paper, a generic novel tool called *InterPretA*, is introduced which builds upon pre-existing process mining and visual analytics techniques to enable the user to perform such process oriented analysis. InterPretA contains a set of options to provide high level conformance analysis of a process from different perspectives. Furthermore, InterPretA enables detailed investigative analysis by letting the user interactively analyze, visualize and explore the execution of the processes from the data perspective.

## 1 INTRODUCTION

Business process management systems and other "process aware" information systems (CRM, ERP, etc.) are gaining popularity across many domains, also in the healthcare sector (Dwivedi et al., 2001). In a healthcare setting, the processes, generally referred to as care pathways, are used to map the flow of patients in order to enable efficient operations in a hospital along with standardizing the treatment plan across patients (Sutherland and van den Heuvel, 2006; Weigl et al., 2012). However, patients in a hospital may have different profiles and hence, different care pathways. Furthermore, the scope of care pathway can cover a broad spectrum from medical guidelines to patient logistics. Thereby, the notion of processes is highly sophisticated and flexible in a healthcare setting. In order to standerdize the overall process, make the steps evident for all the members of the treatment plan, and to improve the overall efficiency, there is a strong emphasis to model and make the steps of the processes explicit and better managed. Alternatively, these steps may be inherently known by the participants of the process, and may not be formally documented anywhere. In such cases only the primary users may be aware of the important steps in the process. Due to the lack of documentation, the standerdization and optimization of the process could become challenging. Moreover, if the resources, e.g. nurses, get replaced, then the new resources may be unaware of the logical flow of the process. The Hospital Information Systems (HIS) are designed to record all the information, such as treatment plan, the interactions with patients and medical staff, exams and treatment procedures, logistics, medical decisions etc., that takes place in a hospital (Graeber, 1997). This information could be effectively used to analyze the process performance and gain insights into the process with the intention of optimizing the care pathways.

*Process mining* acts as a key enabler for analysis of process based systems using event logs (Aalst, 2016). Event logs can be extracted from the HIS of the hospital. On a broad level, process mining can be categorized into process discovery and conformance analysis. As the name suggests, *process discovery techniques* aim at discovering process models automatically using the information from the event log gathered from the HIS. *Conformance techniques*, use a pre-defined process model and map it to the corresponding event log, to determine how well the process is described according to the event log.

If the event logs extracted from the HIS are considered to be accurate depiction of the reality, and if the process model defines the ideal process flow,



Dixit, P., Caballero, H., Corvò, A., Hompes, B., Buijs, J. and Aalst, W.
Enabling Interactive Process Analysis with Process Mining and Visual Analytics.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 573-584
ISBN: 978-989-758-213-4

Figure 1: Conformance analysis view of a real world process model (used in the case study). The sheer size of the process model makes it difficult to draw inferences by merely looking at the whole process in a holistic manner, generating the need for smart interactive analysis. Figure used for illustrative purposes only, to represent the scale and nature of a large healthcare process.

then the analysis performed by the conformance techniques can provide valuable insights about what went right and what went wrong in the process. Conformance analysis can be used to gain insights about the compliance aspect of the process, such as understanding where the deviations occur in a process (Adriansyah et al., 2011; Munoz-Gama et al., 2014). As there is a time notion associated with every step in a process, the conformance analysis could also be used in order to investigate the performance aspect of the process, for example, understanding where the bottlenecks occur in a process. Root cause analysis could be performed using traditional classification techniques to understand the reasons behind the deviations and bottlenecks in a process.

Although conformance results enable compliance and performance analysis, there are still some open challenges which could make the analysis difficult. This is specifically true for healthcare processes, which are extremely flexible and dependent on multiple factors thereby making them very complex. A process discovery on an event log from HIS usually results in complicated and/or huge process model as shown in Figure 1. Most of the current visualizations display the conformance results in their entirety. From a user perspective, displaying the conformance results on a big and complicated process model with many nodes, may turn out be too daunting. Moreover, most of the current techniques are standalone techniques and do not support effective combinations of intra-disciplinary analysis. Furthermore, investigation of the satisfaction of specific protocols and/or KPIs in the process is usually only possible through data specific analysis. That is, it is not possible to interactively perform various compliance and performance analysis directly on a process model.

In order to address the issues discussed above, we propose a tool - *InterPretA (Interactive Process Analytics)* which serves as a one stop solution for performing process-centric analytics using pre-existing conformance analysis techniques. InterPretA has been implemented using Process Analytics[1] plug-in available in the nightly builds of the Process Mining framework - ProM[2] (Dongen et al., 2005). As opposed to traditional approaches, which usually provide metrics represented by some *numbers* which may be difficult for the user to comprehend, in our approach the user can interactively explore the performance and compliance specific issues guided by visual analytic techniques. Firstly, InterPretA enables the user to get different *helicopter* views on the process. That is, the results from conformance analysis could be visualized from different perspectives on the complete process. Next, InterPretA lets the user *interactively* explore the results from conformance analysis for certain fragments/paths in the process. This is especially useful for large process models, wherein the user may want to focus on analyzing only a part of the process. Finally, InterPretA also supports performing the root cause analysis, explaining why the deviations or bottlenecks occur in the process. At all times, the user can also visualize and quantify the information with the help of some graphs, which could be used for exploratory purposes or to answer some questions or for making reports.

## 2 BACKGROUND

The applicability of visual analytics in process mining has been explored in (Kriglstein et al., 2016; Aalst et al., 2011; Mannhardt et al., 2015). However, it is still in its nascent stages and most of the techniques focus on a broad perspective. The process model (along with the results from conformance analysis) is central to the analysis in InterPretA. In this section, we introduce the background of the elements which constitute to the backbone of InterPretA.

### 2.1 Event Log

The HIS, such as electronic health records systems, could be used in order to extract the event logs. *Events* form the basic building blocks of an event log. Every event is associated with a particular case, which is identified by a case ID. An event represents occurrence of an *activity* in a particular case. Events have a

[1] http://svn.win.tue.nl/repos/prom/Packages/ProcessAnalytics

[2] http://www.promtools.org

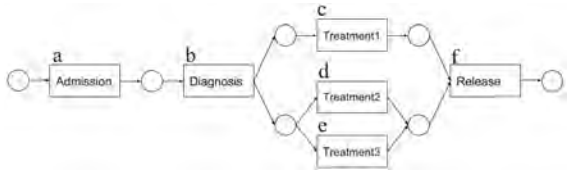



Figure 2: A simple carepathway process, modeled using Petri net. After *Admission*, *Diagnosis* is performed, following which *Treatment1* is performed in parallel with either *Treatment2* or *Treatment3*. Finally the patient exits the carepathway as denoted by *Release* activity.

*transaction type*, such that for every *activity instance*, there may be a schedule event, a start event, a suspend event, a resume event and finally, a complete event. Each event has a *time-stamp* associated with it, which denotes the time of occurrence of a particular event. Furthermore, each event may contain additional event specific *data attributes*, for example the *resource* which performed the particular activity represented by the event. Similarly, each case may have case level data attributes, such as the gender of the patient, age of the patient, etc. The event log is thus a collection of a sequence of events (so-called traces) that represent the individual cases. Event logs extracted from the HIS can be used in a process mining context for performing process discovery; or conformance analysis on a pre-existing or a discovered process model.

## 2.2 Petri Nets

Multiple notations exist to represent process models, for example, BPMN, UML diagrams, EPCs, Petri nets. InterPretA uses Petri nets as a way to represent process models. The selection of Petri nets was inspired by the virtue of the properties supported by Petri nets which enable detailed conformance and performance analysis. Petri nets support modeling of the traditional business process concepts, such as concurrency, choices and sequences (Aalst, 2016). Figure 2 shows an example of a Petri net model, where places (circles) are used for modeling logic (i.e. sequence, choice, concurrency, loops etc.) and the rectangles represent the activities (tasks) of the process.

## 2.3 Alignments

In our approach, we use alignments-based conformance analysis as proposed in (Adriansyah et al., 2011) for guiding the compliance and performance analysis in the context of processes. As discussed above, conformance analysis helps in determining how well a process model fits the reality represented by the event log. This information can be beneficial when determining any compliance issues related to either the complete process or some fragments of the process. Furthermore, this information could also be used in order to determine any performance related problems and analysis. In this sub-section, we briefly discuss the idea behind the alignment strategy that is used for determining the conformance of a model and event log as proposed in (Adriansyah et al., 2011). Often times, the event log may contain noisy and/or incomplete data. Alignments provide a handy way to deal with such data. Hence, instead of relying completely on the event log, we use the information from alignment based conformance analysis as the ground truth. As alignments relate events in the event log to model elements, they are ideal for the process-oriented analysis approach supported in InterPretA. Aligning events belonging to a trace with a process model can result in three types of so called *moves* - synchronous move, move on model and move on log. An example alignment is shown in Table 1.

Table 1: Example of conformance alignment moves using Figure 2. Step 1,2,3 and 5 are synchronous moves. Step 4 and step 6 are move on model and move on log respectively.

| Trace in event log | a | b | c | d | e | >> |
|---|---|---|---|---|---|---|
| Possible run of model | a | b | c | >> | e | f |
| Steps | 1 | 2 | 3 | 4 | 5 | 6 |

- **Synchronous move:** Occurrence of an event belonging to a trace can be mapped to occurrence of an enabled activity in the process model.
- **Move on model:** Occurrence of an enabled activity in the process model cannot be mapped to the current event in the trace sequence.
- **Move on log:** Occurrence of an event in the trace cannot be mapped to any enabled activity in the process model.

Optimal alignments provide a means to match a trace in an event log with a corresponding model run. If for a trace, all the moves are synchronous or invisible model moves, then that trace can be perfectly replayed by the model.

## 2.4 Classification

In literature, classification techniques (Goedertier et al., 2007; Buffett and Geng, 2010; Poggi et al., 2013) have been applied in the field of process mining, in order to address multiple problems. (Leoni et al., 2015) provide a framework for applying classification and correlation analysis techniques in process mining, by using the results from conformance analysis. Traditionally, classification techniques in process mining context are used to perform tasks such as

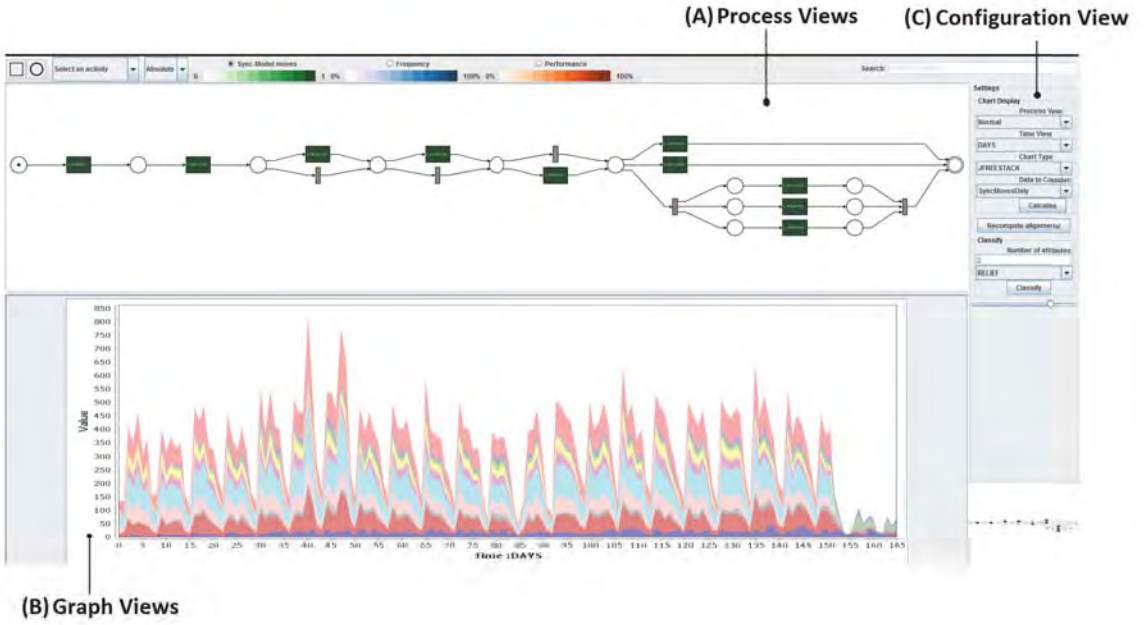



Figure 3: A snapshot of our interface. Section (A) is dedicated to the process views. The user can directly interact with the process model. The visualizations, in Section (B), are triggered by this user interaction and by the configuration settings in Section (C).

abstracting event logs, identifying bottlenecks, understanding and modeling guards in data-aware process models, annotating and clustering cohorts by splitting event logs into sub logs etc. However, many of the applications of classification techniques in process mining focus entirely on the data perspective. As discussed above, in InterPretA, a process model is central to everything. Hence, the classification task is also driven from a process perspective, enabled by the results of the conformance analysis, visualized on the process model.

We use the traditional classification techniques in order to perform root cause analysis, for example, to find which resource caused a delay in the process or what caused a deviation in a process. More particularly, we support the use of pre-existing feature selection techniques, which use data attribute values (either case level attributes or event level attributes) in order to classify based on the desired output, for example, values above or below a certain threshold for performance. Furthermore, well-known ranking techniques are used to determine the rank of each attribute based on the ability of an attribute to classify based on the desired output.

Besides using traditional classification and ranking techniques, we use the context-aware process performance analysis framework recently proposed in (Hompes et al., 2016) to classify the event data using functions that take the process context into account. For example, prefixes of activities in their trace and structural properties of the model can be used to classify activities and cases. Additionally, this technique can help rank attributes and attribute values by means of statistical analysis, as described in sub-section 2.5. For example, cases that lead to statistically significantly different conformance results can be classified together, and features that lead to those classifications can be ranked.

In our approach, the user interactively selects the interesting fragments in the process model and configures the classification parameters. Based on user selection, the alignments are either recomputed, or pre-existing alignments are used, for the classification task. For performing the traditional classification tasks we use Weka, which inherently supports an array of classification algorithms (Hall et al., 2009).

## 2.5 Statistical Performance Analysis

Statistical analysis can be performed after classification in order to discover whether observed differences between different classes are statistically significant. For example, different resources that execute an activity might lead to significantly different execution times, cases for patients of a specific age group might take significantly longer than those for other patients, waiting times might be shorter when an activity is preceded by a certain other activity, etc. When different classes do not lead to statistically significant different values for the chosen performance metric, classes

can be grouped in order to reduce the selection dimensionality for the user. Well-known statistical tests such as analysis of variance are used here.

# 3 PROCESS ANALYTICS

In this section, we discuss the main application analysis enabled by InterPretA. As mentioned above, we focus primarily on the analysis from a process perspective, as compared to analyzing the data from a data perspective only. That is, a process model is central to the analysis enabled by our technique, and the results from conformance and performance analysis are the primary enablers of the data analysis.

Firstly, we discuss the general analysis that can be performed from a process perspective. Traditionally, process based analysis can be used to identify what is functioning correctly in the process and to analyze where the potential problems lie in the executions of process in reality. Process oriented analysis can be broadly categorized into:

- **Compliance analysis:** The focus of compliance analysis is to investigate questions pertaining to auditing, adherence to business rules or protocols etc. It should be noted that InterPretA provides easy support for detecting generic compliance issues that might exist in the process. For more sophisticated compliance analysis, we refer to specific techniques by (Ramezani Taghiabadi et al., 2013; Knuplesch et al., 2013).
- **Performance analysis:** Performance analysis is used to analyze the performance aspect of the process execution. Particularly, the performance based analysis are used to explore issues such as identifying any bottlenecks in the process and the steps needed to optimize the process. The need for optimization of process is discussed in, and motivated from (Mahaffey, 2004).

InterPretA enables both compliance oriented and performance oriented analysis of the process. Evidently, the conformance and performance analysis are rather closely tied; and not actually independent from each other. Firstly, InterPretA supports helicopter views on the process, which provide a high level overview of the behavior of the process based on the conformance analysis. Secondly and more importantly, InterPretA allows the user to interactively explore and perform detailed analysis of the process. In the subsections to follow, we discuss the types of analysis enabled by InterPretA's interface, and the configuration options available that correspond to each type of analysis. We begin with the so-called views, followed by the interactive analysis component of the tool.

## 3.1 Graph Views

The event data are visualized in terms of stacked area charts and stacked bar charts. This view is represented in Figure 3B. We chose this representation because it makes comparisons more natural for the user and allows rapid discovery of outliers. For non-classification based analysis, the X-axis (domain axis) describes the time distribution and the Y-axis describes the frequency of occurrence. The X-axis can be sorted and configured primarily based on two view types: absolute time and relative time. The absolute time view shows the actual occurrence time of a particular activity, based on the conformance analysis results. Furthermore, it is possible to abstract the absolute timescale into categories such as: the day of the week when the event occurred, or the month of the year when the event occurred etc. Alternatively, the user can choose to view the X-axis as a relative time axis. That is, the graph can be plotted corresponding to the occurrence of event *relative* to something. For example, the user can choose to view the distribution of events compared to the start of the case (Figure 4), to which the event belongs, or relative time compared to the execution of another event in the case (Figure 5).

The X-axis of the graph view represents the configurable absolute or relative timescales for the user. The Y-axis, i.e. the frequency axis is also configurable. The user can choose from plotting only synchronous moves, only log moves, or both the synchronous and log moves in the alignment. Viewing both synchronous and log moves essentially shows the information from the complete event log. Figures 4 and 5 provide examples of such configuration settings.

The different views on graphs lead to interesting analysis aspects from the data. For example, one popular type of analysis enabled by our approach is concept drift analysis. Concept drift analysis allows the user to identify how a process changes over a period of time. Considering the absolute timescale, the user can analyze the changes in event patterns over time.

## 3.2 Process Views

We support three types of process views that can be used for interactive analysis by the user (see Figure 3A). These are explained in the following subsections.

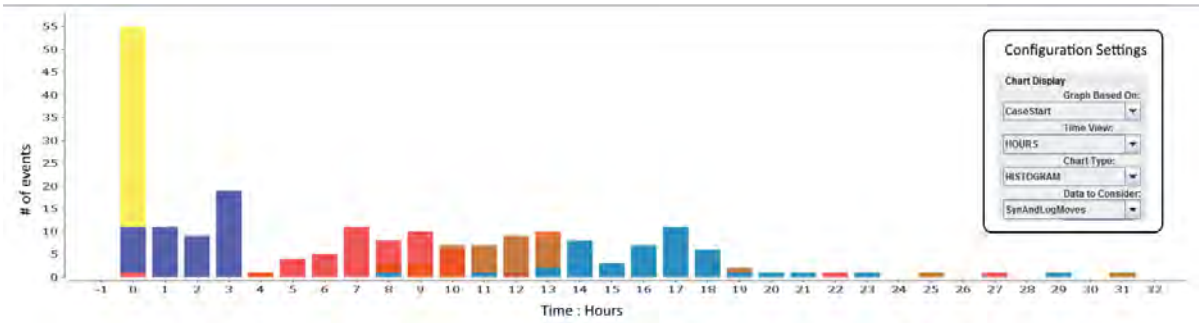



Figure 4: Graph view to show the disribution of *selected* activities from the process model (not shown here), since the start of the case. Both the synchrnous and log moves are considered here, as evident from the configuration settings. The color of each bar in histogram corresponds to every selected activity from the process.

