STUDIES IN HEALTH TECHNOLOGY AND INFORMATICS 285

pHealth 2021

Proceedings of the 18th International Conference on Wearable Micro and Nano Technologies for Personalized Health

Editors: Bernd Blobel
Mauro Giacomini

IOS Press

Smart mobile systems – microsystems, smart textiles, smart implants, sensor-controlled medical devices – together with related body, local and wide-area networks up to cloud services, have become important enablers for telemedicine and the next generation of healthcare services. The multilateral benefits of pHealth technologies offer enormous potential for all stakeholder communities, not only in terms of improvements in medical quality and industrial competitiveness, but also for the management of healthcare costs and, last but not least, the improvement of patient experience.

This book presents the proceedings of pHealth 2021, the 18th in a series of conferences on wearable micro and nano technologies for personalized health with personal health management systems, hosted by the University of Genoa, Italy, and held as an online event from 8 – 10 November 2021. The conference focused on digital health ecosystems in the transformation of healthcare towards personalized, participative, preventive, predictive precision medicine (5P medicine). The book contains 46 peer-reviewed papers (1 keynote, 5 invited papers, 33 full papers, and 7 poster papers). Subjects covered include the deployment of mobile technologies, micro-nano-bio smart systems, bio-data management and analytics, autonomous and intelligent systems, the Health Internet of Things (HIoT), as well as potential risks for security and privacy, and the motivation and empowerment of patients in care processes.

Providing an overview of current advances in personalized health and health management, the book will be of interest to all those working in the field of healthcare today.

ISBN 978-1-64368-226-6 (print)
ISBN 978-1-64368-227-3 (online)
ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

pHEALTH 2021

Studies in Health Technology and Informatics

International health informatics is driven by developments in biomedical technologies and medical informatics research that are advancing in parallel and form one integrated world of information and communication media and result in massive amounts of health data. These components include genomics and precision medicine, machine learning, translational informatics, intelligent systems for clinicians and patients, mobile health applications, data-driven telecommunication and rehabilitative technology, sensors, intelligent home technology, EHR and patient-controlled data, and Internet of Things.

Studies in Health Technology and Informatics (HTI) series was started in 1990 in collaboration with EU programmes that preceded the Horizon 2020 to promote biomedical and health informatics research. It has developed into a highly visible global platform for the dissemination of original research in this field, containing more than 250 volumes of high-quality works from all over the world.

The international Editorial Board selects publications with relevance and quality for the field. All contributions to the volumes in the series are peer reviewed.

Volumes in the HTI series are submitted for indexing by MEDLINE/PubMed; Web of Science: Conference Proceedings Citation Index – Science (CPCI-S) and Book Citation Index – Science (BKCI-S); Google Scholar; Scopus; EMCare.

Volume 285

Recently published in this series

Vol. 284 M. Honey, C. Ronquillo, T.-T. Lee and L. Westbrooke (Eds.), Nurses and Midwives in the Digital Age – Selected Papers, Posters and Panels from the 15th International Congress in Nursing Informatics

Vol. 283 R. Röhrig, T. Beißbarth, J. König, C. Ose, G. Rauch, U. Sax, B. Schreiweis and M. Sedlmayr (Eds.), German Medical Data Sciences 2021: Digital Medicine: Recognize – Understand – Heal – Proceedings of the Joint Conference of the 66th Annual Meeting of the German Association of Medical Informatics, Biometry, and Epidemiology e.V. (gmds) and the 13th Annual Meeting of the TMF – Technology, Methods, and Infrastructure for Networked Medical Research e.V. 2021 online in Kiel, Germany

Vol. 282 I. Verma (Ed.), Universal Design 2021: From Special to Mainstream Solutions

ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

pHealth 2021

Proceedings of the 18th International Conference on Wearable Micro and Nano Technologies for Personalized Health

8–10 November 2021, Genoa, Italy

Edited by

Bernd Blobel

Medical Faculty, University of Regensburg, Germany
Chair, Scientific Program Committee

and

Mauro Giacomini

Department of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Genoa, Italy
Chair, Local Organizing Committee

IOS Press

Amsterdam Berlin Washington, DC

ISBN 978-1-64368-226-6 (print)
ISBN 978-1-64368-227-3 (online)
Library of Congress Control Number: 2021949038
doi: 10.3233/SHTI285

Cover image by Sara Mora, Genoa

Publisher
IOS Press BV
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

For book sales in the USA and Canada:
IOS Press, Inc.
6751 Tepper Drive
Clifton, VA 20124
USA
Tel.: +1 703 830 6300
Fax: +1 703 830 2300
sales@iospress.com

PRINTED IN THE NETHERLANDS

Preface

pHealth 2021 is the 18th conference in a series of scientific events that has brought together expertise from medicine, technology, politics, administration, and social domains, and even from philosophy and linguistics. It opens a new chapter in the success story of this series of international conferences on wearable or implantable micro and nano technologies for personalized medicine.

Begun in 2003 as a Dissemination Activity in the framework of a European Project on Wearable Micro and Nano Technologies for Personalized Health with personal health management systems, pHealth conferences have evolved to become truly interdisciplinary and global events. As comprehensively represented in the conference series, pHealth also covers technological and biomedical facilities, legal, ethical, social and organizational requirements and impacts, as well as the basic research necessary for the enabling of future-proof care paradigms. It thereby combines medical services with public health, prevention, social and elderly care, wellness and personal fitness to establish participatory, predictive, personalized, preventive, and effective care settings. In this way, it has attracted scientists, developers, and practitioners from various technologies, medical and health disciplines, legal affairs, politics, and administration from all over the world. The conference brings together health-service vendor and provider institutions, funding organizations, government departments, academic institutions, professional bodies, and also patients and citizens representatives.

Smart mobile systems, such as microsystems, smart textiles, smart implants, sensor-controlled medical devices, and innovative sensor and actuator principles and techniques, as well as related body, local and wide-area networks up to cloud services, have become important enablers for telemedicine and ubiquitous pervasive health as the next generation of healthcare services. Social media and gamification have added even further knowledge to pHealth as an eco-system.

The OECD has defined four basic areas on which to focus in the new care model: addressing the challenges of big data; fostering meaningful innovation; understanding and addressing the potential new risks; and supporting a concerted effort to un-silo communities for a virtual care future. The benefits of pHealth technologies offer enormous potential for all stakeholder communities, including patients, citizens, health professionals, politicians, healthcare establishments, and companies from the biomedical technology, pharmaceutical, and telecommunications domain, not only in terms of improvements in medical quality and industrial competitiveness, but also for the management of healthcare costs.

The pHealth 2021 conference benefits from the experience of and the lessons learned by the organizing committees of previous pHealth events, particularly 2009 in Oslo, 2010 in Berlin, 2011 in Lyon, 2012 in Porto, 2013 in Tallinn, 2014 in Vienna, 2015 in Västerås, 2016 in Heraklion, 2017 in Eindhoven, 2018 in Gjøvik, 2019 in Genoa, and 2020 in Prague. The 2009 conference raised the interesting idea of having special sessions focusing on a particular topic and organized by a mentor/moderator. The Berlin event in 2010 initiated workshops on particular topics taking place before to the official start of the conference. Lyon, in 2011, launched so-called dynamic demonstrations which allowed participants to demonstrate software and hardware

solutions on the fly without the need for a booth. Implementing pre-conference events, pHealth 2012 in Porto gave attendees a platform for presenting and discussing recent developments and provocative ideas that helped to animate the sessions. The highlight of pHealth 2013 in Tallinn was the special session on European project success stories, and also presentations on up and coming paradigm changes and challenges associated with Big Data, Analytics, Translational and Nano Medicine, etc. Vienna, in 2014, focused on lessons learned from national and international R&D activities and practical solutions, particularly from Horizon 2020, the new EU Framework Program for Research and Innovation. Alongside reports about technology transfer support and building ecosystems and value chains to ensure better time to market and higher impact of knowledge-based technologies, the acceptability of solutions, particularly considering security and privacy aspects were presented and deeply discussed. pHealth 2015, held in Västerås, addressed mobile technologies, knowledge-driven applications and computer-assisted decision support, as well as apps designed to support the elderly and chronic patients in daily and possibly independent living. The fundamental scientific and methodological challenges of adaptive, autonomous, and intelligent pHealth approaches, the new role of patients as consumers and active parties with growing autonomy and related responsibilities, as well as the requirements and solutions for mHealth in low- and medium-income countries were also considered. The 2016 pHealth conference was aimed at the integration of biology and medical data, the deployment of mobile technologies through the development of micro-nano-bio smart systems, the emphasis on personalized health, virtual care, precision medicine, big bio-data management and analytics. The pHealth 2017 event in Eindhoven provided an inventory of the former conferences by summarizing requirements and solutions for pHealth systems, highlighting the importance of trust, and focused afresh on the behavioral aspects of designing and using pHealth systems. A specific aspect addressed was the need for flexible, adaptive and knowledge-based systems, as well as decision intelligence. pHealth 2018 established national and European satellite workshops, so complementing the more theoretical consideration of the majority of the papers with organizational and practical experiences. Borrowing good experiences from former events, pHealth 2018 responded to the national and regional need for advancing healthcare systems and their services to citizens and health professionals. pHealth 2019 placed a particular focus on artificial intelligence (AI) and machine learning (ML) and their deployment for decision support, and ethical challenges and related international manifests were discussed in depth in that context. pHealth 2020 – organized as a virtual event – addressed AI and robots, bio-data management and analytics for health and social care, security, privacy and safety challenges, integrated care, and also the intelligent management of specific diseases including the Covid-19 pandemic. The 2021 edition of the pHealth conference series – again a virtual event – focuses on digital health ecosystems in the transformation of healthcare towards personalized, participative, preventive, predictive precision medicine (5P medicine). The deployment of mobile technologies, micro-nano-bio smart systems, bio-data management and analytics, autonomous and intelligent systems, as well as the Health Internet of Things (HIoT) for personalized health, systems medicine, public health and virtual care are thereby especially considered. The conference also addresses new potential risks for security and privacy as well as safety chances and challenges, trustworthiness of partners and processes, and the motivation and empowerment of patients in care processes. The multilateral benefits of pHealth technologies offer enormous potential for all stakeholder communities, not only in terms of improvements in medical quality

and industrial competitiveness, but also for managing health care costs and, last but not least, improving patient experiences.

The conference is organized under the patronage of the City of Genoa and the Liguria Regional Authority, the University of Genoa and the Department of Informatics, Bioengineering, Robotics and System Engineering (DIBRIS) in particular, and Healthtrophy srl as a University of Genoa's Spin-Off. Following a long-standing tradition, the Working Groups "Electronic Health Records (EHR)", "Personal Portable Devices (PPD)", "Security, Safety and Ethics (SSE)", and "Translational Health Informatics" of the European Federation for Medical Informatics (EFMI) have also been actively involved in the preparation and realization of the pHealth 2021 event.

Neither the pHealth 2021 Conference nor the publication of the pHealth 2021 Proceedings by IOS Press would have been possible without the aforementioned financial and spiritual supporters and sponsors. This also includes the Italian Scientific Society of Biomedical Informatics, the IEEE Engineering in Medicine and Biology Society (EMBS), the Camber of Engineers Genoa, and the European Federation for Medical Informatics (EFMI) and standard-setting organizations such as HL7 International, ISO/TC215 or CEN/TC251.

The editors are also grateful to the Members of the international Scientific Program Committee, but especially for the dedicated efforts of members of the Local Organizing Committee and their supporters for the careful preparation and the smooth operation of the conference.

Bernd Blobel, Mauro Giacomini
(Editors)

This page intentionally left blank

pHealth 2021 Committees and Reviewers

pHealth 2021 Scientific Program Committee

Bernd Blobel, Medical Faculty, University of Regensburg, Regensburg, Germany (Chair)

Eric McAdams, INISA, Lyon, France

Dag Ausen, SINTEF ICT, Oslo, Norway

Mauro Giacomini, Dept. of Informatics, Bioengineering, Robotics and System Engineering (DIBRIS), University of Genoa, Genoa, Italy

William Goossen, Results4Care, Amersfoort, The Netherlands

Lenka Lhotská, Czech Institute of Informatics, Robotics and Cybernetics and Faculty of Biomedical Engineering, Czech Technical University in Prague, Prague, Czechia

Maria Lindén, School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden

Diego Mauricio Lopéz-Gutierrez, Faculty of Electronic Engineering and Telecommunications, University of Cauca, Popayan, Colombia

Andreas Lymberis, European Commission, DG Connect, Brussels, Belgium

Stefan Sauermann, University of Applied Sciences Technikum Wien, Vienna, Austria

Filipe Sousa, Fraunhofer Portugal-AICOS, Porto, Portugal

Marc Schurr, Ovesco Endoscopy AG, Tübingen, Germany

Bian Yang, IIK, NTNU, Gjøvik, Norway

pHealth 2021 Local Organizing Committee

Mauro Giacomini, Dept. of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Genoa, Italy (Chair)

Viola Lanza, Healthropy srl, Savona, Italy

Laura Pastorino, Dept. of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Genoa, Italy

Silvana Quaglini, Dept. of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

Carmelina Ruggiero, Dept. of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Genoa, Italy

pHealth 2021 Reviewers

Sun-Ju Ahn, Sungkyunkwan University, Suwon, Korea

Arriel Benis, Holon Institute of Technology, Holon, Israel

Contents

Personal Health Records, Patient-Centered Data Management, and Cloud Services

Solutions for Personalized Care

Challenges and Solutions in Health Education and Training

mHealth Applications

This page intentionally left blank

Keynote

This page intentionally left blank

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210567

Autonomous Systems and Artificial Intelligence – Hype or Prerequisite for P5 Medicine?

Bernd BLOBEL[a,b,c,1], Pekka RUOTSALAINEN[d], Mathias BROCHHAUSEN[e]
[a] *Medical Faculty, University of Regensburg, Germany*
[b] *eHealth Competence Center Bavaria, Deggendorf Institute of Technology, Germany*
[c] *First Medical Faculty, Charles University Prague, Czech Republic*
[d] *Faculty of Information Technology and Communication Sciences (ITC), Tampere University, Finland*
[e] *Dept. of Biomedical Informatics, University of Arkansas for Medical Sciences, USA*

Abstract. For meeting the challenge of aging, multi-diseased societies, cost containment, workforce development and consumerism by improved care quality and patient safety as well as more effective and efficient care processes, health and social care systems around the globe undergo an organizational, methodological and technological transformation towards personalized, preventive, predictive, participative precision medicine (P5 medicine). This paper addresses chances, challenges and risks of specific disruptive methodologies and technologies for the transformation of health and social care systems, especially focusing on the deployment of intelligent and autonomous systems.

Keywords. Healthcare transformation, pHealth, Artificial intelligence, Autonomous systems, Learning, Knowledge representation, Knowledge management, Ethics

1. Introduction

Personalized medicine individualizes diagnoses and treatments according to the personal health status, genetic, environmental, occupational, and social conditions and context by understanding the pathology of diseases including the individual predisposition to diseases and responsiveness to treatment. For understanding a disease's pathology and undertaking scientifically sound predictions and preventions, we have to explore the mechanisms and processes from molecule up to society, transferring basic sciences and biomedical research into clinical practice, adding precision medicine to the approach. Thereby, all interacting factors and components impacting individuals' health, such as genomes, epigenomes, proteomes, microbiomes, metabolomes, pharmacomes, transcriptomes, cognitive-affective behavioromes, etc., summarized as interactomes, must be considered. The entire approach requires a massive involvement of the subject of care and/or his/her social environment, extending the approach to participative health. The resulting personalized, preventive, predictive, participative precision medicine (P5

[1] Corresponding Author: Bernd Blobel, PhD, FACMI, FACHI, FHL7, FEFMI, FIAHSI, Professor; Medical Faculty, University of Regensburg, Germany; Email: bernd.blobel@klinik.uni-regensburg.de

medicine) allows containing costs despite aging and multi-diseased societies, improving care quality and safety, and managing health and social services consumerism. P5 medicine requires the involvement of multiple domains and disciplines with their own methodologies, languages, terminologies and ontologies, such as systems medicine, biology, physics up to the quantum level, chemistry, bioinformatics, genomics, but also social sciences, public health, etc., resulting in an incredible and continuously growing amount of data and information. Table 1 summarizes objectives, characteristics and methodologies as inevitable prerequisites for transforming health and social care systems [1].

Table 1. Transformed health ecosystems' objectives and characteristics as well as methodologies for meeting them, after [1]

Objective	Characteristics	Methodologies
Provision of health services everywhere anytime	• Openness • Distribution • Mobility • Pervasiveness • Ubiquity	• Wearable and implantable sensors and actuators • Pervasive sensor, actuator and network connectivity • Embedded intelligence • Context awareness
Individualization of the system according to status, context, needs, expectations, wishes, environments, etc., of the subject of care	• Flexibility • Scalability • Cognition • Affect and Behavior • Autonomy • Adaptability • Self-organization • Subject of care involvement • Subject of care centration	• Personal and environmental data integration and analytics • Service integration • Context awareness • Knowledge integration • Process and decision intelligence • Presentation layer for all actors
Integration of different actors from different disciplines/domains (incl. the participation/ empowerment of the subject of care), using their own languages, methodologies, terminologies, ontologies, thereby meeting any behavioral aspects, rules and regulations	• Architectural framework • End-user interoperability • Management and harmonization of multiple domains including policy domains	• Terminology and ontology management and harmonization • Knowledge harmonization • Language transformation/ translation
Usability and acceptability of pHealth solutions	• Preparedness of the individual subject of care Security, privacy and trust framework • Consumerization • Subject of care empowerment • Subject of care as manager • Information based assessment and selection of services, service quality and safety as well as trustworthiness • Lifestyle improvement and Ambient Assisted Living (AAL) services	• Tool-based ontology management • Individual terminologies • Individual ontologies • Tool-based enhancement of individual knowledge and skills • Human Centered Design of solutions • User Experience Evaluation • Trust calculation services

For collecting, managing and using those data, new techniques and methods have to be exploited such as mobile, bio-, nano- molecular technologies, big data and analytics, advanced computing, virtual reality, learning algorithms, etc. An overview on technologies and methodologies enabling P5 medicine is presented in Table 2 [2].

Table 2. Technologies and methodologies for transforming health ecosystems [2]

• Mobile technologies, biotechnologies, nano- and molecular technologies • Big data and business analytics • Integration of analytics and apps • Assisting technologies → Robotics, autonomous systems • Natural Language Processing → Text analytics → Intelligent media analytics • Conceptualization → Knowledge management (KM) and knowledge representation (KR) → Artificial intelligence (AI) → Artificial common (general) intelligence → Intelligent autonomous systems • Security and privacy, governance, ethical challenges, Education → Ethical AI Principles • Cloud computing, cognitive computing, social business	• Edge computing as a "family of technologies that distributes data and services where they best optimize outcomes in a growing set of connected assets“ (Forrester Research) • Virtual reality and augmented reality, thereby blurring “the boundaries between the physical and digital worlds“ (Gartner) • Creation of IoT-Platforms and app-ecosystems • Patient-generated health data ecosystem → multiple, dynamic policies • Web content management → Digital experience management • Data bases → NoSQL technologies → Data warehouses → Graph DBs → Data lakes • EHR extension with genomic data • Specifications → Implementation → Tooling → Testing → Certification

Use of artificial intelligence (AI) technologies for health holds great promise and has already contributed to important advances in fields such as drug discovery, genomics, radiology, pathology and prevention. AI could assist health-care providers in avoiding errors and allow clinicians to focus on providing care and solving complex cases. Further details on health transformation can be found in [3].

2. Methods

Transformation of health and social services according to the P5 medicine paradigm results in highly complex and highly dynamic multi-disciplinary systems, which have to be context-sensitive and cognitive to represent the intended settings in structure and function correctly and consistently. We do not have workforce, skills and power to manage such systems manually anymore. This holds for data search, collection, interrelation, interpretation and processing, but also for designing and managing the underlying complex and domain-crossing processes, not talking about the knowledge representation and management challenges discussed before and in [4]. Furthermore, we cannot place specialists next to every person to be ubiquitously cared for. Therefore, the deployment of robotics and artificial intelligence, or more generally autonomous and intelligent systems (AIS), using machine learning, big data and analytics at different levels is inevitable. Focusing on autonomous and intelligent systems in general regarding their cognitive functions, in this paper we will not address the specific properties of robots physically interacting with its environment.

Intelligence is a concept in cognition theory with four foundational principles: data, information, knowledge, and wisdom [5]. During investigations and observations, organs or sensors collect data as measures or symbols describing the world, that way forming

the structural level of intelligence. By attaching meaning to data, they are then transformed into information for taking decisions, establishing the semantic level of intelligence. Knowledge enables proper actions on the represented system, supervised and evaluated by wisdom. More background information on knowledge representation and intelligence can be found in [2].

The approach to artificial intelligence (AI) as used today was originated in 1950 by Alan Turing [6]. Considering AI, many definitions exist. McKinsey defines AI as "the ability of a machine to perform cognitive functions we associate with human minds, such as perceiving, reasoning, learning, and problem solving" [7]. It represents an interdisciplinary approach including mathematics, logics, cognitive sciences, life sciences, but also computation and engineering to manage processes such as modeling, simulating, understanding, etc. [8]. A simpler definition characterizes AI as ability of machines to simulate human intelligence [9]. The OECD provides in its Recommendation of the Council on Artificial Intelligence a specific AI definition addressing main aspects of this paper as follows. "An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy." [10]. Autonomy and automation levels are non-autonomous (assistive) or full autonomous refined by multiple risks levels, and conditional automation, high automation, and full automation, respectively [11].

AI and automation in healthcare aims at augmentation of capabilities and engagement for care provider and subject of care including education, access to information and services, etc. It enables cooperation, improves staff and patient experiences, but also processes including clinical workflow and scheduling, business efficiency, productivity and cost containment as well as risk analysis. Furthermore, it facilitates faster and more precise decisions of direct and indirect caregivers, administrators and patients including prognosis. Finally, it enables collaborative business intelligence as self-service. Therefore, we can distinguish different levels of intelligence in AI such as assisted intelligence automating tasks like pattern recognition with human input and intervention, augmented intelligence combining existing information with predetermined solutions based on different levels of machine learning, and autonomous intelligence deploying genetic algorithms and evolutionary strategies to act independently from humans [12].

Application domains of AI in healthcare are for example AI in medical imaging, AI in digital pathology, AI in genomics, AI for understanding and predicting the course of a disease, etc. The aforementioned application domains require different levels of AI from machine learning through deep learning up to swarm learning and are closely associated with big data, so representing the different levels of intelligence introduced before [13]. Most of current AI applications follow the weak AI (narrow AI) approach, usually just replicating human intelligence in the specific context such as simple classification, pattern recognition and assistive systems. Here, facial recognition, conversational assistants and chatbots, but also recommendation engines are well-known applications. The concept of artificial general intelligence (AGI) (sometimes also called strong intelligence) aims at mimicking the full range of human cognitive and intellectual capabilities, resulting in autonomous systems [14]. Artificial superintelligence (ASI) goes even beyond this approach by exceeding human intellectual power, almost comprehensively covering all categories and fields of endeavor [15]. Another way to classify AI refers to its implementation (in bracket equalizing the levels with the classification provided before). At lowest level we find reactive AI systems, followed by

limited memory machines (narrow intelligence) using historical experiences to inform future decisions, theory of mind AI systems (AGI) to infer intentions and predict behavior, and finally self-aware AI systems (ASI) [16]. Technologies for enabling AI, summarized in Table 2, are for example robots, virtual agents, computer vision and virtual reality, analytics, machine learning and natural language processing, understanding and generation.

Data science incorporates various disciplines -- for example, data engineering, data preparation, data exploration, data mining, predictive analytics, machine learning and data visualization, as well as mathematics, statistics and software programming. A predominant challenge data science addresses is the elimination of bias in data sets and analytics applications [17].

Data analytics can be provided at different levels. The first level is descriptive analytics, presenting what has happened. The next level is predictive analytics, providing insights regarding what will happen. The highest level is prescriptive analytics, providing foresights by defining how to perform to make things/processes happen. More details on analytics, its adoption model and related issues can be found in [4].

Machine learning as subset of artificial intelligence is at its most basic the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world. Here we distinguish between supervised learning training algorithms with human-labeled data, unsupervised learning inferring some structure from unlabeled data, and reinforcement learning deploying algorithms rewarding overcoming mistakes. In the latter case, Markov decision processes are typically used. Deep learning is a subset of machine learning that processes data using multi-layer neural networks, leveraging learning algorithms that mimic the function of the human brain. Specializations of neural networks are recurrent neural networks (RNNs) used in speech recognition and natural language processing by pattern analysis to predict the next likely object or scenario. Convolutional neural networks (CNNs) deployed in computer vision belong to the same neural network class [18]. Swarm learning is the newest approach using AI at the edge by decentralizing the analysis of data from multiple locations and sharing insights while protecting data sovereignty [19, 20]. When creating new knowledge by properly modeling a system in question, we have to validate the outcome on the real world system and thereafter adopted the knowledge representation if needed [21]. This holds for human-made and AI mediated knowledge representation development processes.

Modern AI applications rely more on learning from data to discover possibly new knowledge that needs verification than just codifying existing knowledge to automate problem solving. Thereby, they are facing two challenges. First, the solution might suffer from the lack of sound explainability compared with an established knowledge framework. Second, it might incorporate bias or errors due to biased or poor data the application has learned from [22]. Possible sources of bias are, e.g., insufficient data, skewed data, limited features, historical bias or unreliable labels, proxies [23].

3. Ethical Challenges of Artificial Intelligence and Autonomous Systems

Any action and relationship in enlightened democratic societies, and especially in health and social care ecosystems have to accommodate legal, moral and ethical principles. Contracts and law define and enforce behavior for maintaining relations, peace, and justice in a society. Ethics provides code and conduct guiding to decide what is good or

bad, and how to act and behave properly. It establishes and defends rules of morality, which frequently go beyond or contradicts the law [4, 24]. Ethical values are strongly impacted by culture, social norms and geographic locations. With the evolution of societies including sciences and technologies, approaches to ethics and its underlying theories in the framework of meta-ethics, normative ethics and applied ethics show evolutionary characteristics as well. Examples of this evolution are Aristotle's and Plato's virtue ethics, Kant's deontological ethics, Mill's utilitarian ethics, and Rawls' justice as fairness ethics [24]. In the special context of autonomous and intelligent systems for health and social care, we have also to mention consequentialism ethics, raising questions about acceptable consequences to the individual, or how to balance personal and benefits. Therefore, it is impossible to implement one global comprehensive standard of ethics. Instead, basic social and ethical principles such as dignity, freedom, autonomy, privacy, equality and solidarity, or the more technological categories like fairness, robustness, explainability and lineage have been established by different organization for different domains. In the context of AIS, such classification is more relevant than a philosophical one. The different sets of principles provided by governmental and non-governmental organizations as well as by vendors will be briefly discussed in the following chapter.

4. Ethical Frameworks

The four traditional bioethical principles are autonomy, beneficence, non-maleficence, and justice. For building public confidence in disruptive technologies, promoting safer practices and to facilitating broader societal adoption, explicability should be added. [25]. For designing and managing AIS solutions, following principles must be considered: Fair Information Principles [26]; Fair Information Practice Principles [27]; Ethical Principles; Big Data Best Privacy Practices according to the aforementioned Federal Trade Commission (FTC) Guidelines [27]. Digital change requires zero trust and a changed role of Chief Information Security Officers (CISOs) / Chief Privacy Officers (CPOs) – from security management to risk management, but also the inclusion of newer ethical initiatives. AI ethics according to IBM's definition aims at optimizing AI's beneficial impact while reducing risks and adverse outcomes for all stakeholders. It also involves identifying, studying, and proposing technical and nontechnical solutions for ethics issues arising from the pervasive use of AI in life and society. Examples of such issues are data responsibility and privacy, fairness, inclusion, moral agency, value alignment, accountability, transparency, explainability, trust, robustness and, awareness of technology misuse. [28, 29].

The WHO normative guidance "Ethics and governance of artificial intelligence for health" requests putting ethical considerations and human rights at the center of the design, development, and deployment of AI technologies for health, that way also fighting digital divide locally (exclusion of populations) and globally (low- and middle-income countries). Challenges to be met are the establishment of key ethical principles for the use of AI for health; the protection of human autonomy; the promotion of human well-being and safety and the public interest; ensuring transparency, explainability and intelligibility; fostering responsibility and accountability; ensuring inclusiveness and equity; and the promotion of AI that is responsive and sustainable [30]. That way, WHO responds on pioneering developments in fields such as genomics, epigenetics, gene

editing, artificial intelligence, and big data, all of which pose transformational opportunities but also risks to global health.

In its Global Initiative on Ethics of Autonomous and Intelligent Systems, IEEE defines the following success factors: Participatory design, consensus, multiple discipline focus, recognition of the socio-technical, and focus on design. [31, 32]. AI principles supporting IEEE Ethically Aligned Design include: Sustainable development, Well-being; Human-centered values; Fairness; Transparency and explainability; Robustness, security and safety; Accountability. The first author is actively involved in several projects of the IEEE 7000 Series "Ethics in Action in Autonomous and Intelligent Systems" [33].

The OECD proposed in the Conference Toward AI Network Society, April 2015, in Japan some years ago already Principles for AI Research and Development such as: Transparency; User Assistance; Control-ability; Security; Safety; Privacy; Ethics; and Accountability. The principles resulted in a related guideline prepared by the conference host Japan [34].

The World Economic Forum published the following Top Ethical Issues in Artificial Intelligence [35] to be addressed: Unemployment; Inequality; Humanity; Artificial stupidity; Racist robots; Security; Evil genies; Singularity; Robot rights.

The European Commission recently released a proposal for a regulation of the European Parliament and the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain EU legislative acts [36]. The project aims at guaranteeing that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values such as data protection, consumer protection, non-discrimination and gender equality.

Google established the following 6 objectives for AI applications: be socially beneficial, avoid creating or reinforcing unfair bias, be built and tested for safety, be accountable to people, incorporate privacy design principles, uphold high standards of scientific excellence [37].

Further ethical frameworks are The Asimolar AI Principles of the Future of Life Institute [38]; the Congress Resolution Supporting the Development of Guidelines for the Ethical Development of Artificial Intelligence [39]; and the BS 8611:2016 Robots and Robotic Devices. Guide to the Ethical Design and Application of Robots and Robotic Systems [40]. More information can be found in [4, 41].

Table 3 summarizes the essence of the different ethical frameworks.

Table 3. Common A/IS principles proposed by different organizations

Guideline Originator	Transparency	Accountability	Controllability	Security	Value Orientation Ethics	Privacy	Safety	Risk	User Assistance
OECD	x	x	x	x	x	x	x		x
IEEE	x	x	x		x	x		x	x
Asilomar	x	x	x	x	x	x	x	x	
US Congress	x	x	x	x		x	x		x
World Economic Forum				x	x			x	

5. Discussion

The different application domains, managing the different objectives of transformed pHealth ecosystems with different methodologies (Table 1), require different technologies (Table 2) including computing technologies and computing power.

At simplest level, we program a computer to perform a specific task without learning needed at computer side. This approach was followed by the deployment of specific or generalized models of statistics including probabilistic reasoning for automated medical diagnosis. An example is monitoring (recording and assessing) vital signs in the context of home care or ambient assisted living (AAL) supported by any type of portable devices such as smart watches, realized by nowadays microprocessors.

Machine learning focuses on prediction of known properties learned from training data, e.g. codifying and mapping input and output data patterns (existing knowledge) as pattern recognition by supervised learning, while data mining focuses on discovery of unknown properties, e.g. discovering hidden data patterns (new knowledge discovery) by unsupervised learning. Machine learning as optimization reduces the loss in training data sets by comparing predictions with current instances. Machine learning as generalization towards deep learning minimizes unknown properties, moving from data model to algorithmic model.

For assessing impacts of the aforementioned individual interactomes on the predisposition to a disease and responsiveness to treatment, a multidisciplinary approach to a complex, dynamic system is inevitable. As the specific structure and behavior of the considered system and its relations is unknown, a deep learning approach with advanced neural networks is necessary, requiring strong computing power.

When considering complex genotype-phenotype interactions for understanding the detailed pathology of diseases, analyzing and predicting the protein structure in the context of cell mutation and treatment in cancer or for developing vaccines and predicting their pharmacology, the deployment of evolutionary (genetic) algorithms on super computers or even quantum computers is necessary. The same holds for integrating individual health aspects and public health strategies. When creating new knowledge, its verification requires interactions with reality and therefore cognitive computing. Cognitive computing systems consider their environment and dynamically process huge amount of data from multiple and variant sources at high speed. For that purpose, they have to be adaptive, interactive, contextual, iterative and stateful [42]. A shift to cognitive computing is occurring in omics-driven biotechnology to enable precision medicine [43]. Cognition goes beyond recognition and includes knowledge and understanding.

The growing complexity, flexibility and dynamics of pHealth ecosystems as well as new technologies make it harder to maintain governance, security and privacy. Therefore, AIS must also be developed and implemented to maintain those important principles.

For modeling transformed health ecosystems, we have to represent the multiple domains contributing to the health and social care process, but also the rules to defining the behavior of the systems, summarized as policy domain. This domain can be refined into sub-policy domains such as process policies, legal policies, contextual policy (conditions and preferences influencing, e.g., privacy decisions) or ethical policies. For correctly and consistently integrating and interrelating multiple domains, they have to be modeled using the ISO 23903 Interoperability and Integration Reference Architecture [44]. This standard allows to model different knowledge spaces by architecturally representing the related domains and representing them through the domains ontologies. This requires representing the ethical domain by an ethical domain ontology, which has

been recently developed by the IEEE 7007 project the first author is member of [45]. In addition, key components of ethical considerations, such a deontic roles, rights and obligations, have been represented in the Document Act Ontology (d-acts) [46, 47]. Figure 1 shows the ISO 23903 base model (a) and its instantiation for the policy domains relevant in the context of autonomous and intelligent systems (b).

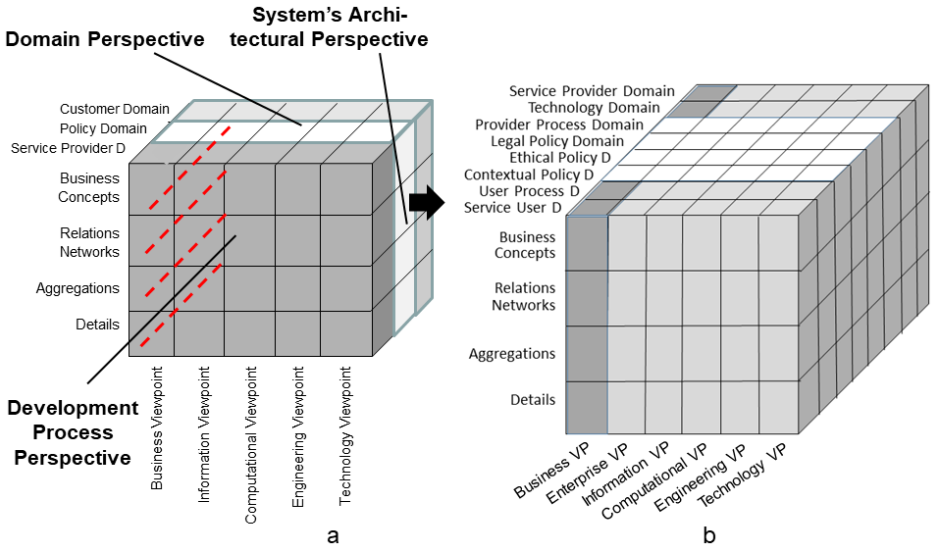


Figure 1. AIS representation, design and implementation enabling advanced interoperability and integration acc. to ISO 23903

6. Conclusion

Artificial and intelligent systems have a huge potential for strengthening the delivery of health and social services everywhere and anytime, achieving universal health coverage including low- and middle-income countries, but it also poses risks to global health and wellbeing. So it can cause or strengthen digital divide between rich and low- and middle-income countries, but also in developed countries. This could be reasoned by available resources, but also by gender, geography, culture, religion, language, or age. For guaranteeing the promised benefits for public health and medicine when designing, developing and deploying AIS, moral and ethical considerations as well as human rights must play a central role [30].

P5 medicine without artificial intelligence and autonomous systems is not feasible, but we have to advance ourselves to do this right. Hereby, objectives, basic principles, limitations, etc., must be carefully considered and defined in their economic, social, political and environmental context.

Innovations in science and technology are always bound to new social, moral and ethical challenges [48]. With AIS, we can do good things, but also wrong things at all, faster and stricter. Ethical issues could be the misuse of personal information or misinformation and deep fakes, but also lacking oversight and the acceptance of responsibility. Moreover, advanced neurotechnology could change behavior or thought patterns, affecting privacy and dignity. Genetic engineering could overcome damaging

genetic mutations, but also create new pandemic viruses. While possessing great potential for human health and the recovery from damaging genetic mutations, there are considerable ethical considerations that surround the editing of the human genome. A crucial aspect of new technologies is they weaponization, as discussed in the context of killer robots and combat drones. Ashley Watters raised the question: "At what point do we trust our technology to fight a war for us?" [48].

Our social, moral and ethical decisions are strongly impacted by our underlying value system. Current core aspects are individual freedom with a tendency to ethical egoism, materialism, profit orientation. Contrary to other domains such as manufacturing, trading, consumption, etc., which can be managed with market economy principles such as supply and demand, profit maximization by cost-benefit minimization, etc., realized according to opportunities and choices, the request for health services is usually defined by objective needs. The global health coverage needs contradicting the aforementioned wrong principles. Health services should not be first seen as business opportunities, but as responsibilities, care and duties practiced. The market- and profit-driven global economy puts us in danger regarding security and safety aspects (availability and safety of products), clearly demonstrated during the recent COVID-19 pandemics. Therefore, the European Union decided to re-organize its economy by reducing dependencies from other regions such as India or China. Another example is the state health and social insurance system as fiduciary duty, introduced in the eighties of the 19th century by the German Emperor as implementation of basic ethical concerns raised in the enlightenment period. Meanwhile, this approach has been partially eroded by opening the field to private players for market orientation and competitions. As another of many further examples from all around the world, the recurring discussion on moves such as Obama Care from an ideological instead of moral-ethical perspective demonstrate misconceptions and lack of virtue and acceptance of responsibility.

Considering the financial pressure health systems face to increase income and reduce expenditures, resulting even in inappropriate diagnostic and treatment procedures just because they are more expensive than traditional practices (e.g. CT and MRI instead of auscultation, or surgery instead of physiotherapy). Recently, the American College of Physicians (ACP) criticized in a paper published in the Annals of Internal Medicine the increasing dominance of the profit motive in medicine [49]. It cites Thomas G. Cooney, Chair, Board of Regents, ACP, with "We need to be sure that profits never become more important than patient care in the practice of medicine."

In consequence, we have to ethically revise and redesign our health ecosystem. A good moral, humanistic, social and ethical, but also transparent and quality-controlled approach to autonomous and intelligent systems in health and social care covers the full spectrum from the individual obliged to act in good faith, serves the society, protect the earth and understands the universe, mirroring the continuum addressed with the P5 medicine paradigm. That way, we would reactivate and advance the humanistic, cognitive and logical principles of the Enlightening period with ethical and moral codes of conduct, thereby establishing core values such as virtue, equality, integrity, solidarity, respect, faith and truth. We have to go clearly beyond the fiduciary duties defining most ecosystems' behavior today. In that context, we could also learn from the Committee on Standards in Public Life in Great Britain, which defined 7 ethical principles [50]. Those principles are selflessness, integrity, objectivity, accountability, openness, honesty, and leadership, acknowledging that principles alone cannot guarantee ethical AI [51].

Learning from our practice and reflecting our principles, autonomous and intelligent systems result in faster and stricter innovations and evolutions of transformed health

ecosystems both in good and in bad faith. It is up to us to reshape our reality to strengthen the benefits and reducing the risks of the new technologies in the ongoing transformation of health and social care systems.

Acknowledgement

The authors are indebted to thank their colleagues from HL7, ISO TC 215 and CEN TC 251 for their kind and constructive support and cooperation.

References

[1] Blobel B, Ruotsalainen P, Lopez DM, Oemig F. Requirements and Solutions for Personalized Health Systems. Stud Health Technol Inform. 2017; 237: 3-21.
[2] Blobel B, Ruotsalainen P, Oemig F. Why Interoperability at Data Level Is Not Sufficient for Enabling pHealth? Stud Health Technol Inform. 2020; 273: 3-19.
[3] Blobel B. Challenges and Solutions for Designing and Managing pHealth Ecosystems. Front. Med. 2019; 6: 83. doi: 10.3389/fmed.2019.00083
[4] Blobel B, Ruotsalainen P. Healthcare Transformation Towards Personalized Medicine – Chances and Challenges. Stud Health Technol Inform. 2019; 261: 3-21.
[5] Makhfi P. Introduction to Knowledge Modeling. MAKHFI.com: www.makhfi.com/KCM_intro.htm
[6] Turing A. Computing Machinery and Intelligence, Mind, October 1950; LIX (236): 433–460, doi:10.1093/mind/LIX.236.433, ISSN 0026-4423
[7] McKinsey. An executive's guide to AI. McKinsey Insights 2020. https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/an-executives-guide-to-ai
[8] Guan J. Artificial Intelligence in Healthcare and Medicine - Promises, Ethical Challenges and Governance. Chin Med Sci J. 2019; 34 (2): 76-83.
[9] General Electric Company. AI in Healthcare: Keys to a Smarter Future. GE Healthcare 2018.
[10] Organisation for Economic Co-operation and Development. Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449, 2019. http://legalinstruments.oecd.org.
[11] Bitterman DS, Aerts HJWL, Mak RH. Approaching autonomy in medical artificial intelligence. Lancet Digit Health. September 2020; 2(9): e447–9.
[12] Defining the Roadmap - The Age of Intelligence: Matching Mind and Machine. AI in Healthcare. Pure Storage, Summer 2018.
[13] Ahuja S, McNamara M. AI in Healthcare – Smart Infrastructure Choices Increase Success. NVIDIA, November 2019
[14] Schmelzer R. Data science vs. machine learning vs. AI - How they work together. TechTarget, Business Analytics, 29 August 2021. https://searchbusinessanalytics.techtarget.com/feature/Data-science-vs-machine-learning-vs-AI-How-they-work-together
[15] Wigmore I. artificial superintelligence (ASI). TechTarget, EnterpriseAI, February 2018. https://searchenterpriseai.techtarget.com/definition/artificial-superintelligence-ASI
[16] Tucci L. A guide to artificial intelligence in the enterprise. TechTarget, EnterpriseAI. 9 July 2021. https://searchenterpriseai.techtarget.com/Ultimate-guide-to-artificial-intelligence-in-the-enterprise
[17] Stedman C. What is Data Science? The Ultimate Guide. TechTarget; EnterpriseAI, August 2021. https://searchenterpriseai.techtarget.com/definition/data-science
[18] Laskowski N. recurrent neural networks. TechTarget, EnterpriseAI, July 2021. https://searchenterpriseai.techtarget.com/definition/recurrent-neural-networks
[19] Behr A. Advancing medicine with AI at the edge. Hewlett Packard Enterprise Company, 11 May 2021. https://www.hpe.com/us/en/insights/articles/advancing-medicine-with-ai-at-the-edge-2105.html
[20] Chernicoff D. What's all the buzz about swarm learning. Hewlett Packard Enterprise Company, 11 May 2021. https://www.hpe.com/us/en/insights/articles/whats-all-the-buzz-about-swarming-learning-2108.html
[21] Doerner H. Knowledge Representation. Ideas – Aspects – Formalisms. In: Grabowski J, Jantke KP, Thiele H (Edrs). Foundations of Artificial Intelligence. Berlin: Akademie-Verlag; 1989.
[22] Fridsma DB. Response of the American Medical Informatics Association on FDA-2019-N-1185; "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning

(AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback". AMIA, 3 June 2019.
[23] DataRobot. Trusted AI 102: A Guide to Building Fair and Unbiased AI Systems. https://www.datarobot.com/resources/trusted-ai-102-fairness-and-bias/
[24] Tzafestas SG. Ethics and Law in the Internet of Things. *Smart Cities* 2018; *1*(1): 98-120.
[25] Ursin F, Timmermann C, Steger F. Explicability of artificial intelligence in radiology - Is a fifth bioethical principle conceptually necessary? Bioethics July 2021; 00: 1-11. doi: 10.1111/bioe.12918. https://onlinelibrary.wiley.com/doi/full/10.1111/bioe.12918
[26] Organisation for Economic Co-operation and Development. OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. http://www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1_1,00.html
[27] Federal Trade Commission, Fair Information Practice Principles. https://www.ftc.gov
[28] Goehring B, Rossi F, Zaharchuk D. Advancing AI ethics beyond compliance. IBM Corporation, April 2020.
[29] The IBM Approach to AI. https://www.ibm.com/se-en/artificial-intelligence/ethics
[30] Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO
[31] Institute of Electrical and Electronics Engineers (IEEE). Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition. IEEE; 2019.
[32] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems-faq-11.22.2020
[33] IEEE Ethics in Action in Autonomous and Intelligent Systems_The IEEE 7000 Series
[34] Japan Science and Technology Agency, National Institute of Advanced Industrial Science and Technology, Council on Competitiveness-Nippon, July 25, 2017.
[35] Bossmann J. Top 9 ethical issues in artificial intelligence. World Economic Forum, Oct 21, 2016. https://www.weforum.org/.../top-10-ethical-issues-in-artificial-intelligence/
[36] Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. https://op.europa.eu/en/publication-detail/-/publication/e0649735-a372-11eb-9585-01aa75ed71a1/language-en/format-PDF/source-205836026
[37] Google. Artificial Intelligence at Google: Our Principles. https://ai.google/principles/
[38] Future of Life Institute. Asilomar AI Principles. https://futureoflife.org/ai-principles/
[39] https://lawrence.house.gov/media-center/press-releases/brenda-lawrence-and-ro-khanna-introduce-resolution-calling-ethical
[40] British Standards Institute. BS 8611:2016 Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems. London: BSI; April 2016.
[41] Blobel B, Ruotsalainen P, Brochhausen M, Oemig F, Uribe GA. Autonomous Systems and Artificial Intelligence in Healthcare Transformation to 5P Medicine – Ethical Challenges. Stud Health Technol Inform. 2020; 270: 1089-1093.
[42] Botelho B. cognitive computing. TechTarget, EnterpriseAI. https://searchenterpriseai.techtarget.com/definition/cognitive-computing
[43] Glasscock J. Cognitive Take on RNA Elevates Biomarker Development, Advances Precision Medicine. Home Magazine December 2019 Vol. 39 No. 12
[44] International Organisation for Standardisation. ISO 23903:2021 Health informatics – Interoperability and integration reference architecture – Model and framework. ISO: Geneva; 2021.
[45] IEEE P7007 - IEEE Draft Ontological Standard for Ethically Driven Robotics and Automation Systems. https://standards.ieee.org/project/7007.html
[46] Almeida MB, Slaughter L, Brochhausen M. Towards an ontology of document acts: Introducing a document act template for healthcare. OTM 2012 Workshops, LNCS 7567, Berlin, New York, Heidelberg, 2012: 420–5.
[47] Brochhausen M, Almeida MA, Slaughter L. Towards a formal representation of document acts and the resulting legal entities. In: Ingthorsson, R.D., Svennerlind, C., and Almäng J (eds.) Johanssonian Investigations. Ontos, Frankfurt, 2013, 120-139.
[48] Watters A. 5 Ethical Issues in Technology to Watch for in 2021. CompTIA, 1 July 2021. https://connect.comptia.org/blog/ethical-issues-in-technology
[49] ACP Says Profit Motive in Medicine May Contribute to a Broken Health Care System. American College of Physicians, https://www.acponline.org
[50] Gilman SC. Ethics Codes and Codes of Conducts as Tools for Promoting an Ethical and Professional Public Services – Comparative Successes and Lessons. World Bank, Washington 2005.
[51] Mittelstadt B. Principles alone cannot guarantee ethical AI. Nature Machine Intelligence November 2019; 1: 501–507.

Invited Papers

This page intentionally left blank

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210569

The International Patient Summary and the Summarization Requirement

Stephen KAY[1]
The Health Informatics Standards Consultancy Ltd

Abstract. The 'patient summary' has an important role in delivering continuity and coordination of a person's health and care. 'patient summary' implementations are pervasive and important to both healthcare providers and to their subjects of care. The digital version of the patient summary, however, often falls short of its intended functionality and its potential value. The requirements of summarization and what they mean for the communication situation in which the summarization of health and care data takes place has been analyzed. The purpose is to understand the limitations and potential of current digital solutions for communicating a 'patient summary'. The International Patient Summary (IPS) standard is a step towards communicating safe, relevant patient summaries for use throughout the world. To meet this grand challenge, the IPS can capitalize upon the inherent capacity and competence of all people to produce and consume summaries.

Keywords. IPS, Standard, Summarization, Patient Summary

1. Introduction

The 'patient summary' plays an important role in the healthcare domain. Patient summaries are used by all specialties and for all health conditions, from emergency to elected care situations, and consequently they are 'present, appearing, or found everywhere' in all types of oral, written, and digital communications within healthcare. Its pervasive nature partly explains why patient summaries are so attractive as a health informatics application. A central well-known concept, and a deceptively simple one at that, sees variants of the patient summary implemented in most, if not all provider systems. The implementation of a patient summary is still contentious, which detracts from its value as a shared and safe resource for healthcare providers and subjects of care alike.

The International Patient Summary (IPS) Standard [1] (ISO IS 27269:2021)[2] is a reference standard for describing and defining the data that a patient summary might contain. The expectations placed on the IPS standard and the importance of understanding the summarization requirement for its sustainability and success are considered. The main thrust of this paper posits the importance of summarization in general for everyday communication before considering its specialized application to the patient summary, to the healthcare domain, and to the applications of Digital Health.

[1] Corresponding Author: Stephen Kay, THISC Ltd., Manchester, UK; E-mail: hi.standards@gmail.com.
[2] ISO IS 27269: 2021 is the fast-tracked version from CEN's EN 17269 [2] that was published in late 2019. This step was taken to ensure that the global requirements were taken into consideration. The process improved the original document by applying editorial comments whilst retaining the same scope of the original. ISO IS 27269 is the single international standard for the reference data model superseding EN 17269.

2. Patient Summary Concerns

2.1. The Problems of a Non-standard Patient Summary

Patient summaries are not overly complex. They use a relatively small amount of healthcare data to complete a defined purpose. Nevertheless, despite this form of communication being common, and the amount of data within the summary being miniscule (c.f., even the smallest of Big Data applications), it has been extraordinarily difficult to reach agreement about what data should be included and what should be discarded from any standard 'Patient Summary'. The purpose of the 'patient summary' is essential for determining what data is relevant for any given situation. The lack of consensus about its purpose, however, has resulted in many, non-standard variants of patient summaries being generated and deployed.

Although non-standard summaries can be sufficient for local needs, it is recognized that these have serious limitations when a patient summary is required to be shared outside of its original context. Different specialists, different conditions, and different purposes, all of which, no doubt, are important, all vie for attention; their many advocates argue for other data to be included within 'their patient summary' either as additional data or as a replacement for data that they regard as less important. This makes any agreement difficult to reach regarding the patient summary's definition and complicates the intended use, making sharing difficult and undermining the clinical needs for brevity and relevance. These problems are significantly amplified when agreement is required to exchange healthcare data cross-border, where different contexts may require a different set of policy rules, legal regulations, cultural and ethical requirements to be taken into consideration.

2.2. Multiple, Eclectic Stakeholders

Multiple stakeholders engaging with a problem is usually a good thing, but in the case of the 'patient summary' the stakeholders' actions have tended to make any solution more difficult than it should be. This is surprising given that a good quality 'patient summary' is held to be in everyone's interest. Governments, healthcare providers, vendors and, not least the individual person in need of care, attest to the importance of the patient summary. In the formative eHealth Network (eHN) guideline, a patient summary is described as an identifiable "dataset of essential and understandable health information at the point of care to deliver safe patient care during unscheduled care [and planned care] with its maximal impact in the unscheduled care". [3]

The lack of agreement about the content of a 'patient summary' and the proliferation of multiple agreements necessary to achieve local success, damage the goal of achieving sustainable interoperability within the healthcare domain. Consequently, the quality of shared data needed for effective continuity of care is put at risk. Any attempt to address these communication problems, however, is exacerbated by the reluctance to make changes to an artefact that plays such a central role within the healthcare domain.

Some of the problems can be laid at the door of the Digital Health 'solutions' on offer, as they focus to manage the diversity of need by meeting each local requirement. The considerable investment made in such applications can also obstruct any necessary change, even if that means the adoption of an open standard. Stakeholder' resistance can arise when alternative ways of working need to be considered and these can be both daunting and costly.

It has been said that "digitizing effectively is not simply about the technology, it's mostly about the people." [4] A truism, but perhaps one which glosses over the fact that 'people' have many different roles to play, e.g., the stakeholders are vital in recognizing the need, in setting the requirements, in reaching consensus about what and how to implement, as well as being the ultimate beneficiaries and critics with respect to their own life and care. Unfortunately, the wide range and large number of existing interested parties offer no guarantee that any standard for a patient summary will be even known by them let alone adopted.

If we assume the stakeholders are doing their level best to share essential clinical data, yet failing, then it follows that the simple concept of a 'patient summary' is probably more complex than first impressions would suggest. Furthermore, the complexity of the ecosystem surrounding the patient summary mounts a considerable technical challenge, albeit one with the potential for immense social and economic benefit. The 'patient summary', however, is not just "mostly about the people"; it is a special case of summarization, which is at the core of what it means to be human. This realization gives substantive hope for achieving a sustainable, global standard with relevance to the patient summary.

2.3. The Expected Value of a Standard 'Patient Summary'

Standardization is a generic consensus process that should help to realize a common purpose, agreed content, and coherent semantics of a standard 'patient summary'. The process is intended to facilitate formal agreements between stakeholders; the outputs from such agreements are, by definition, consensus products. The main objective of open standardization is to simplify matters; to produce just a few agreements, better still to produce a single agreement, to replace the proliferation of numerous ad hoc ones.

The advantages of standardization are many and obvious; they include better governance opportunities, savings, reduced effort, elimination of wasteful duplication, fewer interfaces, and easier maintenance. Unfortunately, there is also a well-known downside; too many standards can be incompatible, conflicting, and competing, they can be too complicated or impenetrable, and are often slow to develop. Standardization Development Organizations (SDO) often fail to identify and engage with the relevant interested parties. They tend to overlook the powerful, wider audiences beyond just the immediate and technical one, and thereby inevitably hinder their work's future acceptance and adoption. An on-going challenge facing the new IPS standard is to maximize the upside of standardization whilst minimizing the downside.

CEN/ TC 251 [5], a regional, European SDO, was contracted by the EC to make an existing guideline for exchanging patient summaries [6] an international standard. Specifically, the initial standardization focus was to normalize the embedded dataset outlined in that guideline. The 'International' prefix was chosen for the Patient Summary (IPS) to strengthen its claim to be something new, targeted at a global scope, to be the epitome of healthcare without borders. The new would-be 'global' IPS standard was positioned to be that standard, one which would solve urgent local and national needs as well as international ones.

2.4. Managing Expectations

This inclusive yet global aspiration contributes to the current interest around IPS. Whilst helpful, it was initially just an 'aspiration' and, by itself, naming is not sufficient to

stimulate engagement or convince stakeholders that it can succeed. From the outset, there was a need to differentiate the proposed standardization of the 'patient summary' from existing offerings. It was (and still is) important to offer something 'new', and to avoid the fate of being dismissed as just one more patient summary amongst many.

However, the history, and the pervasive and ubiquitous nature of the 'patient summary concept' would seem to defeat any attempt to describe IPS as being something 'new'. Kleppe, in her book on Software Language Engineering [7], remarked how "deceptive the difference between the old and new is". Kleppe quotes,

> "(to) make something that was not there before, … is deceptive, because the separate elements already existed and floated through history, but they were never before assembled in this manner. Joining, assembling, the new will always consist of that." [8].

This dual emphasis, on 'purpose' and 'assembly', can be applied to the IPS and serves the standard well; it substantiates the claim of being 'new' and, more importantly, indicates how it will work and be developed in an incremental fashion to be usable and useful.

The commitment of the SDOs is to implement a fully sharable, single solution for the patient summary, one that provides efficient and relevant communication. The decision to base the standard upon an existing guideline on patient summary exchange helped, as did the decision to focus the initial scope of the standard on just the dataset in the guideline [9]. The success of this standardization initiative, however, cannot just be attributed to these decisions, important as they were to its development and important as they remain for its progress in the future. Ironically, it seems that offering something 'new' in 'patient summaries' may largely be attributed to the past (i.e., 'floated through history') and this history is of fundamental importance to the success of the IPS standard going forward.

The IPS unconsciously mimics and utilizes what is known about the way humans communicate with each other through summaries; this human competence in summarization precedes the idea of a computer-based application and even the paper-based precursors used within the health and care domain. The following quote taken from Winston Churchill [10] seems especially applicable to the patient summary and the requirement of summarization, "The farther back you can look, the farther forward you are likely to see."

It is suggested here that a better understanding of the 'summarization' act and its results, i.e., the 'summaries', will provide stakeholders with a rich source of foundational requirements to confirm, to challenge, and to improve the present IPS. In part, this review explains the success of the standard's development and offers a promissory note to support its future dissemination, its adoption, and its use. The underlying foundation of the summarization requirement offers the hope that the IPS will be sustainable, making the project's earliest aspiration a reality.

3. Summarization in Everyday and Professional Communication[3]

Summarization is an integral part of both every day and professional types of human communication. Professional summarization in healthcare, and especially the 'patient summary', are considered here to be domain-specific specializations of the more general concepts.

Figure 1 illustrates some fundamental aspects of generic 'summarization', including the summary production and consumption, the principal actors involved in the process, the summarization situation, external input(s) and output(s) and the summary object itself; all of which are relevant to the healthcare domain and are clearly within the scope of Digital Health applications.

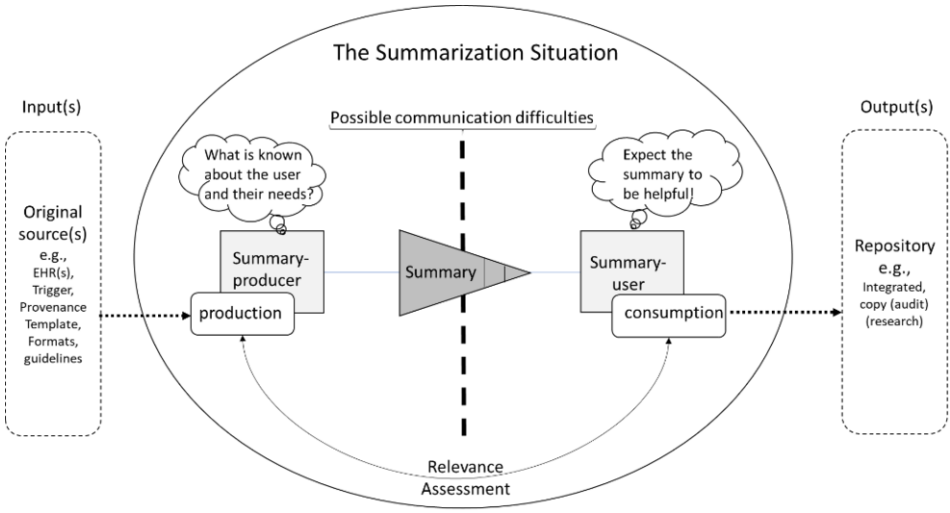


Figure 1. A Simplified Model of Summarization[4]

3.1. Summarization and the IPS

'Summarization' is a practical skill, an integral and versatile part of human communication. As Endres-Niggemeyer comments in her introduction of *Summarizing Information* [11],

> "We all summarize, very often, when reporting about the movie we saw yesterday or the negotiations during a meeting, recoding an accident, or wring a resume' of a stage play at school. Everyday summarizing skills belong to everybody's communication competence."

[3] Section 2 of this paper owes much to Endres-Niggemeyer B. Summarizing Information. Springer-Verlag, Berlin Heidelberg, 1998 [11].

[4] In Figure 1, the single, vertical dashed line represents one or more barriers that the Summary must go through. The barriers may impact the quality of the summary produced, affecting the usefulness.

This human competence is motivated. Sperber and Wilson claim in their book on "Relevance, Communication and Cognition" [12], "that all human beings automatically aim at the most efficient information processing possible. This is so whether they are conscious of it or not." [12, p49].

'Summarization' is a good example; indeed, it may be the best example of such efficiency-driven, unconscious behavior in humans. In essence, it is the reduction of information to its most essential points; it retains what is relevant and discards the irrelevant for the purpose of effective, efficient communication.

The core requirement of summarization is to concentrate on the important points. However, determining what is important enough to include in the summary is non-trivial. What is deemed to be essential or important, or conversely what is deemed inessential or unimportant, is coupled to the idea of relevance assessment and this is part of both the production and consumption processes of the producer and user of the summary.

Summarization has always been deployed in the healthcare domain, i.e., from the first time it was necessary to report, document and/or share clinical data and "to express the most important facts or ideas about something or someone in a short and clear form" [13]. This summarization act goes largely unnoticed but is integral to all note taking, the quality of the note reflecting directly on the training, experience, and professionalism of the clinician. The patient's longitudinal record then is an aggregation of one or more 'summaries' of this kind entered into notes taken from clinician/patient encounters.

The IPS can be regarded as a summarization function applied to the known healthcare history of the patient. IPS defines the core dataset for a patient summary. It uses the eHN guideline as the initial source of requirements but takes into consideration other international efforts, providing a dataset specification for global application. IPS provides an abstract definition of a Patient Summary from which derived models can be made and assessed as being conformant and interoperable.

The IPS, Dataset and the associated business rules, is a norm for what data is required in a patient summary, but the first iteration of the standard explicitly excludes details of workflow and therefore does not detail the summarization process itself. The IPS standard openly states what is not in its present scope:

> "This document does not cover the workflow processes of data entry, data collection, data summarization, subsequent data presentation, assimilation, or aggregation. Furthermore, this document does not cover the summarization act itself, i.e., the intelligence/skill/competence that results in the data summarization workflow." [1, Scope clause]

Except for the first ever interaction with the subject of care, a patient summary will probably be produced with reference to some pre-existing source or sources of existing patient information. This may include extracts from one or more EHRs, possibly clinical guidelines, templates/formats, and even other local patient summaries if relevant. The inputs would also include information about the trigger-event and the agent responsible for starting the process for producing the summary (e.g., an information request from a clinician or a patient managing their own care). This data may be directly available from the sources or part of the provenance meta-data required for any health data exchange.

The main output is the patient summary that is produced within the summarization situation, but there may also be related outputs, that might be considered as being secondary. A patient summary might be only used within the immediate communication

situation, but it is probable that it will also be stored as a persistent copy, at least for the purposes of audit if not research. The extent to which the content can or should be retained in a patient's record or integrated into the recipient's system itself varies and is subject to policy, regulation, and ethics as well as the technological sophistication of the information systems used.

It is possible too that Digital Health applications [14] may include the end-to-end sources as being part of a single communication situation. In that case, the scope of interest will include the Summarization Situation in Figure 1 and the associated inputs and outputs. The computer-based information systems are pervasive, providing an indispensable framework and tools that summarizers use. These information environments impose fixed roles, particularly if producer and user of the summaries are physically remote. The functionality and sophistication of these tools will be of paramount importance to the quality of the patient summaries in the future as efficient health information systems are a major objective of Digital Health. Further on, Summarization, whilst remaining a 'person skill', is likely to be assisted by the application of Artificial Intelligence within the clinical systems of the future.

3.2. The Summarization Situation and the IPS

The 'communication situation' is the context in which the summarization takes place; it is also especially important to the outcome of a summarization effort. Summarization is bound to the communication situation which brings the necessary communication parties together for some purpose:

- The interaction may not necessarily be face to face. For the primary patient summary use case that requires a patient summary to be accessible at the point of care, the parties will generally be location-remote to each other and if so, they will almost certainly be producing/using summaries in different contexts with different constraints dependent on their circumstances.
- A patient summary can also be produced at any time if there is no urgency (for example, a person requesting a recent summary before going on a future business trip), but a more urgent request, perhaps due to an accident, will determine whether timeliness is an important requirement for the consumer.

Summarization, then, is an example of situated communication, wherein time and space can be stretched, and what is possible may reflect the sophistication of the technology being used. Physical distance, for example, often precludes a real conversation and consequently the parties will rely on technical media to send and receive the summary using a messaging paradigm. In these cases, the information system provides a framework for their work and can be part of the production (e.g., the reduction/condensing activities), the exchange, and the consumption (e.g., the presentation/use) of the summary.

The Summarization Situation is particularly important to the outcome of a person's treatment. In one of the IPS scenarios, that of required but unscheduled care, the situation is likely to be urgent and may even be life threatening. This pressurized situation can impact both the summarizers, but it is particularly stressful for the Summary-user. Endres-Niggemeyer explicitly calls attention to the burden on the recipient who is "expected to pick up the content, to restructure it with respect to their own prior

knowledge, to integrate it into their own knowledge structure and finally to use it" [11, p86].

Furthermore, summarization situations may cause communication problems that affect the quality and usability of the summary; some of the communication difficulties are shown in Table 1. These difficulties are generic and are not healthcare domain specific [11]. This congruence gives hope of wider problem-solving collaborations beyond the healthcare domain, which may be mutually beneficial.

Table 1. Summarization situations may cause communication problems

1.	Disturbance from the situation, related perhaps to the urgency of the request on the producer or the stress for the user to assimilate the summary content. The recipient may be unable to determine chronology and timeliness of data content provided.
2.	A lack of common shared knowledge or consensus between the parties on how to handle each other and/or the transferred knowledge. If for some reason the shared, common background becomes too weak, communication problems will increase. With respect to patient summaries, the Summary-user may be unable to establish confidence in the trustworthiness, accuracy, and integrity of data content. Furthermore, specialised conditions, such as Rare Diseases, may require expertise that the Summary-user does not have.
3.	Content related inadequacies, perhaps due to terminology or structural differences. For example, unable to verify context including vital inter-relationships between clinical data content. Given that a patient summary will be a snapshot, extracted from one or more source documents at a point in time, it is vital that the original context is not lost, to ensure faithful and safe communication.
4.	Frustrated reader expectations, because of differences, between how the parties assess relevance. For example, the amount of data received may contain too much to take in, with too much irrelevant data being exchanged (information overload). This may be critical in unplanned care scenarios, where the patient is likely to be a complete 'unknown as far as the attending clinician is concerned.
5.	Insufficient adaptation to the use situation. Policies may not exist to establish confidence in externally sourced content sufficient to allow such content to be fully integrated into their local health record and instead must keep it segregated, managed, and accessed separately.
6.	Interoperability and technology filters; The Summary-user may be overwhelmed by a plethora of conflicting and/or duplicative fragments of data from many sources (information overload).
7.	Different contexts, such as attempting to communicate effectively between different cultures. Differences in usage between unplanned and planned care.
8.	Boundaries and borders, and the associated rules, can impact the summarization process and the content of the summary. Cross-border exchange may mean that the healthcare provider as the Summary-user finds themselves unable to ascertain provenance of data content and discrete elements.

Digital Health solutions will need to counter the socio-technology problems in the Communication and Summarization Situations. The IPS as a reference model of the content can assist by supporting conformant implementations that should be interoperable, but much will depend on the infrastructure and sophistication of the technology in use.

Provenance[5] is a key element to support 'trust' and for providing confidence in the data to be used. Common, shared understanding, consensus and knowledge can be strengthened, and this will help to overcome some of the other problems associated within the complex Summarization situations in the healthcare domain [15].

3.3. The Summarizers and the IPS

The terms 'Summary-producer' and 'Summary-user' are used here to distinguish their roles in the communication situation, albeit both parties are capable of producing and consuming summaries[6]. In the healthcare domain, the Summary-producer and Summary-user would typically be healthcare providers, i.e., healthcare organizations and healthcare professionals [16]. Increasingly, the Subject of Care (SoC) may also have a role to play in validating and/or adding a personal perspective on their own story.[7] If the SoC becomes the mediator, taking response for the summary exchange themselves, many of the existing consent issues, for data and infrastructure to safeguard the summary, are greatly simplified or simply not required.

In the most general case, it is the Summary-producer who determines what goes in or what is left out of the summary. With the application of computers to healthcare, support can also come from third parties (e.g., clinical guidelines from professional bodies [17]), in the form of predefined templates that constrain what is summarized by the Summary-producer. These inputs may have the authority of a de jure standard of international standing or be non-standard, operating exclusively at the local level. Furthermore, the situation in which a summary is to be used is often not precisely known by the Summary-producer, nor the details of what the user requires. Even though the summaries are usually produced for the recipient, it is not necessarily the case that the user will have prior knowledge about the content. The Summary-user is hopeful that the received summary will help them in their current situation. In general, the Summary-producer will not know the specific situation that the attending clinician(s) face, nor the precise problem(s) besetting the patient.

The recipient of the content may be the intended user of the summary, but there is no guarantee that will be the case. In scenarios, where the summary is the result of a health information request, the recipient is likely to be the intended user, but that still means the content might provide surprises for the recipient given the degrees of freedom of a non-standard summary. Either way, it would be wrong to think of the Summary-user as being just a passive consumer of information. The value of the summary to the actual recipient may be judged by "looking at the incoming information and its semantic structure, in relation to the communication needs, or considering it from the viewpoint of the Summary-user" and "Summarizing is an intelligent skill, roughly comparable to translation." [11 p47].

In the first IPS scenarios, it is the would-be recipient or Summary-user that requests or initiates the patient summary. The assumption is that this is an IPS-on-demand,

[5] Provenance is a common requirement for all health data exchange, and ideally the common part should be a separate standard that the IPS can use rather than be embedded within the standard as it is in the first iteration. It may however be necessary to specify summary-specific provenance data, and this will need researching.