#### 3.2.1 Frequency View

The frequency view allows the user to get an overview of the actual occurrence frequencies of the overall process (e.g. Figure 6). The frequencies of occurrence of an activity are obtained from the alignments projected on the process model. The frequency view encodes the frequency of occurrence of activities in the shades of blue, as projected on the synchronous moves in the alignments. Darker blue and lighter blue colors represent extremely frequent activities and infrequent activities respectively. The user can easily identify the common and uncommon activities or fragments in the process, based on the background color of the activities in the process model. The following options are available to configure the view on frequencies:

- **Absolute View:** Calculates the frequencies, based on the absolute numbers. For example, the color of an activity is determined by the absolute number of synchronous moves. The activity with most number of synchronous moves is colored the darkest. This view is similar to the default view of the inductive visual miner (Leemans et al., 2014).
- **Average Occurrence per Case View:** Calculates the frequencies, based on the average occurrences per case. The color of the activity is determined by the number of times the activity has a synchronous move in a case, normalized over the log. Hence, in case of a loop, if the activity has many synchronous moves within a trace, for many cases in the event log, then this activity would be colored the darkest. Similarly, if an activity has no synchronous moves (i.e. no corresponding event in the event log), then it would be colored the lightest.
- **Number of Cases:** Calculates the frequencies, based on the number of cases for which the synchronous move occurred. The color of an activity is determined by the ratio of the number of cases for which synchronous moves occurred over the total number of cases in the event log.

#### 3.2.2 Fitness of Model and Log View

The fitness of model and log view is derived based on the synchronous and model moves, along with the log moves in the process model. This view represents the fitness of event log and process model. The activities are colored in the shades of green. The actual color of an activity is dependent on the ratio of:

$$\frac{\#Synchronous\ moves}{\#Model\ Moves + \#Synchronous\ moves} \quad (1)$$

Hence, darker shade of green for an activity in a model imply that the particular activity had more synchronous moves than model moves, i.e. it is described very well according to the model. As with the *frequency view*, fitness of model and log is configurable to obtain different views. That is, the activities can be colored based on the ratio of absolute occurrences of synchronous and model moves, or the ratio of average occurrences of synchronous and model moves per case, or the ratio of number of cases for which synchronous and model move occurred.

#### 3.2.3 Performance View

The frequency and fitness views enable the user to easily investigate the common and uncommon paths and deviations in the process. However, the performance view allows the user to investigate the stages of the process where bottlenecks occur, or where there is room for improvement. In the performance view, the activities are labeled in the shades of red. The shade depends on the execution time between the immediate previous synchronous activity and the current synchronous activity for every case. The darker shades of red imply that the synchronous versions were executed after a long delay since the completion of the

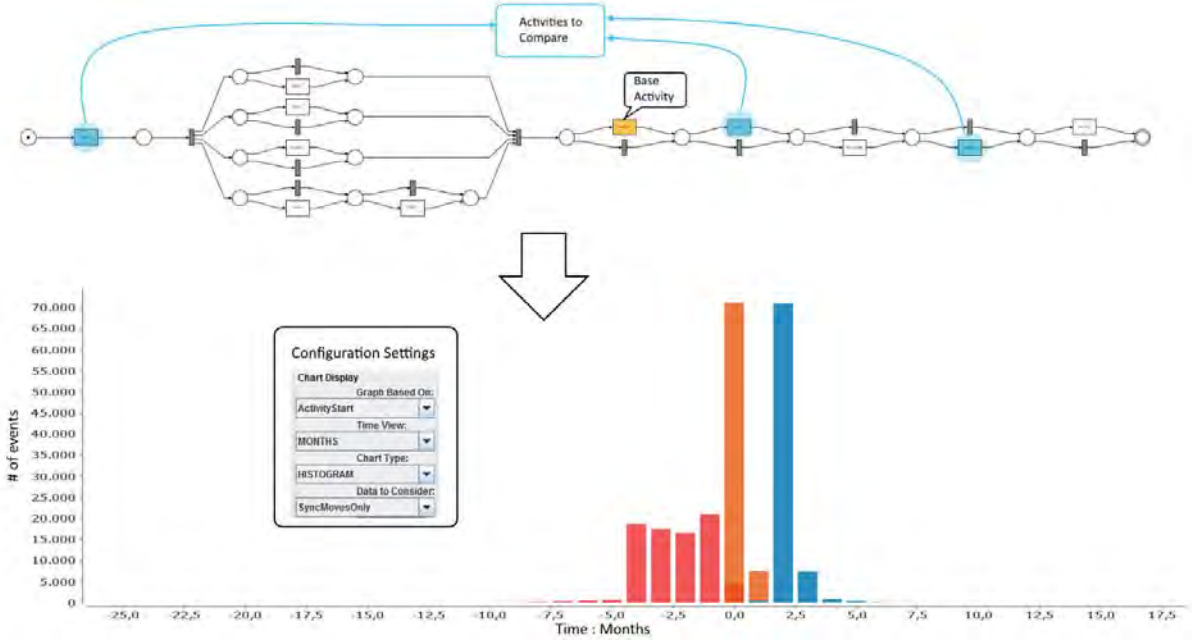



Figure 5: Graph view to compare the distribution of activities with respect to a selected activity. A single activity from the process view is chosen by the user as the base activity, followed by a number of activities which should be compared with the base activity. The histogram shows the time distribution and the number of times the selected activities to be compared occur, *before* or *after* the base activity. The color of each bar in histogram corresponds to each selected activity from the process. It should be noted that the base activity always occurs at time 0 and is not shown in the histogram view.

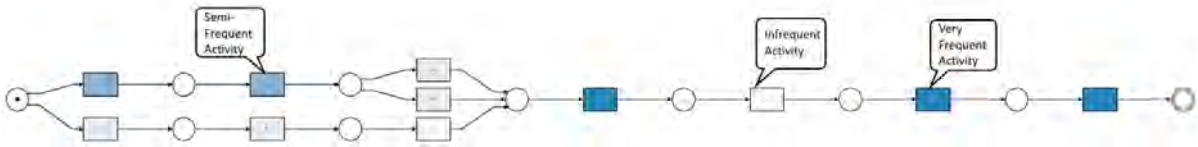



Figure 6: An example of frequency view of the process model. The more frequent activities/paths could be easily visualized in the model. The user can use this information to interactively explore the specific details pertaining to the frequent tasks.

previous synchronous activities. As with the fitness of model and log, the user can configure different views (absolute, average occurrence per case or number of cases) on the performance dimension.

It should be noted that the transactional information of an activity (e.g. events representing lifecycle an activity) can also be used for performance analysis. That is, the lifecycle information of an activity is useful in detecting the actual performance time of an activity. However, we do not show this information on the helicopter performance view as often event logs only contain complete events, i.e. only one event per activity instance. When lifecycle information is present, additional performance statistics are calculated and can be shown.

## 3.3 Interactive Analysis

The graph views and process views enable analysis of the process from a high level. In a typical process analytic setting, the user is also interested in performing a detailed root cause analysis. In this section, we describe how InterPretA supports the user in performing such detailed analysis starting from the high-level views. Five high-level tasks (Schulz et al., 2013) have been identified to perform the interactive analysis.

**Task 1: Deviation Analysis.** The *fitness of model and log view* provides a good overview of how well the data and model fit, and where the possible deviations are located in the process. The next step would be to investigate what causes the deviations in the process. For example, suppose that the conformance analysis indicates that a particular task is sometimes skipped (move on model) in the process. This could be explained by the attributes associated with the case, for example the resources performing the task.

In order to investigate what or who causes deviations in the process, classification analysis

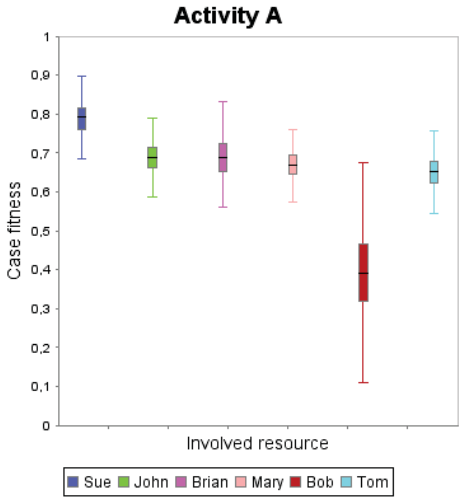



Figure 7: Compliance analysis for resources involved in activities. When resource 'Bob' is involved in activity 'A', the case fitness to the model is significantly lower.

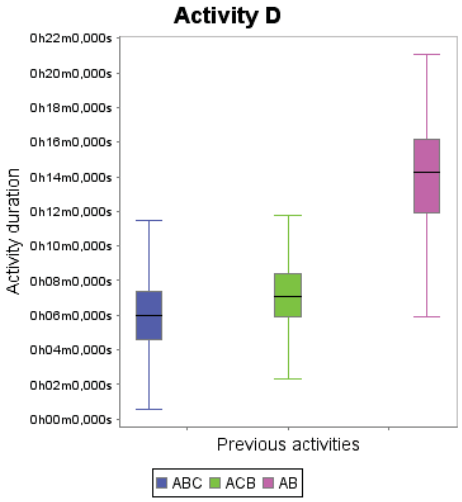



Figure 8: Bottleneck analysis for activities preceding a bottleneck activity 'D'. When activity 'C' is not performed before activity 'D', the duration is significantly longer.

is used. The user can select the desired activity from the activity drop down list, and configure the desired classification settings, to classify what or who causes the synchronous, log and model moves for the particular activity. Here, we make use of the context-aware performance analysis framework proposed in (Hompes et al., 2016) to provide and rank the statistically significant attributes (i.e. the process context). Many combinations of classification and performance/conformance functions are automatically generated from a collection of contexts. Then, statistical testing is automatically performed to verify whether different classes (contexts) lead to statistically significantly different results, e.g. in Figure 7, we see that Bob has a high variability in activity A. For those combinations where this is the case, the specific classification is returned to the user as it might lead to interesting insights. For example, Figure 7 shows an example plot of the fitness of a model, split by the different resources that can perform activity 'A' in the process. From this example, we might deduce that the fitness of a case to the model depends highly on which resource executes activity 'A'. The statistical analysis tells us which resource(s) lead to significantly different results.

**Task 2: Bottleneck Analysis.** The *performance* view provides an overview of the time spent between the activities. For example, it may be the case that a particular path in a process is much slower than the other paths. The user might be interested in finding out, how a particular path with delays is different from other paths in the process. In order to find this out, the user can select the activities from the process model which correspond to the path with delays. Next, the user can perform classification analysis, with the output of classification set to be fitting versus non-fitting cases. The fitting cases in the output would be the ones which have the *synchronous moves* for all the selected activities. Alternatively, the user might be interested in finding out all the cases in the process (or a path in the process) which are executed within a certain time frame. For such analysis, the output for classification could be a threshold for *time taken*, such that the attributes which better define the performance (e.g. above or below the threshold for time taken) would be identified by the classification techniques. Alternatively, we automatically verify whether different contexts lead to significant differences in performance. For example, Figure 8 shows an example where the duration of an activity is significantly different for different prefixes in the trace. From this example, we might conclude that activity 'C' should be made mandatory.

**Task 3: Frequency-oriented Compliance Analysis.** The frequency view on process model gives a good overview of the frequency distribution of activities in the overall process. This view is already useful for answering some compliance questions such as *whether or not an activity occurs at least once for every case*. However, the user might also be interested in interactively investigating some non-trivial KPIs such as the occurrence of a particular activity triggering the occurrence of another activity. In order to enable such analysis, the user can select a set of activities, and select the appropriate configuration settings, to check the co-occurrences and the frequencies of co-occurrence for selected activities. Additionally,

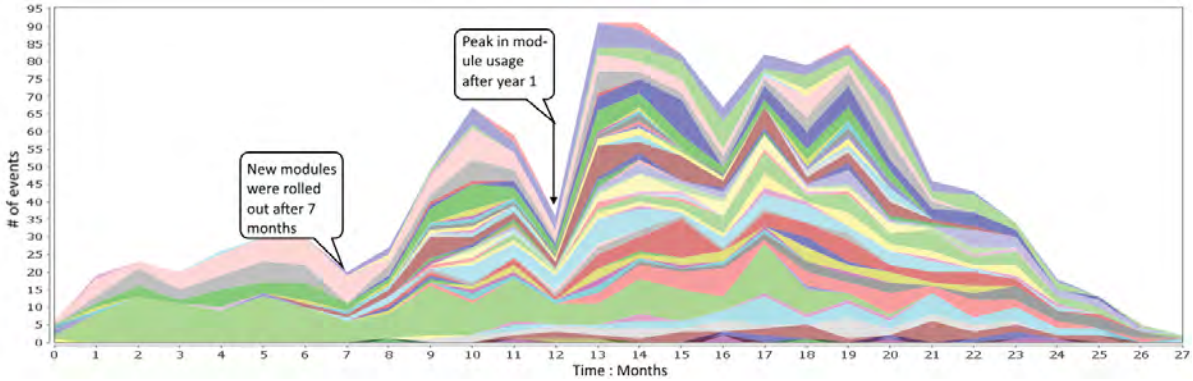



Figure 9: Graph showing the concept drift and changes among different modules in the LUMC dataset. The colors in the graph correspond to an activity belonging to a particular module. For first 7 months, all the activities belonged to one module type. There is a steep fall in the module usage towards year end and a peak in the module usage in year 2.

the user can view differences in performance for different classes.

**Task 4: Performance-oriented Compliance Analysis.** The performance view allows the user to easily investigate the overall time between activities and/or where the bottlenecks may be in the process. However, there could be some time-critical KPI analysis that the user might be interested in. For example, a certain activity 'B' should be performed within *x-hours* after executing an activity 'A'. The user can select the activity 'A', and activity 'B' (among others if needed), to visualize the time span of occurrence distributions of 'B' with respect to 'A'. Different contexts can be tested for their impact significance on the time between the two activities.

**Task 5: Process Fragmentation.** The user might also be interested in exploring certain fragments of a process. That is, instead of considering the alignments on the complete process model, the user might only be interested in investigating how the process behaves, corresponding to only a few activities. In order to achieve this, the user can select the interesting activities, and re-compute the alignments only for such activities. The re-computed alignment would consider the moves only for the selected activities, and all the other moves would either be ignored or considered model moves with zero cost. Based on the fragmented process view, and the alignments based on filtered model, all the previous analysis could be repeated. An example of such task can be derived from Figure 5, wherein the user selects some activities (marked blue). The user may then *re-compute* alignments only for the selected activities.

# 4 CASE STUDY

As a part of evaluating our approach, we perform analysis on a real-life dataset from a diabetes treatment care process, provided by a Dutch hospital - Leiden University Medical Center (LUMC). In order to cope with unstructured processes as discussed in (Fernandez-Llatas et al., 2015), LUMC has proposed and rolled out six specialized modules as a part of its diabetes treatment plan process. Each module can be viewed as a separate process, each consisting of a series of appointments, some of which may be skipped and some of which may occur in any order. The HIS used by LUMC records all the relevant information about the appointments in each of these modules. The event log created from the HIS logs spans over more than 2.5 years, and involves almost 300 cases (patients). The data was anonymized by using pseudonymised patient ids. Using this information, we perform both exploratory analysis, and also answer some compliance and performance related questions. Rather than splitting the log into sub-logs corresponding to each module type, we use the complete event log and use the inductive miner infrequent process discovery algorithm (Leemans and Aalst, 2014) to discover a process model, containing all the modules together.

## 4.1 Analyzing Change of Modules

By plotting the data for all the events, of all the modules, its easy to visualize the patterns of changes in the types of module over time. The domain axis was sorted to contain the absolute time span from lowest to highest, represented in terms of months. From Figure 9, it can easily be concluded that for the first few months (approximately until 7 months), only one

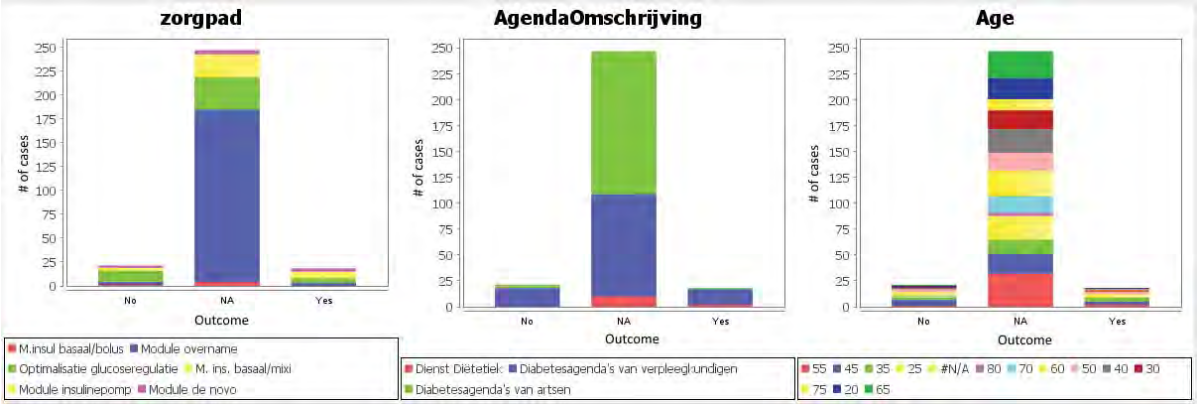



Figure 10: The top three ranked attribute classifiers for the classification analysis based on the answers to the question *Were the personal goals met?*. The top ranked attribute classifier based on the ranking algorithm was zorgpad (i.e. *module type*).

module existed, as all the activities within this time span belong to the same module type. From Figure 9 it can also be seen that this module still persisted over time, but was gradually replaced by the other modules. One more interesting pattern that could be observed from Figure 9, is the decrease in the number of events towards the end of the timeline. A logical explanation for this pattern is that since the appointments (events) for modules are usually scheduled in the future, a more distant future has fewer number of appointments.

## 4.2 Classification Analysis

At the end of each module, the patients are asked to fill in a questionnaire, evaluating the patients treatment plan, understanding and satisfaction of the module. It is interesting to analyze the outcome of the treatment, based on the answers in the survey. For one of such questions, *Were the personal goals met?*, we perform classification analysis with the Information gain classifier, and answers to question (*Yes, No or NA*) as output. The top three ranked attributes which best classify the output are shown in Figure 10. It is however important to note that *module overname* is a basic module (to meet diabetic team members) and hence majority of the patients from such modules have not reached their goals yet, thereby having the corresponding value of NA, as shown in Figure 10. We select the top ranked attribute classifier, found to be the *module type*. For the majority of patients, no value was recorded for this question (value of NA). However, for the patients who did fill in the survey, one prominent outcome, as evident from Figure 11 is that for the module *Optimalisatie glucoseregulatie*, almost twice the amount of patients did not meet their expectations fully. This suggests that another module might have been more suitable or calls for improvement in the module to better manage the patient expectation.

## 4.3 Compliance Analysis - Time Perspective

Next, we focus on one individual module, to perform some preliminary compliance analysis. The expectation is to have every appointment completed within a certain time-frame. We select a particular appointment type from the chosen module and plot a histogram with domain axis showing the time in weeks, since the beginning of the case. Ideally, the chosen activity should be completed within 9-10 weeks since the start of the case. However, as becomes clear from Figure 12, in reality this activity is mostly completed *before* the expected time duration, thereby meeting the desired KPI for majority of the patients.

# 5 CONCLUSION AND FUTURE WORK

In this paper, we introduced a novel tool for enabling interactive process-oriented data analysis. The tool builds upon and brings together existing techniques from process mining, data mining and visual analytics field, to enable interactive process analysis. It supports exploratory analysis through different *helicopter* views on the process. In contrast to existing approaches, it is highly interactive, which could be used to perform root cause analysis for any problems in the process. The application areas of the tool are broadly categorized, and the tool was utilized to analyze a real-life dataset. The tool relies on the traditional ways of representing the data (e.g. histograms) for process analytics.