[6] These general terms also allow AI and human agents to be deployed without discrimination. Also, Endres-Niggemeyer uses the term *Summary-consumer* rather than *Summary-user*, which is the term used in this paper to better align with Digital Health.

[7] In traditional Digital Health settings, it is usual to consider summarization as an interaction between clinical actors. However, increasingly the patient as a Summary-user, and to a lesser extent, the patient as the Summary-producer may have a greater role to play in this type of communication in the future.

although other scenarios can be envisaged such as chronic patients or patients with rare diseases whose planned care may require communication between specialists without the patient in attendance.

The Summary-producer acts on the request and may use the IPS as-is, as a source, as a template or checklist for what to say, as a framework to add annotation to the content. In addition, specialized content not defined by the IPS, may usefully extend the IPS before the summary request is answered. The Summary-user is the attending clinician(s)[8] and depending on the time (i.e., scheduled/unscheduled), the location (i.e., point of care) and the patient state (e.g., conscious, or not), the IPS may be the only relevant healthcare information available to support the treatment of an unknown patient.

Digital Health can play an important role, in the production, exchange, and consumption processes of the patient summary. The summarizers too are reliant upon technology to send, receive and/or present the patient summary, as typically the two parties are in different geographic locations. Further ahead, Digital Health may well consider computational agents as assistants if not replacements for the human Summary-producer and the Summary-user. The Summary-producer would be the earlier and probably the easier option to automate in any given Summarization situation. The advanced technology environment will lead to much more machine-readable data being present and available to the computerized information system.

The advent of summarization assistants does not make the idea of the IPS obsolete as the IPS Dataset will help any agent, human or computational, to select the relevant data for exchange conforming to its headings and business rules. Perception and knowledge of the received content, likely to be the standardized data from the IPS together with the non-standard extensions, would be much more difficult to automate the Summary-user, who, as noted, is more than a passive recipient of information. These future options may sound far-fetched but the requirement for information extraction tools to assist automatic summarization in the healthcare domain will be driven, in part, by the disparity between the diminishing number of healthcare professionals relative to the increase in the world's population [18].

3.4. The Summary and the IPS

Information from any representation can be summarized. Summarization is versatile and can be multi-media. However, the output summary does restrict itself to conveying important information, and in the patient summary especially the data will necessarily be much more constrained and focused than summaries in more general every day communications.

The original input(s) or information source(s) are reduced in the final summary but that does not stop original content from finding its way to the summary; indeed, the information source(s) can determine the main content of the summary:

> "When the source is a well-organized document, the information organization in the summary should conform to the original presentation … If the source information is not well organized, there is a tension between faithfully reflecting the source and producing a well-structured summary." [11, p47]

[8] The Summary-user may be a multi-disciplinary team.

Even so, the intent is that the final summary is usually reduced in size from the original source to be more easily and quickly assimilated. Summarization provides a filter and must strike a balance between offering too little or too much data to the Summary-user. "The source information or input may become important in a follow-up situation, however, when the users want to know in more detail and first-hand what the summary has told them in brief" [11, p47].

The amount of information in a summary is not fixed and can vary (represented by the graduated triangle in Figure 1). Typically, a small summary with few words will convey less information than a larger one all things being equal, but relevance decisions and the needs of the Summary-user to manage the summary are paramount. A single term, perhaps a scientific fact or a labeled diagnosis that carries weight with the Summary-user, may convey much in a very condensed form.

In everyday communication, people often vary what they present in their summarizations, saying the same things in different ways to evoke interest. In professional summarizing, the motivation is different and variety in presentation is discouraged. The variety expressed, or maybe the lack of it, will be mainly determined by the sophistication of the medium being used and the tools that are available to help the clinicians with their tasks.

Typically, a patient summary implementation will support the healthcare provider by limiting redundancy, avoiding repetition, and avoiding a variety of presentation, to convey condensed information. Summaries have a motivation and a goal and are valued for their utility; however, their functionality and success will be ultimately assessed on their ability to inform or help the Summary-user to solve problems.

The success of a summary does not just depend on its content alone, but also on the communication convention, i.e., the Summary-producer and the Summary-user must have a common basis, a largely consistent knowledge base about the intent of the communication for it to succeed. The IPS Dataset contributes to this; one of the original, key principles for the IPS is that core data should be easy to understand by all clinicians, and therefore the content is always to be generally applicable and specialty agnostic, in as far as that is possible. The 'planned care' scenario of IPS, however, may mean that a patient summary for that purpose includes specialist knowledge for example with respect to Rare Diseases and this might require the creation and use of optional IPS Sections related to specialist conditions, which may not be relevant or even understood by clinicians in different communication situations.

The IPS is a specification of a patient summary but many of the properties of a 'summary' are applicable to the IPS itself. For example, it aims for a concise representation and indicates what is the most important data to be processed and what is left out. It can specify multimedia data and the resulting summary can be represented in many forms. The IPS is implementation independent. Whilst IPS uses the common 'document' metaphor to explain its content, it does not require any conformant implementation to represent the data as a document. For example, the actual data can be represented in a messaging paradigm such as a document (e.g., CDA [19]) and as a set of resources (e.g., HL7 FHIR [20]). Furthermore, the IPS may serve as a dashboard, an aide memoir, as a library or source for reusable data blocks, and even be incorporated in its entirety within another type of summary document, e.g., within a discharge summary[9], although it is unlikely that a full summary will be required. However, for any IPS conformant implementation, the purpose of the IPS must be immutable (i.e., it provides

[9] But note the IPS is not a Discharge Summary, which has a distinct and different purpose.

a formal snapshot of the patient's longitudinal data) and it specifies the IPS data conforming with the associated, defined business rules. Other recommended data are also defined in the standard as being a legitimate option for inclusion in the IPS, but they are not mandatory or necessary for every instance.

How the summary is consumed depends greatly upon the situation and the requirements of the Summary-user. It may be entirely consumed in situ and then discarded, or it may also be retained in some persistent form. Endres-Niggemeyer noted the difficulties of identifying the target user, commenting that at that time, "there is no global definition of summaries that are suited to groups of people or individuals in specific use situations." [11, p48]. The International Patient Summary (IPS) Standard is perhaps the first global definition of a summary suited to healthcare providers for specific care situations, albeit it is still in its infancy, and that hypothesis must be tested.

There are still the inevitable disagreements concerning the choice of data blocks and elements within the current IPS Dataset. However, the IPS Dataset standard now provides a global focal point, one that is tangible, subject to greater scrutiny and is flexible, it can be changed by consensus as required. It is some ways on from a blank sheet of paper; it is now a de jure standard and starter set that can provide consistency and an opportunity to manage change in a coherent way, gaining agreement concerning the contents meaning and use.

As with Summary-production, Summary-consumption may also be helped by the information system used by either Summarizer. Personal documents such as a diary, or professional ones such as an engineering logbook, or even medical records are not expected to capture the totality and substance of what is communicated; even if it were practical to record everything, the result would be unhelpful if not completely unmanageable. The typical output from a consulting room visit, for example, is usually just a partial representation of the whole encounter between a clinician and a patient. That is not to denigrate the result in any way, but rather to emphasize the author's intent and skill, enabling them to make a clear record of the important points whilst omitting the unimportant detail.

The amount of information in the IPS is not fixed; structure, content and associated rules are defined within the standard, but the size and volume are not explicitly prescribed. Implicitly however, the intended scope of IPS is to provide the Summary-user with a concise summary. This utility would be one of the main objectives of IPS given that the Summarization Situation is one that may require the recipient to more easily absorb the summary information to treat the patient in the quickest and in the most effective way as possible.

It is possible that the Summary-producer might be required to send all the data that they had as a safety precaution (that is, unless they had some knowledge of stricter requirements). However, this begs the question as to why not send the complete EHR rather than an extract?

3.5. The IPS and the EHR

Figure 1 shows one or more Electronic Health Records (EHRs) as being important, but distinct sources for a patient summary. The IPS should not be regarded as an EHR; it is not a full-blown longitudinal record. IPS is a point in time extract of the important parts of that record, discarding the inessential. IPS is, in general, a much smaller artefact. The European project for exchanging health records across Europe reflects the difference in scale [21]. Consequently, it is also simpler, one that offers a much reduced, condensed

form, which is generated explicitly by request and purposed on delivering relevant data for a patient's safe continuity of care.

Although the IPS is extensible by design [22], the extensions are not intended to create a full EHR. In some cases, where no national or organizational EHR exists, the IPS data definitions can be used as a starter template from which an EHR could in theory be constructed. It is important, however, to emphasize that a full health record application would require much more functionality and data than the IPS and, to restate, it fulfils a fundamental different purpose. If IPS needed another reason for being created, it would surely be to mitigate the burden placed on the EHR would-be user, for example a front cover or dashboard for a record, that may be too large or poorly structured to be of much use to the clinician with no prior knowledge of the patient demanding treatment.

The IPS standard uses the ISO definition from a 2009 technical report [23], i.e.,

> "Health record extract comprising a standardized collection of clinical and contextual information (retrospective, concurrent, prospective) that provides a snapshot in time of a subject of care's health information and healthcare"

The definition makes it clear that the IPS is an 'extract', but an extract can also be a 'summary'. ISO 13940:2015 defines the 'electronic patient summary' as an "electronic health record extract containing essential healthcare information intended for specific uses." [24]. Humans are wired to communicate efficiently, and summarization is one of their key competences to achieve this aim. It is part of the clinician's workflow to summarize, to produce and consume summaries and the summary, as filter, should be expected to ease some of the information burden on clinicians in these tasks.

4. Conclusion

This paper re-enforces just how foundational the summarization requirement is to the healthcare domain as a whole and especially to the 'patient summary'. To summarize is a fundamental requirement within healthcare and the aim of the IPS standard.

Summarization finds ways of representing the important, relevant facts from a patient's entire healthcare history in an efficient manner. The pervasive nature of summarization and its underpinning of what is a basic human competence, lends credence to the IPS ideal of one single standard solution being feasible across the globe.

The value of data, however, is found in its use and, furthermore, "data by itself has no value. It's the ever-changing ecosystem surrounding data that gives it meaning" [25]. Summarization, at some level, underpins all clinical communication, regarding every patient, whatever their health condition, wherever and whenever! The summarization of patient-level information is still a challenge. Patient Summaries can be considered as being clinical tools, and of providing a basic level of clinical decision support (CDS, 2008) [26], one that might be able to productively use the existing free text portions found in today's records:

> "The CDS challenge is to intelligently and automatically summarize all of a patient's electronically available clinical data, both free text and coded, and to create one or more brief (e.g., 1–2 page) synopses of the patient's pertinent past medical history,

current conditions, physiologic parameters, and current treatment(s)."

The IPS is a late start, and a small but necessary contribution towards that goal, providing the data model standard for an IPS to be used for planned and unplanned care and for local and cross-border use. The IPS Datablocks will also provide the reusability to support other applications in a more coherent and consistent fashion. The on-going challenge is to make the IPS and its implementations as optimal and effective for information sharing within healthcare as everyday summarization is for humankind.

References

[1] International Organization for Standardization. ISO 27269:2021, Health Informatics: The International Patient Summary. ISO, April 2021.
[2] European Committee for Standardization. EN 17269:2019 Health Informatics: The International Patient Summary. November 2019.
[3] eHealth Network (eHN). Guidelines on Minimum/non-exhaustive Patient Summary Dataset for Electronic Exchange in accordance with the Cross-border Directive 2011/24/EU, Release 2, 21st November 2016.
[4] Cadzow S. Wood S. The argument in favor of eHealth standardization in ETSI, White paper, 29, September 2018.
[5] CEN TC 251 website, https://www.ehealth-standards.eu/
[6] CEN and European Commission. Specific Grant Agreement SA/CEN/GROW/EFTA/000/2015-6 for developing an International Patient Summary standard. 2015
[7] Kleppe A. Software Language Engineering: Creating Domain-Specific Languages using Metamodels. Addison Wesley, Boston, 2009.
[8] Palmen C. Lucifer, Prometheus Publishers, Amsterdam, 2007, p.261.
[9] Cangioli G, Kay S. The International Patient Summary Standards. HL7 Europe Newsletter, May 2018.
[10] Churchill W. 50 Sir Winston Churchill Quotes to live by BBC America Editors, April 9, 2015
[11] Endres-Niggemeyer B. Summarizing Information. Springer-Verlag, Berlin Heidelberg, 1998.
[12] Sperber D, Wilson D. Relevance, Communication and Cognition, 2nd Edition. Blackwell, Oxford, 1995.
[13] https://www.lexico.com/en/definition/summary. Accessed: Ma, 2020.
[14] epSOS. Smart Open Services for European Patients - Open eHealth initiative for a European large-scale pilot of patient summary and electronic prescription. http://www.epsos.eu/uploads/tx_epsosfileshare/D3.2.2_Final_Definition_Functional_Service_Req_Patient_Summary.pdf
[15] Dickinson G. Clinical Burden. Report to ISO Fast-track process, HL7 International
[16] International Organization for Standardization. ISO 13940: Health informatics — System of Concepts for Continuity of Care (ISO 13940: 2015)
[17] Ritz D. IPS for Computable Clinical Guidelines (CCG). YouTube video for HIMMS 2020.
[18] Cadzow S, Wood S. The argument in favour of eHealth standardization in ETSI. ETSI White Paper No.29. Sept. 2018
[19] HL7 CDA IG. http://www.hl7.org/implement/standards/product_brief.cfm?product_id=483
[20] HL7 FHIR IG. https://build.fhir.org/ig/HL7/fhir-ips/
[21] EC. X-eHealth Exchange, https://ec.europa.eu/digital-single-market/en/exchange-electronic-health-records-across-eu Accessed: May, 2020.
[22] Kay S, Cangioli G, Nusbaum M. The International Patient Summary Standard and the Extensibility Requirement. Stud Health Technol Inform. 2020; 273: 54-62.
[23] International Organization for Standardization. ISO/TR 12773-1:2009 (en) Business requirements for health summary records — Part 1: Requirements. ISO 2009.
[24] www.contsys.org. Electronic Patient summary.
[25] Borgman CL. Big Data, Little Data, No Data: Scholarship in the networked world. MIT Press, 2015, p3.
[26] Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, Campbell E, Bates DW. Grand challenges in clinical decision support. Journal of Biomedical Informatics 2008; 41: 387-392.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210570

Communicating About Mortality in Health Decision Support: 'What and Why and When, and How and Where and Who'

Jack DOWIE[ab1], Mette Kjer KALTOFT[b], Vije Kumar RAJPUT[c]
[a]*London School of Hygiene and Tropical Medicine*
[b]*University of Southern Denmark*
[c]*Stonydelph Medical Centre, Tamworth, UK*

Abstract. The Covid-19 pandemic has only accelerated the need and desire to deal more openly with mortality, because the effect on survival is central to the comprehensive assessment of harms and benefits needed to meet a 'reasonable patient' legal standard. Taking the view that this requirement is best met through a multi-criterial decision support tool, we offer our preferred answers to the questions of What should be communicated about mortality in the tool, and How, given preferred answers to Who for, Who by, Why, When, and Where. Summary measures, including unrestricted Life Expectancy and Restricted Mean Survival Time are found to be reductionist and relative, and not as easy to understand and communicate as often asserted. Full lifetime absolute survival curves should be presented, even if they cannot be 'evidence-based' beyond trial follow-up limits, along with equivalent measures for other criteria in the (necessarily) multi-criterial decision. A decision support tool should relieve the reasonable person of the resulting calculation burden.

Keywords: mortality, life expectancy, survival curves, Restricted Mean Survival Time, Time-to-Event, multi-criteria decision, reasonable patient, preference-sensitive

1. Background

Our setting is the person making an individual health decision, often in conjunction with a healthcare professional, but often at home in the community (where one of their health decisions may be whether or not to contact a health professional). Our underlying motivation and aim is the development of a *generic* decision support *template*, to be the basis of decision support *tools* (DSTs) for *specific* health-related decisions. We take for granted that any health decision is multi-criterial, whether the objective is health promotion or disease prevention - often in the absence of a professional diagnosis - or therapeutic - often after a professional diagnosis. As eponymously indicated, the focus here is on the mortality criterion, a literally vital one, albeit always alongside other criteria such as those relating to morbidity and option burden. These DSTs must facilitate optimal *individualisation* and *personalisation*, in ways to be defined - and hence clearly distinguished.

[1] Corresponding author: Jack Dowie, LSHTM, 15-17 Tavistock Place, London, UK WC1H 9SH; email: jack.dowie@lshtm.ac.uk

This is a huge topic, so we use Kipling's simple taxonomy of 'What and Why and When, And How and Where and Who' to move the discussion along. It will immediately be apparent, however, that the possible answers to all of these are interrelated, indeed in many cases seriously interdependent. Our prime concern is with What and How, as the main determinants of the structure and contents of the decision support template. When, Where, Who, and Why address the background contexts and conditions. Throughout, the proffered answers reflect our *preferences*. Others may, and almost certainly will, have different preferences - and answers - reflecting beliefs and interests different from ours.

2. Who, When, Where

Who for? is the best place to start, since we need to determine and distinguish the parties involved in decision-supporting communication about mortality, and Who splits naturally into 'Who *for*?' and 'Who *by?*'. To us, the communication is *for* the *reasonable person* (RP) making a health decision. A subset of RP's health decisions involve those taken within professional healthcare services, following a decision to engage with them. RP is assumed to require no support in recognising that a health decision is required and acknowledges the mortality implications of their health-related options must be considered. They will therefore positively request and/or expect this within any decision support.

When is RP making a health decision? They are making a health *decision* whenever, after maximum possible *individualisation*, there is no option that has the best or equal best performance rates on all their decision criteria – 'the things that matter to them'. If there is such a 'dominant' option, there is actually no *decision* to make and so no need for *personalisation* to make it preference-sensitive. Individualisation is the process of refining the *option performance rates* for a criterion on the basis of the individual's characteristics. Personalisation is the process of adjusting the inputs to a decision on the basis of the individual person's preferences, i.e. their value judgements. Our preference is for these to be expressed as their *criterion importance weights*.

So, taking life expectancy (LE) as an example of a possible mortality criterion, in individualisation the estimated LE for an individual is progressively refined from an initially coarse (least individualised) one, based only on their sex and age, through adjustments in the light of other epidemiological characteristics (e.g. family history, ethnicity, location) and with reference to personally observable characteristics (e.g. weight, blood pressure), and, finally, to the most individualised one by adding those determinable mainly by healthcare professionals or services (biological markers, medication history). Personalisation may take two forms. In comparing the LE associated with two options, the importance attached to the extension offered by the superior option may not involve simple linear extrapolation: the importance of LE changes may be subjected to *time preference*, usually 'discounting'. Secondly, LE is assigned importance weight relative to other criteria in the decision. The degree of personalisation will reflect the extent to which the elicitation and incorporation of criterion and time preferences are facilitated in (the) decision support.

Who by? For us, the communication is *by* either a healthcare professional or a credible provider-independent source. This reflects our distinction between the RP-*as-person* and RP-*as-patient*, i.e. as 'reasonable patient', legally speaking [1]. That distinction leads us directly into that between *apomediative* and *intermediative* decision support [2, 3] - and so to our preferred answers to Where**.**

Where? Intermediative decision support usually comes in the form of a Patient Decision Aid (PDA), designed to help the clinician and patient decide - *in a clinic or teleconsultation* - what is best for the patient. The options included in a PDA are typically restricted to those which the clinician feels relevant to the patient, and those they are able to prescribe or recommend within a guideline. The delivery of the aid within the encounter is under the clinician's control. In contrast, an apomediative Decision Support Tool (DST) is a 'direct-to-person-*as-person'* resource designed to help the person decide what is best for themselves - *at home or elsewhere in the community,* and usually online. They are developed by a provider-independent team of professional health analysts, operating within consumer law and rights provisions rather than under health professional 'duty of care' commitments, legal and ethical. A clinician may become involved, subsequently, in adding their support to that provided by an apomediative DST that the person has consulted - and will add value to the extent that they have greater relevant knowledge and can enhance *individualisation*. However, the significance of their individualisation improvement will depend on *personalisation* - improving individualisation in relation to a lowly-weighted criterion may be of little importance. Finally, a clinician may actively promote 'hybrid apo-intermediation' by inviting patients to engage with an apomediative aid at home in advance of a clinic consultation. While this transforms the apomediative DST into a clinician-managed one, the resulting encounter will be different from pure intermediation, because of the different answers to What and How, and Why, to which we now turn.

3. Why, What, How

Why? The healthcare literature displays growing acceptance of the need to deal with mortality more directly, and openly, than has been traditional in clinical practice. The Covid-19 pandemic has only accelerated this process, because the effect on survival is central to the comprehensive assessment of option harms and benefits to which the person is legally entitled under a 'reasonable *patient'* standard. In the absence of an overriding reason, this standard makes clear that *some* communication about mortality is required, leading directly to the consequential questions of What and How regarding the contents of the 'some'. At the other extreme, simple consumer demand, again absent some overriding reason, provides sufficient grounds for *some* communication about mortality, either on request as intermediative patient, or on search as apomediative person. The law trumps, as in the case of *lack* of demand, or outright refusal of communication about mortality by the patient, the acquiescing clinician may not obtain informed consent and be at legal risk. [1]

What constitutes an 'overriding reason' becomes crucial. It will almost always take the form of an *ethical* argument of a consequentialist or deontological sort, perhaps endorsed or even mandated by a professional code. How particular ethical reasons play out within the 'reasonable patient' standard is likely to remain legally moot in our view. However, we reject any attempted ethical justification for the suppression or distortion of the best available mortality estimates achievable through *individualisation*. These include ones intended to maintain 'hope' and 'optimism', since a requisite understanding of probability provides a firmer and superior ethical foundation for the *warrantable* hope and optimism that RP is entitled to maintain. RP is not devoid of emotions, moods and feelings, but accepts that the decision support offered (and requested) should be minimally distorted in content and delivery by temporary - entirely 'natural' and 'human'

- affective states. At the same time, they are aware and assured their basic feelings constitute the necessary and valid basis for preference-sensitive decisions, as achieved through *personalisation*.

*What***?** We assume RP requires decision support incorporating the best (individualised) numerical central point estimates for a mortality measure, accompanied by credible intervals. RP accepts that introducing verbal quantifications – high/low, long/short, big/small - will not add value and increase the possibility of bias, cognitive or motivational.

Any mortality measure relates to a single option (alternative/strategy//intervention), and so, in the (genuine, undominated) decisional context, the preferred mortality measure is required *for each of these* options. This creates a key issue, unfortunately too often ignored or overlooked. Are the two (or more) option-specific mortality measures to be processed as a set, producing a *mono-criterial* comparative assessment of the relative mortality impact of the options (usually in the form of a *difference)*? Or are they to be kept in isolation from each other and processed separately, each in conjunction with the equivalent measures for the other criteria in the *multi-criterial* assessment that characterises most health decisions? The answer has serious implications for personalisation, as we now illustrate, in the course of introducing the most familiar example of one of the possible types of answer to What: Life Expectancy, a *summary mortality measure on an unrestricted time scale*.

Assume that the LE for a 70 year old man without an intervention is 81, and that with the intervention it is 84. A widespread practice in making such a binary decision is to *reduce* the two LEs to a single summary measure of mortality impact, by taking their difference. The effect of the intervention is seen as a LE increase of 3 years, maybe phrased as 3 Life Years Gained by the intervention, or Lost by not having it. However, if the person attaches importance to other criteria, such as avoiding morbidity and option burden, they need to compare their personalized overall assessment of a non-dominant intervention e.g. 'LE 84; morbidity 79; burden 65', with that for the no intervention scenario 'LE 81; morbidity 79; burden 75'. Premature reduction of absolute criterion values to their differences, rules out proper personalization through criterion importance weighting.

This fundamental issue remains latent as we go more deeply into the possible answers offered to What, starting with the term 'unrestricted'. Unrestricted means that the mortality measure is calculated over full lifetime, not for any limited or 'restricted' time horizon. LE is unrestricted in being calculated as the average length of life of all those in a cohort, including the years lived by those in the tail of its distribution, e.g. those surviving beyond 100. (Median LE is sometime used to reduce the effect of the more extreme outliers have on the Mean LE, but it is still an unrestricted measure.) While it may seem obvious that we require an unrestricted measure, some of the most used mortality measures are 'Restricted', including the increasingly popular Restricted Mean Survival Time (RMST) metric [4]. Its methodological home is in the wider 'Time-to-Event' (TTE) literature, which embraces events less final than death, especially adverse ones such as heart attack, stroke, or fracture. Many recent papers make the case for TTE metrics as being equal to, if not better than, conventional measures, such as the Hazard Ratio (HR), Relative Risk (RR), or Number Needed to Treat (NNT), in capturing and communicating the evidence-based merits of options to clinicians and patients. Most who advocate TTE measures do so because, in line with those conventional measures, they present trial results in a statistically valid way, even when follow-up falls well short of lifetime, as it usually does.

In briefly documenting the case being made for TTE measures when the event is death, we see that How is being addressed along with What. As explained by Kloecker:

"The RMST difference compares the areas under the 2 survival curves for the intervention and control groups for a specified (restricted) interval. This contrast corresponds to the mean temporal postponement of the outcome in one group compared with the other, with each group-specific RMST quantifying the average delay in the event over the specified time horizon… [Take as example a 70-year-old man.] On the basis of EMPA-REG OUTCOME [a trial with a 4 year follow-up], the Hazard Ratio for all-cause mortality is 0.68. The corresponding RMST difference is 21 days postponement over 4 years. The health care professional could advise the patient that empagliflozin, on average, would prolong his life by 21 days over 4 years" [5] (pp 541, 545).

There is internal debate as to whether RMST is best measured by the dominant 'vertical' method of Kloecker and others [6] or by the alternative 'horizontal' method first deployed by Lytsy in relation to the 4S statin trial [7] and used recently by Bellavia [8]. However, we do not need to enter this debate since, measured either way, restricted follow-up means that in most cases only a limited number of patients have experienced the target event by study closure. In the horizontal measure TTE can be calculated only for those percentiles of the population that have experienced the event (death) - a percentage that will often fall far short of even the median survival length (50^{th} percentile). So, in the Lytsy statin study, the delay of 1 year calculated for the seventh percentile *by definition* applies only to the 7% in the untreated group who we know died within the follow-up period of the study. But we also know, again by definition, that this delay does not apply to the 93% who did not die.

Why use any *restricted* survival measure then? The answer reflects the preference for producing - and having guidelines and clinicians limited to using - 'evidence-based' mortality measures, where evidence is effectively restricted to that produced by the results of RCTs. Apart from the terminal situation, there is therefore only one possible exception: when the trial data permit 'unrestricted' conclusions. Claggett, et al. argue an 'actuarial' approach sometimes makes this possible. Used by them in two conditions [9,10], and by Dorresteijn [11], the 'delay of death' is measured from the person's *age at* diagnosis, not from *time of* diagnosis (irrespective of age). This method requires the underlying study/trial data to have "wide age range in the patient population, as well as a sufficiently large number of events occurring across the age spectrum in order to allow for relatively stable age-specific risk estimates… For any given age, a survival curve was estimated, representing the survival probabilities over time for patients alive at that age and receiving LCZ696 [the control]. A corresponding survival curve was then estimated using data from patients receiving enalapril [the intervention]. This process was repeated to produce estimated treatment effects for each age from 45-80 years old....[9](Supplementary Appendix, p2). The Utrecht group led by Dorresteijn applied this unrestricted lifetime TTE measurement to the results of the Women's Health study in which nearly 28,000 initially healthy women were followed up for a mean of 10 years. Crucially, there were a substantial number of older participants, because "predictions of lifetime models are not limited by the follow-up time of the study but rather by the age distribution of study participants. Therefore, observations in elderly patients are essential for stable long term predictions" [11] (p1).

So, an 'evidence-based' commitment means that the unrestricted lifetime mortality estimates to which RP is entitled can be provided only in the relatively rare cases meeting standard evidential requirements. However, we note that it is not only, or even primarily, the scientific standards for what constitutes 'evidence', that are the problem. Many of

the appropriate standards could be met, if it were not for the domination of short-term funding arrangements and commitments by the public and private supporters of research. For most researchers and groups, a circa 3-year 'restricted mean survival time' is the reality.

No particular restricted TTE is meaningful to a RP. Short of their lifetime, each and every particular restricted time period could be meaningful to a particular RP. There is therefore no way in which one of these can be privileged just because it is the follow-up limit of relevant robust, 'evidence-generating' research. RP is entitled to attach personal meaning to each and every extra year of life as represented in the full survival functions relevant to their decision. In practice and in life there is therefore no alternative to lifetime extrapolation beyond the restricted evidence. There are two possibilities. One is *explicit* extrapolation based on one of many possible principles, e.g. that the proportional hazard ratio calculated from the survival curves continues to apply. The other possibility is implicit, undiscussed, maybe unconscious, extrapolation. However, for the clinician, who, remember, is always engaging with RP, the latter is not a legally secure choice. They will need to supply, and document, their best explicit extrapolation, based on all their resources.

Despite emphasising its virtues, most advocates of RMST agree there is a serious issue with its simple use as an answer to What, because:

"… estimates of RMST differences depend on, and should be interpreted with reference to, the event rate in both groups and the duration of follow-up or, rather, the specified time horizon… Because the RMST difference reflects the difference in the areas under 2 survival curves, the same difference may be obtained from diverse combinations of survival curves. For example, an RMST difference of 1 year over 5 years may be the result of an RMST of either 4 years in the treatment group and 3 years in the control group or 3 years in the treatment group and 2 years in the control group. As with the aforementioned measures of absolute risk reduction, we therefore advise interpretation of the RMST along with the survival curves" [5] (p548).

This is too weak. The much stronger version, and our preferred answer to What, takes us immediately to the two (Kaplan-Meier) survival curves, from which all restricted or unrestricted single summary measures, such as RMST and LE, are derived. Referring to the underlying survival curves (restricted or unrestricted) cannot be merely an *ideal,* or something that *could* lead to *'better appreciation'* [12] (p58). All single summary mortality measures are completely dependent on the two generative survival curves, and the *absolute* probabilities (KM estimators), that compose those curves, are essential to give personal meaning to the difference between them. Any single number mortality metric such as the RMST difference is *reductionist* as well as *relative*; it removes the *two* absolute groundings required in a multi-criterial health decision - which means all health decisions involving benefits and harms. The plotting of survival probabilities and curves on a truncated scale, common in the TTE literature (e.g. from 0.95 to 1 in Kloecker), exacerbates this problem.

Let us *imagine* how Reasonable Person might respond to the RMST TTE measure, having had it explained in the context of the Lytsy statin study:

'So if - but only if - I were to die 6 years from now, after taking statins for that time, I would be dying one year later than I would have, if I hadn't taken statins for that period. I would have 'postponed my death' from 5 years to 6 years by taking statins - *if* I hadn't avoided dying. But it is very likely I will avoid dying in this period according to your chart of the two survival curves. It shows that at the end of the period shown - 6 years - my probabilities of survival are 87% without statins and 91% with them. This delay

calculation, which after considerable effort I do *understand*, might mean something if I knew I was fairly certain to die 6 years from now. However, I am certainly not certain to die in exactly 6 years - or any particular time from now. The difference between the two probabilities which generates the 1 year benefit from statins is 4%. But this difference has no real use for me either, because 7% and 11% would generate the same 1 year 'postponement' - but affect my decision very differently from 87% and 91%. My decision involves trading off any mortality benefit from statins with the absolute probabilities for my other benefits and harms, such as morbidity, adverse events, side effects, and treatment burdens from statins. In fact, it's now clear, no one pair of absolute probabilities on the survival curves (like 87% and 91% at 6 years) is of any use. What has unquestioned personal meaning is my maximum lifetime, so I need to keep in mind the full survival curves, not any summary measure and if some restricted time horizon has meaning for me, I will decide what it is!"

*How***?** The best available estimates of the KM survival curves for the individual, for each option, are our starting point for decision supporting any RP. Restricted, reductionist, relative measures may be introduced, but only subsequently, never initially, and never as substitutes in the belief (we believe unwarranted) that they are easier to understand and communicate than probability-based measures. Throughout the TTE literature it is suggested that summary measures based on a *time scale*, should replace continuous measures based on a *probability* scale.

"It is well known that the established effect measures are associated with some difficulties when used in clinical care for individual decision-making…Probabilistic thinking is difficult. Laymen, patients, and even skilled professionals all suffer from various degrees of statistical illiteracy, making it difficult for many to perform simple arithmetic calculations and to comprehend risk estimates. This predicament is further supported by research showing that the format of the effect measure may influence patients' acceptance of taking a medication as well as doctors' and health authorities' willingness to recommend or prescribe it. This signifies the challenge clinicians face when deciding how to describe treatment outcomes to their patients for the purpose of shared decision-making" [7] (p 905).

We argue the reverse: a time scale measure should be introduced, *if at all*, only as a supplement to the fundamental probabilistic formulation. Neither ease of understanding or communication (*if true*) can justify prioritising a non-probabilistic measure, derived from a probabilistic one, on the grounds that it reduces the difficulty of accepting and dealing with the inherent uncertainty (correctly) reflected in that underlying one. Succumbing to the tempting reductionist relative alternative is a cognitive version of the 'streetlight fallacy', where keys lost in the dark are searched for under the light, because it is easier to see there.

Perhaps surprisingly, many of the most explored issues in relation to How loom much less large for us than others. Our purpose in presenting the survival curves is not to have them interpreted, even if there is some evidence that they can be well understood [13, 14]. Indeed, we positively discourage attempts at interpretation of visual displays of all kinds, including emoticon displays that follow best visualization practice in providing information *on a single criterion*. This is because *calculation* on their basis is essential and multi-criterial calculation is not a task to be imposed on an individual engaging with a decision support tool, whether highly numerate or not. Implying they are capable of it amounts to symbolic violence [15]. Reasonable Person should know what the various data elements represent and what procedure ('algorithm') is being used to integrate them, but not have to attempt synthesizing calculation. The purpose in presenting survival

curves is to *help the RP understand them only to the extent necessary to understand why analytical calculation based on them - in conjunction with equivalent lifetime measures for other criteria - is essential*. And so to understand why the multi-criteria scores resulting from these calculations should be the core output of personalized decision support tools.

References

[1] UK Supreme Court Judgment, Montgomery v Lanarkshire Health Board (2015) [Available at https://www.supremecourt.uk/cases/docs/uksc-2013-0136-judgment.pdf
[2] Eysenbach G. From intermediation to disintermediation and apomediation: new models for consumers to access and assess the credibility of health information in the age of Web2.0. Stud Health Technol Inform. 2007; 129: 162–6.
[3] Dowie J, Kaltoft M.K. The future of health is self-production and co-creation based on apomediative decision support. Med Sci. 2018; 6: 66.
[4] Nemes S, Bülow E, Gustavsson A. A brief overview of Restricted Mean Survival Time estimators and associated variances. Stats 2020; 3: 107–19.
[5] Kloecker DE, Davies MJ, Khunti K, Zaccardi F. Uses and limitations of the restricted mean survival time: Illustrative examples from cardiovascular outcomes and mortality trials in type 2 diabetes. Ann Intern Med. 2020; 172: 541–52.
[6] Eloranta S, Smedby KE, Dickman PW, Andersson TM. Cancer Survival Statistics for Patients and Healthcare Professionals – a Tutorial of Real-World Data Analysis. J Int Med 2020; 289: 12-28.
[7] Lytsy P, Berglund L, Sundström J. A proposal for an additional clinical trial outcome measure assessing preventive effect as delay of events. Eur J Epidemiol. 2012; 27: 903–9.
[8] Bellavia A, Wallentin L, Orsini N, James SK, et al. Time-based measures of treatment effect: Reassessment of ticagrelor and clopidogrel from the PLATO trial. Open Heart 2017; 4: e000557
[9] Claggett B, Packer M, McMurray JJV, Swedberg K, et al. Estimating the long-term treatment benefits of sacubitril-valsartan. N Engl J Med. 2015; 373: 2289–90.
[10] Claggett B, Lachin JM, Hantel S, Fitchett D, et al. Long-term benefit of empagliflozin on life expectancy in patients with type 2 diabetes mellitus and established cardiovascular disease survival estimates from the Empa-Reg OUTCOME trial. Circulation. 2018; 138: 1599–1601.
[11] Dorresteijn JAN, Kaasenbrood L, Cook NR, van Kruijsdijk RCM, et al. How to translate clinical trial results into gain in healthy life expectancy for individual patients. BMJ. 2016; 352: 1–6.
[12] Belot A, Ndiaye A, Luque-Fernandez MA, Kipourou DK, et al. Summarizing and communicating on survival data according to the audience: a tutorial on different measures illustrated with population-based cancer registry data. Clin Epidemiol. 2019; 11: 53–65.
[13] Zikmund-Fisher BJ, Fagerlin A, Ubel PA. Mortality versus survival graphs: Improving temporal consistency in perceptions of treatment effectiveness. Patient Educ. Couns. 2007; 66: 100–107.
[14] Rakow T, Wright RJ, Bull C, Spiegelhalter DJ. Simple and multistate survival curves: Can people learn to use them? Med Decis Mak. 2012; 32: 792–804.
[15] Kaltoft MK, Nielsen JB, Salkeld G, Dowie J. Health informatics can avoid committing Symbolic Violence by recognizing and supporting generic decision-making competencies. Stud Health Tech Inform. 2015; 218: 172-7.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210571

How a Service User Knows the Level of Privacy and to Whom Trust in pHealth Systems?

Pekka RUOTSALAINEN[a,1], Bernd BLOBEL[b,c,d]
[a] *Faculty of Information Technology and Communication Sciences (ITC), Tampere University, Finland*
[b] *Medical Faculty, University of Regensburg, Germany*
[c] *eHealth Competence Center Bavaria, Deggendorf Institute of Technology, Germany*
[d] *First Medical Faculty, Charles University of Prague, Czech Republic*

Abstract pHealth is a data (personal health information) driven approach that use communication networks and platforms as technical base. Often it' services take place in distributed multi-stakeholder environment. Typical pHealth services for the user are personalized information and recommendations how to manage specific health problems and how to behave healthy (prevention). The rapid development of micro- and nano-sensor technology and signal processing makes it possible for pHealth service provider to collect wide spectrum of personal health related information from vital signs to emotions and health behaviors. This development raises big privacy and trust challenges especially because in pHealth similarly to eCommerce and Internet shopping it is commonly expected that the user automatically trust in service provider and used information systems. Unfortunately, this is a wrong assumption because in pHealth's digital environment it almost impossible for the service user to know to whom to trust, and what the actual level of information privacy is. Therefore, the service user needs tools to evaluate privacy and trust of the service provider and information system used. In this paper, the authors propose a solution for privacy and trust as results of their antecedents, and for the use of computational privacy and trust. To answer the question, which antecedents to use, two literature reviews are performed and 27 privacy and 58 trust attributes suitable for pHealth are found. A proposal how to select a subset of antecedents for real life use is also provided.

Keywords pHealth, eCommerce, privacy, trust, antecedents

1. Introduction

According to Lodewjk Bos, pHealth has both personal and personalized Health dimensions and it takes palace in digital environment [1]. Despite that the concept of pHealth is somewhat fuzzy, its focus seems to be how personal health information can be used in patient care and how a person can monitor and manage his or her health and health behavior. pHealth is also a data driven approach where major parts of collected and used personal health information (PHI) is generated by the data subject or patient

[1] Corresponding Author. Pekka S. Ruotsalainen, DSc (Tech.), Adjunct professor, Research professor emeritus, Faculty of Information Technology and Communication Sciences (ITC), Tampere University, Kanslerinrinne 1, 33014 Tampere, Finland; Email:pekka.ruotsalainen@uta.fi

itself or by sensors, wearable technology and motes. Information measured and monitored includes person's vital signs, health behaviors and activities, location, movement, feeling and emotions, social relations, environmental factors and vital signs. Typical pHealth service includes the collection of PHI and the output is processed information (e.g. calculated values and trends), personalized recommendations and guidelines which help the user to monitor and management own health and health behaviors. For data transfer, communication pHealth typically uses existing third party services such as the Internet, wireless networks, and short range digital communication systems. Digital platforms, clouds and application are typically used for data processing, storing and sharing tasks. From this perspective, pHealth, eCommerce, eHealth, mHealth, Internet shopping and social networks look similar. pHealth has also lot of common with the novel 5P medicine approach, as both offer personalized and preventive services for health management. Both approaches also collect wide spectrum of PHI that exceeds radically the content of current regulated EHR [2]. According to Gorini, the future medicine (e.g. 5P medicine) requires patients/person's full psychological and cognitive profile, i.e. information such as health lifestyle, personality, cognitive dispositions, social conditions and psychological state, specific needs and values, habit and behavior patterns, hopes and fears, beliefs, individual characteristics, decision making style, emotional profile, psychological contexts (presence of stress, anxiety, depression) and information about physical, social, and economic environments [3]. In real life, this amount of sensitive personal information cannot be collected by any single organization or service provider. Instead many sources a needed such as social networks (data the person himself/herself discloses), pHealth, mHealth and eHealth applications, eCommerce and Internet shopping services, and public and private health care. That way, they all together form a big data ecosystem. In this ecosystem, pHealth applications can be play a meaningful role as data collector and information source.

Unfortunately, this data driven future of pHealth and 5P medicine raises many new and until now unsolvable privacy and trust problems. The multi-stakeholder natures of the data ecosystem and the huge spectrum of collected and used PHI together make it difficult for the service user to know why, and how much, to trust in pHealth services and the eco-system, and what the actual level of privacy in pHealth is.

There are many reasons, why the service user cannot blindly trust the service provider and expect that necessary privacy safeguards are in place and legal privacy requirements are met in the pHealth ecosystem. First, networks and ecosystems are often multi-stakeholder systems, where commercial stakeholders (e.g. platform operators and non-regulated health service providers) have own business goals, privacy policies and trust behaviors. They often do not see people only as customers. Service users' data is raw material for their products and new businesses [4]. Firms also often do not hold what they have promised in their privacy documents and trust manifestos, and in real life, users have almost no control what data is collected and how it is used and shared [5]. Furthermore, commercial organizations typically expects that people trust them blindly, and organizations' privacy documents are written more to protect them, and they are typically written in a legal language that is difficult to understand. Furthermore, currently widely used security-based privacy protection tools cannot really guarantee privacy.

It is evident that – from service users' point of view – the current situation is unsatisfactory and shall be changed. The authors state that the user of digital services collection and processing sensitive PHI such as pHealth applications need a possibility to evaluate on-line the level of privacy and trust of services and information system behind it. To help the user of pHealth services in decision making on starting to use or

not to use service and how much PHI to disclose, the authors propose the development of an evaluation service solution that is easy to use for a human and that reflects the service user's view point. Two main element of the evaluation system are calculation methods and appropriate input variables (antecedents) used in calculation. In this paper, the authors' focus are antecedents.

2. Privacy and Trust

Information privacy and trust are complex concepts with many approaches and definitions. They are also interconnected in such a way that the amount of positive trust reduces the need for privacy protection. High privacy and trust are prerequisites for successful use of pHealth, eHealth, eCommerce and Internet shopping. Basic privacy types are general privacy and contextual privacy [6]. Widely used information privacy approaches include: privacy as right and control (ability to control); privacy as legal construct; risk based privacy; privacy as contextual integrity; privacy as concerns [7, 8].

Originally, trust was understood to exist between persons, but currently it is accepted that trust also exists between human and organizations, human and computers as well as technology in a general sense. The way trust is understood depends on culture and context. Human trust is a personal trait. Trust is needed in situations where insufficient information is available. Disposition to trust is understood as tendency to trust in others. Other views to trust include subjective probability, belief in trustor's features, attitude, perception and trust as risk and willingness to trust [9, 10].

Trust can be general trust and context- or system-specific trust. In digital information systems, the person (service user) has to trust in organizations, technology, computational features of the information and communication system and computer applications. Computational trust imitates human trust, and at the same time, it enables the service user to estimate the degree of trust. For trust calculation, mathematical model considering changes in trust caused by its antecedents are often used [11].

3. Antecedents for privacy and trust

A widely used approach in eCommerce, Internet shopping and social networks is to assume that the user beliefs that information privacy is guaranteed, and he or she feels that service provider and the network/ecosystem is trusted. Unfortunately, this approach is not true in real life digital information systems. Trust and privacy in pHealth services depends of service providers and computational environments contextual features. Therefore, contextual privacy and trust models should be used instead of general privacy and trust. Contextual privacy and trust require that antecedents used describe contextual features of the service provider and information systems.

To find candidate privacy and trust antecedents for pHealth the authors made a literature review of privacy and trust focused papers published in major journals and covering eCommerce, Internet shopping, social media and eHealth. Because, as discussed in Chapter 1, pHealth uses similar ICT technological solutions and services as eCommerce, Internet shopping and eHealth, the authors expect that privacy and trust antecedents researchers have found valid for them can be also used to evaluate privacy and trust in pHealth.

3.1. Privacy Antecedents

Table 1 presents an aggregated summary of widely used privacy approaches and corresponding antecedent retrieved from a literature review. It is almost impossible for the service user to measure privacy itself. Therefore, and because privacy depends more on cognitions and perceptions than on rational assessments, privacy proxies such as belief, risk, concerns, benefits, perceived harm and other perceptions are widely used as antecedents [6].

Table 1 Privacy approaches and their attributes [12-15]

Privacy approach	**Possible antecedents**
General privacy	Belief, disposition
Privacy as control and restrict access Privacy as individual right	Knowledge of service provider's practices and information system, direct experiences, past experiences, privacy promises audit-trails, privacy seals, information practices, other's proposals perception
Privacy as concerns	Personality, motivations, perceptions, context information, service user's behaviours, technology used data type, others opinions, perceived severity
Risk based privacy	Assessed risk level, perceived risk in technology, perceived concerns, cost/benefit ratio, perceived harm or impact, privacy seal
Privacy as contextual integrity	Context type and its features, type and sensitivity of data, contextual privacy practices, privacy culture
Privacy as legal concept	Legal requirements, compliance analysis

3.2. Trust Antecedents

In a literature analysis performed by the authors, 58 different trust antecedents were found. The authors classified them into seven groups (customer perception, customer experiences, service provider characteristics, features of the service, information based features, infrastructural factors and external and environmental) provided in Appendix A [9, 12, 16-29]. The biggest group, i.e. service provider (vendor, organization or institute) characteristics, contain 24 antecedents. From another review focused on privacy and trust in eHealth, the authors found that in eHealth privacy, reputation and informational factors (e.g. professionalism of information and medical quality of information) are most meaningful antecedents for the user.

4. Challenges in Evaluation of Privacy and Trust in pHealth

The service user's ultimate goal is to measure the level of actual privacy of the service provider and the surrounding ecosystem, and to know why and how much he or she can

trust in a service provider. In real life, there are many things, which make it difficult to reach this goal. A big challenge is the lack of reliable and accurate information of service provider's privacy and trust features and behaviors. Another challenge is that privacy laws (e.g. the EU-GDPR) are high-level documents without information on implementation details. Laws typically balance industry's business needs and national interest against a person's privacy needs, resulting in insufficient privacy. For example, the EU-GDPR enables service providers to define "mandatory cookies" and what the content of legitimate interest is. Researchers have argued that in digital environment laws are insufficient to give the person reasonable power to control what personal information is collected by service providers and organizations, and how this data is processed and disclosed [30, 31].

Many of the antecedents shown in Table1 and Appendix A such as belief, intention, motivation, benefit, ability, honesty, goodwill, harm and reliability are abstract, difficult to conceptualize and measure, and often proxies for perceptions, opinions or even feelings. Widely used others opinions concerning privacy and trust are unreliable and can be misleading. Perceptions such as perceived risks or perceived harm are more opinions than a description of the real life risks and caused harm. Caused by the lack of reliable information and the vagueness of risk and harm concepts, it is an illusion that a pHealth service user can make credible measurement of privacy risks and possible future harm.

In Table 1 and Appendix A, totally 85 antecedents are shown. A subset should be selected to make them practical for human use. For that purpose, a selection criterion is needed. The authors propose the prioritization of antecedents, which values are available (e.g. third partner privacy seal or audit-trail) or measurable. Other antecedents can be grouped as follows: own previous experiences, own perceptions, other proposals, personal opinions or feeling and beliefs.

5. Discussion

Trust is not only a personal trait, it is also a glue between people, organizations, and information systems. In other words, our society requires trust to function. Therefore, trust is a public good that together with information privacy enables a person to safely use pHealth services and disclose PHI [34]. Unfortunately, it is a common practice in today's eCommerce, Internet shopping, eHealth and pHealth to expect that service users automatically belief that service providers are trustworthy, information privacy is guaranteed (i.e. necessary protection tools and protocols are implemented correctly) and service providers keep their promises (e.g. what is promised in privacy policy documents and trust manifestos). Additionally, it is expected that the user accepts service provider's business rules and the collection of PHI without the possibility to define own rules. In digital information systems, hidden collection of person's behavioral information is a dominant practice, and the service user has no real possibility to know how PHI is used inside information systems and to whom data is disclosed or sold. From a user's point of view, this is an unacceptable situation. Therefore, the authors state that pHealth and eHealth service users need a tool to evaluate the level of privacy, and to know in whom to trust before starting to use services and to disclose PHI.

Because there is a big amount of computation privacy and trust solution available, the authors have focused in this paper on antecedents [6, 35, 36]. For the pHealth service

user, the authors propose a method where the service user evaluates the service provider's privacy and trust using computational method and (as far as possible) in real lively measurable contextual antecedents [33]. Furthermore, they conclude that privacy and trust antecedents of eCommerce, Internet shopping and eHealth are also deployable in pHealth. Based on literature review the authors presented a set of candidate antecedents.

The biggest barriers to implement the author's proposal is the lack of measurable and reliable information describing service provider's and surrounding information system's privacy and trust features and behaviors. Another challenge is that currently widely used privacy and trust approaches do not work in digital environments. The disparity in power between the service user and service provider enables the service provider to use own privacy rules and laws allowing this. To solve those problems, it is necessary to redefine current privacy and trust concepts and move them to virtual and digital environments where pHealth takes place. A novel approach to those problems is, e.g., the definition of privacy as personal property and trust as specific legal fiducial duty [37, 38]. New laws and regulations are also necessary to support those new privacy and trust approaches and to force service providers to publish reliable, detailed and measurable information concerning their privacy practices and trust features and behaviors.

6. Conclusions

Current privacy and trust situation in pHealth resembles the ongoing climate change discussion: researchers know what is going on and what should be done, but industry and governments prefer economic grow and profit orientation. A change is inevitable. If current situation persists, there will be no privacy in the future and a complete loss of trust in anyone. Our PHI is monetarized, and person's privacy needs are overridden by business and political objectives, manipulating people to trust blindly. This way leads to an inhuman, immoral, unethical society.

For overcoming the challenges, the authors have proposed the use of computational methods for privacy and trust evaluation and defined a set of suitable antecedents. The next step is the development of practical solutions enabling the services user an on-line evaluation of information privacy and the possible to trust in a service provider.

References

[1] Bos L. pHealth. Stud Health Technol Inform. 2012; 177, 3-13. doi:10.3233/978-1-51499-069-7-3.
[2] RuotsalainenP, Blobel B. Privacy s Dead – Solutions for Privacy-Enabled Collections and Use of Personal Health Information in Digital Era. Stud Health Technol Inform. 2020; 273, 63-74. doi:10.3233/SHTI200616
[3] Gorini A, Caiani G, Pravettoni G. Psycho-cognitive Factors Orienting eHealth Development and Evaluation, in G. Pravettoni, S. Triberti (eds.), P5 eHealth: An Agenda for the Health Technologies of the Future, Springer Open, https://doi.org/10.1007/978-3-030-27994-3_7
[4] Zuboff S. The Age of Surveillance Capitalism, Public Affairs, ISBN 9781781256855.
[5] Aquisti A, Laura Brandimarte L, Loewenstein G. Privacy and human behavior in the age of information, Science Special Section The end of privacy, Science. 2015 Jan 30; 347 (6221): 509-514 DOI: 10.1126/science.aaa1465, http://science.sciencemag.org/.
[6] Smith, H J, Dinev T, Xu H. Information privacy research: an interdisciplinary review, MIS Quarterly December 2011; Vol. 35 No. 4: 989-1015.

[7] Marguilis ST. Privacy as a Social Issue and Behavioral Concept, Journal of Social Issues 2003; 59 (2): 243-261.
[8] Zwick D. Models of Privacy in the Digital Age: Implications for Marketing and E-Commerce. 1999.
[9] Lilien L, Bhargava B. Trading Privacy for Trust in Online Interactions. Purdue University, West Lafayette, U.S.A.
[10] Jøsang A, Ismail R, Boyd C. A Survey of Trust and Reputation Systems for Online Service Provision, Decision Support Systems, 2007; 43(2): 618-644.
[11] Liu X, Datta A, Lim E-P. StereoTrust: a group based personalized trust model. Proceedings of the 18th ACM conference on Information and knowledge management, 2009. doi: 10.1145/1645953.1645958.
[12] McKnigth H D, Choundhury V, Kacmar C. Developing and line Interactions. Validating Trust Measures for e-Commerce: An Integrative Typology, Information Systems Research, 2002 INFORMS. September 2002; 13 (3): 334–359.
[13] Corbitt B J, Thanasankit T, Yi H. Trust and e-commerce: a study of consumer perceptions, Electronic Commerce Research and Applications 2003; 2 (3): 203-215. doi:10.1016/S1567-4223(03)00024-3.
[14] Barth S, D.T, de Jong MDT, Jungerc M, Hartel PH, Roppelta JC. Putting the privacy paradox to the test: Online privacy and security behaviors among users with technical knowledge, privacy awareness, and financial resources, Telematics and Information. August 2019; 41: 55-69. https://doi.org/10.1016/j.tele.2019.03.003.
[15] Kosa TA, El-Khatib K, Measuring Privacy, Journal of Internet Services and Information Security (JISIS), 2011; 1 (4): 60-73.
[16] Arifin, D M, Antecedents of Trust in B2B Buying Process A Literature Review, 5th IBA Bachelor Thesis Conference, July 2nd, 2015, Enschede, The Netherlands. Copyright 2015, University of Twente, The Faculty of Behavioural, Management and Social sciences.
[17] Pennanen K. The Initial Stages of Consumer Trust Building in e-Commerce: a Study on Finnish Consumers, ISBN 978–952–476–257–1 http://urn.fi/URN:NBN:fi-fe2018062126262.
[18] Meziane F, Nefti S. Evaluating E-Commerce Trust Using Fuzzy Logic, International Journal of Intelligent Information Technologies, October-December 2007; 3(4), 2-3, IGI Global, edited by Vijayan Sugumaran, Hershey, USA.
[19] Vega J A, Determiners of Consumer Trust towards Electronic Commerce: An Application to Puerto Rico, Esic Market Economics and Business Journal January-April 2015; 46 (1): 125-14.
[20] Egger, FN. From Interactions to Transactions: Designing the Trust Experience for Business-to-Consumer Electronic Commerce, Eindhoven University of Technology (The Netherlands), 2003, ISBN 90-386-1778-X.
[21] Grabner-Kräuter S, Kaluscha EA. Consumer trust in electronic commerce: conceptualization and classification of trust building measures, Chapter in 'TRUST AND NEW TECHNOLOGIES' edited by Teemu Kautonen & Heikki Karjaluoto, Edward Elgar Publishing 2008, pp. 3-22.
[22] Brændeland G, Støle K. Using Risk Analysis to Assess User Trust – A Net-Bank Scenario, C.D. Jensen et al. (Eds.): iTrust 2004. Springer LNCS; 2004; 2995: 146–160.
[23] Hussin Ab R C, Macaulay, The Importance Ranking of Trust Attributes in e-Commerce Website, 11th Pacific-Asia Conference on Information Systems, 2007.
[24] Salam A F, Iyer L, Palvia P, Sin R. Trust in e-Commerce. Communications of the ACM, February 2005/Vol. 48, No. 2, pp. 73-77.
[25] Beldad A, de Jong M, Steehouder M, How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust, Computers in Human Behavior 26 (2010) 857–869, Elsevier. doi:10.1016/j.chb.2010.03.013
[26] Yan Z, Holtmanns S. Trust Modeling and Management: from Social Trust to Digital Trust, book chapter of Computer Security, Privacy and Politics: Current Issues, Challenges and Solutions, IGI Global, 2007, DOI: 10.4018/978-1-59904-804-8.ch013.
[27] Chen CC, Dhillon G. Interpreting Dimensions of Consumer Trust in E-Commerce, Information Technology and Management 4, 303–318, 2003, 2003 Kluwer Academic Publishers. Manufactured in The Netherlands.
[28] Mayer RC, Davis JH and Schoorman FD. An Integrative Model of Organizational Trust,: The Academy of Management Review, Vol. 20, No. 3 (Jul., 1995), pp. 709-734. URL: http://www.jstor.org/stable/258792. 137-154, ISSN 0718–1876 Electronic Version, www.jtaer.com, DOI: 10.4067/S0718-18762010000200009.
[29] Kimi A, Choobineh J. Trust in Electronic Commerce: Definition and Theoretical Considerations, 1060-3425/98,1998 IEEE.
[30] Xu, H, Teo Hock-Hai, Tan BCY. Predicting the adoption of location-based services: the role of trust and perceived privacy risk

[31] Lankton N K, McKnight D H, Tripp J. Technology, Humanness, and Trust: Rethinking Trust in Technology, Journal of the Association for Information Systems, Volume 16, Issue 10, pp. 880-918, October 2015.
[32] Richards, N M, Hartzog W. Taking Trust Seriously in Privacy Law. (September 3, 2015). 19 Stanford Technology Law Review 431 (2016), Available at SSRN: https://ssrn.com/abstract=2655719 or http://dx.doi.org/10.2139/ssrn.2655719
[33] Ruotsalainen P, Blobel B. Health Information Systems in the Digital Health Ecosystem—Problems and Solutions for Ethics, Trust and Privacy Health 2020, Int. J. Environ. Res. Public 2020; 17(9): 3006; doi:10.3390/ijerph17093006.
[34] Engel C. Privacy as a Public Good, Duke Law Journal, Volume 65, Number 3, December 2015.
[35] Pinyol I, Sabater-Mir J. Computational trust and reputation models for open multi-agent systems: a review Artif Intell Rev 2013; 40: 1–25 DOI 10.1007/s10462-011-9277-z.
[36] Sabater J, Sierra C. Review on computational trust and reputation models. Artif Intel Rev 2005; 24(1):33–60.
[37] Ritter J and Mayer A. Regulating Data as Property: A New Construct for Moving Forward, 16 Duke Law & Technology Review 2018; 220-277.
[38] Waldman AE. Privacy as Trust- Information Privacy for an Information age Cambridge University Press, 2108, ISBN 978-1-316-63694-7.

Appendix A Factors impacting to trust formulation in e-services

Perceptions [16, 17, 18, 19, 21, 20, 22, 23, 24]	**Customer experiences** [17, 25, 26, 27]
- Perceived quality of services - Perceived lack of privacy - Perceived lack of customer control - Perceived risks - Perceived trustworthiness, expertise and credibility - Perceived usefulness and perceived ease of use - Perceived predictability and consistency in the vendor's actions - Perception that vendor is honest and concerned about its customers	- Satisfaction with previous online transactions - Past experiences, Purchase experiences - Satisfaction of the product - Feedback and recommendations - How well the observed behaviour of the system meets their own standards - Past behavior and seller keeps promises

Service provider (vendor, organization, institution) characteristics [9, 12, 13, 16, 19, 20, 21, 23,24, 25, 26, 27, 28, 29, 30]

- Responsibility
- Firm type
- Ability or competence, benevolence, integrity, honesty, fairness, faith
- Absence of guarantees
- Appearance
- Competence
- Credibility
- Dependability
- Goodwill
- Familiarity and Friendliness
- Fiduciary and size
- Motivation
- Predictability
- Performance
- Persistence
- Policies e.g. return policy, privacy policy, quarantine policy
- Potential opportunistic behaviors
- Reliability
- Reputation of the company
- Structural assurance and Situational normality
- Values of the seller
- Vendors' presence (Availability of mailing address
- and telephone numbers)

Features of the service [19, 21, 23, 27]

- Service quality (tangibles, reliability, responsiveness, assurance, and empathy and satisfaction)
- Quality certificate
- Lack of customer control
- Service professionalism
- Product price

Information based features [21, 23]

- After sales service
- Existing data and literature
- Information about product and services
- Lack information regarding the behavior or characteristics of the object of trust
- Lack of information concerning IT-technology and privacy safeguards
- Service users' knowledge

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210572

Towards Personalized Medication

Michal HUPTYCH[a1], Jiri POTUCEK[a,b], Lenka LHOTSKÁ[a,c]
[a] *Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, Prague, Czech Republic*
[b] *MWCare Ltd., Prague, Czech Republic*
[c] *Faculty of Biomedical Engineering, Czech Technical University in Prague, Prague, Czech Republic*

Abstract. The paper describes some aspects of precision medicine and shows the importance of pharmacokinetics and pharmacodynamics for the therapeutic drug monitoring and model-informed precision dosing. A key element in the design of the pharmacokinetics and pharmacodynamics (PKPD) models is relevant literature search that represents an essential step in the procurement and validation of a new drug. Available search engine resources do not offer specific functionalities that are required for efficient and relevant search in reliable literature sources. We present a prototype of such an intelligent search engine and show its results on real project data.

Keywords. Precision medicine, model-informed precision dosing, intelligent literature search, pharmacokinetics, pharmacodynamics

1. Introduction

Fast technological development enabled design and development of new applications across many fields during last two decades. Medicine and health care belong to these fields. Almost all medical devices can be connected to computers and communicate acquired data. Wireless technologies, sensors, wearables and Internet of Things contribute to development of continuous data acquisition and communication. Electronic health record is an inseparable part of this development. Development of new drugs is almost unimaginable without extensive computational resources. Few years ago, we spoke about the concept pf P4 health care - Participatory, Predictive, Preventive, Personalized. Recently the fifth P has been added: Precision. Precision medicine is understood as an approach to patient treatment that considers individual variability in genes, environment and lifestyle for each patient. This approach is already applied in some areas of medicine for targeted treatment, as for example in oncology. With the development of new software tools there opens a great opportunity to apply this approach to wider spectrum of treatments.

The core of the software tools for precision medicine are modules for therapeutic drug monitoring and model-informed precision dosing. Background for their development constitute pharmacokinetics and pharmacodynamics models for individual drugs. Pharmacokinetics and pharmacodynamics represent two phases of the processes

[1] Corresponding Author: Michal Huptych, PhD, Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, Prague, Czechia; Email: Michal.Huptych@cvut.cz

between drug input and emergence of the response on the output. The pharmacokinetic phase represents all events between the input (drug dose administration) and the achievement of drug concentrations throughout the body. The pharmacodynamic phase represents all events between the arrival of the drug at its site of action and the onset, magnitude, and duration of the biological response [1]. The models have to take into account many parameters, including information about the patient, e.g. weight, age, gender, ethnicity, metabolism rate, interaction with other drugs, negative effect on potential co-morbidities, etc. Additional parameters are dose, frequency of administration and route of administration. These parameters represent variables that can be adjusted to optimize the therapeutic effects of a drug for a particular disease condition. It is possible to control the onset, intensity and duration of the positive drug effect, while minimizing any harmful effect. Obviously, for designing such models detailed understanding of human physiology, metabolism, dose-response relationships, dependence of response on drug route, drug elimination and clearance, molecular biology and genetics is needed. That means that such work is strongly interdisciplinary.

In the paper we present briefly development in the area, show the concept of a system for model-informed precision dosing and describe an important tool supporting development of models for drugs for given populations. The aim of this tools is to speed up the development of models and make the work more efficient.

2. Related Work

Last decade showed that the demand for precise drug dosage and therapeutic drug monitoring is continuously growing. There have been performed many clinical studies, in particular in oncology, showing the positive effect of targeted therapy [2, 3, 4]. Other studies include optimizing care of critically ill patients with severe infections [5], optimizing the dose for children [6], physiological regulation of drug metabolism and transport [7, 8]. Several reviews discuss current state and future of precision medicine [9, 10, 11]. Some authors analyze impact on the industry and development of new drugs [12, 13, 14].

Development of software tools for therapeutic drug monitoring and model-informed precision dosing is not an easy task. It requires extensive and deep knowledge from different disciplines, which means gathering the experts willing to collaborate on the development, and also extensive literature search for drug models and their validation. There are not so many tools available as two evaluation studies, performed in the last decade show [15, 16]. In [15], published in 2013, 12 software tools were tested and evaluated using 12 criteria: pharmacokinetic aspects (population and drug, models, modularity, plot, and various) general characteristics (user interface, computing aspects, interfacing, cost, report, and storage) and expertise of authors. The authors of [16] developed with the help of 22 experts more detailed evaluation criteria, finally grouped in eight categories: user friendliness and utilization, user support, computational aspects, population models, quality and validation, output generation, privacy and security, and cost. They evaluated 10 software tools, three out of them were already evaluated in 2013, however, since then new versions were developed with more advanced functionalities. They were MwPharm (upgraded to MwPharm++), MM-USC*PACK© (now BestDose) and T.D.M.S. (now PrecisePK). MwPharm/MwPharm++ achieved in both evaluations one of the highest rankings (first, resp. second). MwPharm++ [17] is continuously complemented with new drug models. In this work it appeared that a tool for supporting

the model development is needed. Thus, in collaboration of the MediWare company with the Czech Technical University in Prague research and development of an intelligent search robot was initiated.

3. Intelligent Literature Search

The MwPharm++ project (software for drug metabolism modelling, in which validated models are implemented) has been expanding into foreign markets over the last 2 years and is currently used in 18 countries worldwide. In connection with the completion of the cloud version of MwPharm Online, the project will upgrade to the category of "software as a service". The project consists of 2 parts:

- a custom algorithm for patient dose adjustment based on a pharmacokinetics and pharmacodynamics (PKPD) model using genetic knowledge related to drug metabolism and
- an ever-expanding database of drug substances that contains PKPD parameters, knowledge of possible polymorphisms, and the literature and factual sources from which these values were determined.

The main task is to specify PKPD models for a selected set of drugs and for the given population (ethnicity is one of the parameters that influences the drug pharmacokinetics and pharmacodynamics). Literature search is an essential step in the procurement and validation of a new drug. The intention of the designed tool is to speed up and improve the process of conducting a literature search by automatically extracting relevant articles using both the tools utilized in natural language processing and by taking advantage of the automatic processing capabilities of the retrieved results. It is worth mentioning that the publishing activity in the area of interest is very intense, with about 10 articles per week. This results in a high number of articles (depending on the number of drugs) related to the problem. With such a large number of papers, it does not make sense to present the results only in the form of a list of publications, but it is necessary to evaluate and structure the results for maximum clarity and the possibility of systematic work. The work to date has established a basic terminological basis for the representation of (abstracts of) articles, the method and conditions for their selection, and the method and possible forms of creating a structured, systematic review for the selected literature.

So far, the available search engine resources (Google, PubMed) have been used to build the database, with subsequent manual selection of suitable papers. As the requests for new drugs and new population groups are increasing, we decided to simplify this process by implementing a robot whose function aims not only at searching and selecting articles, but also at systematically classifying the retrieved literature sources in order to generate a library of papers on the population parameters of the proposed new drug (with the possibility of user annotation). Currently, the system is mainly used for TDM (therapeutic drug monitoring), which involves about 200 drugs out of the total number of drugs in use. As the new cloud-based version of MwPharm Online also allows for first dose reduction in case of renal or metabolic insufficiency of the patient, we are extending its functionality to a total of 1500-2000 drug substances for each population group. Initially, the information needed to determine the PKPD parameters of new drug substances was obtained manually at a cost of approximately 400 EUR / drug. With the new robot, this cost should be reduced to 100 EUR/drug. For the 2 population groups used so far - the Transcaucasian population and the Asian population, this means a saving of almost 2/3 of the costs incurred so far. For example, for the Transcaucasian population

and 2,000 drugs, this is a saving of more than 350 thousand EUR and a similar benefit can be expected for the Asian and other populations. Considering 500 most used medicines for 1 population, the savings compared to the original costs is about 40 - 80 thousand EUR. If we consider that this is an ongoing process where we have to add and update a lot of information on specific medicines every year, this is an ongoing annual saving.

4. Description of the Intelligent Search Engine

As a reference project, we take article retrieval for model validation for a list of drugs provided by the Korean project partner (part of the project No. TF05000020 MwPharmASIA - database extension of drug substances and their MwPharm models for East Asian population and development of NGS diagnostic panel and algorithm for predicting statin pharmacokinetics/dynamics of the Technology Agency of the Czech Republic). This set contains 38 drugs for which we wanted to see if PKPD models can be specified/validated. Furthermore, collaboration with experts in the field of PKPD model building was then used to determine their part of the terms for the text search. Collaboration with these experts was done in an iterative manner to obtain the best quality term sets for text retrieval. Until now, no such standardized sets existed (at least for our subject area).

These documents are used in the form of a dictionary. The dictionary forms the basis of each project created within the application developed. The use of dictionaries is one of the standard methods for text processing by assigning entities to the required (searchable) parts of the text. The user can basically view the dictionary in its three levels. Each dictionary is thus divided into three levels, representing categories, entities and the search words themselves. Primarily, further processing works with entities that represent a defined set of words. However, from the searches carried out, it became clear that for a clearer orientation in the selection result it is advisable to define a higher level, i.e. the category of entities, in certain parts.

After defining the dictionary, the next steps are available. The first step is to collect the articles. We chose the PubMed database as the primary source of articles for this prototype because it contains articles that have been subjected to peer review. Thus, the resulting extraction contains articles that can be considered valid in terms of expertise. The PubMed database has an advanced search system that allows searching for combinations of terms using the logical AND and OR operators.

The query is created automatically either based on entities in mandatory categories or (if no category is marked as mandatory) based on entities in all categories. The complication is that the search

- is either very rigid and not many relevant papers are included in the resulting selection,
- or the search rule is loose and the search results in thousands, tens of thousands or even hundreds of thousands of articles.