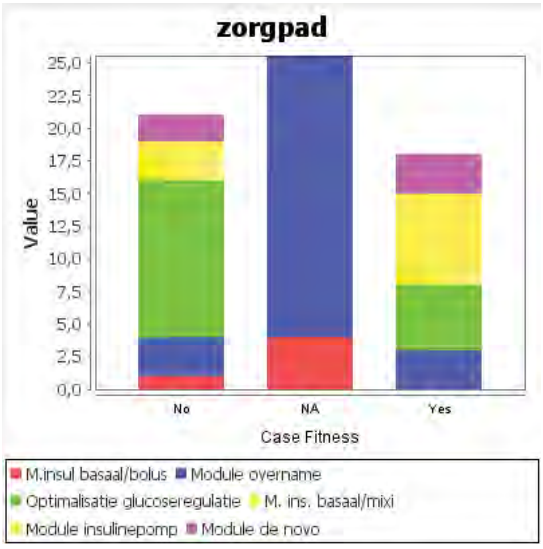



Figure 11: Zoomed in version showing responses for Yes and No, for the top ranked classifier zorgpad (*module type*) based on the answer to the question *Were the personal goals met?*.

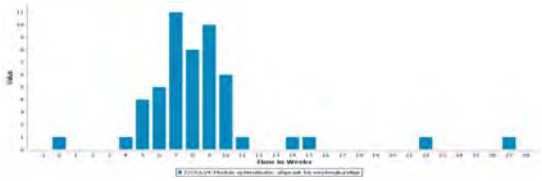


Figure 12: Compliance analysis for completetion of the activity.

In the future, we aim to support more interactive data exploration, for example from the plotted graphs views. Currently the tool is limited to read-only data plots in terms of histograms/stacked charts. Drill down and roll up analysis could be introduced, to support more data-oriented task analysis. Furthermore, the impact on the current process could be visualized based on the interaction of the user with the data from the graph view. This could also lead to understanding the configurability of the process based on certain cohorts. Another research direction would be to introduce more classification and/or correlation strategies. Currently, we consider only one classifier attribute at a time. One obvious next step would be to also consider the combined impact of multiple attributes on the output variable for classification. Here, statistical analysis would also be a beneficial addition.

## ACKNOWLEDGEMENTS

The authors would like to thank LUMC for providing the problems and the access to the data. The authors especially thank Drs. Marielle Schroijen and A.J. (Thomas) Schneider, both from LUMC, for their efforts, thorough help and timely feedback throughout the analysis. Furthermore, the authors would like to thank Philips research and Technical University Eindhoven for funding this research.

## REFERENCES

Aalst, W. M. P. v. d. (2016). *Process Mining - Data Science in Action, Second Edition*. Springer.

Aalst, W. M. P. v. d., Leoni, d. M., and ter Hofstede, A. H. (2011). Process mining and visual analytics: Breathing life into business process models.

Adriansyah, A., Dongen, B. F. v., and Aalst, W. M. P. v. d. (2011). Towards robust conformance checking. In *Business Process Management Workshops*, volume 66 of *Lecture Notes in Business Information Processing*, pages 122–133. Springer Berlin Heidelberg.

Buffett, S. and Geng, L. (2010). Using classification methods to label tasks in process mining. *Journal of Software Maintenance and Evolution: Research and Practice*, 22(6-7):497–517.

Dongen, B. F. v., de Medeiros, A. K. A., Verbeek, H. M. W., Weijters, A. J. M. M., and Aalst, W. M. P. v. d. (2005). The prom framework: A new era in process mining tool support. In *International Conference on Application and Theory of Petri Nets*, pages 444–454. Springer.

Dwivedi, A., Bali, R., James, A., and Naguib, R. (2001). Workflow management systems: the healthcare technology of the future? In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, volume 4, pages 3887–3890. IEEE.

Fernandez-Llatas, C., Martinez-Millana, A., Martinez-Romero, A., Bened, J. M., and Traver, V. (2015). Diabetes care related process modelling using process mining techniques. lessons learned in the application of interactive pattern recognition: coping with the spaghetti effect. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2127–2130.

Goedertier, S., Martens, D., Baesens, B., Haesen, R., and Vanthienen, J. (2007). Process mining as first-order classification learning on logs with negative events. In *International Conference on Business Process Management*, pages 42–53. Springer.

Graeber, S. (1997). The impact of workflow management systems on the design of hospital information systems. In *Proceedings of the AMIA Annual Fall Symposium*, page 856. American Medical Informatics Association.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18.

Hompes, B. F. A., Buijs, J. C. A. M., and Aalst, W. M. P. v. d. (2016). *A Generic Framework for Context-Aware Process Performance Analysis*, pages 300–317. Springer International Publishing, Cham.

Knuplesch, D., Reichert, M., Ly, L. T., Kumar, A., and Rinderle-Ma, S. (2013). *Visual Modeling of Business Process Compliance Rules with the Support of Multiple Perspectives*, pages 106–120. Springer Berlin Heidelberg, Berlin, Heidelberg.

Kriglstein, S., Pohl, M., Rinderle-Ma, S., and Stallinger, M. (2016). Visual Analytics in Process Mining: Classification of Process Mining Techniques. In Andrienko, N. and Sedlmair, M., editors, *EuroVis Workshop on Visual Analytics (EuroVA)*. The Eurographics Association.

Leemans, S. J. J., F. D. and Aalst, W. M. P. v. d. (2014). Discovering block-structured process models from event logs containing infrequent behaviour. In *Business Process Management Workshops*, pages 66–78. Springer.

Leemans, S., Fahland, D., and Aalst, W. M. P. v. d. (2014). Process and deviation exploration with inductive visual miner.

Leoni, M. d., Maggi, F., and Aalst, W. M. P. v. d. (2015). An alignment-based framework to check the conformance of declarative process models and to preprocess event-log data. *Information Systems*, 47:258 – 277.

Mahaffey, S. (2004). Optimizing patient flow in the enterprise. *Health management technology*, 25(8):34–37.

Mannhardt, F., de Leoni, M., and Reijers, H. (2015). The multi-perspective process explorer. In *Proceedings of the BPM Demo Session 2015, Co-located with the 13th International Conference on Business Process Management {(BPM} 2015), Innsbruck, Austria, September 2, 2015*, pages 130–134. CEUR Workshop Proceedings.

Munoz-Gama, J., Carmona, J., and Aalst, W. M. P. v. d. (2014). Single-entry single-exit decomposed conformance checking. *Information Systems*, 46.

Poggi, N., Muthusamy, V., Carrera, D., and Khalaf, R. (2013). Business process mining from e-commerce web logs. In *Business Process Management*, pages 65–80. Springer.

Ramezani Taghiabadi, E., Fahland, D., Dongen, B. F. v., and Aalst, W. M. P. v. d. (2013). *Diagnostic Information for Compliance Checking of Temporal Compliance Requirements*, pages 304–320. Springer Berlin Heidelberg, Berlin, Heidelberg.

Schulz, H.-J., Nocke, T., Heitzler, M., and Schumann, H. (2013). A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2366–2375.

Sutherland, J. and van den Heuvel, W. (2006). Towards an intelligent hospital environment: Adaptive workflow in the or of the future. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, volume 5, pages 100b–100b.

Weigl, M., Müller, A., Vincent, C., Angerer, P., and Sevdalis, N. (2012). The association of workflow interruptions and hospital doctors' workload: a prospective observational study. *BMJ quality & safety*, 21(5):399–407.

# Towards Process Mining of EMR Data
## *Case Study for Sepsis Management*

Gert-Jan de Vries[1], Ricardo Alfredo Quintano Neira[2,1], Gijs Geleijnse[1], Prabhakar Dixit[1,3] and Bruno Franco Mazza[4]

[1]*Philips Research - Healthcare, High Tech Campus 34, Eindhoven, The Netherlands*

[2]*Industrial Engineering Department, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brazil*

[3]*Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands*

[4]*Intensive Care Unit, Hospital Samaritano, São Paulo, Brazil*

*gj.de.vries@philips.com*

Keywords: Process Analysis, Sepsis.

Abstract: Imagine you have cold shivers and a racing heartbeat and high fever. Clear thinking is impossible! Ceiling lights flash by as you are rushed to the emergency department (ED). You feel your body is getting even sicker. Doctors are doing their utmost to treat this acute and threatening condition, while they work piece together all small parts of evidence to set the diagnosis and start targeted treatment. In this situation, the clinical staff depends on a clinical pathway protocol to streamline communication and deliver care according to the latest medical evidence. Today, such clinical pathways are mainly executed and tracked using paper. Hence, there is ample opportunity for technology in a supportive role. Automated process analysis can help improve these processes of delivering standardized care beyond their current level. In this paper, we provide insight into the steps required to perform process mining to EMR data in the challenging domain of sepsis treatment and provide learnings from our preliminary analysis of these data using process mining techniques.

## 1 INTRODUCTION

Sepsis is a potentially life-threatening complication of an infection, where inflammatory responses throughout the body are triggered, which can lead to damage of multiple organ systems, causing them to fail. Sepsis is a condition with a very big impact on patient condition, and has high mortality rates. It is also characterized by high annual incidence rates, e.g., in the US 3-10 in 1000 people are hospitalized with sepsis (Kempker and Martin, 2016). The associated healthcare costs are also high; in 2011 it accounted for $20.3 billion, which is 5.2% of total US hospital costs, therewith the most expensive condition treated (Torio and Andrews, 2013).

The management of sepsis is complicated by the difficulties of detecting the condition. Recently, the community adopted a new definition of sepsis and a strategy for screening was proposed (Singer et al., 2016). As we evaluate our methods on data collected before 2016, this paper focuses on the method commonly accepted until that date, where screening for Systemic Inflammatory Response Syndrome (SIRS) symptoms is used to evaluate starting the treatment for sepsis. Hence, we adopt the 1992 definition from the American College of Chest Physicians / Society of Critical Care Medicine (Bone et al., 1992): "Sepsis is the Systemic Inflammatory Response Syndrome (SIRS) to the presence of infection". A patient is screened positive for SIRS if two or more of the following criteria are met:

- Temperature $>$ 38°C or $<$ 36°C
- Heart rate $>$ 90/min
- Respiratory rate $>$ 20/min or $PaCO_2 < 32$ mmHg (4.3 kPa)
- White blood cell count $>$ 12000/mm$^3$ or $<$ 4000/mm$^3$ or $>$ 10% immature bands

Patients are considered to be septic when the SIRS criteria are satisfied in combination with a suspected or established infection. As the SIRS criteria are not specific, many patients meeting the SIRS criteria will, however, not have or develop sepsis (Lord et al., 2014). When sepsis is complicated by organ dysfunction, it is called severe sepsis, which can turn into septic shock when hypotension persists despite fluid resuscitation. Mortality rates vary strongly per



Vries, G-J., Neira, R., Geleijnse, G., Dixit, P. and Mazza, B.
Towards Process Mining of EMR Data - Case Study for Sepsis Management.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 585-593
ISBN: 978-989-758-213-4

geography, but are known to increase with the three levels of sepsis: up to 30% for sepsis, 50% for severe sepsis and 80% for septic shock over the timespan of 1 year (Jawad et al., 2012). A multi-center study in Brazilian Intensive Care Units (ICUs) showed rates of 34.7%, 47.3% and 52.2% at 28 days, respectively (Silva et al., 2004).

In 2002 the Surviving Sepsis Campaign (SSC) was launched as a global campaign to reduce mortality due to sepsis. The guidelines they published, along with the updates made over the last years, are now widely adopted in clinical practice (Dellinger et al., 2004; Dellinger et al., 2008; Dellinger et al., 2013). The SSC provided a care bundle that comprises the following steps:

**To be completed within 3 hours of presentation:**

- Measure lactate level
- Obtain blood cultures prior to administration of antibiotics
- Administer broad spectrum antibiotics
- Administer 30mL/kg crystalloid for hypotension or lactate $\geq$ 4 mmol/L

**To be completed within 6 hours of presentation:**

- Apply vasopressors (for hypotension that does not respond to initial fluid resuscitation) to maintain a mean arterial pressure (MAP) $\geq$ 65 mmHg
- In the event of persistent hypotension after initial fluid administration (MAP $<$ 65 mmHg) or if initial lactate was $\geq$ 4 mmol/L, reassess volume status and tissue perfusion and document findings
- Remeasure lactate if initial lactate elevated.

As these guidelines provide a recommendation, hospitals implement these guidelines adapted to their standards of care. These guidelines are translated into clinical pathways (CPs), which are "complex intervention[s] for the mutual decision making and organisation of care processes for a well-defined group of patients during a well-defined period" (European Pathway Association, 2016). In this field, Key Performance Indicators (KPIs) are often described in terms of adherence to such guidelines.

During the interpretation and translation process, the guidelines are made actionable for the clinical staff: tasks and responsibilities are defined and a communication structure around the pathway is put in place. CPs are implemented in different areas of health care, such as acute care (e.g., for chest pain in the emergency room (ER), stroke diagnosis), integrated oncology care and chronic disease management (e.g., coordination of care for heart failure patients). Often, the clinical pathway of a patient is managed using a paper sheet. However, this leads to double data entry as the status needs to be recorded in the Health IT system as well as on paper. Moreover, the current phase on the pathway, when represented on paper, is only available to those accessing the sheet, typically at the bed side.

In this research, we are interested in solutions to monitor the status of the patient in a clinical pathway by analyzing data from the Electronic Medical Record (EMR). To this end, we model the clinical pathway in a computer interpretable format. The events in this model are associated with data from the EMR. However, there might not always be in one-to-one correspondence. For example, events such as *patient transferred to ICU* or *Vital Signs measured* may be associated with only single time-stamped entries in the patient record, but as we will see in the following, events such as *Blood Volume Expansion performed* might be more complicated to retrieve.

Process mining is the technique to extract information from event logs (van der Aalst, 2011). In general, the scientific field is concerned with two research areas: process discovery and conformance checking. Process discovery deals with identifying a model that describes the behavior as observed in a set of event logs. In process discovery, the challenge is to identify a model that is not too general but is also not overfitting the behavior as encountered in the set of event logs. In conformance checking, on the other hand, a collection of event logs is compared with a reference model with the aim to research whether the observed behavior is matching the expected behavior. In particular, common deviations or delays in processes can be analyzed. As many KPIs are based on times-to-event or performing actions in a certain order, results from conformance checking can be used as input to KPI analysis.

Applying process mining to event logs obtained from EMRs is known to be a challenging topic, as was concluded in several studies in the application of process mining techniques domains such as oncology, surgery, cardiology, diabetes and clinical images (Rojas et al., 2016). Already in 2008, Mans et al. describe explorations to discover process models from hospital data on stroke care (Mans et al., 2008). Nonetheless, these past attempts were performed on relatively straightforward clinical processes with homogenous patient populations, or incorporated prospective data collection. To the best of our knowledge, two other studies looked into applying process mining for sepsis, however results are limitedly published (Mannhardt et al., 2016; Mcgregor et al., 2011).

In general, process mining techniques can only be

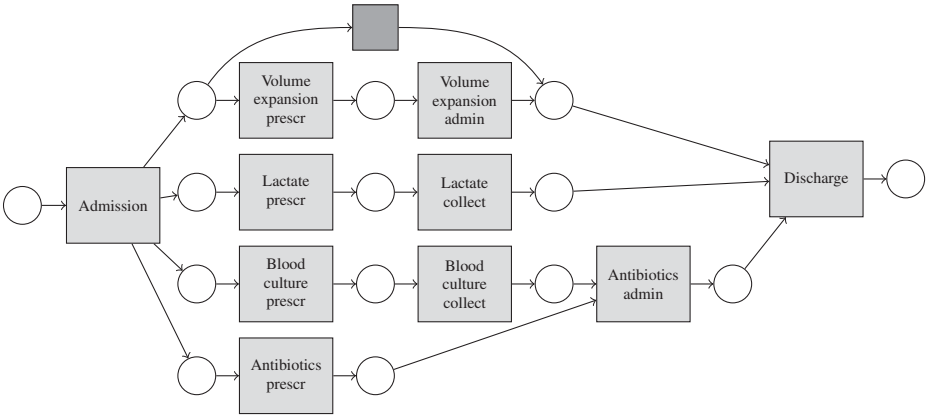



Figure 1: Petri net of the simplified model representing the clinical pathway for sepsis management, used for conformance checking.

applied if the event log describes a recognizable process. The event log for a patient from an EMR will contain events related to various processes and actions. The vast majority of these raw events will not be directly related to the care according to the reference pathway, but rather reflect routine sub-processes that the staff follows in usual care. Hence when analyzing the event log, a projection needs to be created of events that describe actions and data related to the pathway. In this paper, we are interested in exploring the potential of applying process mining techniques on a complete patient record from an EMR. We draw learnings from the modeling, data extraction and process mining steps.

## 2 METHODS

The data used in our study have been obtained from Hospital Samaritano, a private, Brazilian hospital having over 300 beds. For extracting the data from the Health IT system we identified the database tables and attributes that could represent the sepsis treatment activities from the ICU and emergency department (ED) processes. For the selection of sepsis hospitalizations, we considered hospitalization registries that had at least one diagnosis or death related to sepsis using an ICD, 10th edition (ICD-10) code list for sepsis (Canadian Institute for Health Information, 2009). Also, we included patients that were assigned a prescription template related to sepsis. The ICD-10 codings we selected, were validated by 3 Brazilian specialists and the sepsis selection method was validated by the physician responsible for the deployment of the sepsis protocol in the hospital. We extracted 4516 sepsis hospital encounters for a period of two years. To protect the identity of patients and caregivers, we pseudonymized the patient data. Important aspect with respect to the process analysis is that dates were shifted with a fixed amount of days per patient-encounter. Hence, the relative times between events per hospital admission were not altered. The data analysis was conducted with approval of the institution's ethics committee.

As indicated in the introduction, the raw data from the EMR requires interpretation or abstraction to bring it to the level of event analysis suitable to derive meaningful insights. The ultimate aim would be the analysis in terms of KPIs, however that would require more validation and comparison to the formal quality assessment procedure, which is beyond the scope of this paper. To this end, we focussed on the important elements in the first three hours of care as described in the SSC care bundle: Lactate measurement, obtaining blood cultures, antibiotics administration and volume expansion. The first two are relatively easily obtained from the data as they refer to procedures that are directly ordered as such and registered in the EMR. The antibiotics are retrieved in a similar method, using a long list of antibiotics names and active components. Volume expansion is, however, not directly registered as such in the EMR, but required interpretation of sequences of low-level events of administrating fluid. To this end, we collected all administrations of volume expanders and implemented a windowed thresholding that searches for sufficient fluid administration ($\geq 90\% * 30$mL/kg in 4 hours) such that it can be considered fluid administration with the purpose of volume expansion. For each of these four elements of care we collect the times of request and times of administration or collection, which gives 8 event types. To mark the start and end of event traces, we also include the moment of admission and discharge, yielding a total of 10 different event types.

In order to avoid that incomplete timestamps,

that only contain the date of the event, would negatively influence the results, we corrected timestamps of '00:00:00' in appropriate cases as we found that these timestamps referred to the midnight before the actual event must have happened. To allow for a more complete conformance checking, we chose to correct these timestamps as follows: if the event related to collection or administration and if the corresponding prescription event was present, we corrected the timestamp to one second after the corresponding prescription event. By doing so, we corrected the order of events, however it should be noted that these timestamps should still be considered imprecise and were thus excluded from any time-to-event analysis.

Our explorative analysis started with retrieving the times to event for each of the care elements as a step towards measuring KPIs. Note that we used the time of presentation (admission) as time zero to be able to measure all other timestamps relative to this time of admission. After that, we used the ProM software (van der Aalst et al., 2007) to perform conformance analysis of the model outlined by the SSC care bundle. To this end, we constructed the model as a Petri net, displayed in Figure 1, that represents the different steps that can happen concurrently, and the (time-wise) dependency between obtaining the blood cultures and administration of antibiotics. While this model might seem an oversimplification of clinical reality, it does contain all the critical steps outlined in the SSC care bundle (see Introduction) and provides a first step towards more elaborate pathway models. In the process of conformance checking, the event traces in the event log are aligned with the model and a distance is calculated for the alignment. We used the 'standard distance function' that assigns a cost of 0 in case log and model agree and 1 in case of disagreement (move on log or move on model) (van der Aalst et al., 2012).