With a broader coverage, which is the reference in a case like this, the numbers of articles from a PubMed search are very high, and even automatically clustering them according to the occurrence of each class of interest would lead to an unprocessable result. Therefore, the step of selecting articles from PubMed by our independently developed system is quite relevant.

To illustrate the complexity, we show an example: As of December 23, 2019, there are 14980 articles on PubMed corresponding to our query, of which 14912 abstracts are in the internal working database - where the texts are already there after tokenization and lemmatization. The same number (14912) was screened and tagged, and 414 articles were finally selected as best meeting the dictionary-defined requirements. These 414 articles are being worked on further. It is obvious that the reduction of number of articles that have to be reviewed by human experts is significant.

Part of the file tagging is the calculation of the score of each article. For this purpose, a weight is added to each defined entity. However, the user does not directly define the numerical value of the weights, but gives a hierarchy of entities, whereby any number of entities can be at the same level.

The user is not required to define the weights and can leave all entities with the same weight (weight 1). However, s/he can use the weights even if none of the dictionary categories are defined as mandatory and the level of importance of the article is then defined only by the resulting score of the article. In our benchmark example, a combination of both approaches has been chosen, i.e. we have defined 3 categories as mandatory and the weights of each entity are defined at the same time.

If a dictionary is defined and the supporting documents are ready, the user can specify browsing and tagging of articles by dictionary. Here, the result of the search through the PubMed database interface serves as input to our extraction algorithm. The final selection of the articles is based on the dictionaries defined, the weights of each class (entity) defined in the dictionaries, and the algorithm for arranging the results. The output of this part of the processing is a primary ordered list of entities from each category for each of the selected articles. This tagging is available to the user in three different forms of representation.

The first of the possible representations is a tabular form. This output is easy to see for many users, but it is true that the results of the article tagging are stored in Microsoft's Excel format. It is therefore possible to open them in this program and make any further modifications to the output. A far more meaningful representation of the results is the tree structure, which is formed by creating classes for combinations of entities (there can be, but also one single entity) from each category. An example of detail of such an output is shown in Figure 1, in which it can be seen that the order of the categories is population, drugs, pharmacological information, characterization and the last one is the article itself.

Figure 1. Detail of the resulting tree

However, this order is not fixed and the user can modify the order of the categories. The order of the entities has a relatively strong influence on the clarity of the result tree, so it is a good idea to choose it appropriately.

In any case (for any dictionary), the article itself is part of the result. In the tree representation, a link to the PubMed database is added to this tree node, so the user can open the article abstract by double-clicking.

The last form of representation of the results from dictionary tagging is a graph. This form of representation can be very useful for identifying common articles and selecting those for download and study. However, it is true that this form of representation needs to be handled very judiciously to make the resulting graph workable. Therefore, in order to give the user the best possible experience when reading the results of the selection in graph form, the user has two options for editing the input data (labelling results). The first is the option to select the entities to be omitted from the graph. Thus, the user can define that s/he wants, for example, only articles containing the entity Korean and no others. This does not only mean removing the nodes of the other populations, but a very significant reduction in the total number of nodes, because the entity Korean does not connect to many entities from other categories (there are no links connecting them) and so these entities (nodes of the graph) also drop out. The second option is then to select the links themselves. Again, the individual articles in the leaf are linked to the url and a double-click leads to the selected article in the PubMed database.

However, the results of tagging articles using dictionaries have several drawbacks. The biggest of these is the fact that there is no simple coincidence of words within the text. A dictionary is thus a very quick and convenient tool in terms of definition by experts, but it is advisable to complement the results given by it with some more comprehensive form of text representation. In this project, we made two assumptions:

- we are interested in the relationship of entities, or their proximity in the text, and the filtering of the text into classes is of secondary interest to us (the dictionary has proven itself in this respect);
- we work primarily with short texts (abstracts), for which forms of representation such as term_frequency/inverse_document_frequency are common.

We have therefore chosen the standard text representation in the form of word n-grams. In this representation, the text is sequentially split word by word into a combination of n words. The n can be chosen as a number from 2 (bi-gram), 3 (tri-gram), 4 (quatro-gram) and even higher. However, there is of course a logical limitation in the number of words for which text sampling make sense. We have chosen to create 2-gram, 3-gram and 4-gram models and the user has the option to choose which n-gram model to use to label the selected articles. An example of a window listing 4-grams defined based on dictionary-defined articles is shown in Figure 2.

The process of marking and clustering cells is then divided into the following steps:

- Each n-gram has stored with it which entities represent it and which articles contain it. Here it is very important to note that n-grams are made up of the words of articles, but these are represented by entities in our system. Thus, two different n-grams can be described by the same entities.
- A transformation is made where entities are the key and n-grams and articles are subsets of them. This step leads to the clustering of the n-grams. This reduces the dimension of this representation (example from the reference set: the original 13624 4-grams are represented by 334 entities or combinations of entities)

- In the next step, a transformation is performed where entities are atomized (until now they could represent n-grams in combinations) and assigned the value 1 if they are represented in the given n-gram.
- The final step is to convert the previous representation into articles, i.e., each n-gram is replaced by the articles it represents. Since one article can be represented by multiple n-grams, the occurrences of each entity are summed. To reinforce the role of the combination of entities within an n-gram, this is weighted - here, if an entity occurs repeatedly in an article but alone, its value within the vector of that article is increased by +1. If an entity occurs in combination with another entity, the value within the vector for both is increased by +10, in the case of three entities it is increased by +100, and in the case of 4 (we are working with a 4-gram in the description) it is increased by +1000.

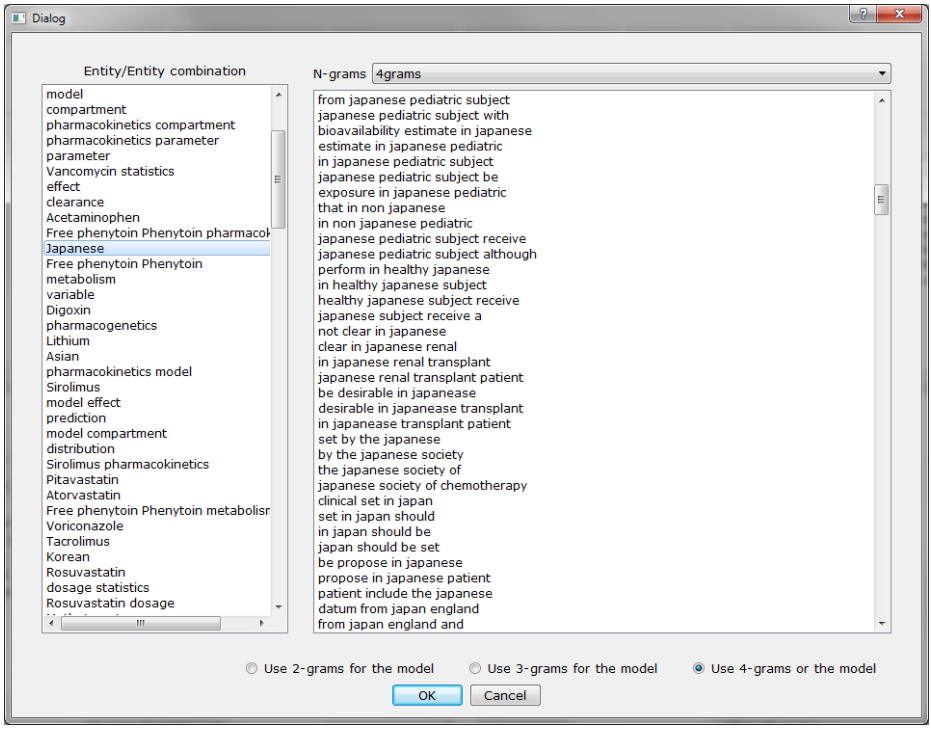


Figure 2. Overview of 4-grams extracted from the articles selected by the dictionary

By this procedure, we have obtained a vector representation of articles that takes into account combinations of entities in n-grams. We can now perform the standard hierarchical clustering process, using cosine similarity as the similarity metric. The result of this process is represented within the application in two ways. The first is a tree representation, an example of its detail is shown in Figure 3.

The second is the dendrogram representation of the clustering result. In this form of representation, the user can again define which entities will be included in the comparison and can then choose which articles they want to see in the display. The items

in the article list are ordered according to the ordering (clustering) in the dendrogram and again serve as links to the PubMed database for the selected article.

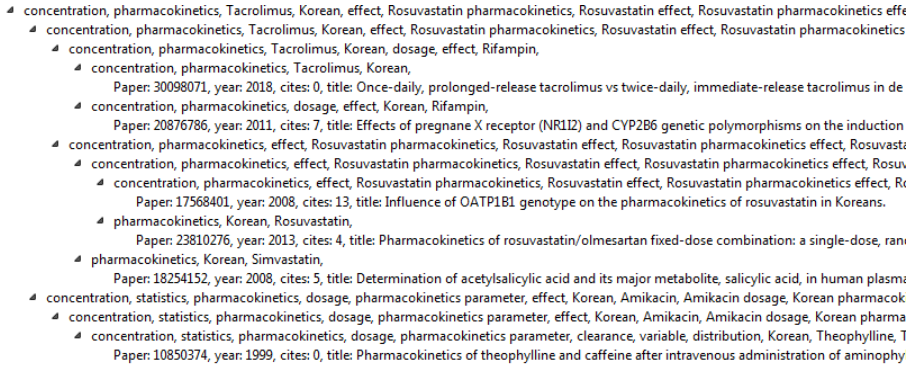


Figure 3. Detail of a tree generated from the clustering results for a model based on 4-grams from the reference example

5. Conclusion

The proposed system allows for functional search, sorting and visualization of literature resources with output storage that allows to refine one's own library of articles. The system is in the prototype stage, which enables all the declared functionalities. Currently, the next phase of evaluation of the system is underway and the evaluation of functionalities and elements that should undergo further development for the most optimal use of the tool.

The studies presented in the Section 2 Related work show that the demand for model-informed precision dosing software tools will grow and that it may positively influence the whole area of precision medicine. Thus, all software tools that can support the process of development of new drug PKPD models are and will be useful and effective contribution.

Acknowledgements

Described research is supported by the project No. FV40426 “Supplementary Platform for Physiologically Based Pharmacokinetic Modeling (PBPK)” funded by the TRIO program of the Ministry of Industry and Trade of the Czech Republic.

References

[1] Rosenbaum SE. (ed.). Basic Pharmacokinetics and Pharmacodynamics. 2nd edition. Wiley, 2017
[2] Polasek TM, Ambler K, Scott HS, et al. Targeted pharmacotherapy after somatic cancer mutation screening. F1000Res. 2016;5:1551
[3] Darwich AS, Ogungbenro K, Hatley OJ, et al. Role of pharmacokinetic modeling and simulation in precision dosing of anticancer drugs. Transl Cancer Res. 2017;6(S10):S1512–S1529
[4] Rini BI, Garrett M, Poland B, et al. Axitinib in metastatic renal cell carcinoma: results of a pharmacokinetic and pharmacodynamic analysis. J Clin Pharmacol. 2013;53(5):491–504

[5] Tängdén T, Ramos Martín V, Felton TW, et al. The role of infection models and PK/PD modelling for optimising care of critically ill patients with severe infections. Intensive Care Med. 2017;43(7):1021–1032
[6] Vinks AA, Emoto C, Fukuda T. Modeling and simulation in pediatric drug therapy: application of pharmacometrics to define the right dose for children. Clin Pharmacol Ther. 2015;98(3):298–308
[7] Morgan ET, Dempsey JL, Mimche SM, et al. Physiological regulation of drug metabolism and transport: pregnancy, microbiome, inflammation, infection and fasting. Drug Metab Dispos. 2018;46:503–513
[8] Peck RW. The right dose for every patient: a key step for precision medicine. Nat Rev Drug Discov. 2016;15(3):145–146
[9] Polasek TM, Shakib S, Rostami-Hodjegan A. Precision dosing in clinical medicine: present and future. Expert Review of Clinical Pharmacology. 2018; 11(8): 743-746. https://doi.org/10.1080/17512433.2018.1501271
[10] Elemento O. The future of precision medicine: towards a more predictive personalized medicine. Emerg Top Life Sci. 2020 Sep 8;4(2):175-177. doi: 10.1042/ETLS20190197.
[11] Akhoon N. Precision Medicine: A New Paradigm in Therapeutics. Int J Prev Med. 2021 Feb 24; 12: 12. doi: 10.4103/ijpvm.IJPVM_375_19. eCollection 2021.
[12] Allerheiligen SR. Impact of modeling and simulation: myth or fact? Clin Pharmacol Ther. 2014;96(4):413–415
[13] Dugger SA, Platt A, Goldstein DB. Drug development in the era of precision medicine. Nat Rev Drug Discov. 2018 Mar; 17(3): 183–196. doi: 10.1038/nrd.2017.226
[14] Peck RW. Precision Dosing: An Industry Perspective. Clin Pharmacol Ther. 2021 Jan; 109(1): 47-50. doi: 10.1002/cpt.2064.
[15] Fuchs A, Csajka C, Thoma Y, Buclin T, Widmer N. Benchmarking therapeutic drug monitoring software: a review of available computer tools. Clin Pharmacokinet 2013 Jan;52(1):9-22. doi: 10.1007/s40262-012-0020-y
[16] Kantasiripitak W, Van Daele R, Gijsen M, Ferrante M, Spriet I, Dreesen E. Software Tools for Model-Informed Precision Dosing: How Well Do They Satisfy the Needs? Front. Pharmacol., 07 May 2020, |https://doi.org/10.3389/fphar.2020.00620
[17] Mediware. MwPharm++ Therapeutic Drug Monitoring Tool, retrieved from: https://mediware.cz/en/mwpharm-documentation

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210573

eHealth Turning Points as Forced by the Covid-19 Dramatic Experience

Francesco PINCIROLI[1]
Dipartimento di Elettronica, Informazione e Bioingegneria – Politecnico di Milano

Abstract. In this ongoing fall of the year 2021, many disciplines are frightened by the Covid-19 situation. A generalized sense of Scientific and administrative impotence, - in keeping the pandemic under real control, - is felt widely in Society. In this Invited Lecture the author reminds us of the blows suffered, recalls pertinent elements present in our social organization, browses selected eHealth experiences and proposes an open agenda of actions to allow the eHealth to help the population segments better, and individuals as well.

Keywords eHealth, pHealth, Covid-19

1. Foreword

In appreciating and accepting the opportunity of an Invited Lecture at pHealth2021, I take the liberty, and the responsibility as well, of sharing with the audience of our community the major professional eHealth reflections that have come to my mind during the Covid-19 not-yet-concluded events. My default point of view in living in such a period is that of a "senior citizen and retired Professor of Bioengineering and eHealth", who spent his entire academic career in the field [1-8]. I take this liberty in the hope of being of some help in letting our Community focus and discuss possible turning points to be considered while continuing to attempt to understand, develop and design new directions that our field will take, or even be forced to take. In putting my liberty into practice, I believe that my main – and more or less only - reference is my long-lasting academic background. Therefore, my expected result is a document to be placed on our community table for discussion. We need this discussion to be rapid, otherwise we will just be overwhelmed by the development of events largely external to us.

2. The Blows

I perceived some of the events related to the Covid-19 situation as real blows. Here they are.

In the multi-steps eHealth pipeline from data to knowledge, our Covid-19 professional drama started with so many people dying rapidly, in so many different Countries, with no credible data, and no time to verify the number of deaths. The footage of rows of military trucks transporting the coffins of the deceased – at night to an

[1] Corresponding Author: Francesco Pinciroli FIAHSI, Prof. Emeritus, Dipartimento di Elettronica, Informazione e Bioingegneria – Politecnico di Milano, Milano, Italy Email: francesco.pinciroli@polimi.it

unknown location from the medium-size Italian city of Bergamo, in the Lombardy region, since there were not sufficient resources to cremate them: these powerful images shown on the TV News hit my mind and constricted my heart. To me, this scene also included the perception of failure in engineering, a failure where the much-appreciated ICT for Healthcare gave zero-help. The high transmission speeds reached by our ICT infrastructures over recent decades were unable to provide effectiveness. We were not able to give help in a situation of contagion. This blow to my engineering background was even greater than the one I had had a few months earlier, when on August 14, 2018, the Morandi Bridge, the main highway bridge in the center of Genoa, collapsed.

We have in our history the historical success of about 50 years of eHealth: from Medical Informatics helping statistics, to ICT for health, even adding fantastic just-in-time skills, allowing people to be saved in the event of sudden heart attacks. Later, we started to do Data Mining, sustained by the large abundance of data produced free-of-charge. Then came eHealth, which includes easy-to-be-immediately reshaped display facilities. All of this was Science. But the Covid-19 period started without data, apart from those counted for infected, dead and hospitalized subjects. If data are not generated, we do not have any eHealth segment to play with. Without data, our Scientific Community is outside the circle helping our Society.

Immediately after the emergence of Covid-19, from month to month we have seen virologists on TV addressing the TV audience with messages meant to be or be perceived as different from one another, even if none of them was fake. Despite our attitude to label the eHealth approach "comprehensive", the methods and the technologies we are familiar with have not proved helpful in letting the audience, and the virologists, come up with better truths themselves.

3. Remarks about our Social Organization

During the Covid-19 period I matured and/or renewed some obvious observations, that I consider pertinent. They do not have the sudden power of a slap. Nevertheless, they can be even more relevant in the process of building the future of our Community. Here they are.

A judge, a priest, a game player, a leader, a physician: these were necessary roles made and identified since the beginning of mankind. However, scientists, and even engineers came later, after some tools had become available, tools to be used in the management of appropriately selected and significant data and materials, and to obtain results that can be repeated. So, as scientists, we are late comers in the Society, while Covid-19 is an ancestral player in Society's evolution. Nevertheless, Society believes in us because we do not change the knowledge we manage; once released, it is stable. It does not include mistakes.

We have to recognize and accept the substantial, intrinsic inability of scientists to manage the population governance at a national level. I mean that all of us, engineers and related disciplines, have a background that grew up under the umbrella of those experimental approaches where somebody else is allowed to repeat your experiments, and he/she is supposed to obtain the same results. Instead, the population governance is requested to be, at any level, the capacity of finding an ongoing baricenter among inhomogeneous needs and/or desires, and to do this, keep society in quiet relationships between citizens. So, to be a governor, a significantly different background is required from the scientific background. Having recognized this, our eHealth Community holds

the role to continue to do eHealth, in support of the Country's governance. The Covid-19 event also triggered rapid changes in yesterday priorities, to tackle the lack of knowledge and to provide practices for the need for governance.

Although each of us is a scientist, none of us is able to persist in behaving fully as a scientist when he/she is in front of the TV. Here we are just citizens, exposed in full to the technicalities of the communication-for-persuasion arena. I think about a car advertisement on TV where, for years, they did not include any information about speed, acceleration, brakes, gas consumption, and other historically technical quantities. Instead, they outline the emotion that the car generates in the owner's soul. And, even worse, the advertisement on TV for hearing aids just says that they cannot be seen by anyone you meet. Not a single technical detail, about how and why our auditory functions might be improved, is forwarded to the audience.

Still, in recent years, my experience in doing home eHealth saw a hospital-dependent variety of accreditation modalities. The public hospitals in my Region attempt to follow a common modality. But each private hospital has its own way. This occurs even if it has been accredited, i.e., verified by the Regional Healthcare Authority as being able to provide the citizen with the same level of care as a public hospital. So, in my role of citizen/patient, for clinical examinations, I choose public or private according to the shortness of the waiting lists. If I wish to put together all the clinical data I have generated in different places, I have to enter each hospital website with its/my credentials. Clinical and healthcare databases are not aligned. At my age, my accreditation modalities occupy about an entire page of my agenda. I do the grouping on my PC. I never received alerts – not from any public or private hospitals, nor by the Regional Healthcare Information System – about any bad – and even hacking - maneuvers attempted on the websites where some of my healthcare data are resident.

4. Browsing on eHealth Experiences

In the middle of September 2021, Italian newspapers published the new Laws on the "Green Pass", since this Covid-19 document is a top subject. Its eventual mandatory level and the society segments to be submitted to such a policy are under a wide political and social debate. Nevertheless, under an eHealth point of view, that focuses on how to collect the data only to release it to any citizen, and how to deliver it to the recipient, the "Green Pass" is not a problem. Let me just mention that all the members of my family have it, including my grandchildren, whose ages range from 21 to 12. I easily downloaded my "Green Pass" after having received an SMS including the instructions. This message arrived after I had had my second injection of the vaccine. The eHealth procedure is extremely simple and effective. We are in front of an eHealth case of success.

We can go back 50 years, in order to start helping Statisticians again to do their jobs better on populations, to compare sub-groups of populations, to foresee evolutions, etcetera. Similarly, we can start again in helping physicists to plan their instruments better to detect, identify and rank this new Covid-19 virus and its variants too. Within Pulmonary Intensive Care Units, we can empower the storing capabilities for bio signals. In doing so we facilitate the pointing out of a better taxonomy of what Covid-19 does on our pulmonary system. But this is for the future. At present people are dying, and dying fast, with low connections to the contents of the hypothesized "Personalized Life-Long Healthcare and Clinical Record" our eHealth was "almost-ready" to provide.

For some years, our eHealth Community has had to worry about the wide spread of Fake News. Even worse for my mentality, I had to learn from psychologists that, in navigating healthcare information on the Internet, citizens are inclined to follow a personally sustainable and step-by-step pathway towards the capture of elements capable of supporting their prejudices. So, by doing eHealth, it is not enough to produce new knowledge and to make it easily available to potential recipients. We must also be concerned about convincing such recipients that the knowledge we offer is a good one, eventually providing such persons with methods and tools for unmasking fakes. And what our community would do when faced with citizens highly inclined to give value to sources that only resonates with their inclinations/prejudices? Can "Blah … this is not a matter for eHealth!" be an answer? Probably not, as it can devalue much of the comprehensive eHealth approaches and results. We are facing the persisting absence of suitable and instant-to-be-used confidence indicators for data sources, tuned to the basic understanding of the envisaged users.

Historically, as citizens, when looking for clinical and healthcare information, we use "physical signs". For instance, when at the newspaper shop I buy the "weekly healthcare focus" added to many newspapers, I foresee that I'll read it easily. I won't have any problems in understanding what I read. However, what I read will be nothing more than a basic generic introduction, even when I find myself interested in the title of the article. Normally, something opposite occurs when, as a citizen, I enter the PubMed website, where I can do queries based on the scientific and specific name of my symptoms and diseases. This behavior of mine will lead me to select scientifically valuable and focused information sources, but I do not have the sufficient background knowledge to understand what I read. Unfortunately, by entering the generic Internet arena, we do not have significant symbols available, even if something starts to exist. For instance, as a citizen, I believe that the websites of some well-known and credited hospitals provide a reasonably trustworthy symbols. Of course, they must be tuned to the asymmetric medical lexicons belonging to specific segments of the population.

The main gate in my courtyard, in the central district of Milan, overlooks a narrow sidewalk. When I exit through the gate, I look left and right. I do so "normally", with the expectation that the sidewalk could be occupied by pedestrians. My body's sensor and actuator systems are tuned to the expected speed range of a human walking. For a couple of years I have had to change this assumption of mine. I started doing so from the moment when, on exiting the gate, I was lucky not to be hit by an electrical scooter running on the sidewalk. This fact was for me a "lesson learned", from a sudden experience, fortunately for free. Such racing was not (not yet?) illegal, because we were at the very beginning of the emergence of the electrical scooter market. In other sudden cases or situations, the "lessons learned" can be costly, even physically. And the sudden Covid-19 situation has changed our best practices, with such fast timing that we know it is impossible to be adopted by our eHealth Standardization Bodies. This makes the discovery of new Best Practices in eHealth more important than what it usually is.

In recent years, the experience doing home banking has seen the methods of accreditation widen, up to the current way consisting usually of username + password + OTP code, the latter received on the smartphone is only valid for a few minutes. In truth, the expansion came from something bad which occurred while using earlier more limited methods. Regarding these bad things which, in the end, happened to someone else, my bank never told me any details. Nevertheless, they sent me warnings such as "Never update your credentials by answering an email and using the included links! We never use this modality!". Moreover, they often add "please carefully verify, letter by letter,

our exact name as the correct sender of an email to you." And, when I receive something strange, I take a picture of it on the PC screen and I send it to the contact person linked to my bank account. Of course, I only do this action from within my section on the bank's website. To reach my section, I would have had to use my updated credentials in full. Subsequently, usually, the contact person sends me a reply, within the protected area, where they thank me for having alerted them. Let me label what I have just described as "Lessons learned from the user by interacting over time with system/platform management". Both progressive and interactive accreditation modalities should be recommended during eHealth contacts between the patient and healthcare platforms.

Obviously, I share the sense of deep sadness in the face of the many millions of deaths by Covid-19. However, given the audience that we are, let me synthetize that worse was caused by a lack of knowledge. At present we see that every Country is trying to amend the situation by uncovering some lines of reactions. For sure they ask for investments, that the governments of our Countries envisage as "recovery funds". A generally accepted line along which to use such money is "Innovation", which is widely recognized as including the "digital transformation". The eHealth infrastructures do not show any difficulty in being recognized as part of the "digital transformation". In addition, eHealth and Telemedicine confirm the promise of getting "things done even at a distance", helping to counteract the close contacts still necessary during the Covid-19 contagion. So, it seems certain that eHealth and Telemedicine can count on a significant increase of funds available for their programs. It matters little that, despite the efforts made in the recent decades, Telemedicine did not exceed the threshold to become widely accepted in everyday practice. With Covid-19, the modality of acting from a distance has emerged as extremely necessary. It is a life-saving necessity. It is no longer a cost/effectiveness approach, as it was before.

There are reasons why eHealth never considered the general practitioner's office as a significant and structured source of data [7]. In such an environment few standards are used, and the generated data are not a lot. Furthermore, the information generated by patient-physician interactions often belongs to conversational language. The evasion of "sending the patient to a specialist" is an open door for GP behavior, sometimes also due to elements generically belonging to the area of Defensive Medicine. But the Covid-19 situation does not leave this scenario unchanged. The awareness of the relevance of the timely recognition of emerging symptoms and the alarming risk of the patient causing a risky contagion, even within his/her family, are the major changes. Also, because, probably in every Country, the current dimension of the hospital departments proved largely insufficient for Covid-19 patient populations. In front of them, eHealth will never be able to decrease the chances of contagion by taking food to the door of the apartment where a patient is obeying his/her quarantine duties. Nevertheless, eHealth can think about actions dedicated to the GP outpatient environment with more conviction than what has been done in the past, also counting – since this is new and facilitating, - on high attention and interest from the government.

As happens in any significantly established segment of the knowledge, and even the market, our community has generated a certain number of associations over the decades. Each of them is now faced with what is generated by the Covid-19 situation. The activities of some of them are published on their websites, usually under a branch that has a name that includes the term Covid-19. The associations do so unavoidably bearing in mind their own survival, in order to not be overwhelmed by the generalized perception of the weaknesses of their own tools and infrastructures in the face of so many deaths. The high number of scientific publications found by PubMed when we do an "eHealth +

Covid-19" query is certainly significant. [By September 15, 2021, PubMed counts some 6365 results.] But all such activities come from the attempt to extend to the appearance of Covid-19 the research approaches already designed before it, when those researches were conceived, i.e. well before Covid-19. Our community must promote the design of entirely-new research.

The inclination of simply adding "Covid-19 oriented" to previously conceived and existing programs and approaches does not ignore commercial products [9]. In reality, the availability of "recovery fund" resources may exacerbate the situation. If this happens "too much", we will shortly face the risks of losing both government money and our credibility. We have to avoid this happening. We have to be the first and effective referees of the appropriateness of the emerging eHealth products said to be of help in the Covid-19 pandemic.

Due to the Covid-19 situation, we need to anticipate changes in the curricular education. Let me think about the Biomedical Engineering pathway, taking into account the four historical pillars of Mechanics, Electronics, Chemistry and Information Science. I believe that the students' educational paths would be strengthened in Biostatistics, should include basics notions in Epidemiology and in Clinical Laboratory, and Lab-on-a-Chip machines. Other seminar additions to include would be Patient Rights and Healthcare Data Protection, basic Psychology of the Internet, users of Clinical and Healthcare Information Sources. Research on Instant Scoring Tools of Information Sources and on Instant Data Science would also be promoted. Symmetrically, for the curricula in Medicine and Life Sciences, we should become able to define the contents to offer distance seminar courses in "Instant eHealth".

5. Discussion

As after any war, we respect the dead and we have to continue to work hard for the care of the survivors, particularly when some of them still need to be treated for diseases other than Covid-19. eHealth is a mature adult in its 50's, i.e., it is an experienced entity, that still has a lot of power to single out innovation and to keep it tuned to the real needs of the entire population [10]. So, our Community will continue to work hard. Our role as "late comers" awaits the additionally new and robust scientific knowledge that we will help to discover.

A question which we will never know the exact answer to is: how many more deaths would have occurred in the current not-yet-ended pandemic period if we had not had the normally running ICT-for Health infrastructures available, even if we say every day that they should work better?

6. Conclusions

According to what has been comprehensively caused in healthcare by Covid-19, there are some lines of development that we need to start caring about more than in the past [9]. Given my foreword and after the reflections I have shared with you here above, I endeavor to suggest a possible open agenda of courses of action.

Find indicators and/or "markers" of trust for healthcare and clinical information sources that can be used "instantly" by users, including "rapid tests" that the user can do on the source to establish whether he/she can believe it or not.

Take the Standardization Bodies involved in setting up procedures intended for "rapid best practices"

Revise the contents of the educational pathways of any level, centered on and/or including eHealth.

The General Practitioner office should be considered, studied and helped more than in the past.

We as individuals, as well as through our eHealth Associations, have to remain as close as we can to the Healthcare Government Offices and Officials, including the level of Ministries, especially in the role of eHealth auditors, and also informed promoters.

We have to act by staying intelligently alert and by being respectful professionals, skilled designers and rigorous scientists.

References

[1] Pinciroli F, Anderson J (ed.) Changes in Healthcare Instrumentation due to Microprocessor Technology. Amsterdam: North Holland; 1981, pp 326
[2] Meester GT, Pinciroli F (ed.) Databases for Cardiology. Dordrecht: Kluwer Academic Publishers; 1991, pp. 418
[3] Pinciroli F (ed.) Informatica di Base per la Medicina. Torino: UTET; 1992,pp 661
[4] Cristiani P, Pinciroli F, Stefanelli M (ed.) I Sistemi informativi Sanitari. Bologna: Pàtron Editore; 1996, pp 300
[5] Pinciroli F, Masseroli M (ed.) Elementi di Informatica BioMedica. Milano: PoliPress; 2005, pp 382
[6] Pinciroli F, Bonacina S (ed.) Applicazioni di Sanità Digitale. Milano: PoliPress; 2009, pp 390
[7] Marceglia S., Ferrante S., Bonacina S., Pinciroli F., Lasorsa I., Savino C., Pozzi G., Domains of Health IT and Tailoring of Evaluation: Practicing Process Modeling for Multi-Stakeholder Benefits. Stud Health Technol Inform. 2016; 222: 63-76
[8] Quaglini S, Cesarelli M, Giacomini M, Pinciroli F (ed.) eHealth-Medicina Digitale. Bologna: Pàtron Editore; 2017, pp 362
[9] Scott RE, Mars M. Editorial Covid-19 and eHealth: A Promise of Peril Paradox?, J Int Soc Telemed eHealth 2021;9:e1
[10] Demiris G (ed.) E-Health: Current Status and Future Trends. Stud Health Technol Inform. 2004; 106, pp160

Data and Knowledge Management and Methodologies

This page intentionally left blank

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210575

Sentiment Analysis on USA vs. New Zealand on Health and Safety Mandates During Early Stages of COVID -19 Pandemic

Joshua DALES[a], Farhaan MIRZA[a], and Amr ADEL[a & b]
[a] *Auckland University of Technology*, *Auckland New Zealand*
[b] *Yoobee Colleges, School of Information Technology, Auckland, New Zealand*

Abstract. The Coronavirus pandemic has surprised the world and social media was extremely used to express frustrations and development of the cases found. Social media tools, such as Twitter, show a comparable impact with the number of tweets related to COVID -19 indicating remarkable development in a limited ability to focus time. The purpose of this paper is to investigate the impact of Coronavirus on the United States of America (USA) and New Zealand (NZ), and how that is reflected in a sentiment analysis through the examination of American and New Zealand tweets. We have gathered tweets from a March 2020 – August 2020 and used sentiment extraction on the tweets. The major finding of this sentiment extraction is the fact that the overall average sentiment over the 5-month period stayed in a negative range in the USA and NZ. This paper aims to analyze these trends, identify patterns, and determine whether these trends were caused by the COVID -19 pandemic or outside sources. One trend that was analyzed was the spike of COVID -19 results in relation to the number of protests occurring in the USA.

Keywords. Sentiment Analysis, COVID -19, Coronavirus, Data Request, Twitter.

1. Introduction

COVID -19 is a vital area to investigate in this current time. This will be analyzed through a Sentiment Analysis of 10,000 tweets during the period of March to July. Sentiment analysis is the process of organizing and grouping large amounts of text together based on certain facts contained in the text. The text is categorized into groups based on feelings such as positive, neutral, and negative. By grouping the text into several categories, it can provide a greater understanding of the material. The topic area for this research is «Sentiment Analysis of COVID -19 and how the understanding of it changed general sentiment towards the virus». The motivation of this research focuses on the sentiment analysis on twitter data because it gives us the option to see how sentiments towards COVID -19 grew or reduced over a period. This research has the potential to aid future researchers in understanding how normal people will react to a pandemic. The findings of this research are useful for assessing a specific research gap. The research gap is a social comparison between America and New Zealand. Governmental branches in America have made several attempts to suggest a health and safety mandate that were created to reduce the spread of COVID -19. Section 3 focuses on the research methodologies and how the data was obtained and validated.

Section 4 goes into the research findings and identifies themes, patterns, and areas for analyzing. Section 5 is a discussion part focusing on possible reasons for why the USA has such high numbers of COVID -19 cases. Furthermore, it analyses why New Zealand does not have such high cases. Finally, Section 5 goes into a comparison and identifies what New Zealand did that the USA might not have been able to do. Section 6 focuses on the Conclusion of this research.

2. Literature Review

Previous studies have used sentiment analysis & social media to discuss how public has reacted to public health issues. For example, one existing solution that has attempted to use data mining techniques to understand the sentiment of collective outspoken citizens is the research done by Darliansyah et al in their report of Long-Term trends in public sentiment in Indian demonetization policy. It is unlikely that the results found in analyzing data contributing to COVID -19 will be as clear cut as Demonetization trends, but through this we can begin to understand how large audiences from a variety of countries can react to drastic change. That report analyzed the sentiments of 41,724 tweets by Indian citizens through a period of 9 months; what they found was that only 33% of the data was positive with another 32% being negative and only 34% of the data being neutral to the changes.

Another study reported on similar areas of study and conducting sentiment analysis to understand greater audiences. A report done by Barkur, et al. (2020) [10]. Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India, shows evidence that in India, that many citizens responded positively to the government lockdowns, roughly 25000 tweets show a positive lean-to going into a lockdown. What this report shows is that depending on the country-of-origin reactions to a countrywide lockdown will differ greatly. The primary negative sentiment towards the imposed lockdown was one of fear, a sentiment that was shared by 10,000 other citizens.

A third report conducted by Matosevic and Bevanda (2020) [9] that studied a Sentiment analysis of tweets about COVID -19 disease during the pandemic. What this report shows is that in countries that handled COVID -19 in safe way their citizens tweet emotion rating was “Trust” whereas in countries that were struggling with COVID -19 their populations tweets tended to range on the “fear” side of the emotion spectrum. This report studied outgoing tweets from USA, UK, Spain, Italy, Sweden and Germany, this report made a distinction between politicians and the public and determined that the public and politicians shared similar emotions when tweeting about COVID -19.

Through understanding the report by Davison (2017) [11] we can understand how the public’s opinion shifts. Through this understanding, we can identify and analyze how data opinions shift over time. This is due to how more evidence or facts become relevant. Based on this report, we can theorize that in the beginning of the COVID -19 pandemic the public was not affected, nor did they think it was a problem. But as COVID -19 started to spread globally and more information came out public opinion would shift, and they would become more interested in stopping the spread. It can be hypothesized that such sentiments would be positive in the face of policies that would be implemented to slow the spread of COVID -19. Additionally, negative sentiment would be based on those of the public who either do not believe COVID -19 is a problem or by others who are not happy with the policies and what change they would bring.

Through the reviewed literature this paper was able to recognize and understand how sentiment analysis can aid in understanding how a countries sentiment towards a specific topic can shift as new information is introduced. Evidence provided by Davison (2017) [11] reports that opinions shift over time as more knowledge and facts are provided. This trend is also observed in Barkur, et al. (2020) [10] where a sentiment analysis was conducted in regards to the opinions of the citizens of India where more information was provided the sentiment shown was seen in a positive light. Furthermore, this trend is seen in the report by Matosevic and Bevanda (2020) [9] who did a sentiment analysis on USA, UK, Spain, Italy, Sweden, and Germany. This report shows that where the pandemic was better handled there were more tweets that gave sentiment results of "Trust" and "Anticipation" whereas Spain a country that is not handling the pandemic well has a lot of tweets the express "fear" more prevalently than any other emotion. Finally, what can be noted from these scholars is that the less the public knows about an event can cause misinformation to spread.

This information is helpful in the way that it will be able to reflect on the differences between America's response to COVID -19 and compare it to New Zealand's response to COVID -19 during the 5-month period of March-July of 2020. Additionally, this research will assist in understanding the benefits of following a strict health and safety mandate when facing a global pandemic versus not following any health and safety suggestions.

This research aims to investigate a research gap in sentiment analysis of countries afflicted with COVID -19. By comparing the difference of following government mandates versus not following government mandates. This gap is relevant due to the overwhelming numbers of deaths and cases that have appeared in America. This paper aims to understand how the American people react to government enforced mandates and analyze any outstanding circumstances that occurred throughout the 5-month period. Following that, this paper will analyze why sentiments toward COVID -19 raised or fell during this period.

3. Methodology

In this study, twitter data was captured using a python script called "GetOldTweets3" created by Mottl (2019) [8] which allows for the capture of historical data. After the tweets were captured, that data was run through the tool "Rapid Miner" and a sentiment analysis was applied to the extracted tweets. Due to budget constraints only 10,000 tweets over a period of 5-months were analyzed. The tweets are categorized as very positive, positive, neutral, negative, and very negative tweets.

3.1. Data Acquisition

The areas that this study will research are identifying trends in data mined from twitter based on hot topic key words. By analyzing data pulled from twitter based on these keywords, we can notice trends and identify certain statistics. From these statistics this study, we will investigate what caused certain sentiments to form in the following months. The months that were selected were March 2020 to July 2020, was done based on the number of tweets that could be analyzed and categorized by the Rapid Miner. Each month was restricted to only pull 2000 tweets giving us a total of 10,000 tweets over a 5-month period.

The data for this investigation was gathered through a python script that is known as “GetOldTweets3”, due to the way that Twitter gathering historical twitter data is incredibly difficult. “GetOldTweets3” worked around these problems and allowed for large twitter data extraction. The code has some limitations, but it allows for the large amount of twitter data to be extracted. The process for twitter data extraction is demonstrated in Figure 1.

Figure 1. Data Request Process

To fully analyze the data, the raw twitter data had to go through a “cleaning” process. The main point of this process was to get all the data into one excel file with the correct date attached to the correct twitter data. When all the data was in one file, the sentiment extraction process could begin. The sentiment analysis program analyzed the data and outputted what sentiment level each tweet carried.

3.2. Sentiment Analysis

Following the cleaning process, RapidMiner was used to analyze the data. The data was analyzed through a specific add-on known as the “Operators Toolbox” which splits the incoming data into positive and negative results. From these results we can identify several key factors such as, ‘0’ being mainly neutral remarks, remarks that were ranked between 1 to 10 and -1 to - 10 being positive and negative.

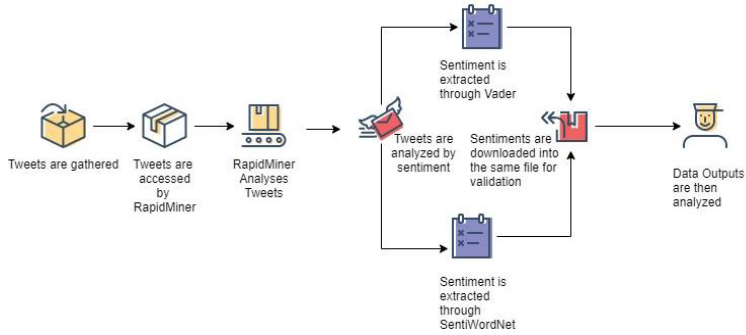


Figure 2. Proposed Architecture of Sentiment Extraction and Validation

3.3. Data Validation

Data validation is the process of ensuring that the data that is being used for analytics is accurate as mentioned by SafeSoftware (2020). Furthermore, data validation process as shown in Figure 2 is a critical move in the data workflow as it ensures that the data that is being used is accurate. The data went through this process firstly through a sentiment analysis tool known as Operator's Toolbox which has access to three sentiment analysis models. The data went through the data analysis model Vader first. The Vader model aligned text based on their content and gave a sentiment score based on the content contained in the text. To validate Vader's results this report used a second sentiment analysis tool from the Operator's Toolbox known as SentiWordNet which gave out similar results with some deviation. By validating the twitter sentiment data, this investigation can be assured that all data used based on the tweets have been correctly analyzed by the software to provide accurate analysis based on the results.

4. Findings

The data that was captured through the months of March to July 2020 in Figure 3 show us that the sentiment towards coronavirus depends on the sentiment analysis type and tends to be "low" nearing the neutral in March before descending in July to being mainly negative. The data sets when spread out over a 5-month period tell us that a major turning point can be viewed between June and July, this point is constant in both Vader and SentiWordNet analysis. This means that an event occurred during these two months. Another event can be observed occurring more visibly in the Vader analysis due to the decrease in sentiments between April and May. The observed events in the data findings can be related to two major events that occurred in the USA during this time. The events that occurred were anti-mask and anti-lockdown protests during April-May and Black Lives Matter protests during June-July, it is during these events that we can see fluctuations in the data.

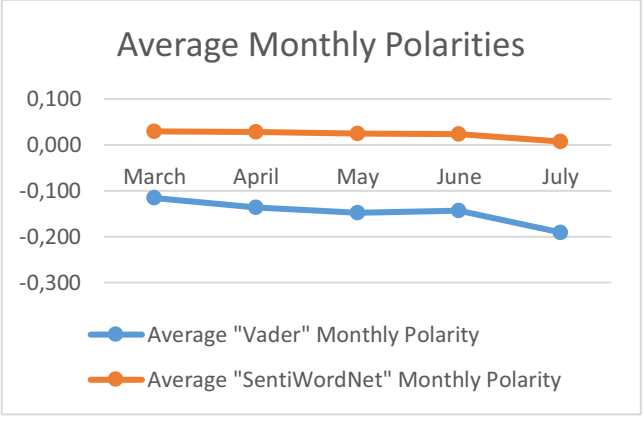


Figure 3. Average Monthly Polarity

4.1. Twitter Data

As seen in the table below, tweets have been analyzed based on their contents and categorized by their sentiment value. Based on the data that this research has gathered; several themes can be identified from Figure 3 such as that during the month of March, we can note that there is a steady decrease in the average polarity results. From this steady decrease we can identify that there must have been several events that occurred between May and July that caused a steady decrease of the overall polarity results. Furthermore, from the data results we can identify that a significant event occurred in March to April that caused the negative polarity results before another event occurred in May to cause a sharp decrease in results.

Table 1. *A table of tweets detailing different types of polarity*

Tweet	SentiWordNet Polarity	Vader Polarity	Average Polarity	Polarity Type
So hypocritical, when you KNOW damned well that #5G is the culprit, not #coronavirus itself. 5G directly affects human cells with radio waves, so that underlying conditions are escalated. This is a KILLING TECHNOLOGY! It needs to be stopped!	-3.9	-5.9	-4.9	Very Negative
I am DONE being nice about corona virus . If you think I wasn't very nice before, just wait. WEAR A FUCKING MASK YOU FUCKING MORONS. DEAD SERIOUS. YOU ARE KILLING PEOPLE WHILE BEING A BITCH. ASHMA ISNT AN EXCUSE NOT TO WEAR ONE NEITHER IS GLASSES. I WOULD MORE THAN HAPPY TO FIND	-11.3	-10.4	-10.9	Very Negative
Mother and brother of NYC hoops legend Sebastian Telfair, reportedly both have died from COVID -19. #COVID 19#CoronaVirus #COVID _19	-0.1	-0.7	-0.4	Negative
Uhh do you not read mystery books? Did you not watch the promise neverland? Literally ANYTHING. You can have a theory on. For example the whole 5G coronavirus theory. None of these idiots went out here to a 5G tower and tested it for radiation but that theory still exists.	.3	-.9	-.3	Negative

5. Discussion

5.1. America Discussion

The major aims that this paper have identified are, how did the American population react to Health and Safety mandates as well as lockdowns initiated by the government? Did any outstanding events occur during the 5-month period that could cause a spike in the sentiment analysis? What caused sentiments about COVID -19 to rise or fall during this period?

The major areas of debate and negative polarity results are tweets about mask wearing and lockdowns. During the months of April to May, there were many reports of Anti-Mask and Anti-Lockdown protests that occurred. According to Aratani (2020) [4] many of the governmental leaders have been providing mixed messages about face masks. Democrats have been more vocal about wearing them and many democratic governors made it mandatory to wear face masks. In contrast Republican leaders have been more hesitant to mandate masks.

Additionally, Aratani (2020) [4] notes that a lot of protests face masks because they call them "Muzzles" or see them as their freedom being eroded. One more vocal leader raised anger at the face mask being a way of "Throwing God's wonderful breathing system out the door" Aratani (2020) [4] reports. Even though both the CDC (2020) [6] and The World Health Organization (2020) [5] have evidence of masks and lockdowns slowing and removing COVID -19 many Anti-Maskers and Lock-downers dislike or even rebuke the thought of wearing a mask or going into lockdowns. According to

McKelvey (2020) [7] who interviewed a health professor at Morgan State University quotes " We're seeing politics and science literally crashing"

It is through this reason that the rise of COVID -19 cases can be observed to be rising in Figure 4. It is through this that we can acknowledge the reason for sentiment value of April and May being at -0.1 due to the rise of COVID cases. Further evidence of this is observed in table 1, where a sample of tweets can be seen about masks and lockdowns. It is through this table that we can observe how anti-mask/lockdown sentiments could have affected the rise of new COVID -19 cases in America.

The data gathered from the sentiment analysis provides a timeline of rises and falls in how people are feeling which they express through tweeting. From these results we can note that major events are related to the low sentiment analysis scores recorded. The results that were provided have determined that several events that occurred during the 5-month period of March to July appear to be related to several infection rate increases. The events that have been identified are the Black Lives Matter, Police Brutality, Anti-Mask and Anti-Lockdown Protests that sparked between the months of March and July of 2020. This evidence is seen in Figure 4.

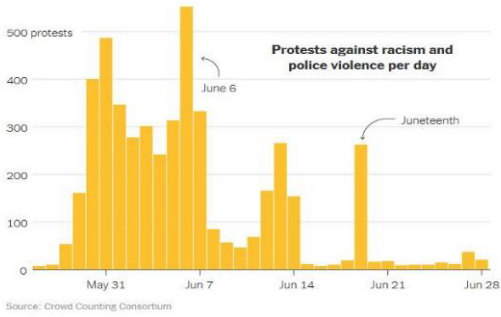


Figure 4. Number of Protests May 31st - June 28th[1]

Another event that occurred that led to mass infection is due to the protests involving Anti-Mask and Anti-Lockdowns. One of the more vocal leaders raised anger at the face mask's being a way of "Throwing God's wonderful breathing system out the door" Aratani (2020) [4] reports. From the sentiment analysis we can extrapolate from the results is that due to the mixed messages given by the various politicians. According to Aratani (2020) [4] many of the governmental leaders have been providing mixed messages about face masks. Democrats have been more vocal about wearing them and many democratic governors made it mandatory to wear face masks. In contrast Republican leaders have been more hesitant to mandate masks.

5.2. New Zealand's Response to COVID -19

In Comparison to how the USA has handled the COVID -19 crisis New Zealand initiated a nationwide lockdown. This lockdown allowed the people of New Zealand to reduce the infection rate to zero in a quick succession with minimal repercussions to setting a nationwide lockdown (Kunzmann, 2020). Comparably in America it would be harder to

[1] L. Buchanan, Q. Bui, and J. Patel, The New York Times, Crowd Counting Consortium.

initiate a nationwide lockdown according to Watson (2020) [3] "is that because there is no federal lockdown authority or a quarantine authority there is no standard or precedent that they can impose". Furthermore, according to Watson's (2020) [3] interviews, President Trump can only shutdown incoming travel at best Mr. Trump can recommend that people stay home. Additionally, according to the Federal Emergency Management Agency (FEMA) "States and Cities are responsible for announcing any type of restrictions or safety measures" What this means is that states are cities are given the ability to be pass restrictions and other measures on their own time, this means that if the state officials do not agree with the recommendations, they can choose whether or not to initiate such changes.

6. Conclusion

This paper has discussed how the sentiment of Americans changed between the months of March to July. Over this 5-month period of change, the paper has analyzed and observed what major events caused these sentiment changes and how they would increase or decrease the rate of COVID -19 infection. One of the main goals of this research was to discuss why America had such a high rate of infection compared to New Zealand and analyze possible reasons for America's high rates of infection. Some of the main reasons for America's high rate of infection is due to the citizens and their Presidents lack of care in regards to the pandemic by not following the health and safety measures suggested by health professionals. Another reason for the high rate of infection is due to two main factors, the murder of George Floyd which sparked Black Lives Matter Riots and the Anti-Mask and Anti-Lockdown protests. Due to these main points and the fact that American states were unable to put proper lockdown procedures in place, can be credited for the high rate of infection in the country. In conclusion, when it comes to dealing with a pandemic, it is ideal to quickly close off the country's borders and initiate a nation-wide lockdown.

References

[1] Shear MD, Mercosh S. (2020, April 17) Trump Encourages Protest Against Governors Who Have Imposed Virus Restrictions. New York Times, Updated April 29, 2020 https://www.nytimes.com/2020/04/17/us/politics/trump-coronavirus-governors.html
[2] Darliansyah A, Wandabwa HM, Naeem MA, Mirza F, Pears R. (2019) Long-Term Trends in Public Sentiment in Indian Demonetisation Policy. CCIS 2018;932:65-75
[3] Watson KTrump probably can't impose a nation lockdown or quarantine. Here's why. CBS News 6 April 2020. https://www.cbsnews.com/news/national-lockdown-quarantine-president-powers/
[4] Aratani L. How did face masks become a political issue in America? The Guardian 29 June 2020, https://www.theguardian.com/world/2020/jun/29/face-masks-us-politics-coronavirus
[5] World Health Organization. (2020). Advice on the use of masks in the context of COVID -19: interim guidance, World Health Organization, 5 June 2020. https://apps.who.int/iris/handle/10665/332293. License: CC BY-NC-SA 3.0 IGO
[6] CDC. Considerations for Wearing Masks. CDC 7 August 2020. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fprevent-getting-sick%2Fcloth-face-cover.html
[7] McKelvey T. Coronavirus: Why are Americans so angry about masks? BBC 20 July 2020. https://www.bbc.com/news/world-us-canada-53477121
[8] Safe Software. Data Validation. (2020). https://www.safe.com/
[9] Mottl, D. Get Old Tweets 3. Pypi. 2019. https://pypi.org/project/GetOldTweets3/#description

[10] Matošević G, Bevanda V. Sentiment analysis of tweets about COVID-19 disease during pandemic, 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2020, pp. 1290-1295, doi: 10.23919/MIPRO48935.2020.9245176.
[11] Barkur, G, Vibha, Kamath GB. , (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. Asian J Psychiatr. 2020 Jun;51:102089. doi: 10.1016/j.ajp.2020.102089.
[12] Davison WP. Public Opinion. Encyclopedia Britannica (2017, March 17). https://www.britannica.com/topic/public-opinion

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210576

The Building Blocks of Information Are Selections - Let's Define Them Globally!

Wolfgang ORTHUBER[a,1]
[a] *Department of Orthodontics, UKSH, Kiel University, Germany*

Abstract. Digital information consists of sequences of numbers that are selections. So far, these are defined by context. We can globalize this by using an efficient global pointer (UL) as "context". The article explains new globally identified and defined "Domain Vectors" (DVs) for transporting digital information. They have the structure "UL plus sequence of numbers", where UL is an efficient identifier and global pointer (link) to the unified online definition of the sequence of numbers. Thus, the format of the number sequence and its meaning is defined online. This opens up far-reaching new possibilities for the efficient exchange, comparison and search of information. It can form the basis for a new global framework that improves the reproducibility, search, and exchange of data across systems, borders, and languages.

Keywords: Online definition, Domain, Domain Vector, DV, Reproducibility, AI

1. Introduction

We need a viable long-term solution to interoperability problems [1] and data silos. Serious problems arise from the variable local definition of digital information. There is a lack of global reproducibility [2] of digital information representation, with all the resulting complications.

This article therefore starts from the basic building blocks of (digital) information, namely selections represented by sequences of numbers. It is shown that an efficient global adaptation of a selection depending on the application (e.g., rough medical diagnosis) is possible. This is done by online definition of a suitable digital number sequence that represents application-relevant information bijectively (one-to-one). A globally defined and identified data structure is derived for the efficient, reproducible, comparable and searchable transport of digital information.

2. Precise Definition of Information

If we want to transport "information" globally and independently of language, we must define it precisely and globally. Set theory provides the precise and very fruitful basis for defining mathematical objects, and mathematical methods are used and applied for dealing with digital information. Therefore, it is only consistent to define the information itself by using the same basis as for other mathematical objects, namely sets and

[1] Corresponding Author: Wolfgang Orthuber; E-mail: orthuber@kfo-zmk.uni-kiel.de.

functions on these sets. The selection of an element from a set describes an elementary class of functions on a set. A combination of selections is again a selection - from a more complex set. We note that information consists of building blocks, all of which are selections from sets of possibilities. A combination of information is again a selection - from a more complex set of possibilities. So we observe:

Information is selection (from a common ordered set or "domain"). (1)

This is consistent with reality: "Information" is the (transportable) result of any well-defined physical experiment. This result is a selection from the (common ordered) set of possible experimental results. As a special case, the bits of digital information are also selections. They encode sequences of numbers, all of which are selections.
Now we see why the precise definition of information is important: (Transportable) information selects from a set that must be the same (common) to all participants in the conversation. This set is often an unconscious (early learned) prerequisite for conversation, e.g., language vocabulary. Because of its importance, the ordered common set of possibilities from which information (1) selects is abbreviated called "domain" [2, 3] (as the domain of the definition of the function "information"). The domain is ordered so that we can select (address) its elements, for example by numbers.

2.1. Language Vocabulary as a Domain

Vocabulary of language(s) is a preferred domain for information exchange because our senses and brain are adapted to it. As a result, international character sets such as UNICODE have gained acceptance [4], defining a mapping between characters of multiple languages and numbers for the digital representation of individual language vocabulary. This made it possible to continue the familiar linguistic conversation by digital means. Words from the language vocabulary are small building blocks with broad applicability - if we want to describe a complex issue ORGINFO, we "only" need to combine enough words to get a certain meaning. The number of possible combinations increases exponentially with the number of words. There are many possibilities of word-based representation of the same original meaning ORGINFO. Moreover, words are often open to different interpretations. For example, the same temperature, even in the same environment, may be described as "cold" by one person and "not cold" by another. Therefore, the linguistic representation of information is usually more or less ambiguous and vague [2]. There is no one-to-one mapping or bijection between the original information ORGINFO and its digital representation DIGINFO. This is a serious shortcoming of language-based representation, since bijection would be important if we want to search original information ORGINFO via its digital representation DIGINFO. Strictly speaking, the "language vocabulary" domain is a detour: original information ORGINFO from nature is not language-based, but must be translated into language. This is then transported and translated back at the receiver. Sender and receiver of information are different. This translation into the domain "language vocabulary" and back is not a bijection and thus not reproducible, which causes many problems [2]. Therefore, it is important to remember that besides language vocabulary, it is also possible to use "adapted domains" that are adapted to the application [3] and allow efficient and reproducible communication. We "only" need to ensure that senders and all receivers of information know the adapted domains in order to use them.

2.2. Application-Specific Adapted Domains

Many situations require a more or less precise exchange of information. In such a case we use a quantitative description and adapt the definition of the numbers to the application. This is often necessary even for simple applications, as shown in **Table 1**, or for professional applications, e.g. in medicine, industry, business, science. There can be many important parameters in a given situation (as original information ORGINFO). Which parameters are important? It depends on the rough situation for which the parameters should provide an accurate description. The parameters are determined from the (total) available information. This is called "feature extraction".
For demonstration purposes, Table 1 shows an example simple feature extraction for the "Ellipse" application. Obviously, this is much more efficient than carrying information about each pixel. In addition, features can be defined to be of direct interest, such as for comparison and search.

Table 1. Application "Ellipse", example feature extraction: angle in 0..45 degrees, height, width, red, green, blue in relative units 0...255 . The domain of these 6 numbers is sufficient, but not the domain of "language vocabulary". Phrases of language like "deep blue, very wide and flat ellipse" or "dark, thin and long ellipse" etc. are not reproducible and not precise enough for comparing and searching such objects.

Ellipse	Angle	Height	Width	Red	Green	Blue
	0	3	3	255	0	0
	0	3	10	255	0	255
	0	3	20	0	255	0
	0	2	20	0	0	255
	15	2	5	0	0	255

2.3. Online Definition of Digital Information for Global Communication

We want to exchange the parameters of an adapted domain (as shown in Table 1) in a globally reproducible and comparable way. As mentioned earlier (2.1), for this we not only need a one-to-one mapping (bijection) between these parameters and their digital representation DIGINFO (typically by "quantification"), we also need to ensure that this mapping is globally consistent. This is efficiently achievable through an online definition. The online definition is unique and can be localized uniformly by an efficient global pointer. This global pointer is called "UL" below (2).

Selection within the adapted domain is done by a sequence of numbers. We can consider sequences of numbers as building blocks of digital information. Since the existence of the Internet, it would have been possible to provide a standard and convenient software for online definition of number sequences and their domains [3]. These domains are language independent and globally valid! Their size grows exponentially with the count of defined numbers (dimensions). Existing online definitions can be reused and nested in new definitions. Thus, we can globally define huge multidimensional domains (huge spaces as "sets of possible messages"). The online definition can be sharp and very detailed (multilingual, even with multimedia parts). It is automatically globally unique. Each online definition can contain links and additions, e.g., the definition of a distance function for similarity comparison and similarity search [5]. The data structure for the globally defined information transport is called "DV" [2, 3] as an abbreviation for "Domain Vector" and has the form:

DV: UL plus sequence of numbers (2)

"UL" is a "Uniform Locator" that refers to the (unique) online definition of the sequence of numbers and is also (directly or indirectly) a unique identifier. The online definition contains machine-readable information about the format of the number sequence and about the domain from which it selects. The UL has similar tasks as a URL, but the UL can be optimized to meet the requirements with maximum efficiency.

Notes on the UL format: as shown earlier [2], the UL can be represented by a hierarchical sequence of numbers. The binary format of the numbers can be based on self-expanding positive integers starting with a half-byte [6]. The compact UL refers to a unique online definition that can be very detailed and explanatory. There is no reason to waste bits in a DV which is designed for efficient information transfer. For example, consider the following DV:

UL plus 0, 3, 3, 255, 0, 0 (3)

As prior knowledge, we only know the format of the UL in the DV, for example, a hierarchical sequence of numbers N1, N2, N3, 0, as proposed in [2]. Here, N1 is an integer pointer within an official table of web addresses of online presences with collections of online definitions, N2 is the number of the user within that online presence, and N3 is the number of the definition of this user. In this way, we know the online definition of the following sequence of numbers. In this case, it contains the information about the first ellipse given in **Table 1**.

In the DV (3), the quantified features (as identified numbers) are directly comparable and searchable. Such number sequences can be defined for any application! Further examples have been published [2][3][6][7][8] and search of information has been demonstrated [9]. When defining the numbers, care must be taken to ensure that they represent the relevant information bijectively (one-to-one). Properly defined DVs with adapted domains change reproducibly if and only if the relevant original features change. More and more adapted domains can be defined online. For example, adapted domains can be defined for diagnoses from ICD-10 [10] or SNOMED-CT [11] as well as for diagnostic devices and diagnostic procedures. In medical reports, DVs with such adapted domains can be used to reduce non-reproducible parts step by step. The more adapted domains are defined online, the more tools are available for a globally reproducible, searchable *and* precise individual digital description.

Correctly (with adapted domain) defined DVs realize:

- globally the same (online) definition for (bijective) one-to-one conversion of original information (application-specific relevant features) into digital information (number sequence).
- globally the same uniform labeling ("UL" = efficient link to the unique machine-readable online definition) in front of the transported digital information (number sequence).

DVs are deliberately designed to transport the online defined digital information (sequence of numbers) as efficiently as possible. Existing online definitions can be reused and nested in new online definitions (e.g., the 3 components of color in **Table 1**).

3. Discussion

It has already been recognized that global definitions are important for certain quantitative data. Therefore, LOINC [12] was introduced. However, the generalization has not yet been realized. So far, it is not common to consider information in general as a selection from a domain (see (1)) and conclude that the global definition of domains of information is important. Technically, this can be realized very efficiently by online definitions of number sequences (for the selection in their domain). DVs (2) are optimized for transport of such globally defined number sequences and universally applicable. Existing work can be continued in online definitions. For example, any ontology, nomenclature, and terminology such as ICD-10 [10], SNOMED-CT [11], and LOINC [12] could be efficiently implemented in a single online definition. In the case of LOINC, for example, the first defined numbers could carry the LOINC code followed by one or more numbers containing measurement data. All DVs transporting this information (from the same online definition) would have the same UL. The online definition is machine-readable, which can also be used for convenient editing of DVs.

The domain "language vocabulary" (see Section 2.1) is widespread because it is obvious that it can be used as a domain of information (as a common set of possible messages). Thus, standards such as Unicode [4] have been introduced for (indirect) selection of elements (words) of the language vocabulary. Therefore, language-based information is globally searchable. This may suffice if we are searching for words in a particular language. But the above considerations about the domain "language vocabulary" make it clear that language-based information is not reproducible and often inadequate, even for transport of simple natural information or for applications, as shown in Table 1. It is not possible to correct such deficiencies subsequently. Even sophisticated linguistic and semantic analysis or AI cannot. In this respect, we need to rethink the acquisition and storage of more or less complex original data from nature, e.g., data from medical findings. If we convert them early into a language-based representation of information for communication, we will lack the required accuracy (Table 1) and reproducibility. But what can we do if the human brain is not adapted to accurately handle high-dimensional original data ORGINF?

We can first focus on the most important features and build on them in a targeted way: We can search and select an online definition of DVs with adapted domain that is suitable for the application or situation (e.g. ICD-10 diagnosis, see below). Optimally, the online definition would also provide appropriate software for reproducible feature extraction and automatic conversion of the original data relevant in this situation to their digital representation as DVs. Thereafter, the reproducible collection, comparison and search of such precise data could become everyday routine. Instant and individual global statistics for decision support would also be possible, more than an elaborate scientific study today. The DVs are also adaptable as *precise* source information for *global* machine learning and AI. The exemplary DV (3) clarifies also the efficiency of the DV data structure. Using directly editable formats for transport of these data (3), we would need several lines of code [8]. This is not only less efficient. In particular, the combinatorial freedom of directly editable formats (like XML, JSON, Turtle) contradicts the reproducible digitization of information. To improve the situation, many rules were published (e.g., in FHIR [13], LOD [14]) to make the transformation of the original information ORGINFO into its digital representation DIGINFO more reproducible. In the process, many problems came to light. Obviously, the optimal conversion depends on the application and the situation, e.g., medical diagnosis.

This could be one of the first applications of DVs. The online definitions of adapted domains (2.2) can be adapted to more and more applications and situations and provide rules and software for reproducible digitization. These are immediately accessible, globally valid, and can be created for more and more topics (e.g., given by a medical term from ICD-10 [10] or SNOMED-CT [11]). Using RFC4648 [15] and suitable editors, we can integrate DVs into existing editable formats, automatically incorporating FAIR principles [16]: Each DV is globally identified via UL and is thus "findable". The online definition clarifies the content of the DV (this can be done descriptively and extensively) and makes it editable and "accessible". The identification of the DVs (by UL) makes them "interoperable" and "reusable." The online definitions themselves are also reusable and can be embedded and nested within other online definitions via link. Thus, DVs can be online defined as globally consistent and comprehensive building blocks of digital information.

4. Conclusion

The proposed DVs can be viewed as globally defined building blocks of digital information. Online definitions of DVs can be adapted to the application or situation to globally determine a one-to-one mapping from the important features of the original information (ORGINFO) to the digital representation (DIGINFO) as DV. The DV data structure is well-defined and has further objectifiable advantages, e.g. in terms of global reproducibility and efficiency of information transport. Its introduction is recommended.

References

[1] Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digital Medicine 2019; 2(1):1-5.
[2] Orthuber W. Reproducible Transport of Information. Studies in Health Technology and Informatics Vol. 281; IOS Press 2021: 3-7. https://ebooks.iospress.nl/doi/10.3233/SHTI210109
[3] Orthuber W. Information Is Selection-A Review of Basics Shows Substantial Potential for Improvement of Digital Information Representation. Int. J. Environ. Res. Public Health 2020, 17(8), 2975.
[4] Allen JD, Anderson D, Becker J, Cook R, Davis M, Edberg P, et al. The unicode standard. Mountain view, CA, 2012.
[5] Zezula P, Amato G, Dohnal V, Batko M. Similarity Search: The Metric Space Approach. Vol. 32. New York, USA Springer Science & Business Media, 2006.
[6] Orthuber W. How to make medical information comparable and searchable. Digit Med 2020;6:1-8.
[7] Orthuber W. Global predefinition of digital information. Digit Med 2018;4:148-56.
[8] Orthuber W. Online definition of comparable and searchable medical information. Digit Med 2018;4:77-83.
[9] Orthuber W. Demonstration of Numeric Search in User Defined Data. Available from: http://www.numericsearch.com [Accessed on 2021 Jun 08].
[10] World Health Organization. ICD-10: International Statistical Classification of Diseases and Related Health Problems: Tenth Revision; 2016.
[11] Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Studies in health technology and informatics 2006; 121: 279.
[12] McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, et al. LOINC, a universal standard for identifying laboratory observations: A 5-year update. Clin Chem 2003;49: 624-33.
[13] Benson T, Grieve G. Principles of health interoperability: SNOMED CT, HL7 and FHIR. Springer, 2016.
[14] Bauer F, Kaltenböck M. Linked open data: The essentials. Edition mono/monochrom, Vienna, 2011.
[15] Josefsson S. The base16, base32, and base64 data Encodings. RFC 4648, October 2006; 1-18.
[16] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data 2016; 3(1):1-9.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210577

ContSOnto: A Formal Ontology for Continuity of Care

Subhashis DAS[a,1] and Pamela HUSSEY[b]
[a] *CeIC, ADAPT, School of Computing*
[b] *CeIC, ADAPT, School of Nursing & Human Sciences, Dublin City University, Ireland*

Abstract. The global pandemic over the past two years has reset societal agendas by identifying both strengths and weaknesses across all sectors. Focusing in particular on global health delivery, the ability of health care facilities to scale requirements and to meet service demands has detected the need for some national services and organisations to modernise their organisational processes and infrastructures. Core to requirements for modernisation is infrastructure to share information, specifically structural standardised approaches for both operational procedures and terminology services. Problems of data sharing (aka interoperability) is a main obstacle when patients are moving across healthcare facilities or travelling across border countries in cases where emergency treatment is needed. Experts in healthcare service delivery suggest that the best possible way to manage individual care is at home, using remote patient monitoring which ultimately reduces cost burden both for the citizen and service provider. Core to this practice will be advancing digitalisation of health care underpinned with safe integration and access to relevant and timely information. To tackle the data interoperability issue and provide a quality driven continuous flow of information from different health care information systems semantic terminology needs to be provided intact. In this paper we propose and present *ContSonto* a formal ontology for continuity of care based on ISO 13940:2015 *ContSy* and W3C Semantic Web Standards Language OWL (Web Ontology Language). *ContSonto* has several benefits including semantic interoperability, data harmonization and data linking. It can be use as a base model for data integration for different healthcare information models to generate knowledge graph to support shared care and decision making.

Keywords. EHR, Interoperability, Semantic, Ontology, OWL

1. Introduction

The global crisis caused due to the ongoing pandemic, has largely altered the functionality of various service industries including healthcare sectors. Healthcare sector issues include but are not restricted to a lack of conformance with standards use or vendor lock in due to use of proprietary Electronic Healthcare Records (EHRs) software and systems which are unable to exchange data within and across government organizations. However, few Artificial Intelligence (AI) companies are promising that AI based solution will solve this issue by providing intelligent allocation of resources among care facilities such as beds, doctors, and patients. Recently there is a proliferation of white

[1] Corresponding Author. Subhashis Das, CeIC, ADAPT, School of Computing, Dublin City University (DCU), Dublin 9, Ireland; E-mail: subhashis.das@dcu.ie.

papers reports and publications reflecting the buzz around Artificial Intelligence (AI) based healthcare [1, 8]. AI however can only be realised with standard data exchange which is machine understandable at the same time capable of aggregate data from various sources [15].