# 3 RESULTS

The cohort extracted for the present analysis, using the inclusion criteria outlined in the previous section, consisted of 4516 patients. 4442 patients entered the hospital via the ED and were selected for the subsequent analysis. These patients have a median age of 37.7 years, 51.5% were male, median length of stay (LOS) was 5 hours, and 2.5% died in hospital. Further descriptive statistics can be found in Table 1.

## 3.1 Events

In total there were 37635 events extracted for the 4442 patient admissions. 4204 events had a timestamp of 00:00:00. The vast majority (4162) of these events were the collection of lactate. For 3700 events we could correct the timestamps using the aforementioned procedure, another 504 could not be corrected (no corresponding request event could be found) and were removed. Note that for the time-to-event analysis, we excluded all the 4204 events with imprecise timestamps.

Table 2 shows the number of events retrieved from the EMR. We observe that all event types are highly represented in the database, with at least 85% (lactate collection) and more than 95% of cases for the other obligatory event types. Volume expansion is less often represented, however this is also considered an optional (conditional) event, as specified in the model (Figure 1).

Figure 2 shows the histograms of time-to-event (from the moment of presentation) for each of the prescription and administration/collection events. Note that that the rightmost bars in the histograms (at 3 hours from presentation) contain all samples with times $\geq$ 3 hours. We observe that the vast majority of events happen within the first hour after presentation, with modes being 16 minutes for prescription of antibiotics and volume expansion, and 17 minutes for lactate and blood cultures. For administration/collection, the mode are 15 minutes for lactate, 19 for volume expansion, 21 minutes for antibiotics and 38 for blood cultures. The following fractions of prescription events happen outside of the window of 3 hours: Lactate (5.4%), Antibiotics (5.2%), Blood cultures (4.9%), Volume expansion (14.7%). Note that for lactate collection, the number of events found is much smaller than for the others due to the inaccurate timestamping mentioned earlier.

Conformance analysis using ProM yielded the results presented in Figure 3. The number of prescrip-

Table 1: Descriptive statistics of the patient population.

| Name | Valid *N* | *N* (%)<br>Median [25th-75th] |
|---|---|---|
| Age (year) | 4442 | 37.5 [26.0-56.3] |
| Male | 4442 | 2295 (51.7%) |
| LOS (hour) | 4439 | 5.0 [3.4-75.2] |
| Died in hospital | 4442 | 113 (2.5%) |
| Initial diagnosis | 4442 | |
| Missing | | 77 (1.7%) |
| Infections / Parasites | | 1041 (23.4%) |
| Respiratory | | 1631 (36.7%) |
| Abnormalities | | 953 (21.5%) |
| Other | | 740 (16.7%) |

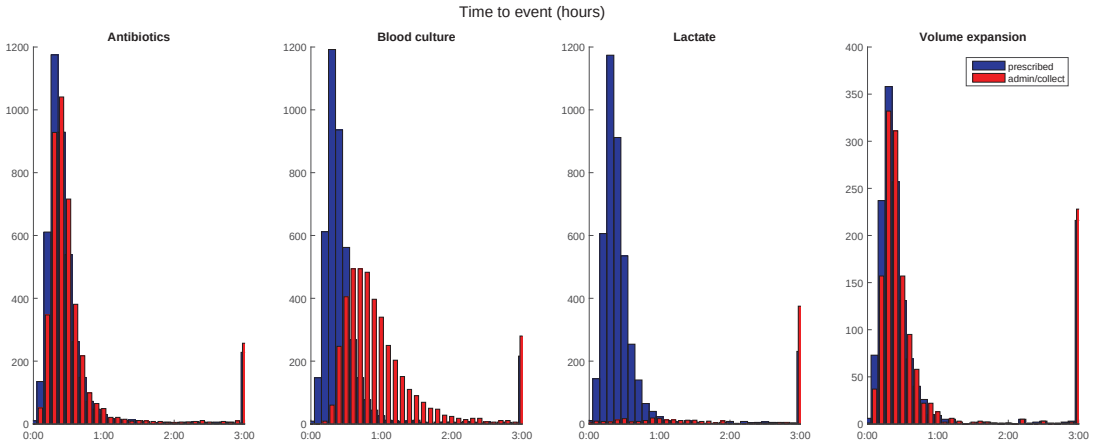



Figure 2: Distributions of time-to-event (from the moment of presentation) for the four care elements. The blue histograms represent the time to prescription, the red histograms represent the time to administration or collection. The horizontal axes represent time (hh:mm), the vertical axes represent counts.

tion events conforming to the model are the same as the number of (valid) events found in the time-to-event analysis. For the administration/collection events, we see different numbers of conforming events as compared to the numbers of events found in the time-to-event analysis. This is because the conformance checking does not only take into account presence of the events in the log, but also whether the order is according to the model. Here we see, for example, that there are 4352 blood culture prescriptions found in correspondence to the model versus 90 not; similarly, 1229 volume expansions that are in correspondence with the model. Note that volume expansion is, following the guidelines, an optional step if certain conditions are not met.

Table 2: Numbers of events found.

| Event name | $N$ (%) |
|---|---|
| Admission | 4442 (100.0%) |
| Discharge | 4431 (99.8%) |
| Blood culture prescr | 4355 (98.0%) |
| Blood culture collect | 4339 (97.7%) |
| Antibiotics prescr | 4324 (97.3%) |
| Antibiotics admin | 4309 (97.0%) |
| Lactate prescr | 4231 (95.2%) |
| Lactate collect | 3772 (84.9%) |
| Volume expansion prescr | 1465 (33.0%) |
| Volume expansion admin | 1463 (32.9%) |

Volume expansion is only managed when clinically indicated (see also Figure 1).

If we now connect these numbers to the earlier found number of events logged (Table 2), we can derive, e.g., that for $3772 - 3751 = 21$ lactate collections there was a log-entry, however not in the order prescribed by the model. Similarly, we can see that for volume expansion there are $1465 - 1461 = 4$ prescriptions that are logged, however not in the way anticipated by the model. For antibiotics administration we observe many ($4309 - 252 = 4057$) not conforming events, which turned out to be caused by an order mismatch with the blood culture collection (i.e., antibiotics administered before blood cultures were collected). Potential reasons for these mismatches will be discussed in the next section.

# 4 DISCUSSION

In our analysis, we have first looked into time-to-event analysis, which looks at the number of events logged and can derive various statistics from the timestamps of these events. Although this can give a good insight into how processes are executed on average, and identify outliers with respect to time-to-event, it does not take into account correct order of events. Using process mining, and conformance checking in particular, we can also study the order in which events occur and study deviations in more detail. One particular challenge that we tried to address here, is that EMRs are general tools to support overall clinical processes and that fields in the EMR can be used for multiple conditions and are pathway aspecific by design. Often patients are concurrently evaluated and treated for a variety of conditions, and there is often little or no evidence of which data entries relate to which diagnosis; this relation has to be inferred. Also, it is important to stress that not all patient care activities are documented in the EMR.

Before reflecting on the results obtained, we would like to emphasize that this experiment of gath-

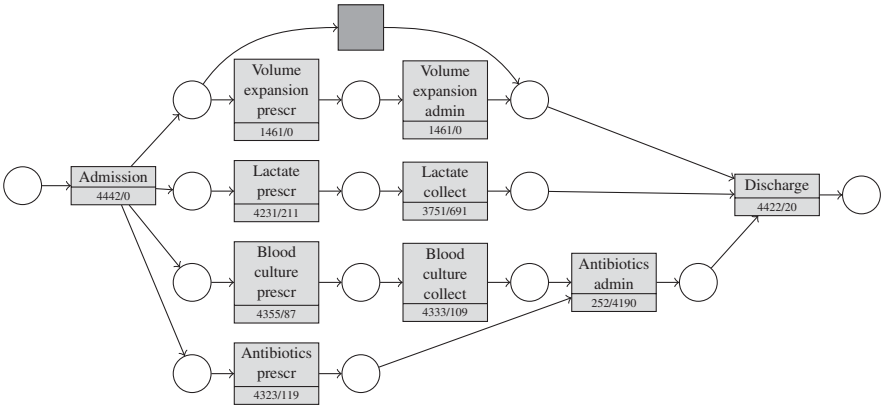



Figure 3: Output of conformance analysis in ProM, showing per event type the number cases that conform to that step in the model versus that do not.

ering KPI information directly from EMR data without a thorough, manual, quality analysis is likely to provide an underestimation of guideline adherence compared to reality. This is due to the following list of potential causes for our analysis not picking up adherent behavior:

- Not logged: Action has been performed but not logged
- Incorrect timestamping: Action has been performed but with incorrect or imprecise timestamp
- Incomplete querying: The query used for interpreting the EMR data can miss certain cases

Hence, we should not interpret the outcomes of our current analysis as quality measure for the care performed before carrying out a more thorough quality analysis. We are also reporting intermediate results, and therefore cannot draw conclusions on the KPIs themselves, but our focus is to share the challenges relating to process mining on ”real-life” EMR data.

Although the blood volume expansion does only happen when clinically indicated, the relatively low number of blood volume expansion events, might suggest that our interpretation of the EMR data is not completely covering the different ways these events are reflected in the EMR, rather than they are often not prescribed, or that they are prescribed, but not logged. Further analysis is required to analyze the volume expansion management of these sepsis patients. In any case, the quality of the logging influences the results. Bose et al. distinguish 27 classes of quality issues with event logs (Bose et al., 2013). In our data, we observe presence of the following classes of issues: missing events, missing timestamps, incorrect timestamps and imprecise timestamps. The first category has been reflected upon already, the missing, incorrect and imprecise timestamps typically reflect clinical reality as it is simply not possible to 100% accurately timestamp all events. Imprecise timestamping can be observed in the lactate collections where often only date information was information. Incorrect timestamping might be observed in, e.g., many antibiotics administration events that are found not conforming to the model (4195 out of 4304). This is further substantiated by the notion that the clinical staff at the hospital, at which the study was performed, is all well aware of the fact that antibiotics influence the results of the laboratory measurements from the blood samples. It might well be that there are differences in the actual time of performing the event versus the moment of logging in the EMR, or alternatively that we made incorrect assumptions in the interpretation of raw data. Further verification with the hospital's quality assurance process is required to find the reason of this mismatch.

The treatment of sepsis in the ED is a particularly challenging environment as the condition is life threatening and quick responses are required, which we anticipated to potentially lead to problems in process mining with respect to the aforementioned quality issues. Despite that, we observed high levels of presence of events in the eventlog: at least 85% for all obligatory events. The inherently diverse group of patients with sepsis poses a challenge to process analysis techniques. We have shown that for a relatively simple model, we can successfully apply process mining techniques, with the ultimate aim of measuring KPIs. This provides a good outlook in the possibilities to also analyze the sepsis pathway at a finer grain. It remains, however, topic of research what the optimal level of detail in the process modelling is for a given purpose. The heterogeneity of sepsis patients might become more prominent in more detailed analysis and require some form of clustering before performing process analyses on the subgroups. Patients can be clustered on patient or on process characteris-

tics (see, e.g., (de Medeiros et al., 2007)).

One particular issue that we faced when interpreting the EMR data was that we observed the need to interpret the purpose of actions performed from the event logs rather than purely the actions themselves. As an example, the administration of fluid in itself can happen for a multitude of reasons, however in order to interpret whether volume expansion was performed, we had to monitor whether a certain amount of fluid was prescribed in a certain amount of time. Similarly, for antibiotics we would like to know that they were prescribed and administered for the purpose of managing sepsis, however this intended purpose is not stored with the medication prescriptions. One way of obtaining more information on the purpose of certain actions performed is through careful analysis of clinical notes, where typically the intend of the medical staff is reflected. This will, however require the use of natural language processing (NLP) techniques to be able to extract structured information from these unstructured text data. Important to note in this respect is the lack of ground truth in such analysis of EMR data; the only evidence of what happened with the patient is the data in the EMR. Hence, the interpretation of raw EMR data should be given sufficient attention.

# 5 CONCLUSION AND FUTURE WORK

We have shown that we can successfully use process mining to follow selected events derived from the main KPIs for the sepsis pathway purely from EMR data. However, no conclusion should be drawn about the actual quality of care or adherence to these guidelines before verification with the clinical quality assurance process. It should be noted that it required a great effort in data preparation to create the event log and time-consuming manual quality checks to interpret the EMR data in terms of the concepts required for the pathway analysis. Using process mining techniques, we can analyze beyond the pure presence or absence of events and also address correct versus incorrect order with respect to a model that represents best practice. Applying these techniques on a dataset gathered at a large Brazilian hospital, we could analyze the data in terms adherence to the guidelines provided by the SSC. The reason for deviation in order of administering antibiotics and collecting blood cultures, however requires further research. In general, further follow up with the quality department would be required to quantify the accuracy of our assessment in comparison to the formal quality process that is in place in the hospital at hand. This actually highlights a big limitation of the data driven analysis of processes in general: it is impossible from event data alone to distinguish whether event logs are missing due to actions not being performed, performed actions not being logged or logged actions not being picked up by the data extraction and interpretation. For that reason, results should always be interpreted with care and at least a randomized sample should be analyzed through a formal quality assessment process in order to quantify the accuracy of the overall data-driven analysis results.

Although our analysis shows high levels of availability of time stamps (at least 85% per obligatory event type), there is room for improvement. The quality of the event log generated from the EMR data could be further improved by better support from the data entry module to allow for more accurate and timely data entry and the use of structured reporting over free-text notes. It should be noted, though, that this will remain difficult in busy environments such as the ER, where top priority is to provide acute care to the patient. It might require a change in the workflow to improve the precision of timestamps of time critical events such as lactate collection.

Our present analysis is limited to a relatively small and simple model to reflect sepsis care. Nevertheless, this model allows already for analysis in terms of various clinical KPIs. Future work includes the extension of the model used for conformance analysis in order to assess the clinical pathway in further detail. In our future aim of extending the model to cover more detailed steps in the sepsis care pathway, we expect that more elaborate data interpretation might be required. While many steps have already been taken to digitize hospital data in structured fields, rich information can also be found in non-structured text fields such as clinical notes. The analysis of such data will require NLP approaches to reliably retrieve structured information from text. Being able to analyze adherence to such a more detailed model would open up further analysis of conformance to and deviations from the best practice. The application of process discovery techniques can also provide a bottom-up view of the process as it is performed by the clinical practitioners. A root cause analysis into the reasons for deviation could help to further improve the guidelines and standard of care for sepsis.

# ACKNOWLEDGEMENTS

We would like to thank the clinical and technical staff at Hospital Samaritano for their help in collecting the

event data, as well as our colleagues from Philips Research for making this work possible.

## REFERENCES

Bone, R. C., Balk, R. A., Cerra, F. B., Dellinger, R. P., Fein, A. M., Knaus, W. A., Schein, R. M. H., and Sibbald, W. J. (1992). Definitions for Sepsis and Organ Failure and Guidelines for the Use of Innovative Therapies in Sepsis. *Chest*, 101(6):1644–1655.

Bose, R. P. J. C., Mans, R. S., and van der Aalst, W. M. P. (2013). Wanna improve process mining results? In *2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pages 127–134.

Canadian Institute for Health Information (2009). In Focus. A National Look at Sepsis. *Ottawa, Ont: CIHI*.

de Medeiros, A. K. A., Guzzo, A., Greco, G., van der Aalst, W. M. P., Weijters, A. J. M. M., van Dongen, B. F., and Saccà, D. (2007). Process Mining Based on Clustering: A Quest for Precision. In Hofstede, A. t., Benatallah, B., and Paik, H.-Y., editors, *Business Process Management Workshops*, number 4928 in Lecture Notes in Computer Science, pages 17–29. Springer Berlin Heidelberg.

Dellinger, R. P., Carlet, J. M., Masur, H., Gerlach, H., Calandra, T., Cohen, J., Gea-Banacloche, J., Keh, D., Marshall, J. C., Parker, M. M., Ramsay, G., Zimmerman, J. L., Vincent, J.-L., and Levy, M. M. (2004). Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. *Intensive Care Medicine*, 30(4):536–555.

Dellinger, R. P., Levy, M. M., Carlet, J. M., Bion, J., Parker, M. M., Jaeschke, R., Reinhart, K., Angus, D. C., Brun-Buisson, C., Beale, R., Calandra, T., Dhainaut, J.-F., Gerlach, H., Harvey, M., Marini, J. J., Marshall, J., Ranieri, M., Ramsay, G., Sevransky, J., Thompson, B. T., Townsend, S., Vender, J. S., Zimmerman, J. L., and Vincent, J.-L. (2008). Surviving Sepsis Campaign: International guidelines for management of severe sepsis and septic shock: 2008. *Intensive Care Medicine*, 34(1):17–60.

Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., Sevransky, J. E., Sprung, C. L., Douglas, I. S., Jaeschke, R., Osborn, T. M., Nunnally, M. E., Townsend, S. R., Reinhart, K., Kleinpell, R. M., Angus, D. C., Deutschman, C. S., Machado, F. R., Rubenfeld, G. D., Webb, S. A., Beale, R. J., Vincent, J.-L., Moreno, R., and Surviving Sepsis Campaign Guidelines Committee including the Pediatric Subgroup (2013). Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Critical Care Medicine*, 41(2):580–637.

European Pathway Association (2016). E-p-a definition of care pathway. http://http://e-p-a.org/.

Jawad, I., Lukšić, I., and Rafnsson, S. B. (2012). Assessing available information on the burden of sepsis: global estimates of incidence, prevalence and mortality. *Journal of Global Health*, 2(1):010404.

Kempker, J. A. and Martin, G. S. (2016). The Changing Epidemiology and Definitions of Sepsis. *Clinics in Chest Medicine*, 37(2):165–179.

Lord, J. M., Midwinter, M. J., Chen, Y.-F., Belli, A., Brohi, K., Kovacs, E. J., Koenderman, L., Kubes, P., and Lilford, R. J. (2014). The systemic immune response to trauma: an overview of pathophysiology and treatment. *Lancet (London, England)*, 384(9952):1455–1465.

Mannhardt, F., de Leoni, M., Reijers, H. A., and van der Aalst, W. M. P. (2016). Decision mining revisited - discovering overlapping rules. In Nurcan, S., Soffer, P., Bajec, M., and Eder, J., editors, *Advanced Information Systems Engineering: 28th International Conference, CAiSE 2016, Ljubljana, Slovenia, June 13-17, 2016. Proceedings*, pages 377–392. Springer International Publishing, Cham.

Mans, R., Schonenberg, H., Leonardi, G., Panzarasa, S., Cavallini, A., Quaglini, S., and van der Aalst, W. (2008). Process mining techniques: an application to stroke care. *Studies in health technology and informatics*, 136:573.

Mcgregor, C., Catley, C., and James, A. (2011). A process mining driven framework for clinical guideline improvement in critical care. In *In LEMEDS Workshop*.

Rojas, E., Munoz-Gama, J., Sepúlveda, M., and Capurro, D. (2016). Process mining in healthcare: A literature review. *Journal of biomedical informatics*, 61:224–236.

Silva, E., Pedro, M. d. A., Sogayar, A. C. B., Mohovic, T., Silva, C. L. d. O., Janiszewski, M., Cal, R. G. R., de Sousa, E. F., Abe, T. P., de Andrade, J., de Matos, J. D., Rezende, E., Assuno, M., Avezum, A., Rocha, P. C. S., de Matos, G. F. J., Bento, A. M., Corrła, A. D., Vieira, P. C. B., Knobel, E., and Brazilian Sepsis Epidemiological Study (2004). Brazilian Sepsis Epidemiological Study (BASES study). *Critical Care (London, England)*, 8(4):R251–260.

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G. R., Chiche, J.-D., Coopersmith, C. M., Hotchkiss, R. S., Levy, M. M., Marshall, J. C., Martin, G. S., Opal, S. M., Rubenfeld, G. D., van der Poll, T., Vincent, J.-L., and Angus, D. C. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8):801–810.

Torio, C. M. and Andrews, R. M. (2013). National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011: Statistical Brief #160. In *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. Agency for Healthcare Research and Quality (US), Rockville (MD).

van der Aalst, W., Adriansyah, A., and van Dongen, B. (2012). Replaying History on Process Models for Conformance Checking and Performance Analysis. *Wiley Int. Rev. Data Min. and Knowl. Disc.*, 2(2):182–192.

van der Aalst, W. M. P. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, Place of publication not identified, 2011 edition edition.

van der Aalst, W. M. P., van Dongen, B. F., Günther, C. W., Mans, R. S., de Medeiros, A. K. A., Rozinat, A., Rubin, V., Song, M., Verbeek, H. M. W., and Weijters, A. J. M. M. (2007). ProM 4.0: Comprehensive Support for Real Process Analysis. In Kleijn, J. and Yakovlev, A., editors, *Petri Nets and Other Models of Concurrency ICATPN 2007*, number 4546 in Lecture Notes in Computer Science, pages 484–494. Springer Berlin Heidelberg.

# Translating a Clinical Workflow into a Computer-executable Model
## *User Needs Discovery, Challenges and Lessons Learned*

Ana Leitão, Jennifer Caffarel and Erin Stretton
*Royal Philips - Research, High Tech Campus 34, Eindhoven, The Netherlands*
*ana.leitao@philips.com*

Keywords: User Requirements, Healthcare IT, User Confrontation, Clinical Pathways.

Abstract: Getting to technical requirements from user input is already a hard task in an environment where the workflow processes are very well defined. When trying to extract a unique process from the users of such a variable work environment such as a healthcare institution can be very challenging. In this paper, we share our experience with extracting user requirements from clinical users by presenting the specific example of transforming workflows into models that can then be used as part of an IT solution to support workflow guidance. Here we present not only some of our main challenges when approaching different institutions and professionals with different roles, but also some of the methods we find most useful to establish communication and extract as much relevant information possible. In the end we explain some of the differences between a workflow as explained by the users and a computer–executable model and how to make the connection between the two.

## 1 INTRODUCTION

In 2004 Ash et Al. said that *"we should strive to have a national system of Electronic Health Record (EHR) that can share information on any patient in any health care setting"*. While some institutions have the financial capabilities to have the latest technology which allows them to automate and make a wide range of tasks digital, there are still institutions who depend very highly on paper, especially when it comes to documenting clinical processes and pathways.

From our learnings this persistence on the use of paper to document clinical processes and pathways within a hospital is not only driven by lack of resources and financial capabilities to go digital but in many cases, paper is seen as an easier alternative to implement than modifying the IT system to support such tasks. Also, the IT systems are still having negative impact on the work of the clinicians by increasing the documentation time and being incompatibility with clinical workflow. This leads to a higher amount of interruptions in the medical work and system-introduced errors in patients care. (Ologeanu-Taddei, R. et al., 2015; Jamoom, E. W. et al., 2016)

Another limitation of paper supported processes is that although this may make the administrative task of collecting most relevant information in one location easier, they generally do not aid the clinical users to adhere to the recommended steps in a pathway in real-time – they mainly serve as a reminder of which data to collect so that the management can at a later date evaluate how well the care was delivered. And while with an EHR this collection of information is easier and more automated, there are still gaps on the usage of these that limit the step by step tracking of clinical events. One of our research aims is to investigate how we can proactively support the clinical staff to adhere in real-time to clinical pathways, with a greater focus on delivery of care than on care delivery documentation. We want to do this by going beyond the simple digitization of the paper process.

Even though it is possible to identify processes in healthcare, these are different from the ones that could be found in an industrial manufacturing environment (Mans et al., 2015). In the healthcare environment, users are dealing with the care of patients, who are not predictable machines with predictable failure modes. Therefore enforcing a process can be an extremely challenging task, let alone having a system trying to follow the steps making up that process. It is extremely difficult to identify all the steps healthcare professionals follow in a process performing retrospective data analysis



Leitão, A., Caffarel, J. and Stretton, E.
Translating a Clinical Workflow into a Computer-executable Model - User Needs Discovery, Challenges and Lessons Learned.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 594-600
ISBN: 978-989-758-213-4

since EHRs are rarely pre-programmed to support clinical workflows in a structured manner. Therefore any evidence we collect from these, represent point measures of the process either indicating partial activities in the process or post-documentation information. Our research challenge was therefore to understand how the users incorporate the clinical processes in their routine, how they interact with the systems, and how they could optimally be supported with technology in the future to support them in real-time to adhere to clinical pathways as they deliver care to patients. There are no specific guidelines to support extraction of requirements from end-users in such situations, with most publications in the area focusing on modelling of clinical processes, but not giving insights onto the differences one might expect across hospitals or recommendations how to translate user specifications to technical requirements. (Böckmann and Heiden, 2013; Dadam et al., 2000; Latoszek-Berendsen et al. 2010; Staccini et al. 2001; Oosterhout et al. 2005)

In this article we share some of our experiences and methods used to go from an expressed wish from users to actual user and technical requirements, using the specific example of transforming workflows into models that can then be used as part of an IT solution to support workflow guidance.

In the specific study that we are using as example we have confronted 4 different institutions with different levels of understanding and complexity regarding the same type of care delivery pathways and associated workflows. We had a total of 51 participants in which 13 where physicians, 21 where nurses (some with coordination roles within the departments), 11 where quality staff and 6 had other roles within the Hospitals. The Hospitals involved have between 200 and 350 beds each and cover most specialties inside the institution. Associated to this they all have implemented an EHR but only one was considered a paperless Hospital. Our expectation is that models can allow us to create a unique translation of realities, bringing together the workflows of the different institutions. Therefore, we focus on the specification of requirements for a technology solution to assist the adherence to a Clinical Pathway that involves multiple stakeholders, sometimes across departments; and for a variety of care settings.

# 2 USER CONFRONTATION

## 2.1 Preparation Phase

Our team conducted user confrontations to validate hypotheses, developed from literature reviews and existing knowledge on the topic, on what users would want in a workflow product. When researching how to design these products to fit the needs of a specific population of healthcare professionals, doing many iterations of user confrontations is important. The aim with the first user confrontations is to understand what aspects of the problem cause the user the most hindrance, and from there prioritize aspects of the future product and create a research roadmap. For each iteration after this, the scope of the confrontations becomes more granular to specify wishes for components of the proposed product. Preparation for these confrontations consists of 3 phases: (i) identifying the activities to be done with the clinicians and what type of professionals need to be included, (ii) preparing legal and regulatory documentation with our company and the hospitals with which we will collaborate, (iii) contacting hospitals and agreeing on an agenda.

(i) **Identifying activities and participants**: For every round of user confrontations, we need to identify:
   a. What information we need to retrieve from the activities (proving or disproving our hypotheses) [more on this topic in the next sub-section]
   b. Who we should involve in the confrontations in terms of professional roles

Then, we plan activities that help us gather this information. For each activity, we usually assign 2 researchers per activity, for 1-4 participants. This way one person is able to take notes and another leads the exercise. When performing user confrontations, a risk that is run is that of collecting the opinion of too few users and using these opinions to generalise to the overall population of users. One way to address this, if resources are limited, is to organise group activities which enable multiple users' viewpoints to be collected in one session.

(ii) **Legal and Regulatory:** For conducting user confrontations, there are legal and regulatory agreements that need to be made on the institutions involved, covering the interviewers and the interviewees. It is important to keep this in mind when planning the confrontations. Translating the relevant documents such as Non-Disclosure Agreements and Consent

Forms into the local language and allowing time for reading and signing by the institution could take several weeks. We strive to send the Participation Information Letter and the Informed Consent in advance of the visit, to ensure the participants have had time to read it and consider their involvement.

(iii) **Preparing the agenda:** When planning the activities it is important to keep in mind time. Healthcare professionals will typically have 30-45 minutes to spend in the activities. Lastly, the agenda request must be made to the hospital. In this request, there should be an overview of the project, goals of the interviews, and request for specific users for specific amounts of time (for individual interviews and group exercises).

## 2.2 Exercises Used by the Team

In this paper, we focus on the exercises carried out in the first round of confrontation sessions with the users. For this round, our main aim was to derive the users' main needs and challenges when it comes to supporting the implementation of Clinical Pathways in practice. As we were still in our project definition phase, our scope was large: we wanted to learn more about the topic from the users (general information on how this is done in practice), from a variety of users (from administrative to clinical and managerial staff) and from all phases of Clinical Pathways (from the creation of pathways, the use at the point of care to the evaluation reporting). With this in mind, we devised a number of activities:

- **Interviews** to understand the users' and institutions' realities and scope the landscape of Clinical Pathways as it stands today in clinical practice
- A **model building exercise**, where we asked the participants to build a pathway of their choice and take us through their routine to help us identify elements common to various pathways as well as routine bottlenecks and deviations from the set pathway
- **Confrontations** of our work and assumptions:
  - Review of a generic model we created, to get some validation on the modelling of the pathways
  - Feedback on user interface designs for use in supporting clinical and administrative workflow to verify some assumptions we made on information needs for clinical users when it comes to needing to track and care for a patient on a pathway
  - Interactive activity enabling participants to prioritise importance of pre-assumed needs, as well as creation of unthought-of needs
- **Questionnaires** to derive quantitative results related to the concepts explored with the users.

Whenever possible, we asked the users to walk us through their way of working, showing us how they interact with the clinical systems and indicating when they would do this at various points of the care.

We had a variety of one to one and group settings. Considerations when setting up group confrontations are the following:

- Size of group: 4-6 participants for 2 facilitators is an appropriate size, larger groups may have to be separated in sub-groups
- Roles within group: hierarchy and personality may influence the interaction dynamics within a group, for overall opinion on a concept, we prefer to separate the groups according to role; mixing of roles can work well when complimentary perspectives about different aspects of a topic are sought
- Facilitation skills: the facilitator of the activity should direct the involvement of participants when necessary, to ensure a fair representation of all participants in the discussions
- Discussion material: having material to discuss (such as a concept video, screen designs, a conceptual poster) can help the conversation along, as a starting point or as a way to focus the participants on the topic matter.

Overall, we derived a lot of useful information from the users, which ranged from scoping a landscape of realities and challenges from various hospitals with varying levels of maturity when it comes to implementing Clinical Pathways; all the way to having a much clearer picture of the roadmap we needed to create in order to meet the most pressing needs of the users.

Some pitfalls we encountered were:

- Broad scope and limited time meant that some topics could not be deeply explored
- Tight planning meant that not all exercises could be conducted with all users
- Questionnaire not specific enough to provide significant added value on top of qualitative results
- Unexpected changes in times and personnel available for participation in the activities.

Our recommendations include:

- Be clear within the team on the objectives of each exercise

- Dry-run the activities with colleagues or proxy users not involved in your project to ensure instructions and questionnaires make sense before finalisation
- Double check translations (back translation if possible) to ensure that the meaning is retained. This might seem quite obvious but it is often dismissed specially when using official translators. It is important to make sure the interpretation is the same for all readers no matter the language.
- Perform debriefing within the team as frequently as possible, at least at start and end of each day, and if possible, in between activities especially in the first days to ensure that the activities can be refined as the interviews progress.

Finally, keep in mind that structured activities and questionnaires are important to ensure focus is not lost, but ability to improvise and follow the participants' train of thought in a semi-structured interview format can often be invaluable to discover user needs the team had not thought of or planned for.

## 2.3 Challenges of Interaction with the Users

The first challenge of interacting with users on an international level is communication. It is imperative that the professionals that are participating in the discovery activities, fully understand what is presented so that your questions can be answered in the end of the exercises.

To facilitate the understanding and better communication we try to provide all the material and conduct all the activities in the language of the users, whenever possible.

Also, it is important to keep in mind that the realities differ among institutions so the speech should be adapted to the reality of each institution and professional. You should always take into consideration the following factors:

- **Technical resources**. Not all institutions have the same resources, such as, imaging machines, beds or medications. This has a very high impact on how the tasks are done, meaning that the same step in a process can be executed in different ways and sometimes even include a third party institution who provides the resource. A good model based system can help not only to optimize the existing resources but also to find the best workflow using the resources available at each institution.
- **People**. Not only is there variation in the availability of staff, but also in the interaction between different types of professionals among the different institutions. As an example, in some hospitals strict hierarchy may be the norm (e.g. in one institution nurses may be empowered to put a patient in a clinical pathway whereas in another this may only be done by a clinician). This has a big impact not only on the identification of who should be involved in a task but also on the attribution of authority and permission for decision making. This is so far the hardest factor that can affect not only the way the exercises are done during the user interactions but also can have a big impact on how an IT solution will be used in the institution. If you are looking to create a solution which could be used in different institutions it is important to identify all the potential users, how they interact and who will be the potential main users (which can include different type of professionals)
- **Impact of geographical, organisational and legislation factors on the level of maturity of clinical processes.** By association, institutions that are involved with certification organizations and medical societies usually have very clear ways of working which are based on best practices. This is also very closely related to differences in implementation of health services between different institutions, regions and countries. In countries where there are little or no public healthcare institutions except for the primary care facilities, most Hospitals and private institutions will rely on certifications to distinguish themselves from others. In the case of countries where the Healthcare service is very well managed by the government and advances, chances are that standardization and certification processes are stimulated if not required by the government to guarantee the minimum quality of services.
- **Knowledge**. It is easy to assume that different types of professionals have different levels of knowledge. While that is true on a high level and most people have greater knowledge on their roles rather than on that of others, it is good to not only explore their roles, but also how they interact with and perceive the roles of others in the organisation. When approaching users from different institutions, cities or even countries, we must take in consideration their knowledge not only regarding technology (how

familiar are they with the latest technology and how they use it in their work) but also on the content level. As mentioned before, the involvement of the professionals with the latest news on best practices will also define how prepared they are to understand the concept to be discussed during the user interactions. And as rewarding as it may be to involve Key Opinion Leaders who are very much up to date with state of the art and best practices in the area you want to discuss, it can be even more insightful to talk with users who have less or little knowledge of the theoretical aspects so that you can understand the real practical issues the end-users are actually confronted with. Nonetheless, it is good practice to assume the users know very little about the topic, and be prepared with a good but not restrictive definition of the concept you want to present.

## 3 FILTERING KNOWLEDGE INTO REQUIREMENTS

The main challenges we have had are:

- How to transform information collected from users into requirements usable by the technical team?
- How to ensure that the collected needs and corresponding requirements are in some way weighted to reflect the input from the variety of users we interacted with?

To address the above, we employed a number of methods, which included use of:

- Raking/Rating systems where possible, e.g. when confronting 3 user interface designs, beyond asking for specific feedback, we also asked the users to classify the designs from their preferred one to least preferred one; for the pre-assumed needs list, we asked the users to rate each requirement as necessary, nice to have, or not necessary
- Quantitative data analysis wherever possible; e.g., for the pre-assumed needs list, we calculated a weighted average across the groups and used this to rank the requirements in order of importance, which gave an objective perspective on the relative importance of the rated requirements
- Consensus technique whereby we analysed the results of the interviews by first having a round of insights extraction from our notes at an individual interviewer level, before coming together to share findings and debate if the insights resonated with one or more interviewees before including this as relevant insights for our results.

Concerning Clinical Pathways, a main insight that was drawn from our study with users is how to bridge the technical viewpoint and the user viewpoint: there are really two aspects to workflow modelling. One level are the workflow elements needed for user interaction to support them in their daily work; the other level are those workflow elements which are essential to the user to follow a Clinical Pathway, but may not be relevant to model in technical terms, either because it is not measurable or is difficult to model, e.g. due to lack of evidence in the clinical IT systems.

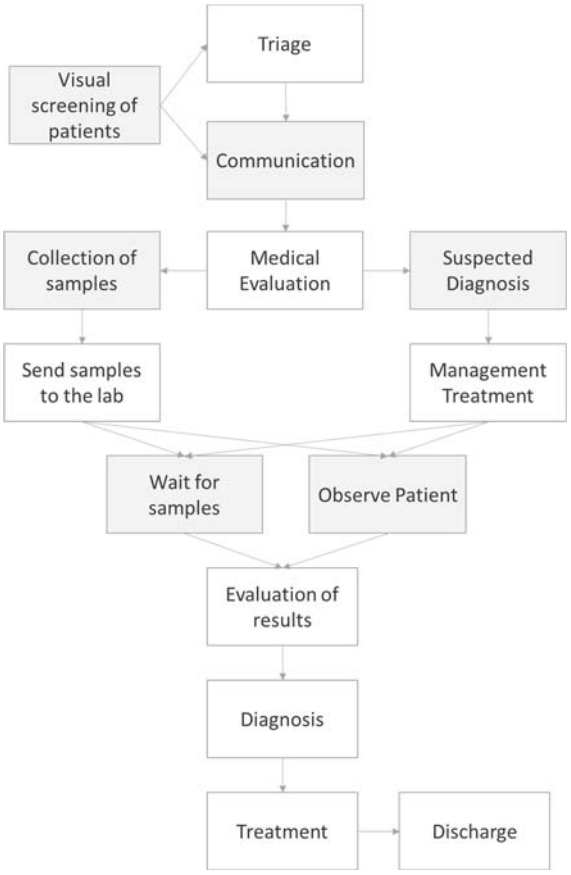



Figure 1: Simplified clinical process example as it would be described by the user. The boxes in grey represent the activities in the clinical process which are essential for the users to carry out the process but only essential to the people carrying out the task and not the model; or not captured in the IT system because non-measurable or not included in the IT documentation of the process.

Taking the example of a high level description of a clinical process as described by a user such as the one in Figure 1, we can identify 5 steps identified in grey, of such type, e.g., the communication interaction whereby the conversation process is more important than the actual data exchanged. In the same example we have the visual triage which is not measurable since it is done mostly following the professionals' instinct and experience and it is not associated with any record in the EMR; or activities such as "collection of samples" which are not captured in the EMR because so far, when the EMR is used mainly for documentation of patient medical data, there was no need to capture such process-related information.

The same model can be translated into a machine executable model, including only the steps that can be found or recorded using the EMR, which would look more like the model presented in Figure 2. Here we can see loops appear in the place of a step by step flow. While the user feels the need to represent every step of the process as being unique, when mapping these to the EMR the distinction loses relevance. For example we can say that "Medical Evaluation" and "Evaluation of results" are the same task since these are represented by a new iteration of a clinical note in the EMR.

Another big difference between a model described by a clinical user and a technical model as the one of Figure 2 is the detail and grouping of steps. We can, for example, remove the "Suspected Diagnosis" step described by the user as this is usually included only as part of the clinical note. Also, steps that are in distinct areas of the EMR and can be done in different contexts outside the flow described by the user can be represented as sub-processes. For this we have the example of the "Diagnostic sub-process" or the "Treatment sub-process" which can be done in a different order or sequence than the one of Figure 1 when used in a different patient or clinical context.

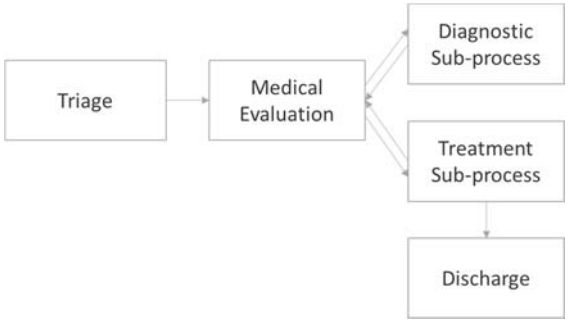



Figure 2: Simplified clinical process example as it would be used in the backend.