To tackle this data exchange issue across health systems in this paper we propose a formal ontology of Continuity of care (*ContSOnto*). ContSOnto as an emerging research area consisting of the extension of healthcare ontology to the continuity of care domain. This field is positioned at the confluence of health informatics, nursing informatics, process modeling, and artificial intelligence. Gulliford et al. (2006) [7] describe "Continuity of care as a process which is concerned with the quality of care over time". There are two aspects to this perspective. One is based on the patient's experience of a 'continuous caring relationship' with the healthcare professional. Another one is based on a system of care where seamless service is needed to provide care through integration, coordination and interoperable information systems. WHO (2018) [14] defines Continuity of care as reflects the extent to which a series of discrete health care events is experienced by people as coherent and interconnected over time and consistent with their health needs and preferences. To develop our ContSOnto model we engaged with healthcare professionals as well as standards bodies who originally were involved in development work of ISO 13940:2015 ContSys. Earlier Horizon 2020 project Hospital at Home (H@H) project [12] proposed a conceptual model of social care to be included in the care system but model does not provide an outline for any real implementation after five years. Although those H@H conceptual model is not based on Resource Description Framework (RDF) it therefore hinder the main objective of data exchange with other healthcare systems. ContSOnto model align with the European ISA recommendation on new European Interoperability Framework (EIF) [3] and describes how ContSOnto is conforming with EIF level as depicted in Figure1.

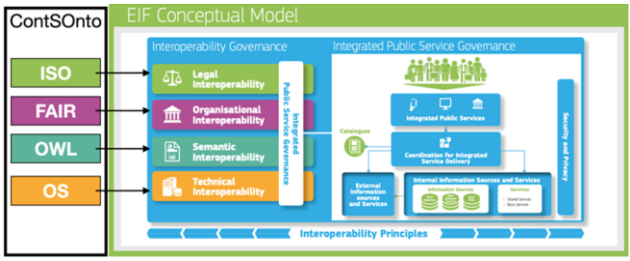


Figure 1:ContSOnto Alignment with New EIF

The International Standard Organization (ISO) provides a legal framework for intergovernmental interoperability. The acronym FAIR equates to data that is Findability, Accessibility, Interoperability, and Reusable. The FAIR principle allowed ContSOnto organizational Interoperability, as ContSOnto model is based on Web Ontology Language (OWL) which enables the most needed se- mantic interoperability ecosystem [2]. In addition, the process of ContSOnto is focused on developing using open source (OS) software and Open protocol (Open API), thus ContSOnto is neutral and does not rely on any proprietary software. This paper is structured as follows: In Section 2, we describe the overall methodology, in Section 3 results and implementation, and we conclude in Section 4 with a discussion on future work.

2. Methodology

ContSOnto development methodology is based on two main features. One is requirements analysis to verify what is needed to have for a care model so that it fulfills mod- ern's days software (i.e. app) needs described in Section 2.1. Other aspect is to develop a model based on ontological decisions as stated in OntoClean methodology [6] described in Section 2.2.

2.1. Requirements

How the ContSys standards are approached, designed, constrained, or extended is based on a formal logical model. The ContSys ontology model therefore needs to be mapped explicitly to Resource Description Framework (RDF) formalism as per W3C Semantic Web Standards. Without such a model to operate from, ContSys Ontology will lack the semantic and structural consistency required to make ContSys computable and generate knowledge graphs. Priorities are indicated using MoSCoW terms (MUST, SHOULD, COULD, WON'T). **1) ContSys Ontology Mappings (MUST):** We shall define lossless bi-directional transformations from ContSys UML instances to OWL/RDFS ontology representations and vice versa. **2) Complete ContSys Coverage (MUST):** The RDF representation of ContSys Unified Modeling Language (UML) element instance data shall be capable of expressing all legal ContSys instances that make use of any valid ContSys sub-set, including extensions. An RDF instance data representation that is limited to only a subset of possible ContSys instances is not acceptable. **3) Monotonic with Modifier Extensions (MUST):** ContSys RDF data with modifier extensions shall be "consistent" for RDF reasoning, i.e., the semantics of the RDF must be monotonic even in the presence of modifier extensions. **4) Vocabulary Bindings (MUST):** The ContSys ontology shall support vocabulary bindings to code, Coding and Codeable Concept - including dealing with extensible value sets and multi-code system value sets. (SHOULD) The ContSys vocabulary representation should be able to leverage existing semantic web terminology representations (e.g., SNOMED-CT). **5) Enforce Constraints (SHOULD):** The ContSys ontology should enforce constraints that are representable in OWL/RDF whenever possible, e.g., schema constraints, regular expressions, etc. **6) Annotation In- formation (SHOULD):** In the RDFS/OWL Ontology representation, should expose at least minimal annotation information for display in an ontology editor for use by humans. **7) Top-level alignment (SHOULD):** ContSys Ontology should be aligned with one top-level ontology. **8) RDF Quality (MUST):** Transformations into RDF must meet software quality checks including ontological closure. The RDF instance which is transformed from contsys UML must be capable of being opened without further modification by widely available tools including Protégé.

2.2. Formal Ontology

Gruber (1993) [4] defined ontology as a "formal, explicit specification of a shared conceptualization". Ontology provides a shared vocabulary, which can be used to model a domain of discourse that is, the type of objects, and/or concepts that exist, and their properties and relations. As per Guarino (1998) [5] Ontology is "a set of logical axioms designed to account for the intended meaning of a vocabulary". In this definition, Guarino emphasized the role of logic as a way of representing an ontology. Need for

building a formal ontology for contsys highlighted by Martınez-Costa et al. (2015) [11]. We believe that ontology has an important role to play in the general task of managing diverse information.

The purpose of defining a Resource Description Framework (RDF) representation of ContSOnto is not only to enable ContSOnto to be exchanged in an RDF format such as Turtle, JSON-LD but also to ground the semantics of ContSOnto data in RDF, for use with ontologies and other RDF data. Since the ContSOnto data model is losslessly assembled, any component of the data model can be used in conjunction with RDF. The semantics are well kept regardless of source format. We choose Web Ontology Language (OWL) for formalization language. It is built upon the World Wide Web Consortium's (W3C) XML standard for objects called the RDF. OWL provides the benefit of reasoning using Description Logic (DL). More precisely we choose OWL 2 for modeling. It has five main advantages than the previous version such as property chains; richer data types, data ranges; qualified cardinality restrictions; asymmetric, reflexive, and disjoint properties; and enhanced annotation capabilities [9].

2.3. Ontology Alignment

Top-level ontologies provide domain-independent conceptualization, relations, and axioms (e.g., categories like Event, Mental Object, Quality, etc.) in order to standardize upper-level of a model thus enable linking con ontology with other freely available ontology such Link data vocabulary (LOV)[2] and Biomedical Ontology by NCBO[3]. In ContSOnto we use the top-level ontology Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [5] as a middle-out solution between degree of formalization and complexity, contributing to an effective practical solution. In spite of the benefit of top-level ontologies, their alignment and use is not trivial and requires some expert effort. The EU project Advancing Clinico-Genomic Trials (ACGT) [13] as well as other healthcare projects emphasis on need and benefit from top-level alignment. **Figure 2** depicted Class hierarchy of ContSOnto ontology. And **Figure 3** showcase partial view of ContSOnto class visualization using WebProtégé tool and upper part of the figure in green such as *mentalObject*, *stative*, *event* are DOLCE classes and other are domain specific class taken from ISO 13940:2015 ContSys.

Figure 2. Class Hierarchy

[2] https://lov.linkeddata.es/dataset/lov/

[3] https://bioportal.bioontology.org/ontologies

3. Results and Implementation

The resulting formal ontology is available online on National Center for Biomedical Ontology (NCBO) Bioportal *ContSOnto* and full Ontology documentation on GitHub.

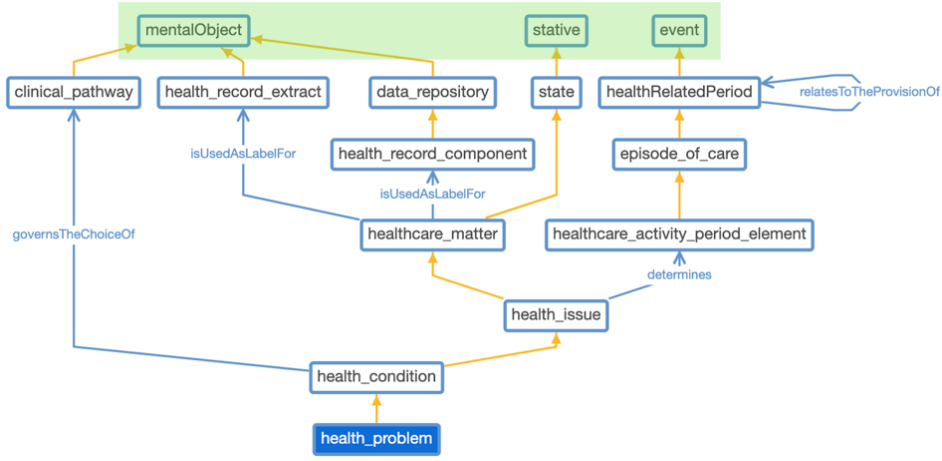


Figure 3. ContSOnto Alignment with DOLCE Top-level ontology (DOLCE classes are in grey)

repository with permanent URI *http://purl.org/net/for-coc*. In its current version, it is based on ISO 13940:2015 ContSys. It consists of 21888 triples. A total of 153 OWL Classes and 144 OWL Properties have been defined. ContSOnto has total 961 axiom with 415 logical axioms and 305 declarative axioms. Expressiveness of ContSOnto model is *ALCHQ(D)* as per description logic (DL) scale.

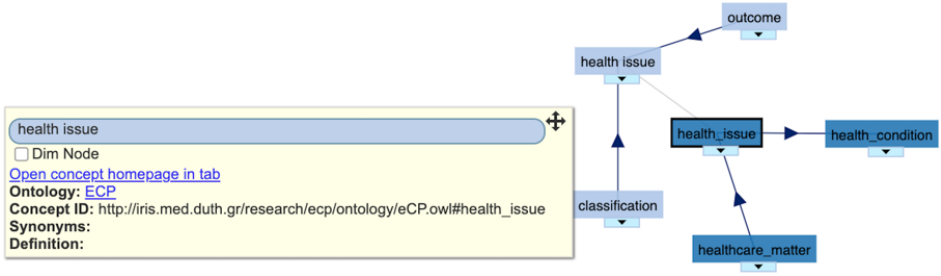


Figure 4. Neighborhood relation with ECP ontology using Bioportal web service

4. Discussion

The pandemic has presented many challenges globally, and health researchers, policy analysts and decision makers are reporting worrying results on predictive models for 2020-2021. Whereas connection among different healthcare settings still has a long path to progress, In this direction, ContSOnto can be seen as a base model which will provide scope for wider collaboration. For example SNOMED International will publish ICNP Reference Sets and an associated ontology in September 2021, as the CeIC is an ICNP

R&D Center[4] future progression of nursing sensitive data to advance patient centered integrated care models is under consideration. Initial research on development of Nursing Knowledge Graph (NKG) and our progress in this domain is published and available to view from Journal of Nursing Scholarship [10].The benefits of using NCBO Bioportal is that we can leverage its online annotation and semantic matching facilities to discover other related models available on the Bioportal using Neighborhood matching. Figure 4 above provides one such example, which showcase associated between ContSOnto's *Health issue* (node in dark blue) with ECP ontology (node in light blue).

Acknowledgement

This research has received funding from the European Union's Horizon 2020 re- search and innovation programme under the ELITE-S Marie Skłodowska-Curie grant agreement No. 801522, by Science Foundation Ireland and co-funded by the European Regional Development Fund through the ADAPT Centre for Digital Content Technology grant number 13/RC/2106 P2 and DAVRA Networks.

References

[1] Bhattacharya S, Pradhan KB, Bashar MA, Tripathi S, et al. Artificial intelligence enabled healthcare: A hype, hope or harm. Journal of family medicine and primary care. 2019 Nov;8 (11): 3461.
[2] Blobel B. Interoperable EHR systems–challenges, standards and solutions. European Journal for Biomedical Informatics. 2018; 14(2): 10-9.
[3] European Commission. New european interoperability framework, 2017.
[4] Gruber TR. A translation approach to portable ontology specifications. Knowledge acquisition. 1993 Jun 1;5(2):199-220.
[5] Guarino N, editor. Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy. IOS press; 1998.
[6] Guarino N, Welty C. Evaluating ontological decisions with OntoClean. Communications of the ACM. 2002 Feb 1;45(2):61-5.
[7] Gulliford M, Naithani S, Morgan M. What is' continuity of care'?. Journal of health services research & policy. 2006 Oct 1; 11(4): 248-50.
[8] Han C, Rundo L, Murao K, Nemoto T, Nakayama H. Bridging the gap between AI and healthcare sides: towards developing clinically relevant AI-powered diagnosis systems. InIFIP International Conference on Artificial Intelligence Applications and Innovations 2020 Jun 5 (pp. 320-333). Springer, Cham.
[9] Hitzler P, Krötzsch M, Parsia B, Patel-Schneider PF, Rudolph S. OWL 2 web ontology language primer. W3C recommendation. 2009 Oct 27; 27(1): 123.
[10] Hussey P, Das S, Farrell S, Ledger L, Spencer A. A Knowledge Graph to Understand Nursing Big Data: Case Example for Guidance. Journal of Nursing Scholarship. 2021 May; 53(3): 323-32.
[11] Martínez-Costa C, Kay S, Oughtibridge N, Schulz S. ContSys under Ontological Scrutiny. InDigital Healthcare Empowering Europeans 2015 (pp. 999-999). IOS Press.
[12] Pecoraro F, Luzi D, Ricci FL. An integrated model to capture the provision of health and social care services based on the ContSys and FHIR standards. EJBI. 2017; 13(1): 17-26.
[13] Stenzhorn H, Schulz S, Boeker M, Smith B. Adapting clinical ontologies in real-world environments. Journal of Universal Computer Science. 2008; 14(22).
[14] WHO. Continuity and coordination of care, 2018.
[15] WHO. Ethics and governance of artificial intelligence for health who guidance, 2021.

[4] https://www.icn.ch/what-we-doprojectsehealth-icnpabout-icnpicn-accredited-centres-icnpr-research-development/icn

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210578

Predicting the Aortic Aneurysm Postoperative Risks Based on Russian Integrated Data

Iuliia LENIVTCEVA[a,1], Sofia GRECHISHCHEVA[a], Georgy KOPANITSA[a] , Dmitry PANFILOV[b] and Boris KOZLOV[b]
[a] *ITMO University, 49 Kronverskiy prospect, 197101, Saint Petersburg, Russia*
[b] *Cardiology Research Institute, Tomsk National Research Medical Center of the Russian Academy of Science, Tomsk, Russia*

Abstract. This article describes the results of feature extraction from unstructured medical records and prediction of postoperative complications for patients with thoracic aortic aneurysm operations using machine learning algorithms. The datasets from two different medical centers were integrated. Seventy-two features were extracted from Russian unstructured medical records. We formulated 8 target features: Mortality, Temporary neurological deficit (TND), Permanent neurological deficit (PND), Prolonged (> 7 days) lung ventilation (LV), Renal replacement therapy (RRT), Bleeding, Myocardial infarction (MI), Multiple organ failure (MOF). XGBoost showed the best performance for most target variables (F-measure 0.74-0.95) which is comparable to recent results in cardiovascular postoperative risks prediction.

Keywords. Postoperative risks, aortic aneurysm, integrated data, predictive modeling, feature extraction, machine learning

1. Introduction

Thoracic aortic aneurysm (TAA) is a dilatation of the aorta to more than 150% of normal diameter [1] in ascending, descending aorta, or aortic arch. TAA has a 1-year mortality rate up to 75% [2]. The causes of death include not only aortic rupture, but also such complications as myocardial infarction, renal insufficiency, bleeding, stroke, etc. [3]. These risks are often compounded by several cardiovascular comorbidities which complicates the decision making. The prediction of complications is one of the ways to reduce patient's risks. Machine-learning (ML) offers an approach for risks prediction to address patient's state [4]. It uses routine clinical data to create risks prediction models. The overview of current cardiovascular postoperative models for risks prediction is shown in Table 1. About 80% essential medical data are stored in free-text medical records [5] which limits the number of data available and complicates the prediction models development.

[1] Corresponding Author, Iuliia Lenivtceva, ITMO University, 49 Kronverkskiy prospect, 197101 Saint Petersburg, Russian Federation; E-mail: lenivezzki@gmail.com.

The aim of this work is to develop a model for postoperative risks prediction for patients with TAA based on data from two Russian medical institutions concerning both structured data and free-text medical records.

Table 1. Recent algorithms, estimations, and data

Study	Algorithm	AUC-ROC	Data	Target
Lee, 2018 [6]	XGBoost	0.78	Open heart and TAA surgery	Acute kidney injury
Zhong, 2021 [7]	XGBoost	0.93	Coronary artery bypass surgery, aortic valve replacement and other heart surgeries	30-day mortality, septic shock, liver dysfunction, and thrombocytopenia
Allyn, 2017 [8]	Model ensemble	0.78	Elective heart surgery	Postoperative mortality
Fernandes, 2021 [9]	XGBoost	0.88	Intraoperative open heart surgery data	Postoperative mortality
Coulson, 2020 [10]	Logistic regression	0.78–0.85	Open heart surgery	Acute kidney injury

2. Method

2.1. Data and Features

The predictive model was developed based on two datasets. The first dataset includes structured data from Tomsk National Medical Research Center of the Russian Academy of Sciences and the second dataset includes unstructured medical records from Almazov National Medical Research Center (St. Petersburg, Russia). Tomsk dataset contains 97 structured records for 97 patients with data on aortic operations. Almazov National Medical Research Center dataset contains 56 929 text documents (2008 – 2019) for 343 TAA operations and 319 patients. We formulated 8 target features: Mortality, Temporary neurological deficit (TND), Permanent neurological deficit (PND), Prolonged (> 7 days) lung ventilation (LV), Renal replacement therapy (RRT), Bleeding, Myocardial infarction (MI), Multiple organ failure (MOF). In total, 63 input features were formulated for risks prediction. We organized these features in groups: anthropometric information (6 features), concomitant diseases (8 features), laboratory tests (5 features), coronary angiogram (4 features), echocardiography (8 features), computed tomography (14 features), intraoperative information (15 features), combined surgeries (3 features).

2.2. Feature Extraction

We extracted input and target features from Almazov National Medical Research Center text records; the Tomsk National Medical Research Center of the Russian Academy of Sciences data were already structured. All the data were anonymized by the source medical institutions. Textual data preprocessing included several steps: data cleaning, lemmatization, stop-words and rare words removal, sentences segmentation, POS-tagging, negation detection and removal, tokenization, and vectorization (TF-IDF). To realize these steps, we used the following Python packages: pymorphy2 (to work with Russian language), NLTK, spaCy. Data filtering was organized both by keywords search (for each feature) and by applying shallow algorithms: Support Vector Machine (SVM),

Random Forest (RF), Logistic Regression (LR), and k-nearest neighbors (k-NN). Time frames were considered for data filtering. For instance, preoperative features were extracted from the documents before the operation. The features are extracted using the list of patterns and rules. Feature extraction accuracy was evaluated on 200 manually processed textual records. After the feature extraction step, two datasets were integrated based on feature names.

2.3. Predictive Model

Firstly, data were prepared for modelling: normalized in an interval [0,1], processed strong-correlated features (Pearson correlation), removed features with more than 60% gaps, otherwise missing values were imputed using k-NN method from sklearn package. Secondly, four algorithms were used for feature selection: univariate feature selection based on chi-squares, recursive feature elimination, decision tree ensemble and Lasso regression. Each algorithm selects 10 features and direct them to majority voting. From 8 to 11 features are selected for each target. Eight models were built to predict eight targets and a set of selected features was created for each target. Some target features can also be used as input features for other targets. We tested three algorithms for modelling: 1) LR; 2) RF; 3) XGBoost. SMOTE was used for integrated dataset balancing. The results are estimated by AUC-ROC, F-measure, and Accuracy scores, using 20-fold cross-validation. We also compared the performance of the developed models before (97 operations from Tomsk dataset) and after (440 operations, integrated dataset) extending structured dataset with extracted features.

3. Results

3.1. Data Description

Table 2 shows the percentage of missing values in extracted data.

Table 2. The percentage of missing values (only features that have missing values)

Feature	Missing, %	Feature	Missing, %
Circulatory arrest time	69.6	Cardioplegic arrest time	51.6
Cardiopulmonary bypass time	44.0	Postoperative creatinine	17.9
Aortic arch diameter	12.3	Sinuses of Valsalva diameter	12.3
Surgery duration	10.9	Blood loss	6.2
Height	4.7	Body mass index	4.7
Body surface area	4.7	Ascending aorta diameter	4.1
Left ventricle ejection fraction	3.8	Postoperative hematocrit	3.5
Weight	3.2	Age	2.9

3.2. Predictive Modelling Results

XGBoost strategy in combination with SMOTE yields the best results for most targets. Table 3 represents the best results for each target.

Table 3. The results for predictive modelling

Target	Strategy	Accuracy	AUC-ROC	F-measure
Mortality	XGBoost + SMOTE	0.915	0.928	0.872
TND	XGBoost + SMOTE	0.799	0.846	0.744
PND	XGBoost + SMOTE	0.850	0.932	0.845
Prolonged LV	XGBoost + SMOTE	0.927	0.988	0.948
RRT	XGBoost + SMOTE	0.975	0.986	0.950
Bleeding	RF + SMOTE	0.925	0.987	0.933
MI	RF + SMOTE	0.957	0.986	0.953
MOF	XGBoost + SMOTE	0.903	0.941	0.885

The comparison results before and after structuring are represented in Table 4.

Table 4. Comparing the models' performance before and after structuring textual data

Target	AUC-ROC (before)	F-measure (before)	AUC-ROC (after)	F-measure (after)
Mortality	0.845	**0.928**	0.852	**0.872**
TND	0.828	**0.846**	**0.839**	0.744
PND	0.929	**0.932**	**0.931**	0.845
Prolonged LV	0.911	**0.988**	0.909	**0.948**
RRT	0.771	**0.986**	0.784	**0.950**
Bleeding	0.893	**0.987**	0.889	**0.933**
MI	0.833	**0.987**	0.821	**0.953**
MOF	0.835	**0.941**	0.816	**0.885**

Table 5 shows top-5 most important input features for target prediction.

Table 5. Most important features for targets

Target	Top-5 important features	Target	Top-5 important features
Mortality	1) MOF 2) RRT 3) Fresh frozen plasma, units 4) Deep hypothermia 5) Age	TND	1) Entry site of aortic dissection at the sinotubular junction 2) Descending aortic dissection 3) Left internal carotid artery stenosis (50-75%) 4) Resternotomy for bleeding 5) Ascending aortic dissection
PND	1) Diameter of sinus of Valsalva 2) Entry site of aortic dissection at sinotubular junction 3) Diameter of aortic arch 4) Prolonged LV 5) Right coronary artery stenosis	Prolonged LV	1) Red blood cells, units 2) Fresh frozen plasma, units 3) Bleeding 4) Left internal carotid artery stenosis (<50%) 5) RRT
RRT	1) MOF 2) Postoperative creatinine 3) Extension of aortic dissection down to iliac and/or femoral arteries 4) Fresh frozen plasma, units 5) Prolonged LV	Bleeding	1) Previous MI 2) Aortic valve replacement 3) Height 4) Fresh frozen plasma, units 5) Retrograde dissection
MI	1) Left coronary artery stenosis (>75%) 2) Red blood cells, units 3) Drainage blood loss 4) Dissection of abdominal aorta 5) Left coronary artery stenosis (<50%)	MOF	1) RRT 2) Right coronary artery stenosis 3) Fresh frozen plasma, units 4) Red blood cells, units 5)Previous cerebrovascular accident

4. Discussion

This work is dedicated to the development of the predictive model based on the integrated medical data. For this purpose, we used two datasets from real medical institutions which contain heterogeneous data for patients with TAA operations. For integration purposes the textual data were processed to extract essential features. The extracted features were validated based on the accuracy score on the test sample. We dropped 6 features due to the differences in data formats storage, diagnostic methods for different institutions and due to the missing values. For instance, circulatory arrest time is a feature that characterizes duration of circulatory arrest in minutes, however, for Almazov National Medical Research Center data, it is often possible to extract information about circulatory arrest only as a binary feature – if a procedure was done or not. As a result of the exploratory data analysis some features such as weight (correlated with two other features), circulatory arrest time, cardioplegic arrest time, and cardiopulmonary bypass time were removed due to the large number of missing values (see Table 2) as the use of imputing techniques can affect the quality of the predictive model. To develop a predictive model three machine-learning algorithms were used: 1) logistic regression; 2) XGBoost; 3) random forest. XGBoost algorithm in combination with SMOTE showed the best results for most targets (see Table 3). It also shows comparable results to other studies in predicting postoperative cardiovascular complications (Table 1). The developed predictive model has a high potential for the thoracic aortic surgery risks prediction. Although it should be noted that from the clinical point of view the impact of several parameters in the predictive model is obscure. However, number of them has logical explanation. For example, direct relation of the aortic diameter at the sinuses of Valsalva to temporal neurological deficit is unclear. To find the answer one need to solve a logical chain. Large aortic root is an indication for its replacement. Naturally, it prolongs cardiopulmonary bypass time and, in turn, increases neurological deficit risks.

Our study has some limitations. However, we integrated data from several datasets, the number of patients and operations is relatively small and needs to be extended. We also faced with the imbalance problem during the study, which is usual for medical data [11]. In such situations machine-learning algorithms tend to classify the data into predominant class. To address this problem, we used SMOTE for data balancing and F-measure as a metric which is less sensitive to data imbalance. However, the work with imbalanced medical datasets is still an issue. One more limitation relates to the data losses during the integration process. There is a need not only to compare and map the logical data structures and contents, but also diagnostic methods and treatment approaches in different institutions as it may influence the data collected and stored. However, despite all the mentioned limitations the study showed that data structuring and integration helps to extend the dataset and improve the quality of the predictive model.

5. Conclusion

In this study we developed a model for postoperative risks prediction for patients with TAA based on data from two Russian medical institutions concerning both structured data and free-text medical records. Our study showed that heterogeneous data integration improves the performance of predictive model. Future studies may address current

limitations of the study such as relevant synthetic patients' generation and model validation in a medical practice.

Acknowledgments

This work is financially supported by National Center for Cognitive Research of ITMO University.

References

[1] Capuñay C. Carrascosa P. Chapter 19 - Magnetic Resonance Angiography. In: Ţintoiu IC, Ursulescu A, Elefteriades JA, Underwood MJ, Droc I (Eds.) *New Approaches to Aortic Diseases from Valve to Abdominal Bifurcation*, Academic Press, 2018, pp. 207–214.

[2] Mohr JP, Benavente O, Barnett HJM. Chapter 16 - Spinal Cord Ischemia. In: *Stroke (Fourth Edition)*, Fourth Edi., Mohr JP, Choi DW, Grotta JC, Weir B, Wolf PA (Eds.) Philadelphia: Churchill Livingstone, 2004, pp. 423–436.

[3] Orandi BJ, Dimick JB, Deeb GM, Patel HJ, Upchurch GR. A population-based analysis of endovascular versus open thoracic aortic aneurysm repair. *J. Vasc. Surg.*, vol. 49, no. 5, pp. 1112–1116, May 2009, doi: 10.1016/j.jvs.2008.12.024.

[4] Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can Machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*, vol. 12, no. 4, p. e0174944, Apr. 2017, doi: 10.1371/journal.pone.0174944.

[5] Kong HJ. Managing unstructured big data in healthcare system. *Healthc. Inform. Res.*, vol. 25, no. 1, pp. 1–2, 2019, doi: https://doi.org/10.4258/hir.2019.25.1.1.

[6] Lee H, et al. Derivation and Validation of Machine Learning Approaches to Predict Acute Kidney Injury after Cardiac Surgery. *J. Clin. Med.*, vol. 7, no. 10, p. 322, Oct. 2018, doi: 10.3390/jcm7100322.

[7] Zhong Z, Yuan X, Liu S, Yang Y, Liu F. Machine learning prediction models for prognosis of critically ill patients after open-heart surgery. *Sci. Rep.*, vol. 11, 2021, doi: 10.1038/s41598-021-83020-7.

[8] Allyn J, et al. A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: A decision curve analysis. *PLoS One*, vol. 12, no. 1, pp. 1–12, 2017, doi: 10.1371/journal.pone.0169772.

[9] Fernandes MPB, Armengol de la Hoz M, Rangasamy V, Subramaniam B. Machine Learning Models with Preoperative Risk Factors and Intraoperative Hypotension Parameters Predict Mortality After Cardiac Surgery. *J. Cardiothorac. Vasc. Anesth.*, vol. 35, no. 3, pp. 857–865, 2021, doi: 10.1053/j.jvca.2020.07.029.

[10] Coulson T, et al., Predicting Acute Kidney Injury After Cardiac Surgery Using a Simpler Model. *J. Cardiothorac. Vasc. Anesth.*, vol. 35, no. 3, pp. 866–873, 2021, doi: 10.1053/j.jvca.2020.06.072.

[11] Zhong Z, Yuan X, Liu S, Yang Y, Liu F. Machine learning prediction models for prognosis of critically ill patients after open-heart surgery. *Sci. Reports 2021 111*, vol. 11, no. 1, pp. 1–10, Feb. 2021, doi: 10.1038/s41598-021-83020-7.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210579

An Unsupervised Approach to Structuring and Analyzing Repetitive Semantic Structures in Free Text of Electronic Medical Records

Varvara KOSHMAN[a,1], Anastasia FUNKNER[a], Sergey KOVALCHUK[a,b]
[a] *ITMO University, Saint Petersburg, Russia*
[b] *Federal Almazov North-west Medical Research Centre, Saint Petersburg, Russia*

Abstract. Electronic Medical Records (EMR) contain a lot of valuable data about patients, which is however unstructured. There is a lack of labeled medical text data in Russian and there are no tools for automatic annotation. We present an unsupervised approach to medical data annotation. Morphological and syntactical analyses of initial sentences produce syntactic trees, from which similar subtrees are then grouped by Word2Vec and labeled using dictionaries and Wikidata categories. This method can be used to automatically label EMRs in Russian and proposed methodology can be applied to other languages, which lack resources for automatic labeling and domain vocabularies.

Keywords. syntactical parsing, natural language processing, electronic health records, node2vec, automatic text labeling, graph algorithms

1. Introduction

It has been shown that the use of textual content of EMRs for training language models significantly improves model performance [1]. However, it is currently hardly feasible to include it when working with Russian, because of very few labeled datasets available. One reason for it is that manual labeling requires time and great effort by domain experts. Also, the idea of automatic annotation faces a difficulty that there are no ready-to-use medical terminologies in Russian. A specific syntactic structure with free word order missing subject naming and omitting conjunctions makes automatic annotation difficult. An attempt to extract deterministic characteristics from EMRs in Russian was proposed by A. Funkner [2]. However, results contained many incorrect and unnecessary constructions and it was concluded that a syntactic parsing should be used for discovering sentence structure. Morphological parsing was applied to improve labeling EMR data in Arabic in a task of assigning English labels to textual data in Arabic [3]. Arabic and Russian have in common that there are no resources for language processing and no structured medical databases like PubMed for English. They use Wikipedia to link entities in Arabic with their corresponding English entities. The application of a Wiki-based approach to Russian was studied by Sysoev A., who used Russian Wikidata graph

[1] Corresponding Author: Varvara Koshman ITMO University, 49 Kronverksky Pr., St. Petersburg, 197101, Russia; E-mail:207955@niuitmo.ru

for training word embeddings to improve performance of entity linking [4]. J. . Raiman suggested using Wikidata graph's parental relations as categories for entity linking [5]. However, aforementioned methods have not been utilized yet together for annotation of EMRs. The aim of this paper is a design and development of a method for automatic detection of repetitive semantic constructions in unstructured EMR text data.

2. Methods

The idea behind grouping similar semantic constructions is putting together similar symptom terms, word abbreviations and drug names, including those not present in the knowledge base. With this done, some of the words and phrases get relevant labels as members of a labeled group. To this end, the structure of a sentence is analyzed with morphological and syntactic parsers; cosine distance between word embeddings is used as a similarity metric between words in parsed trees. Figure 1 shows the detailed method schema.

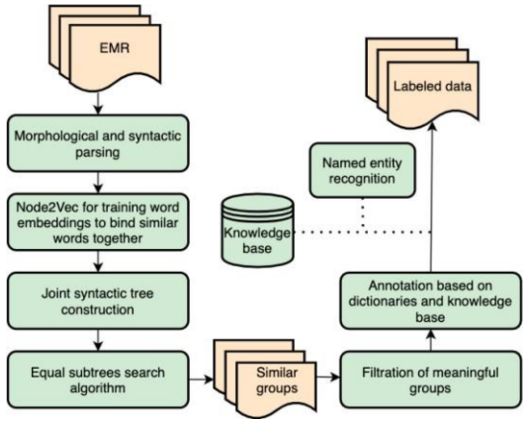


Figure 1. Method schema of sequential modules with EMR as input and labeled groups as output.

2.1. Morphological and Syntactic Parsing

As a morphological and a syntactic parsers BERT-based models implemented by DeepPavlov [6] were used. Models used in this research were trained on UD Russian SynTagRus corpus (version 2.3) and were not additionally trained on medical data.
First paragraph.

2.2. Node2Vec on Syntactic Trees

We use Node2Vec to train a CBOW model. Node2Vec [7] is commonly used with graphical structures [8,9], its strategy of random sampling helps to preserve hierarchical relations between nodes in word embeddings. The graph was created by connecting the roots of all syntactic trees with a virtual node with syntactic relations as weights. Stop words were removed for training and all words were converted to their normal forms. Node2Vec was executed with non-normalized probabilities p=2, q=3, with five random walks per root and five words in one walk at most. When q is higher than p, the algorithm's behavior is similar to local search. Such behavior is more beneficial when

dealing with syntactic relations. The resulting vector space contains embeddings trained on medical data and 50k embeddings pre-trained on Russian fiction dataset.

2.3. Algorithm for Search of Similar Subtrees in a Tree

Our algorithm for grouping similar subtrees in a tree is inspired by equal subtree search [10], but extended to the version with a dependency on multiple node's heights, not a single one, i.e. it allows a subtree (a phrase) occur in different parts of a tree (a sentence) instead of a fixed position. This is especially useful for languages with free word order like Russian. The main idea is shown in Figure 2.

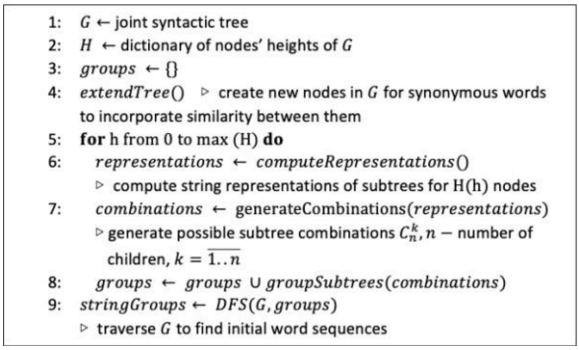


Figure 2. Main idea of similar subtree search module

2.4. Annotation Process

Some group names were decided manually based on the specific context of EMR: 'Event', 'Time stamp', 'Drug', 'Sign and symptom', 'Disease' and 'Physician specialization'. The belonging to these groups is decided by simple rules. A phrase gets an 'Event' label if it contains a verb in passive voice indicating that some action was applied, 'Time stamp' – if there is temporal information contained. For assigning 'Drug', 'Sign and symptom', 'Disease' dictionaries of labels were prepared. Our dictionary of drugs contains a parsed set of names listed in Vidal.ru reference book (6360 names), data for dictionaries of diseases (4657 names), sign and symptoms (355 names) and physician specializations (41 names) was crawled from Russian medical web sites.