In the end we are left with only the steps which can be detected from or triggered using specific activities of the EMR. And while this might bring some value in terms of process evaluation using the EMR, it is not so useful if we are trying to support and stimulate the users to follow guidelines or processes using an abstract process model where the same type of task can have different meanings and relevance.

We believe that a model that reflects the habits and routines of the professionals and not just the steps / recommendations of the protocols / guidelines / pathways is the key to make a process support tool operational and usable in clinical practice. That is, a model which guides the users into doing the right thing using more than just the steps that are recorded in the EMR but also including those necessary for their own routines. Such an ideal model based solution would be the one that is capable of providing the support for the human only tasks mentioned in Figure 1, that usually have no place for representation in the EMR (e.g. nurse calls the lab to check status of sample analysis), even if they are not driving the reasoning of the process. This support can be given not only by making the association between the modelling tools with Clinical Decision Support Systems (CDS) but also organizational tools just like communication, schedule assistance tools or others. A severe limitation of modelling clinical processes (whether prospectively or derived from process mining) is the ability to derive representative models despite some essential activities not being represented in the event dataset.

Concerning those activities that are not possible to model due to lack of evidence in the IT system, these are essential to be aware of as this may imply:

- an incorrect (incomplete) representation of the process when performing process discovery. Which consists in applying process mining algorithms to an event log based on the information from the Hospital's EHR database, to discover a process model. (van der Aalst, W., 2016)
- a necessary change to the IT system which may have an impact on the workflow when trying to derive a process model for real-time tracking of process activities.

The latter has implications that go further than the mere addition of a few requirements to the IT solution: if the additional events cannot be captured automatically, this will imply additional input from the users affecting their workflow and potentially adding burden to the overall process. If the workflow is affected, this would also call for other

measures such as co-creation with the users and change management leading up to and during introduction of the technology to ensure acceptance and good uptake of the solution.

## 4 CONCLUSIONS

Deriving clinical processes based on data available in EHRs is a challenge for a number of reasons: different hospitals are likely to implement similar processes in different ways due to different resources available and local constraints; not all process activities may be directly extractable from the data, due to lack of documentation or impossibility to capture in a structured format; any additional process-related data which needs to be acquired may be seen as an additional burden on the users and may impede the actual process we are trying to support. When extracting knowledge from users to determine relevant events from data or to derive process models, one must be aware of the different realities of each setting and user's role, and try to capture the overall process by approaching the various stakeholders that often work together to make the entire clinical process a reality.

It is really important to find a balance between the tasks that need to be represented and shown to the user and the tasks that can be automated relieving burden from the user. For a good workflow support system we do not necessarily need to present all the steps of the process to the user nor represent in the model all the intermediate steps that are taken by the user. More than a good model, you will need extra support systems that can fill the gaps and fix the bottlenecks of the workflows.

## ACKNOWLEDGEMENTS

We would like to thank the participants in our research study for time and valuable input, as well as our colleagues from our Research project for making this work possible.

## REFERENCES

Ash, J., Bates, D., 2004. Position Paper. *"Factors and Forces Affecting EHR System Adoption: Report of a 2004 ACMI Discussion" Journal of the American Medical Informatics Association* 2005; 12:8–12.doi: 10.1197/jamia.M1684.

Böckmann, B., Heiden, K. 2013. *"Extracting and Transforming Clinical Guidelines into Pathway Models for Different Hospital Information Systems." Health Information Science and Systems* 1:13.doi:10.1186/2047-2501-1-13.

Coiera, E., 2000. "When Conversation Is Better Than Computation." *Journal of the American Medical Informatics Association* 7 (3): 277. doi:10.1136/jamia.2000.0070277.

Dadam, P., Manfred R., Klaus K. 2000. *"Clinical Workflows — The Killer Application for Process-Oriented Information Systems?" In BIS 2000: 4th International Conference on Business Information Systems, Poznań, Poland, 12–13 April 2000, edited by Witold Abramowicz and Maria E. Orlowska, 36–59. London: Springer London.* http://dx.doi.org/10.1007/978-1-4471-0761-3_3.

Jamoom, E. W., Heisey-Grove, D., Yang, N., Scanlon, P., 2016. *"Physician Opinions about EHR Use by EHR Experience and by Whether the Practice had optimized its EHR Use" Journal of Health & Medical Informatics 7 (4) OMICS International.*, ISSN: 2157-7420.

Latoszek-Berendsen, A., Tange, H., van den Herik, H., Hasman. A. 2010. *"From Clinical Practice Guidelines to Computer-Interpretable Guidelines. A Literature Overview." Methods of Information in Medicine 49* (6): 550–70. doi:10.3414/ME10-01-0056.

Mans, R., van der Aalst, W., Vanwersch, R., 2015. *Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes;* Springer.

Ologeanu-Taddei, R., Morquin, D., Domingo, H., & Bourret, R., 2015. *"Understanding the acceptance factors of an Hospital Information System: evidence from a French University Hospital." AMIA Annual Symposium Proceedings,* 2015, 1001–1007.

Oosterhout, E., Talmon, J., Clercq, P., Schouten, H., Tange, H., Hasman, A., 2005. *"Three-Layer Model for the Design of a Protocol Support System." MIE 2003 74* (2–4): 101–10. doi:10.1016/j.ijmedinf.2004.04.023.

Staccini, P., Joubert, M., Quaranta, J., Fieschi, D., Fieschi. M., 2001. *"Modelling Health Care Processes for Eliciting User Requirements: A Way to Link a Quality Paradigm and Clinical Information System Design." International Journal of Medical Informatics 64* (2–3): 129–42. doi:10.1016/S1386-5056(01)00203-9.

van der Aalst, W., 2016. *"Process Mining: Data Science in Action",* Springer.

# Process Improvement in an Oncology Day Hospital: Design of a Traceability and Indicator System

Eduardo Cañada[1], Enrique Casado[1] and Ana M. Bernardos[2]
[1]*Hospital Universitario Infanta Sofía, San Sebastián de los Reyes, Madrid, Spain*
[2]*Universidad Politécnica de Madrid, Madrid, Spain*
*{eduardo.canada, enrique.casado}@salud.madrid.org, anamaria.bernardos@upm.es*

Keywords: Process Improvement, Traceability, Oncology Day Hospital, Information System.

Abstract: Day hospitals (DH) are organizational structures that enable the supervision of patients who must go through diagnosis methods or treatments taking several hours, but not requiring inpatient hospitalization. Oncology Day Hospitals (ODH) are a particularly complex subset of DH, due to the variety and type of pathologies that are treated in them, the characteristics of cytostatic drugs, the involvement of different hospital units and professional profiles, the number of stages of the care procedure and the cost. In this paper, we describe the design of a traceability and indicator system for ODH, which aims at improving the performance and quality of service, providing three-folded benefits for patients, practitioners and hospital managers. The system is currently being tested in a public hospital in the Autonomous Community of Madrid. Their users perceive that they have access to a much more accurate fingerprint of everyday workflow, thus facilitating the design of improvement actions.

## 1 INTRODUCTION

Since their introduction in the 70s, Oncology Day Hospitals (ODH) have played a key role in the treatment of cancer: it is in this organizational structure where the oncology patients receive their chemotherapy treatment, with specifically tailored drugs. ODH are target of continuous improvement measurements: the disease impact in patients' quality of life, the nature of cytostatic drugs, the usually long visit protocols, the number of professionals involved in the treatment workflow and the cost of the attention require that ODH operations are revisited and optimized to guarantee safety, efficiency and quality of service. For example, in the last years, a particular effort has been done to deploy traceability systems for cytostatic drugs over all the clinical workflow, to provide maximum protection in drug administration. The works of Queralt et al. (2015), Kergosien et al. (2011) or Sini et al. (2011) focus on this issue, while the preparation and manipulation of the drugs themselves is also a matter of interest (Masini et al., 2014). From a more holistic approach, some other experiences, such as the one described by Galligioni et al. (2009), examine the hindrances and benefits of the use of specific tools to manage electronic oncological patient records.

This paper describes the process improvement analysis that has been carried out to have an accurate fingerprint of the activity and performance of the ODH at Hospital Universitario Infanta Sofía. Hospital Universitario Infanta Sofía is a public hospital in the Autonomous Community of Madrid (Spain), active from 2008. In 2015, 35515 care sessions where handled in its Day Hospital (SIAE, 2015); 7085 of those were oncology sessions (approximately 20%).

As a result of the procedural analysis, a Traceability and Indicators System (TIS) for the ODH has been designed; it is composed by a real-time visualization interface and a business intelligence tool (dashboard). On one hand, the visualization interface retrieves real-time timestamps at the different stages of the ODH operation workflow, so it can provide real-time data and alerts for health workers and managers, at the same time that facilitates the integration of information services for patients. On the other hand, the business intelligence tool enables the retrieval of a complete set of activity, performance, quality of care and procedure indicators that aims at providing information to design continuous improvement strategies.



Cañada, E., Casado, E. and Bernardos, A.
Process Improvement in an Oncology Day Hospital: Design of a Traceability and Indicator System.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 601-608
ISBN: 978-989-758-213-4

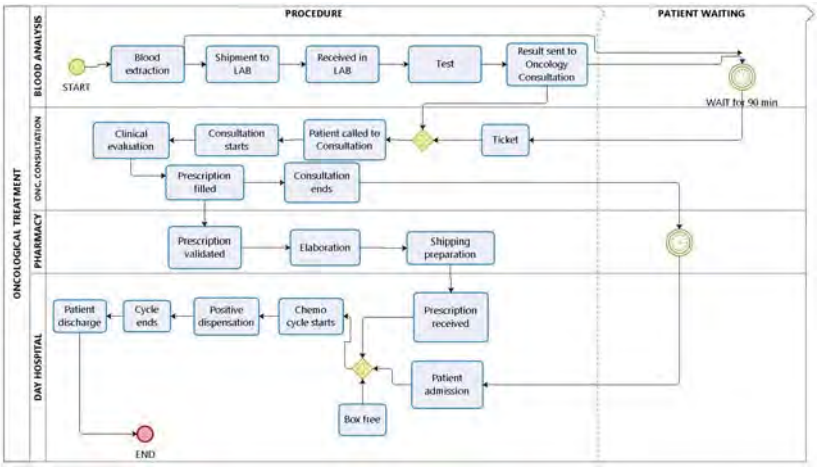



Figure 1: ODH workflow for a patient.

The TIS has been built on already available information that was not being retrieved in an easy-to-visualize way. The system aims at delivering three-folded benefits: for the patient, to reduce and lighten the time at the ODH; for the clinician, to facilitate the retrieval of real-time information about the workload and improved agenda management thanks to the existence of accurate and reliable indicators; for the manager, to facilitate the design of actions to raise of the perceived quality of service and to handle bottlenecks and process problems, so the impact of the investments agenda can be better evaluated.

In this context, the paper is structured as follows. Section II describes the operation workflow at the ODH and the subsequent TIS' functional and non-functional requirements. Section III describes the deployed architecture and the customized real-time visualization interface. Section IV describes the indicators. Finally, Section V concludes the work, stating the impressions gathered about the system and defining further steps.

# 2 OPERATIONAL CONTEXT

## 2.1 Workflow Analysis

A standard visit of a patient to the ODH is as follows: on arrival, the patient goes directly to take the 1) *preliminary blood tests*. After that, they head to the 2) *medical consultation*, where the practitioner examines the overall situation and *prescribes the cytostatic drugs*. The cytostatic drugs are then 3) *prepared* at the Pharmacy Service and *sent to the ODH for dispensation.* When the cytostatic drugs are prepared and delivered at ODH and the needed resources (seat, bed, pump, etc.) are ready, the 4) patient is *admitted to the ODH* for the 5) *drug administration.* When the treatment is over, the administrative staff in charge proceeds to the 6) *patient discharge*, who leaves the hospital. Figure 1 summarizes the full workflow.

In practice, during all the process, it is needed to do a follow up of the patients, personnel, resources and drugs. This tracking involves four Hospital Units: ODH, Laboratory, External Consultations and Pharmacy. The professional profiles participating in the workflow are oncologist (FAC), pharmacologists (PHAR), health staff (NUR), administrative staff (ADM) and ancillaries (ANC).

As our objective is to build a traceability and indicators system (TIS) that may report a real-time picture of the ODH operation and a full view of the service performance, it is important to analyze which milestones may be automatically retrieved through the APIs provided by the commercial information systems that are already deployed in the Hospital. In the particular case of Hospital Universitario Infanta Sofía, these systems include the Hospital Information System (Selene from Siemens), the Pharmacy Information System (FarmaTools from Dominion) and the Laboratory Information System (LIS-ServoLab from Siemens). Apart from those, an Appointment and Queue Tracking System for consulting rooms (AQ-Quenda from Plexus Technologies) is also in use.

Table 1 summarizes all the events in the operation workflow. As the reader will notice, there are events that are not being registered (e.g. when the patient enters the hospital or the NUR takes blood), and some others that are still being registered in manual way, thus their timestamps cannot be directly integrated into the TIS.

## 2.2 System Requirements

On this workflow analysis, the functional requirements for the TIS have been defined (i.e. what the system has to do in practice):

- Show the status (situation in the encounter workflow and particularities) of all the scheduled patients, both for regular visits and ODH.
- Include the non-scheduled patients that may appear during the day.
- Execute the automatic capture of information from all the hospital IS that provide information about the patient whereabouts and health records.
- Facilitate the input of relevant information in manual mode, in case it is not recorded at the HIS.
- Provide user management, so managers, clinicians, administrative staff may access to the information that is useful for them.
- Define specific rules for alerts and generate visual feedback when these rules are not satisfied.
- Generate alert messages for patients and clinicians to manage specific situations, both manually and automatically.
- Provide a dashboard of indicators that may provide an overview of the activity and resources occupancy.
- Generate and submit a daily report.

Table 1: Events in the ODH workflow.

| Unit | Event | Info. Sys. |
|---|---|---|
| ODH | PAT enters the hospital | - |
| ODH | PAT goes to ODH services for blood extraction. | - |
| ODH | NUR prints tags. | HIS |
| ODH | NUR takes blood. | - |
| ODH | NUR puts blood sample in the pneumatic tube. | - |
| LAB | LAB receives blood sample. * | HIS |
| LAB | LAB does blood analysis. | - |
| LAB | LAB provides the report. * | LIS+HIS |
| EXC | PAT takes the turn ticket. * | AQ |
| EXC | FAC calls PAT when LAB report ready. * | AQ+HIS |
| EXC | FAC provides subsequent appointments and lab requests. * | HIS |
| EXC | FAC fills in pharmacy prescription. * | PhIS |
| EXC | FAC fills in a form at HIS, updates the PAT record and copies the report in an unassigned note. The note is always assigned to protocols (colon, pulmonary, tumor committee). * | HIS |
| EXC | FAC finalizes consultation. * | HIS |
| PHAR | PHAR checks the prescriptions through the IS, following a stage-gate predefined process. * | PhIS |
| PHAR | The treatment is verified. * | PhIS |
| PHAR | PHAR prints the Report of Preparation to Administer. This report is signed by NUR and PHAR and it is taken to the clean room. | Manual |
| PHAR | Once the drug is ready, an ANC is called. | Manual |
| PHAR | ANC collects preparation at FAR. | Manual |
| ODH | ODH receives preparation. | Manual |
| ODH | ADM registers the patient and prints the identification bracelet. | HIS |
| ODH | PAT is admitted in ODH. * | HIS |
| ODH | NUR assigns a seat. | HIS |
| ODH | The administering process starts, guided by the pharmacy system for secure administration. * | HIS |
| ODH | PAT discharge. * | HIS |

Events with asterisk are those that can be automatically retrieved from IS. PAT: patient. LAB: laboratory. EXC: External Consultation. ADM: Admissions. In the very specific case of ODH, patients go straight to the blood draw service without going through the reception desk.

Regarding the non-functional requirements (how the system has to be), the TIS must perform satisfactorily with respect to:

- Availability: it must be accessible and easy to configure from any connected workplace at the hospital, not needing any specific additional software.
- Concurrency: its performance must to be not penalized by the simultaneous use from different workplaces.
- Security: the system must manage and control every access and keep trace of them.
- Performance: the system response must be real-time (not above 3 seconds).
- Usability: the system must be easy and comfortable for the users. Not more than 10 minutes training should be necessary for the users to work with the system.

## 2.3 Design Methodology

The methodology utilized to design the system follows an iterative approach, in which iterations are composed by analysis, development and testing phases. In this case, the main users of the systems are practitioners and managers, so their permanent contribution on three prototyping stages that have been necessary to come out with a first stable version has been crucial.

# 3 SYSTEM ARCHITECTURE

The core of all the hospital information systems is the HIS (Selene from Siemens), which is used in Emergency Care, Hospitalization, External Consultations and Day Hospital. Through it, any patient appointment or request is managed (diagnostic tests, subsequent consultations, inter-consultations, follow-up notes, etc.). The HIS facilitates the elaboration of forms, the generation of reports and the visualization of the patient's Electronic Health Record.

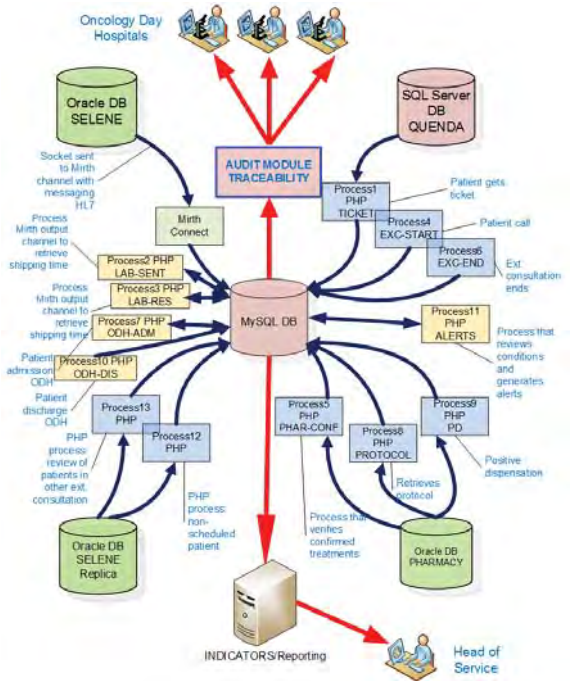



Figure 2: System architecture.

Departmental applications are deployed to cover the specific needs of a given Department or Service. For example, FarmaTools is the departmental tool for Pharmacy (drugs purchase, store distribution, and in this case, cytostatic drug prescription, verification and management). The appointment system (Quenda) avoids voice calls and guarantees privacy overall the hospital consultations. It enables to put in order the waiting rooms and provides indicators for consultation management (arrival time to the hospital, consultation call time, finalization time). The traceability and indicator system for ODH connects to the APIs provided by these three tools to automatically retrieve the data of interest.

Figure 2 shows the TIS architecture, which is in practice deployed over Linux in a virtual machine. The traceability and indicator system is composed by several modules developed in PHP and HTML; these modules retrieve real time information from the mentioned systems. For integration with HIS Selene, HL7 messaging is used, through a channel in the integration engine MirthConnect. For Quenda, FarmaTools and Selene's mirror DB, direct access to the databases is implemented (SQL Server and Oracle DB in the last two cases). The TIS is composed by four main elements: 1) a MySQL database, 2) a main module in PHP/HTML that shows the current state of the ODH patients, 3) eleven processes that update the information in the MySQL database through *cron* programmed jobs (Table 2) and 4) the indicators module.

Table 2: Description of *cron* jobs.

| Job | Description | Provider |
|---|---|---|
| TICKET | Patient takes the turn ticket on arrival the hospital. | Quenda DB |
| LAB-SENT | Patient goes to ODH for blood test. Blood draw is notified. | HL7 ORU^R01, HIS Selene |
| LAB-RES | Lab result is received. | HL7 |
| EXC-START | Timestamp when the oncologist calls the patient for consultation. | Quenda DB |
| PHAR-CONF | Timestamp for confirmation of the drugs for the day, enabled by the oncologist. When PHAR is notified with the treatment confirmation, the pharmacologist in charge verifies the order and submits it to the technicians, who start working on the preparation. | FarmaTools Oracle DB |
| EXC-END | Timestamp when the oncologist finishes the consultation. | Quenda DB |
| ODH-ADM | After leaving the consultation, the patient goes to the ODH and admission verifies the appointment and provides the identification wristband. | HL7 ADT01 |
| PROTOCOL | The drug protocol is obtained from the PHAR database, together with the number of components to administer to the patient. | FarmaTools Oracle DB |
| PD | When the administration of a component is completed, it is registered in the positive dispensation module. | FarmaTools Oracle DB |
| ODH-DIS | When the administration is completed, the patient is discharged. | HL7 ADT^03 |
| ALERTS | It checks if the alerts' conditions are fulfilled and generate the defined alerts if so. | MySQL DB |


Figure 3: a) Main dashboard for traceability. From the fourth column on, it is possible to find the timestamps for the 10 jobs in Table 2. The 11th job generates the alerts, marked as blocks in orange. In the top right corner of the Figure, a shortcut to filter the available agendas is provided.

Everyday at 7:00 am, a *cron* job initiates the patients' registry for the day, by using the agendas of the ODH and the monographic agendas (related to the oncology specialties) or nominative ones (related to specific practitioners). *Cron* jobs are executed each 5-10 minutes (configurable time), updating the data for real-time visualization and feeding the indicators' database.

Access and interface personalization is carried out through IP control. ODH or Consultation users can access the ODH TIS through any browser in an authorized computer. The main interface, in Figure 3, is daily initialized and dynamically completed throughout the day, it starts "growing" when the ODH activity begins. The interface provides visual alerts (in orange) indicating when issues occur in the workflow. These alerts are informative up to now, being generated e.g. when "the laboratory results are taking too long" or "the patient is not in the next step of the workflow". In the future, it is desired that these alerts may trigger automatic or supervised responses, such as "recall results from the laboratory" or "send a notification to the patient".

In the right side of the real-time interface, there are several options that enable to parametrize the TIS, showing information from a) all the agendas, only from the ODH agendas or the Oncologists Consultations agendas, b) detailed information about the patient and c) day indicators.

## 4 REPORTING

Apart from the visualization interface, the TIS provides a business intelligence tool that summarizes a set of relevant key indicators, which can be classified into four different groups: a) activity, b) performance, c) procedural and d) quality of care indicators.

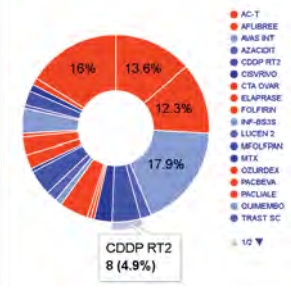
*Activity indicators* reflect the day dynamics at the Oncology Service (both at ODH and consultations).

These six indicators are directly accessible from the interface in Figure 3 and include aspects such as snapshot of number of active patients in ODH and patient distribution per agenda (in ODH and consultations), protocol, visit type and treatment type. Table 3 shows an example of a protocol indicator, in particular the one in which each chemotherapy protocol for a given cancer type is retrieved (an example of protocol is FOLFOX – oxaliplatin, fluorouracil, folinic acid - for gastric cancer).

Table 3: Activity indicator example.

| **Activity indicator no. 4** | |
|---|---|
| Scope | ODH, Oncology Consultations and Pharmacy. |
| Definition | Patient distribution per protocol. |
| Info. source | Selene Replica, FarmaTools |
| Formula | Query that retrieves the total number of patients in the agendas of a) ODH (non-scheduled patients included) and b) Oncology Consultations, classified by protocol. |
| Goal | To be defined by the evaluator / manager. |

Table 3: Activity indicator example (Cont.).

| **Activity indicator no. 4** | |
|---|---|
| Responsible | Exploitation: Hospital Management.<br>Evaluation: Oncology Service Head |
| Visualization |  |

The rest of indicators are retrievable from a specific interface. There are nine *performance indicators*, which aims at providing a view of the service efficiency. This group of indicators include e.g. the average number of admissions per hour, the appointments' status (cancellations, completed, not registered), the number of non-scheduled patients, the resource use or the real duration of the treatment. The information available about an example of performance indicator, the average duration of the patient's stay at ODH, is showed in Table 4.

There are six *quality of care indicators* that are focused on compiling information about how the patient's perception may be. These indicators include issues such as the delay of the patient with respect to the appointment for blood extraction, the delay at the oncology consultation, the time between the patient's appointment and the call time in the oncology consultation or the delay in the start of the treatment, etc. An example is available in Table 5.

Finally, there are some specific *procedural indicators* that are focused on measuring the evolution of procedure errors (e.g. admissions without discharge time, number of never ended consultations, number of non-called patients, etc.).

Table 4: Performance indicator example.

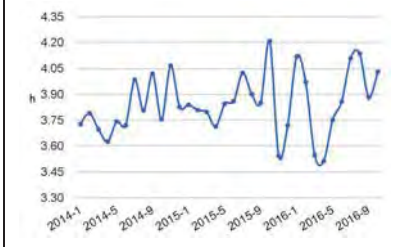
| **Performance indicator no. 8** | |
|---|---|
| Scope | ODH |
| Definition | Average stay duration at ODH. |
| Info. source | Selene Replica, MySQL indicators database |
| Formula | Query, average duration of admitted patients in ODH. |
| Goal | To be defined by the evaluator / manager. |
| Responsible | Exploitation: Hospital Managmnt.<br>Evaluation: Oncology Service Head |
| Statistics | Mean: 3.85 h.; Standard deviation:0.17; Min: 3.51; Max: 4.21. |
| Visualization |  |

Table 5: Example of Quality of care indicator.

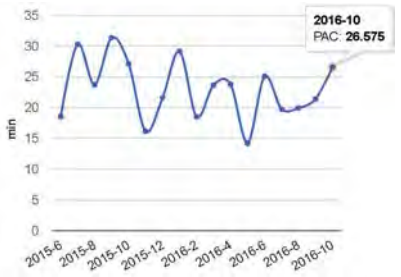
| **Quality of care indicator no. 4.** | |
|---|---|
| Scope | ODH |
| Definition | Delay between the patient's scheduled appointment time and the call time to consultation. |
| Info. source | Selene Replica, Oncology Consultation Agendas. |
| Formula | Query, mean delay between the scheduled time and the real call time to consultation, filtered for days, months or years. |
| Goal | 0 minutes. |
| Responsible | Exploitation: Hospital Managmnt.<br>Evaluation: Oncology Service Head |
| Statistics | Mean: 22.97 m.; Standard deviation: 4.77; Min: 14.21; Max: 31.25. |
| Visualization |  |

Figure 4 shows the components for the extract, transform and load (ETL) process necessary to generate the indicators. All these data are stored in the MySQL database and queried from a PHP service that uses Google Chart libraries to generate the final interface. Everyday, a *cron* job captures the main dashboard screen of the traceability system and emails it to the designated receivers (e.g. Director of the Oncology Service).

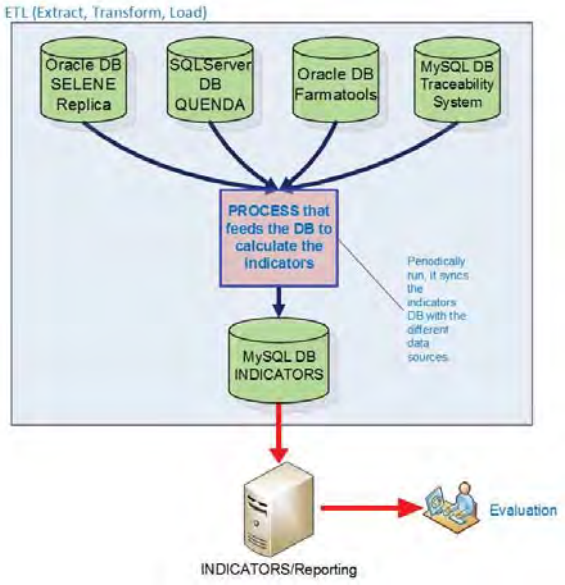



Figure 4: Extract, Transform, Load process for indicators. DB stands for database.

## 5 CONCLUSIONS

The Traceability and Indicators System described in this paper facilitates the retrieval of significant information in an ODH, with the purpose of improving daily operation and patient satisfaction. Traditionally, part of the information retrieved by the TIS has been manually recorded and processed to obtain indicators, and only significant deviations, detected. The TIS has been designed upon request of the Oncology Service Management to have a better view of the ODH workflow, so specific organizational and technical actions can be designed on a solid informational input.

The TIS has been designed in tight collaboration with practitioners and administrative staff, with the objective that it can fulfil the real requirements of day-to-day operation in the Oncology Service of a hospital, providing added value and avoiding problems in the organizational change that the availability of this tool may suppose.

The TIS relies on information that can be automatically retrieved from the existing systems in the hospital. This entails that the deployment does not require complex technology or significant investments, while providing relevant information of the workings of the oncology day hospital. The design process has been useful to identify information gaps (e.g. part of the processes that are not automatized, still done in manual ways), so their improvement can be added to the strategic agenda of investments.

The system has been technically evaluated against the design requirements in Section 2.2. Although it fulfils them reasonably well, there are several aspects that need to be improved. For example, the management of both alerts and notifications still need to be better implemented, and the set of indicators, polished. Additionally, although the available presentation interfaces have been incrementally improved with the help of the users, different presentation formats should be benchmarked. This can also be applied to indicators; it is also necessary to systematically analyse the causes of the detected deviations to feed the TIS.

Up to now, practitioners and nurses have partly tried the system in oncology consultations, ODH and Service Head offices, but their experience with the TIS is still brief to extract conclusive comments. In any case, from their feedback it is possible to say that:

- The system gathers real-time information in a reliable way, although it is still necessary to make some procedural changes to better contextualize some data. For example, due to specific instructions given to ODH patients, the time when these patients take the turn ticket for consultation is stored, but the entry time at the hospital is not (ODH patients are told to go directly for blood draw service).
- It is necessary to provide better traceability at resource level (seat or bed), with the purpose of optimizing its occupation. This traceability is manually done and may not be accurate enough due to human errors. A technical solution involving RFID or barcodes is currently being designed to address this issue, and it will be integrated at the TIS when ready.
- Alerts in the real-time visualization interface need to include specific management options, to track if the alert has been handled. Up to now, they are merely informative and no action is triggered from them.
- One of the possibilities to manage alerts is to make them trigger SMS both to patients or care professionals, as there is a corporate existing platform for this purpose. Other options, such as instant messaging, involve technology and organizational changes and may delay the integration with the service. In any case, it is necessary to study to which extent these SMS may be effective and useful for their recipients (misleading or spam effects). In particular, it is necessary that involved users express their

opinion about the best means and configuration of alerts.

An important issue is related to human factors: to assure the full exploitation of the visualization tool for real time patient-flow management purposes wide acceptance within the health team is needed. All health staff members need to understand the ultimate instrumental goal, focused on improving delivery of quality of care and not staff monitoring. Specific communication initiatives are needed to guarantee that this fact is correctly understood. In this sense, the involvement of navigator nurses in key steps control has been shown of great help. In the next months, it is expected that the tool is integrated in the daily activity of the ODH and improved following the users suggestions.

Taking into consideration this TIS, another tool is currently being designed to monitor the workflow in the 200 hospital consultations, so delay times and bottlenecks can be identified.

## ACKNOWLEDGMENT

Authors want to acknowledge the Hospital Universitario Infanta Sofía Management for the support to this initiative.

## REFERENCES

Queralt M., Frías M.C., de Pedro, V., Ramírez, J., Mesalles, M. 2015. Sistema de trazabilidad integral de administración de citotóxicos. *Boletín de Prevención de Errores de Medicación de Cataluña, Vol. 12,* no.4.

Kergosien, Y., Tournamille, J.F., Laurence, B. and Billaut, J.C., 2011. Planning and tracking chemotherapy production for cancer treatment: A performing and integrated solution. *International journal of medical informatics, 80(9),* pp.655-662.

SIAE (2015). Información estadística de hospitales. Estadística de Centros Sanitarios de Atención Especializada (National statistic information about hospitals). Available at: *https://www.msssi.gob.es/estadEstudios/estadisticas/estHospiInternado/inforAnual/homeESCRI.htm.*

Sini, E., Torresani, M., Locatelli, P., Restifo, N. and Facchini, R., 2011. Information and communication technologies enable a comprehensive governance of patient safety in chemotherapy. An Italian framework project. *In 24th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 1-7). IEEE.

Masini, C., Nanni, O., Antaridi, S., Gallegati, D., Marri, M., Paolucci, D., Minguzzi, M. and Altini, M., 2014. Automated preparation of chemotherapy: quality improvement and economic sustainability. *Am J Health Syst Pharm, 71(7),* pp.579-85.

Galligioni, E., Berloffa, F., Caffo, O., Tonazzolli, G., Ambrosini, G., Valduga, F., Eccher, C., Ferro, A. and Forti, S., 2009. *Development and daily use of an electronic oncological patient record for the total management of cancer patients: 7 years' experience. Annals of Oncology, 20(2),* pp. 349-352.

# Short Paper

# ICT: Health's Best Friend and Worst Enemy?

Egon L. van den Broek
*Department of Information and Computing Sciences, Utrecht University, PO Box 80.089, 3508 TB Utrecht, The Netherlands*
*http://www.human-centeredcomputing.com/*
*vandenbroek@acm.org*

Keywords: ICT, Health Problems, Stress, Black Plague, Affective Computing, Emotion, Personality, Coping.

Abstract: I propose a paradigm shift for health care, as there is an urgent need for i) continuous (semi-)automatic medical checkups, ii) cost reduction, and iii) cure for the 21st century black plague (i.e., stress-related diseases) are very much needed. To realize this ICT's Paradox has to be solved. On the one hand, ICT can cause i) musculoskeletal problems, ii) vision problems, iii) headache, iv) obesity, v) stress disorders (e.g., burn out), vi) metabolic issues, vii) addiction (e.g., to games, social media, and Internet), viii) sleeping problems, ix) social isolation, and x) an unrealistic world view. On the other hand, ICT claims to provide these problems' solutions. Consequently, health informatics needs to adopt a holistic approach, improve its fragile theoretical frameworks, and handle the incredible variance we all show. As a remedy, I propose to take up the challenge to next-generation models of personality, as they are a crucial determinant in people's stress coping style.

*Your worst enemy*
*Becomes your best friend,*
*once he's underground.*
– Euripides –

## 1 INTRODUCTION

Our health care system is not neither functioning effectively nor effectively. "Many of the changes intended to improve or reform health care are resulting in increased costs, increased work, and little (if any) improvement in health." (Bartol, 2016) For example, the much discussed electronic health records were meant to improving care; but, show to be complicated and inefficient. Medical doctors, nurse practitioners as well as patients are forced to check more boxes, use more technology, and produce more data, without health care improving or costs declining (Bartol, 2016; Stylianou and Talias, 2017).

Health care does not need to reform or transform, it needs to be recreated from the bottom-up, it needs a paradigm shift! "Attempts at reforming and transforming health care have been like repairing or fixing up a 50-year-old car, adding newer equipment and modern technology to try to make it better. The problem is that we do not consider if the original vehicle is really what is needed today." (Bartol, 2016) "Health care systems around the world are both highly institutionalized and highly professionalized." (Ferlie et al., 2016). Their processes are directed to patients and their symptoms, using procedures and medication. Hereby, the focus on on treating the ill, ignoring the healthy. So, resources are spend on health care's high spenders, where we ignore the healthy.

There are, at least, three reasons to support this call for this paradigm shift:

i) Continuous (semi-)automatic medical checkups (Jarvis, 2016) and support for healthy living should become part of common health care;

ii) Extent healthy people's health could reduce health care costs significantly; and

iii) Stress-related diseases are rapidly becoming the dominant class of illness.

Next-generation medical check-ups could benefit from the many types of health care data (Stylianou and Talias, 2017), including (Fang et al., 2016):

i) human-generated (e.g., notes, email, and paper documents);

ii) machine-generated monitoring;

iii) financial transactions;

iv) biometric data (e.g., genomics, genetics, x-ray, and electrocardiogram, ECG);

v) social media; and



Broek, E.
ICT: Health's Best Friend and Worst Enemy?.
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017) - Volume 5: HEALTHINF*, pages 611-616
ISBN: 978-989-758-213-4

vi) publications.

A range of apps already exist that conduct limited next-generation medical check-ups, using such data. However, often they are not clinically validated. Examples include e-coaches that support you with sleeping (Beun, 2013), running, and eating to reduce diabetics (Georga et al., 2014). However, many of these apps use no or only basic biometric sensors (cf. (van den Broek, 2010)). So, there is a world to win for unobtrusive and wearable technologies, when shown to result in reliable signal acquisition and, subsequent, analysis.

In October 2016, the American Medical Association (AMA) shared a view point on health care cost reduction: "Chronic diseases account for 7 of 10 deaths in the United States, and treatment of these diseases accounts for more than 85% of US health costs, including those incurring exceptional treatment costs. About half of health expenditures in the United States are accounted for by 5% of patients, yet many chronic conditions are preventable. ... Perhaps the most powerful chronic disease prevention approaches are those that can motivate and facilitate health-promoting choices in daily life."(Dietz et al., 2016). I propose to adopt Mandl and Kohane (2016)'s patient-control, where patients themselves can be asked to collect data. They selected eight reasons for pursuing patient-controlled data:

i) complete data: A more holistic view of patients;

ii) data sharing for coordinated care: Patients as vehicle for data sharing;

iii) as foundation for next-generation medical check-up apps;

iv) support of diagnostic journals (e.g., specific with genetic disorders);

v) data donation for research purposes;

vi) patients as reporters;

vii) an additional pairs of eyes; and

viii) social networking.

I would like to stress the importance of including not-yet-patients, as also healthy people may become patients and, consequently, their contributions are at least as valuable.

Ten years ago, Cary L. Cooper stated: "We're talking now I think about the 21st century black plague. I see stress as the main source of disease or the trigger for disease in the 21st century developed world." (Newby, ABC Catalyst, 2007). In their handbook, Lundberg and Cooper (2011) provided a rich source of evidence for this strong statement. A few months ago, the European Occupational Safety and Health Agency (EU-OSHA) suggested a possible solution: "Software exists that allows the emotions of a computer user to be monitored remotely - this could even be promoted by the developers as a way of detecting early signs of stress in employees, ...". This would aid all three reasons as it could be part of a next-generation medical check-ups, should realize a significant health care cost reduction, and focusses on stress. The development of such software is considered to be part of health informatics' subfield affective computing. Affective computing can be defined as "the scientific understanding and computation of the mechanisms underlying affect and their embodiment in machines" (p. 10) (van den Broek, 2011). Despite its unquestioned potential, affective computing's inherent interdisciplinary complexity limits its progress in performance, as I already denoted in a series of articles presented at this conference (van den Broek et al., 2009; van den Broek et al., 2010a; van den Broek et al., 2010b; van den Broek et al., 2010c; van den Broek et al., 2011). More specifically, affective computing's complexity primarily lays in:

i) its need for a holistic approach (note. this is not a new idea at all; cf. (Follmer, 2016; Schmitz and Wolkenhauer, 2016);

ii) the fragile theoretical frameworks from medicine (e.g., incl. physiology and neuroscience) and psychology it has to rely on (e.g., (Kalisch et al., 2015; Jarvis, 2016); and

iii) the incredible, continuous variance that characterizes our world (Follmer, 2016; Schmitz and Wolkenhauer, 2016).

Moreover, ICT solutions such as affective computing both has their positive as well as its negative sides, as we will discuss in the next section.

## 2 ICT: HEALTH'S BEST FRIEND AND WORST ENEMY?

Even before the age of smartphones and tablets, Joan Stigliani (1995) identified six main health problems, related to computer usage. Since then, more than two decades elapsed in which ICT usage intensified, nowadays using tablets, smartphones, smartwatches, and hardly desktop PCs anymore. Consequently, starting with Stigliani (1995)'s original list, I composed a new list of the 10 main ICT-related health problems:

i) stress disorders (e.g., burn out) (Åborg and Billing, 2003);

ii) musculoskeletal problems (Åborg and Billing, 2003; Gowrisankaran and Sheedy, 2015), including Repetitive Stress Injury (RSI) [1];

iii) vision problems (Salibello and Nilsen, 1995; Gowrisankaran and Sheedy, 2015);

iv) headache (Salibello and Nilsen, 1995; Gowrisankaran and Sheedy, 2015);

v) obesity (de Jong et al., 2013; Schmiege et al., 2016);

which can be complemented by:

vi) metabolic issues, such as vitamin deficiencies (Palacios and Gonzalez, 2014) and diabetics (de Jong et al., 2013);

vii) addiction (e.g., to games, social media, and Internet (Zhou et al., 2017);

viii) sleeping problems (Beun, 2013; Nuutinen et al., 2014);

ix) social isolation (Cacioppo and Hawkley, 2003; Liu and Baumeister, 2016); and

x) an unrealistic world view (e.g., resulting in depression) (Donnelly and Kuss, 2016; Wood et al., 2016).

Note that, compared to Stigliani (1995)'s original list, this list includes both more indirect ICT-related and more mental health problems.

If any ICT branch should be health's best friend, it is health informatics. In solving its challenges, health informatics relies on both clinical experience and knowledge of public health systems and organizations, while conducting experiments, interventions, and scalable approaches. (Kulikowski et al., 2012). Par excellence, ICT's health informatics, should contain the identified computer-related health problems. From an ethical point of view, the ICT community should even consider this as one of its main priorities.

When going through scientific literature, health informatics seems to have solved all ICT-related health problems. For example, musculoskeletal problems can be prevented using persuasive technology (Wang et al., 2014), the problem of obesity is approached similarly (Wang et al., 2014), as are headache (Minen et al., 2016), diabetics (Georga et al., 2014), sleeping problems (Beun, 2013), and social isolation (Chen and Schulz, 2016). So, it seems to be a case of "One size fits all" (Suomi, 1996). However, many solutions are fragile, random control trails are absent or conducted at a small scale, and solutions are at a gadget level, instead of at the level of aimed clinical solutions. Many roads can be followed to remedy this practice. The problem lies is the increasing tendency to just see what the computer shows. (van den Broek, 2012)

In the next section, I will pose one critical concept for health informatics, complementary to the prerequisites defined before (van den Broek et al., 2009; van den Broek et al., 2010a; van den Broek et al., 2010b; van den Broek et al., 2010c; van den Broek et al., 2011): personality. Par excellence, this concept illustrates affective computing's threefolded complexity, as depicted in Section 1. Moreover, health informatics is also the claimed road towards next-generation personalized medicine (Poon et al., 2013). How can this be, if clients personalities are ignored?

## 3 PERSONALITY

The urge to completely redesign our health care system relies on a both crucial and often ignored determinant: personality. For each of the three reasons for this call for a paradigm shift, I will explain why personality is an essential part of the equation:

i) Next-generation medical check-ups are personalized. However, solely the medical check-up does not help; in particular, healthy people have to be persuaded to start, improve, or maintain a healthy living style. Tailored health messages are needed, next-generation customized (semi-)automatic ICT-based communication (Kreuter et al., 2012).

ii) I proposed to implement Mandl and Kohane (2016)'s patient-control paradigm and extend it to people-health control to realize cost reduction; that is, ask people themselves to collect their data. This requires sincere cooperation from people, as they are asked to contribute to their own electronic health records.

iii) The 21st century black plague is directly linked to people's coping style and, hence, personality (Zhou et al., 2017). "Coping is often defined as efforts to prevent or diminish threat, harm, and loss, or to reduce associated distress. Some prefer to limit the concept of coping to voluntary responses; others include automatic and involuntary responses within the coping construct. ... Personality does influence coping in many ways, however, some of which occur prior to coping. Even

[1]Note. Stigliani (1995) mentioned RSI as separate entry. However, essentially it is a musculoskeletal problem and, hence, placed here under this entry.

> prior to coping, personality influences the frequency of exposure to stressors, the type of stressors experienced, and appraisals" (Carver and Connor-Smith, 2010).

With a foundation provided for personality's key role in the proposed paradigm shift, I will now sketch its history, challenges as well as a solution approach.

In this article, at a functional level, I will take a methodological perspective to personality and describe it in terms of its research traditions (or approaches) (Pervin, 2003):

i) clinical (e.g., including Freud's work);

ii) statistical (e.g., the big-five (BF) model); and

iii) experimental (e.g., including the work of Wundt, Pavlov, and Skinner).

Clinical approaches to personality allow a holistic approach, while observing a great variety of phenomena, using self-reports and behavioral observations. As such they satisfy two of the three dimensions of complexity (i.e., holistic approach and explain the huge variance). However, clinical approaches to personality fail in the third dimension of complexity: generation of solid theoretical frameworks, as reliable observations and tests of hypotheses are complicated if possible at all. Statistical approaches to personality focus on individual differences, using trait questionnaires instead of self-reports. This practice enables statistical analysis of the relation between personality traits and other variables. However, this approach suffers from studies on non-representative samples (e.g., students) and a lack of both generalizability and specificity. The statistical approach can provide theoretical frameworks; but, fails to take a true holistic approach and simplifies reality and, hence, is unable to explain the existing real world variance in personality traits. Experimental approaches to personality rely on laboratory studies on cause-effect relationships, as such they are the opposite of clinical approaches. They violate the holistic approach and cannot explain the huge variance of the real world. However, they comply to the third dimension of complexity: generation of solid theoretical frameworks.

Models and theory used in affective computing are heavily skewed towards the statistical approaches (cf. (Vinciarelli and Mohammadi, 2014)). To some extent this makes sense as the statistical approaches are most straightforward to model (e.g., using machine learning or general linear models). Although understandable from a pragmatic stance, the above analysis of the three research traditions of personality makes it hard to justify this position. I propose to reconsider and improve the clinical approach to personality, such that computational models can be build based on it. This requires a well-argued combination of lab and field data and, most likely, a merge of quantitative and qualitative data (Fine and Elsbach, 2000; McCusker and Gunaydin, 2015).

# 4 CONCLUSION

Health care is vibrant and yet conservative, a highly complex field of science, engineering, and practice. As is argued in Section 1, its foundations should be reconsidered. ICT, and in particular, health informatics, can play a crucial role in this process. However, the stakes are high, including potentially big losses as well as big gains, as is discussed in Section 2, which makes such endeavor even more challenging. With paradigms such as personalized medicine and mobile health, the client is put central. However, so far the client's personality has been ignored (see Section 3). I pose that the clinical tradition of personality research should be embraced by health informatics, realizing that this would require a firm step back to enable the so much needed steps forward.

As is illustrated via this article's relatively lengthly list of references, an interdisciplinary, holistic approach was taken. Nevertheless, many issues remained untouched. For example, social and cultural (Kaplan, 2017) and psychosocial issues (Kun, 2016) should be considered on top of the evident privacy and security concerns (Blobel et al., 2016). An analysis of the current state-of-the-art of electronic health records would have been appropriate as well (Wuyts et al., 2012; Stylianou and Talias, 2017). Moreover, a discussion on computational techniques for personality models at a functional level would have been in place (cf. (Huys et al., 2016; Adams et al., 2016)). However, this article's list of references provides a good starting point.

Health informatics is struggling, already since its recent conception (Kulikowski et al., 2012; Nelson and Staggers, 2018). Omnipresent computing power, big data, rapidly improving sensor technology, and our extended lives, have put it in the top list of society's health agenda. This is a promising trend; but, as posed, significantly more mass is needed to change the field's paradigm and see all humans as clients, instead of only those who are already patients.

# ACKNOWLEDGMENTS

I thank the two anonymous reviewers, who provided excellent comments on an early draft of this article. Moreover, I thank Winnie Teunissen for carefully

proof reading of the pre-final version of this article. Last, I thank Gert-Jan de Vries for his invitation to present this article at his special session.

## REFERENCES


Åborg, C. and Billing, A. (2003). Health effects of 'the paperless office' – evaluations of the introduction of electronic document handling systems. *Behaviour & Information Technology*, 22(6):389–396.

Adams, R. A., Huys, Q. J. M., and Roiser, J. P. (2016). Computational psychiatry: towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery & Psychiatry*, 87(1):53–63.

Bartol, T. (2016). Recreating healthcare: The next generation. *The Nurse Practitioner*, 41(11):10–11.

Beun, R. J. (2013). Persuasive strategies in mobile insomnia therapy: alignment, adaptation, and motivational support. *Personal and Ubiquitous Computing*, 17(6):1187–1195.

Blobel, B., Lopez, D. M., and Gonzalez, C. (2016). Patient privacy and security concerns on big data for personalized medicine. *Health and Technology*, 6(1):75–81.

Cacioppo, J. T. and Hawkley, L. C. (2003). Social isolation and health, with an emphasis on underlying mechanisms. *Perspectives in Biology and Medicine*, 46(3 Supplement):S39–S52.

Carver, C. S. and Connor-Smith, J. (2010). Personality and coping. *Annual Review of Psychology*, 61:679–704.

Chen, Y.-R. R. and Schulz, P. J. (2016). The effect of information communication technology interventions on reducing social isolation in the elderly: A systematic review. *Journal of Medical Internet Research*, 18(1):e18.

de Jong, E., Visscher, T. L. S., HiraSing, R. A., Heymans, M. W., Seidell, J. C., and Renders, C. M. (2013). Association between TV viewing, computer use and overweight, determinants and competing activities of screen time in 4- to 13-year-old children. *International Journal of Obesity*, 37(1):47–53.

Dietz, W. H., Douglas, C. E., and Brownson, R. C. (2016). Chronic disease prevention: Tobacco avoidance, physical activity, and nutrition for a healthy start. *Journal of the American Medical Association (JAMA)*, 316(16):1645–1646.

Donnelly, E. and Kuss, D. J. (2016). Depression among users of social networking sites (SNSs): The role of SNS addiction and increased usage. *Journal of Addiction and Preventive Medicine*, 1(2):#107.

Fang, R., Pouyanfar, S., Yang, Y., Chen, S.-C., and Iyengar, S. S. (2016). Computational health informatics in the big data age: A survey. *ACM Computing Surveys (CSUR)*, 49(1):Article No. 12.

Ferlie, E., Montgomery, K., and Pedersen, A. R. (2016). *The Oxford Handbook of health care management*. Oxford, UK: Oxford University Press, 1st edition.

Fine, G. A. and Elsbach, K. D. (2000). Ethnography and experiment in social psychological theory building: Tactics for integrating qualitative field data with quantitative lab data. *Journal of Experimental Social Psychology*, 36(1):51–76.

Follmer, E. (2016). *What we do to fit in: personality, coping, and Person-Environment fit*. PhD thesis, Graduate College, The University of Iowa, Iowa City, IA, USA.

Georga, E. I., Protopappas, V. C., Bellos, C. V., and Fotiadis, D. I. (2014). Wearable systems and mobile applications for diabetes disease management. *Health and Technology*, 4(2):101–112.

Gowrisankaran, S. and Sheedy, J. E. (2015). Computer Vision Syndrome: A review. *Work*, 52(2):303–314.

Huys, Q. J. M., Maia, T. V., and Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, 19(3):404–413.

Jarvis, C. (2016). *Physical examination and health assessment*. St. Louis, MO, USA: Elsevier, Inc., 7th edition.

Kalisch, R., Müller, M. B., and Tüscher, O. (2015). A conceptual framework for the neurobiological study of resilience. *Behavioral and Brain Sciences*, 38:e92.

Kaplan, J. (2017). Artificial Intelligence: Think again. *Communications of the ACM*, 60(1):36–38.

Kreuter, M., Farrell, D., Olevitch, L., and Brennan, L. (2000/2012). *Tailoring health messages: Customizing communication with computer technology*. LEA's Communication Series. Mahwah, NJ, USA: Routledge, Taylor & Francis Group / Lawrence Erlbaum Associates, Inc., reprint edition.

Kulikowski, C. A., Shortliffe, E. H., Currie, L. M., Elkin, P. L., Hunter, L. E., Johnson, T. R., Kalet, I. J., Lenert, L. A., Musen, M. A., Ozbolt, J. G., Smith, J. W., Tarczy-Hornoch, P. Z., and Williamson, J. J. (2012). AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. *Journal of the American Medical Informatics Association*, 19(6):931–938.

Kun, L. (2016). Introductory remarks to the special issue on social implications of technology, in honor of the late dr. Lodewijk Bos. *Health and Technology*, 6(1):1–4.

Liu, D. and Baumeister, R. F. (2016). Social networking online and personality of self-worth: A meta-analysis. *Journal of Research in Personality*, 64:79–89.

Lundberg, U. and Cooper, C. L. (2011). *The science of occupational health: Stress, psychobiology, and the new world of work*. Chichester, West Sussex, UK: Wiley-Blackwell / John Wiley & Sons Ltd.

Mandl, K. D. and Kohane, I. S. (2016). Time for a patient-driven health information economy? *The New England Journal of Medicine*, 374(3):205–208.

McCusker, K. and Gunaydin, S. (2015). Research using qualitative, quantitative or mixed methods and choice based on the research. *Perfusion*, 30(7):537–542.

Minen, M. T., Torous, J., Raynowska, J., Piazza, A., Grudzen, C., Powers, S., Lipton, R., and Sevick, M. A. (2016). Electronic behavioral interventions for headache: A systematic review. *The Journal of Headache and Pain*, 17(1):Article 51.

Nelson, R. and Staggers, N. (2018). *Health informatics: An interprofessional approach*. St. Louis, MO, USA: Elsevier, Inc., 2nd edition.

Newby, ABC Catalyst, J. (06/09/2007). Workplace stress: Stopping the juggernaut. URL: http://www.abc.net.au/catalyst/stories/s2025212.htm [Last accessed on January 5, 2017].

Nuutinen, T., Roos, E., Ray, C., Villberg, J., Välimaa, R., Rasmussen, M., Holstein, B., Godeau, E., Beck, F., Léger, D., and Tynjälä, J. (2014). Computer use, sleep duration and health symptoms: A cross-sectional study of 15-year olds in three countries. *International Journal of Public Health*, 59(4):619–628.

Palacios, C. and Gonzalez, L. (2014). Is vitamin D deficiency a major global public health problem? *The Journal of Steroid Biochemistry and Molecular Biology*, 144(Part A):138–145.

Pervin, L. A. (2003). *The Science of Personality*. New York, NY, USA: Oxford University Press, Inc., 2nd edition.

Poon, C. C. Y., Wang, M. D., and Bonato, P. (2013). Editorial: Special issue on health informatics and personalized medicine. *IEEE Transactions on Biomedical Engineering*, 60(1):143–146.

Salibello, C. and Nilsen, E. (1995). Is there a typical VDT patient? a demographic analysis. *Journal of the American Optometric Association*, 66(8):479–483.

Schmiege, S. J., Gance-Cleveland, B., Gilbert, L., Aldrich, H., Gilbert, K. C., and Barton, A. (2016). Identifying patterns of obesity risk behavior to improve pediatric primary care. *Journal of Specialists in Pediatric Nursing*, 21(1):18–28.

Schmitz, U. and Wolkenhauer, O. (2016). *Systems Medicine*, volume 1386 of *Methods in Molecular Biology / Springer Protocols*. New York, NY, USA: Humana Press / Springer Science+Business Media.

Stigliani, J. (1995). *The computer user's survival guide: Staying healthy in a high tech world*. Sebastopol, CA, USA: O'Reilly Associates, Inc.

Stylianou, A. and Talias, M. A. (2017). Big data in healthcare: A discussion on the big challenges. *Health and Technology*, 7:[in press].

Suomi, R. (1996). One size fits all – or does it? *Behaviour & Information Technology*, 15(5):301–312.

van den Broek, E. L. (2010). Beyond biometrics. *Procedia Computer Science*, 1(1):2505–2513.

van den Broek, E. L. (2011). *Affective Signal Processing (ASP): Unraveling the mystery of emotions*. PhD thesis, Human Media Interaction (HMI), Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands.

van den Broek, E. L. (2012). *Affective computing: A reverence for a century of research*, volume 7403 of *Lecture Notes in Computer Science*, pages 434–448. Berlin/Heidelberg, Germany: Springer-Verlag.

van den Broek, E. L., Janssen, J. H., Healey, J. A., and van der Zwaag, M. D. (2010a). Prerequisites for Affective Signal Processing (ASP) – Part II. In Fred, A., Filipe, J., and Gamboa, H., editors, *Biosignals 2010: Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, pages 188–193, Valencia – Spain.

van den Broek, E. L., Janssen, J. H., van der Zwaag, M. D., and Healey, J. A. (2010b). Prerequisites for Affective Signal Processing (ASP) – Part III. In Fred, A., Filipe, J., and Gamboa, H., editors, *Biosignals 2010: Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, pages 432–435, Valencia – Spain.

van den Broek, E. L., Janssen, J. H., van der Zwaag, M. D., Westerink, J. H. D. M., and Healey, J. A. (2011). Prerequisites for Affective Signal Processing (ASP) – Part V: A response to comments and suggestions. In *Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, pages 301–306, Rome, Italy.

van den Broek, E. L., Janssen, J. H., Westerink, J. H. D. M., and Healey, J. A. (2009). Prerequisites for Affective Signal Processing (ASP). In Encarnação, P. and Veloso, A., editors, *Biosignals 2009: Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, pages 426–433, Porto – Portugal.

van den Broek, E. L., van der Zwaag, M. D., Healey, J. A., Janssen, J. H., and Westerink, J. H. D. M. (2010c). Prerequisites for Affective Signal Processing (ASP) – Part IV. In Kim, J. and Karjalainen, P., editors, *Proceedings of the 1st International Workshop on Bio-inspired Human-Machine Interfaces and Healthcare Applications – B-Interface*, pages 59–66, Valencia – Spain.

Vinciarelli, A. and Mohammadi, G. (2014). A survey of personality computing. *IEEE Transactions on Affective Computing*, 5(3):273–291.

Wang, S.-C., Jiang, C., and Chern, J.-Y. (2014). Promoting healthy computer use: Timing-informed computer health animations for prolonged sitting computer users. *Behaviour & Information Technology*, 33(3):295–301.

Wood, M. A., Bukowski, W. M., and Lis, E. (2016). The digital self: How social media serves as a setting that shapes youths emotional experiences. *Adolescent Research Review*, 1(2):163–173.

Wuyts, K., G, V., Scandariato, R., Joosen, W., and Dumortier, J. (2012). What electronic health records dont know just yet. a privacy analysis for patient communities and health records interaction. *Health and Technology*, 2(3):159–183.

Zhou, Y., Li, D., Li, X., Wang, Y., and Zhao, L. (2017). Big five personality and adolescent internet addiction: The mediating role of coping style. *Addictive Behaviors*, 64:42–48.

# AUTHOR INDEX

Proceedings of BIOSTEC 2017 - Volume 5: HEALTHINF

10th International Joint Conference on Biomedical Engineering Systems and Technologies

http://www.biostec.org

LOGISTICS PARTNER

EVENT MANAGEMENT SYSTEM

by SCITESOFT

IN COOPERATION WITH

PROCEEDINGS WILL BE SUBMITTED FOR INDEXATION BY



ISBN: 978-989-758-213-4