Although dictionaries cover an extensive amount of information in each group, they do not contain synonyms and most common abbreviations for domain terms. This is crucial, as they are very often used in EMRs. Besides, 'disease' and 'Sign and symptom' are too general labels and make it impossible to distinguish groups of related terms. Also, medical procedures, medical organizations, body organs and etc. need to be included. As a solution a medical database was created based on the data gathered and structured from different medical databases by Wikidata. Using SPARQL queries and public MediaWiki API medical Wikidata entities were fetched. An entity is considered medical if it has a link to a medical database as a property. 32 databases that were considered most relevant for EMRs were chosen from the initial list [11]. The resulting number of medical entities was 18.9k entities and 17.1k synonyms for them. Parental relations between entities in the Wikidata graph can be used to define categories for each entity. Among parental relations we picked 'instance of', 'subclass of' and 'part of' as the most defining [5] and used entities they point to for labeling. Figure 2 represents database schema. For example,

according to Wikidata entity for 'electrocardiogram' has medical property 'P486' (MeSH descriptor ID), has synonyms 'EKG', 'ECG' and categories 'medical testtype' (instance of), 'medical test' (subclass of), 'electrophysiology' (part of). This waya mention of 'EKG' gets a 'medical test type' label as the closest parental relation.

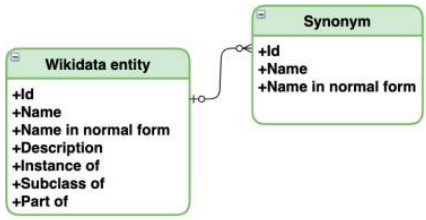


Figure 3. Medical database schema

Though the prepared database has a medical domain, there are a few hundreds of names that point to multiple entities. As these ambiguous cases are quite rare, it was decided to define a rule that prefers those entities that are closer to the context of a corpus being annotated. Concretely, a skip-gram model was trained with Node2Vec on a database graph and a forest of initial syntactic trees (Node2Vec parameters: p=1, q=2, number of walks per root=3, walk length=5) and a decision between possible entities was made in favor of the one with the highest cosine similarity score.

3. Results

3.1. Data

Experiments were conducted on a corpus of 5k sentences with time expressions in Russian of anonymized EMRs of patients with acute coronary syndrome, who were under observation in Almazov National Medical Research Centre in 2010-2015.

3.2. Extracted Groups

Our algorithm extracted nearly 11k groups in total. Figure 4 shows bar charts with frequency statistics of size of groups: commonly groups are small and in most consist of up to 10 repeated phrases. The maximum repeat length was limited to five words. Table 1 represents several examples of repeats in groups and Table 2 – synonymouswords found with Node2Vec.

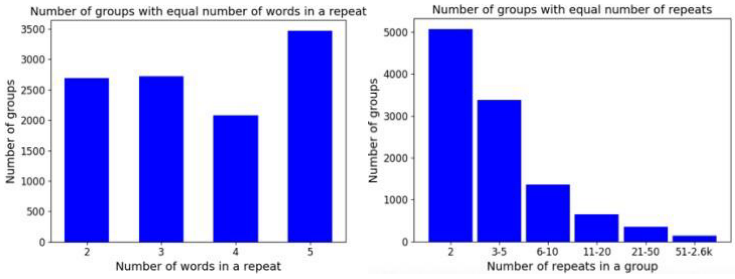


Figure 4. Bar charts with descriptive statistics of the annotated groups

Table 1. Examples of repeats in groups with labels

Group Russian	Group English	Label Russian	Label English
в больницув поликлинику в МСЧв ЦРБ	in hospital in policlinic in medical unit in hospital	медицинская организация, больница, общественный институт	medical organization, hospital, public institution
ухудшение состояния	deterioration	характеристика заболевания	disease characteristic
перелом бедра	hip fracture	анатомическая структура, кость, болезнь	anatomical structure, bone, disease

Table 2. Examples of words similar by cosine distance defined by Node2Vec

Word Russian	Word English	Synonyms Russian	Synonyms English
жена	wife	дочь, сестра, мать	daughter, sister, mother
мрт	mri	томография, флюорография, экг	tomography, fluorography, ecg

3.3. Annotation

Using dictionaries 6.8k out of 11k got annotated. Labelling with Wikidata increased the annotated number of groups to 8.6k and number of groups with more than one label grew from 1.2k to 4k. Figure 5 shows 30 most common labels.

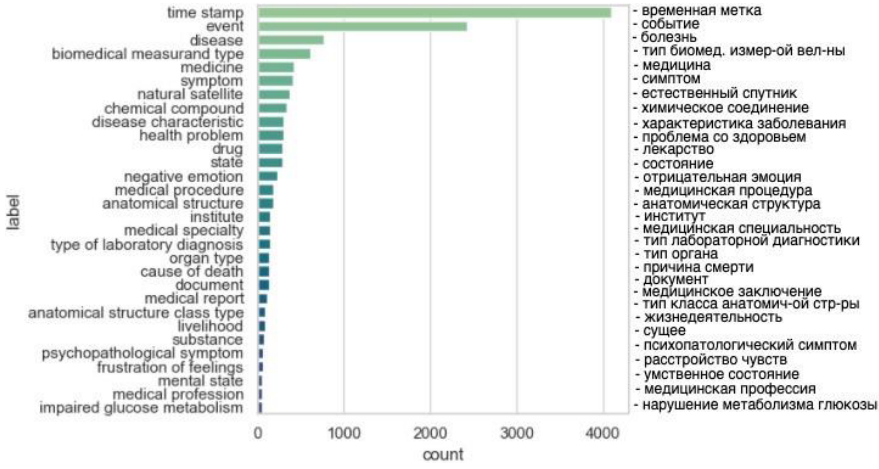


Figure 5. 30 most common labels sorted by their frequency in a result set

4. Discussion

The method we developed succeeded in joining semantically close phrases: some common abbreviations (for example, for medical organizations and lab tests), word reductions (for example, ‘department’ and ‘dep' in Russian) and minor typos. Diseases, organs, body parts, geographical places were grouped by the system. To the best of our knowledge, this is a first attempt of grouping medical free text by semantic similarity before automatic annotation intending to cover more words.

5. Conclusions

The Key contributions of this work are a design of a new methodology for automatic annotation of EMRs, a proposed method for finding similar subtrees in a tree, asuccessful application of a classic Node2Vec algorithm to syntactic trees and a creationof a medical Wikidata-based database for labeling in Russian. The whole pipeline can beadapted to other languages by changing the language-specific preprocessing module, bychanging a language code a corresponding database can be created. For Russian a graphicinterface was implemented for annotating new datasets with statistics representation.

Current limitation of a method is a limit in number of words in a sentence equal to 25. Our algorithm gets slower on longer sentences, as it gets more string combinations to compute and compare on big trees. In the nearest future it is planned to avoid this limitation and provide a fast performance for all sentences. The tool can generally increase the number of labelled datasets available, which can be used later by researches in machine learning problems related to the medical domain. This, in turn, can broad the scope of problems and save time for both domain experts preventing them from huge manual work and for researches who get their data labeled quickly.

Acknowledgement

This research is financially supported by The Russian Scientific Foundation, Agreement #17-15-01177.

References

[1] J. Liu, Z. Zhang, N. Razavian, Deep EHR: Chronic Disease Prediction Using Medical Notes, *Proceedings of the 3rd Machine Learning for Healthcare Conference* (2018).
[2] A. A. Funkner, S. V. Kovalchuk, Time Expressions Identification Without Human-Labeled Corpus for Clinical Text Mining in Russian, *Computational Science - ICCS 2020* (2020), 591-602.
[3] A. Bouziane, D. Bouchiha, N. Doumi, Annotating Arabic Texts with Linked Data, *The 4th International Symposium on Informatics and its Applications (ISIA)* **4** (2020), 1-5. Available from: http://dx.doi.org/10.1109/ISIA51297.2020.9416543
[4] A. Sysoev, I. A. Andrianov, Named Entity Recognition in Russian: the Power of Wiki-Based Approach,
[5] *Proceedings of the International Conference "Dialogue 2016"* (2016).
[6] J. Raiman, O. Raiman, DeepType: Multilingual Entity Linking by Neural Type System Evolution, *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)* (2018).
[7] Y. Kuratov, M. Arkhipov, Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language, *Proceedings of the International Conference "Dialogue 2019"* (2019).
[8] A. Grover, J. Leskovec, node2vec: Scalable Feature Learning for Networks, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), 855–864.
[9] Available from: https://doi.org/10.1145/2939672.2939754
[10] M. Kim, S. Baek, M. Song, Relation extraction for biological pathway construction using node2vec, *BMC Bioinformatics* **19** (2018). Available from: https://doi.org/10.1186/s12859-018-2200-8
[11] F. Shen, S. Liu, Y. Wang, L. Wang, A. Wen, A. H. Limper, H. Liu, Constructing Node Embeddings for Human Phenotype Ontology to Assist Phenotypic Similarity Measurement, *IEEE International Conference on Healthcare Informatics Workshop (ICHI-W)* (2018), 29-33.
[12] M. Christou, M. Crochemore, T. Flouri, C. Iliopoulos, J. Janoušek, B. Melichar, S. Pissis, Computing all subtree repeats in ordered ranked trees, *String processing and information retrieval. Lecture Notes in Computer Science* **7024** (2011), 338–343.
[13] H. Turki, T. Shafee, M. A. H. Taieb, M. B. Aouicha, D. Vrandečić, D. Das, H. Hamdi, Wikidata: A large-scale collaborative ontological medical database, *Journal of Biomedical Informatics* **99** (2019). Available from: https://doi.org/10.1016/j.jbi.2019.103292

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210580

Statistical Inference for Clustering Results Interpretation in Clinical Practice

Alexander KANONIROV[a,1], Ksenia BALABAEVA[a], Sergey KOVALCHUK[a,b]
[a] *ITMO University, Saint Petersburg, Russia*
[b] *Almazov National Medical Research Center, Saint Petersburg, Russia*

Abstract. The relevance of this study lies in improvement of machine learning models understanding. We present a method for interpreting clustering results and apply it to the case of clinical pathways modeling. This method is based on statistical inference and allows to get the description of the clusters, determining the influence of a particular feature on the difference between them. Based on the proposed approach, it is possible to determine the characteristic features for each cluster. Finally, we compare the method with the Bayesian inference explanation and with the interpretation of medical experts [1].

Keywords. Explainable artificial intelligence, interpretable machine learning, clustering interpretation, statistical inference, clinical pathways, k-means

1. Introduction

Machine learning is at the center of numerous domains in science and innovation. More and more human lives depend on their decisions. In placing so much responsibility on algorithms, we need to have complete confidence in how they work. "The problem is that a single metric, such as classification accuracy, is an incomplete description of most real-world tasks." [2] Determining trust in individual predictions is an important problem when the model is used for decision-making. For instance, when using machine learning for medical diagnosis, predictions cannot be acted upon on blind faith, as the consequences may be catastrophic.

The awareness of this problem has led to a rapid increase in the number of scientific papers on the interpretation of machine learning algorithms. For example, Tim Miller [3] gives the following definition of interpretability: «Interpretability is the degree to which a human can understand the cause of a decision» or «Interpretability is the degree to which a human can consistently predict the model's result». The higher the interpretability of a model, the easier it is for someone to comprehend why certain decisions were made. A model has better interpretability than another model if its decisions are easier to comprehend for a human than decisions from the second model.

[1]Corresponding Author: Alexander Kanonirov, ITMO University, Birzhevaya Liniya, 4, Saint Petersburg, Russian Federation; E-mail: alexander.kanonirov@yandex.ru

There's a big number of works dedicated to interpretation in the tasks of classification and regression [5-9]. However, only several works are on the topic of clustering results interpretation [10,11]. Paper [10] uses dendrograms and works only with hierarchical clustering and [11] suggests three-step clustering frameworks which work only for single feature clustering. The fundamental drawback of present methods is that they are model-specific and can only be used to explain a single model.

2. Materials and Methods

This work develops the ideas of interpreting clustering results approach described in articles [1, 4]. This article uses Bayesian inference for the post-hoc interpretation of clustering provided by the K-Means algorithm. Differences and similarities between clusters are investigated by sampling and comparing the posterior distribution of features.

In the presented work, we propose an alternative way to compare and explain clusters based on statistical inference. As an input for interpretation, we get the matrix X (n × m) with n observations and m features and corresponding cluster labels – y, that we get from clustering modeling using K-means. In order to describe the clusters, we compare the distributions of m features between clusters by statistical hypotheses testing. See the pseudocode of the procedure below.

Various statistical tests exist for formally testing statistical hypotheses. To select the proper test, the algorithm divides the data into categorical and continuous. If the data is continuous, then the algorithm checks if the data obeys normal distribution law. After that, a user has to select whether the compared groups are independent and how many groups are compared. Based on all the above, the algorithm decides which test to use to interpret continuous data (see Figures 1, 2).

Let Continuous be a set of continuous features and Categorical be a set of categorical features, where m – the number of features.

```
FOR each feature f_i, i ∈ [0, m]:
    IF f_i ∈ Continuous THEN
        1. Check whether f_i obeys the normal distribution law
            IF f_i is normally distributed THEN
                1. Asking the user if the groups are dependent (YES/NO)
                    IF answer = YES THEN
                        IF groups > 2 THEN
                            1. Apply ANOVA for Repeated Observations
                        ELSE
                            1. Apply Student's t-test for paired samples
                    ELSE
                        1. Asking the user if a comparison with a given value (YES/NO)
                            IF answer = YES THEN
                                1. INPUT value
                                2. Apply Z-test
                            ELSE
                                IF groups > 2 THEN
                                    1. Apply One-way ANOVA
                                ELSE
                                    1. Apply Student's t-test for independent samples
```

Figure 1. Pseudocode for comparing procedure using normally-distributed continuous features.

Similarly, the Figures 1-2 show the algorithm if the features have an abnormal distribution or are categorical.

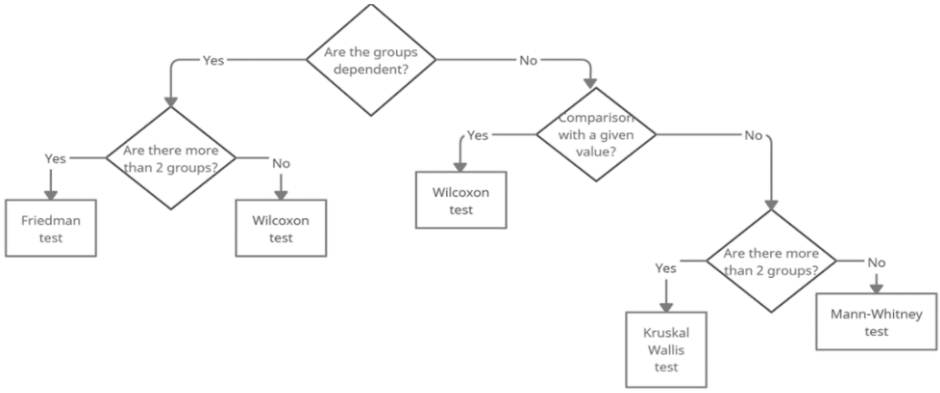


Figure 2. An interpretation algorithm based on statistical inference for not normally distributed continuous variables

If the data is categorical, then the algorithm only needs to check whether the compared groups are independent and how many groups are compared. Based on this, it is decided which test to use to interpret categorical data (see Figure 3). After the test is selected and applied to the data, the algorithm displays in a form understandable for a medical expert, which features have a significant effect on the difference between the compared groups, and also indicates the features characteristic for each group. We will talk in more detail about what the data that the algorithm outputs and their interpretations mean in the next chapter.

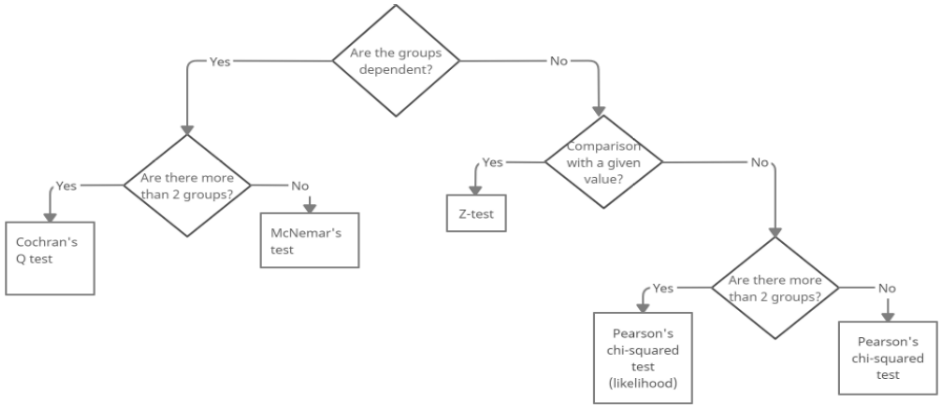


Figure 3. An interpretation algorithm based on statistical inference for categorical variables

3. Results and Discussion

The dataset consists of 3312 observations. By observation, we understand a clinical episode, a single hospitalization of a patient with a Diagnosis of Acute Coronary Syndrome. The initial feature set included the clinical pathways - a sequence of departments a patient passes during hospitalization. In order to improve the interpretation, the feature set was extended with additional information from the electronic health records. All features are boolean. For instance, surgery, death outcome, stroke, stenting, coronarography, rehabilitation, clinical death, cardiogenic shock, rehospitalization.

The data presented were interpreted using Bayesian inference and a medical expert in [1]. Next, we compare these results with the results obtained by the algorithm described in the previous chapter. As a comparison metric, we use the percent of a coincidence for each cluster, calculated as a proportion of a number of features in the intersection to the number of features provided by the doctor or Bayesian inference. The main and most interesting results of the experiment are presented in Table 1.

Table 1 The results of the statistical inference interpretation algorithm and its percentage of agreement with the results of the Bayesian interpretation algorithm and the results of the interpretation of the medical expert.

Cluster	Num Obser-vations	Bayesian inference (BI)	Statistical inference	Doctor's Interpretation (DI)	% of explanation match with BI	% of expla-nation match with DI
1	116	'icu', 'rehospitalization', 'stenting', 'nevrology_dep', 'delayed_surgery', 'additional_surgeries'	'age', 'minutes_before_first_operation', 'stenting', 'serious_condition', 'icu', 'revascularization', 'rehospitalization', 'nevrology_dep'	'rehabilitation', 'rehospitalization', 'additional surgeries'	66.7	66.7
2	821	'outcome better', 'stenting', 'stroke', 'no surgery'	'num_operations', 'age', 'minutes_before_first_operation', 'no_surgery', 'outcome better', 'rehospitalization', 'revascularization'	'optimal path', 'transfer_from_stationar', 'rehospitalization', 'revascularization'	25	50
3	460	'outcome death', 'transfer_from_inhospital, 'no surgery'	'outcome_death', 'revascularization', 'no_surgery'	'outcome death', 'comorbidity', 'coronarography', 'no surgery, 'revascularization'	66.7	60
4	193	'cardiogenic shock', 'coronarography', 'rehospitalization', 'wheelchair_transporting', 'nevrology_dep_rehosp', 'vessel_surgery_dep', 'more_surgeries	'num_operations', 'minutes_before_first_operation', 'no_surgery'	'no surgery'	0	100
9	287	'outcome death', 'cardiogenicshock', 'seriouscondition', 'vessel_surgery_dep', 'transfer_from_stationar', 'no_surgery'	'age', 'operation', 'coronarography', 'clinical_death', 'vessel_surgery_dep', 'no_surgery'	'outcome death', 'coronarography', 'no_surgery	50	100
10	203	'outcome better', 'coronarography', 'urgent_operation', 'rehabilitation', 'cardio_dep', 'vessel_surgery_small', 'optimal_path'	'outcome better', 'coronarography', 'rehabilitation', 'cardio_dep', 'vessel_surgery_small'	'rehabilitation'	71.4	100

Cluster 1. Both methods found that patients in the cluster were more likely to be re-hospitalized and undergo additional surgery during treatment. However, the algorithm,

based on statistical inference, determined that these patients needed additional rehabilitation. All of the above was also confirmed by a medical expert.

Cluster 2. Both algorithms determine a positive outcome for patients without surgery. But the doctor claims that patients from this cluster follow the optimal clinical path, which is not supported by algorithms in any way.

Cluster 3. Both algorithms establish death without surgery as the characteristic outcome for this group of patients. There were also no significant differences found in other characteristics suggested by the medical expert.

Cluster 4. An algorithm based on statistical inference revealed the features characteristic of the cluster - the presence or absence of surgical intervention. An algorithm based on Bayesian inference has shown that patients can have more than one operation. According to the doctor, patients here tend to get conservative treatment, without surgery.

Cluster 5. Both algorithms identified patients with complications characteristic of this cluster, as well as a favorable outcome for them, which emphasizes the algorithm based on Bayesian inference, indicating the sign of the optimal clinical pathway. In this group, the surgeon admits patients with complications and the necessity of additional post-surgery recovery treatment in other departments.

Cluster 6. Algorithms have revealed that stenting surgery is typical for patients, but for some reason, it is postponed. The medical expert mentions a complex diagnostics process for patients in this group.

Cluster 7. Both approaches in interpretation revealed multiple complications and the death of patients. However, unlike the medical expert and Bayesian inference, the algorithm based on statistical inference did not reveal that postponed operations are a characteristic feature of this group.

Cluster 8. The main diagnosis for both interpretations is stroke, but the algorithm based on statistical inference also indicates some complications for patients in the form of cardiac shock. All of the above was also confirmed by a medical expert.

Cluster 9. Algorithms indicate problems with blood vessels in patients, as well as a possible death.

Cluster 10. Both approaches determined that patients in this group are undergoing rehabilitation, therefore, they are likely to have a favorable outcome.

4. Conclusion

In conclusion, we would like to say that the goal of the study has been achieved: an approach has been developed to interpret the clustering results using statistical inference. This method allows you to get an idea of the clusters by determining the influence of a particular feature on the difference between them, on the basis of which it is possible to determine the characteristic features for each cluster.

When comparing the two approaches, no significant contradictions were found in the interpretation. In some cases, both algorithms are also capable of complementing each other's work, which is confirmed on the basis of the conclusions of a medical expert.

These are the primary results of our experiments, but we can already talk about a fairly good accuracy and availability of interpretation, so we are confident that this work will serve well in the application of machine learning models in clinical practice and other fields.

Acknowledgment

This research is financially supported by The Russian Scientific Foundation, Agreement #17-15-01177.

References

[1] Balabaeva K, Kovalchuk S. Post-hoc Interpretation of clinical pathways clustering using Bayesian inference. *Procedia Computer Science*. **178** (2020), 264-273.
[2] Doshi-Velez F., and Been K. Towards A Rigorous Science of Interpretable Machine Learning. *no. Ml:* **1–13** (2017), http://arxiv.org/abs/1702.08608.
[3] Miller T. Explanation in Artificial Intelligence: Insights from the Social Sciences. *arXiv Preprint* (2017), arXiv:1706.07269.
[4] Balabaeva K, Kovalchuk S. Clustering Results Interpretation of Continuous Variables Using Bayesian Inference. *Studies in Health Technology and Informatics*. **281** (2021), 477-481
[5] Molnar C, Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. URL: https://christophm.github.io/interpretable-ml-book/index.html. (Last update: 21.05.2020)
[6] Ribeiro M, Singh S., Guestrin C. Why Should I Trust You?: Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA:1135–1144. (2016).
[7] Lundberg S, Su-In L. A Unified Approach to Interpreting Model Predictions. *Neural Information Processing Systems Conference 2017*. (2017).
[8] Sundararajan M, Najmi A. The many Shapley values for model explanation. *arXiv Preprint* (2019), arXiv:1908.08474
[9] Wachter S, Mittelstadt B., Russell C. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*. **31(2)** (2018)
[10] Randriamihamison N., Vialaneix N., Neuvial Pierre. Applicability and Interpretability of Hierarchical Agglomerative Clustering With or Without Contiguity Constraints. (2019), arXiv:1909.10923
[11] Yang C., Shi X et al. I Know You'll Be Back: Interpretable New User Clustering and Churn Prediction on a Mobile Social Application. (2018), 914-922.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210581

A Modeling Framework for Decision Support in Periprosthetic Joint Infection Treatment

Vasiliy N. LEONENKO[a1], Yulia E. KALIBERDA[a] and Vasiliy A. ARTYUK [b]
[a] *ITMO University, St. Petersburg, Russia*
[b] *Russian Scientific Research Institute of Traumatology and Orthopedics "R.R. Vreden", St. Petersburg, Russia*

Abstract. In this paper, we present a framework, which aims at facilitating the choice of the best strategy related to the treatment of periprosthetic joint infection (PJI). The framework includes two models: a detailed non-Markovian model based on the decision tree approach, and a general Markov model, which captures the most essential states of a patient under treatment. The application of the framework is demonstrated on the dataset provided by Russian Scientific Research Institute of Traumatology and Orthopedics "R.R. Vreden", which contains records of patients with PJI occurred after total hip arthroplasty. The methods of cost-effectiveness analysis of treatment strategies and forecasting of individual treatment outcomes depending on the selected strategy are discussed.

Keywords. Modeling framework, periprosthetic joint infection, decision support, Markov model, decision tree

1. Introduction

In today's aging society with high demands on mobility, arthroplasty is increasingly performed for both elderly and younger patients. The widespread use of replacing the affected joints with artificial ones increases the need for revision arthroplasty (re-THR). Today, periprosthetic joint infection (PJI) is one of the leading indications for revision surgery and by far the most ominous complication in artificial joint patients [1]. Periprosthetic infection is associated with high morbidity and requires complex treatment strategies including multiple surgical revisions and long-term antimicrobial treatment, because the implant as a foreign body increases the pathogenicity of bacteria and the presence of biofilm makes the diagnosis and treatment problematic [2].

A scientific approach to medical practice requires a search for evidence of the efficacy and safety of existing and promising surgical methods for treating PJI. The optimal tool in solving this problem is evidence-based medicine, allowing comparison, generalization and wide practical application of the data obtained [3]. Thus, the need for a clinical and economic analysis of PJI treatment from the standpoint of evidence-based medicine seems to be very relevant. The investigations related to the creation of cost-

[1] Corresponding Author: V.N. Leonenko, ITMO University, Kronverksky Pr. 49A, St. Petersburg, 197101,Russia; E-mail: vnleonenko@itmo.ru

effective approaches to PJI treatment [4, 5] mention the issues connected with the corresponding analysis due to lack of quality studies.

In this paper, we present a framework, which aims at facilitating the choice of the best strategy related to the treatment of PJI. The application of the framework is demonstrated on the dataset provided by Russian Scientific Research Institute of Traumatologyand Orthopedics "R.R. Vreden", which contains records of the patients with PJI occurred after total hip arthroplasty. The methods of cost-effectiveness analysis of treatment strategies and forecasting of individual treatment outcomes depending on the selected strategy are discussed.

2. Data

The analyzed data set contains the records of 571 patient who were subjected to a revision total hip replacement (re-THR) in the period of 2000-2020. The patient records were collected in two different ways. A large part of the disease histories (a so-called retrospective group of 405 patients) was taken from the archives, whereas the remaining part (prospective group, 166 patients) was filled in real time by adding data of the patients undergoingtreatment starting from 2014. All the patients were being observed for the possible PJI relapse till the end of 2020. In the retrospective group, the following treatment methods have been applied: resection arthroplasty (RA), revision operation with the preservation of endoprosthesis (re-THR-PE) and two-stage revision total hip replacement with the two consecutive interventions separated by more than 2 months. In the prospective group, new treatment methods were presented, namely, one-stage re-THR and partial re-THR. Also, in the prospective group, the waiting time between surgeries for two-stage revisions did not exceed 2 months. The patients in the prospective group who underwent two- stage re-THR were divided into two subgroups based on their waiting time: 2-3 weeks and 6-8 weeks correspondingly. In many cases, in addition to PJI-related surgeries, additional operations were required due to the relapse of PJI or other issues (postoperativewound hematomas, spacer dislocations, etc.). The recorded data we worked with contains15 different types of operations, which were divided into three groups: operations, which have no connection with PJI (e.g., endoprosthesis (EP) installation + spacer removal, EPinstallation (no spacer), etc.); first case of PJI or PJI relapse (e.g., debridement, debridement + spacer installation); PJI relapse (e.g., debridement + spacer reinstallation, spacer removal + support osteotomy).

3. Framework Description

3.1. Decision Tree

As a continuation of our previous research efforts [6], we developed an algorithm for creating and verifying detailed decision trees, which in their turn serve as a base of an imitational discrete-event model. The decision tree is comprised of the transitions between the states, which are attributed to different medical interventions. Each transitionsignifies the change in patient functional capacity and is associated with the treatment costs. The time passed between the transitions is not explicitly taken into account.

The framework builds trees for a given treatment method based on the subsamples of the corresponding patient records. The calibration procedure algorithm uses a recursive approach and generates tree branches with the transition probabilities as an output. To assess how the results of the sample analysis might be generalized to an independent dataset of patients, we calculate mean values of the transition probabilities along with their interval assessments with the help of bootstrap aggregating technique. A fragment of a decision tree for partial re-THR is shown in Figure 1.

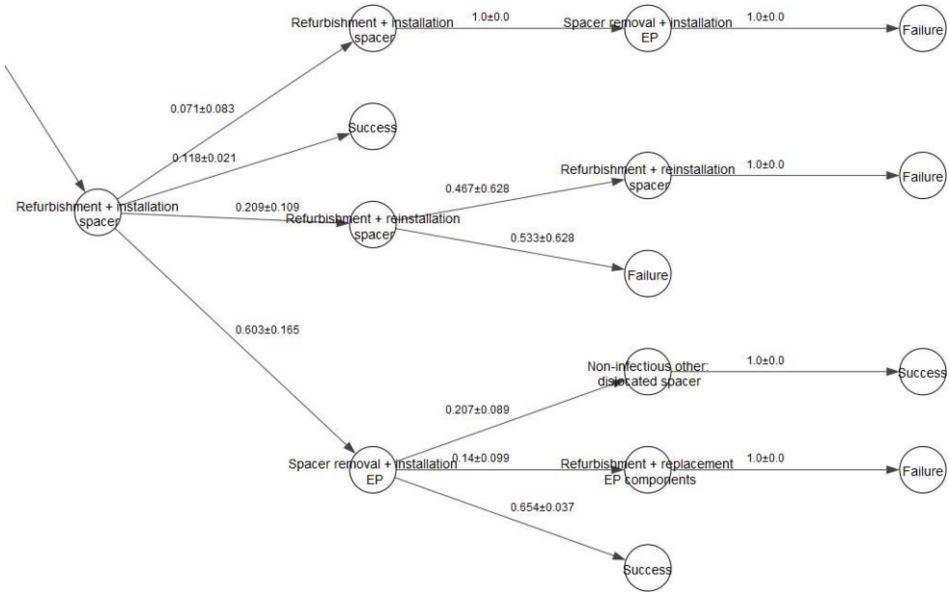


Figure 1. A fragment of the decision tree with confidence intervals for transition probabilities, partial re-THR

The decision trees based on the mean probabilities constitute a base for the discrete event imitational model. Each simulation run consists of generating an individual patient trajectory via Monte-Carlo methods according to the probabilities calculated during the cross-validation procedure. The output of the model includes generated individual trajectories and probability distributions for the treatment states calculated via repetitive simulation runs.

The decision tree has the advantage of detailed description of trajectories, which makes it possible to calculate the statistics of changes in expected expenses and the overall QALY (quality-adjusted life-years, a generic measure of disease burden), related to particular interventions. Our framework supports the assignments of two associated parameters, related to the impact of the intervention, to each branch of the tree. The first parameter is the intervention cost in rubles taken from the list of costs of intervention types. The second parameter is the gained utility of the patient during a fixed timeframe calculated in QALY units. The quantitative outcomes of the treatment in terms of health-care costs and QALY units gained by the patient might be derived from the decision tree using the following formula: $C = \sum_i p_i c_i$, where p_i is the probability of selecting the branch, obtained by cross-validation, c_i is the impact measured in either of the two units.

The interval assessment of C can be calculated using the same formula with left and right boundaries for p_i used instead of their mean assessments.

The resulting values of C_{rubl} and C_u might be used to calculate the costs of one QALY unit and compare them between the treatment strategies. QALY units and costs for particular tree branches might be derived from the external sources, for instance, the data of the Center for the Evaluation of Value and Risk in Health which was used in thesimilar calculation from [7]. In the current study, however, we relied on the data providedby Russian Scientific Research Institute of Traumatology and Orthopedics "R.R. Vreden". The operation costs were taken from the disease histories, and the QALY units were assessed based on the EQ-5D indices for each particular patient measured between the subsequent operations [8]. The results of QALY assessment are given in Tables 1, 2.

Table 1. Average QALY for treatment methods in the retrospective group according to the decision trees

Treatment method	Re-THR-PE	Two-stage > 2 months	RA
QALY	1.537	2.363	0.675

Table 2. Average QALY for treatment methods in the prospective group according to the decision trees

Treatment method	Two-stage 2-3 weeks	Two-stage 6-8 weeks	One stage	Partial
QALY	1.124	0.634	1.035	1.053

3.2. Markov Model

The disadvantage of a decision tree approach is that the tree might become extensively big and intractable in case of big samples of patients, long observation time or big variety of possible interventions. Also, it does not account for the time passed between the model states, which sometimes can be crucial for the accurate cost–effectiveness analysis. As an alternative approach, we developed a Markov model with generalized states. The generalized intervention types assumed in the model are created according to the intervention classification described in Sec. 2. We distinguish PJI-related interventions (PJI or PJI relapse), interventions not related to PJI, and, as a separate intervention type, a second stage intervention for two-stage treatment methods ('Endoprosthesis installation + spacer removal'). The states of the simulated patient are the following: (a) PJI (waiting for the treatment), (b) second stage (no PJI, waiting for the spacer removal in two-stage treatment methods), (c) additional surgeries (waiting for the treatment of a non-PJI issue), (d) observation (no PJI), (e) death. The last state is the absorbing state. The simulation starts from the state 'PJI'. The situations of a first PJI case and a recurrent PJI are not distinguished due to the lack of corresponding data in the records. The time in the model is discrete, with the time step equal to one month.

The transitional probabilities are calculated based on the available patient records using the same bootstrapping procedure as for the decision tree case. An example of the calibrated model is shown in Figure 2. The model makes it possible to generate individual patient trajectories with the consideration of time, which is useful for the calculation of the expected amount of observation time and the hospitalization time — the latter greatly influences the overall treatment costs.

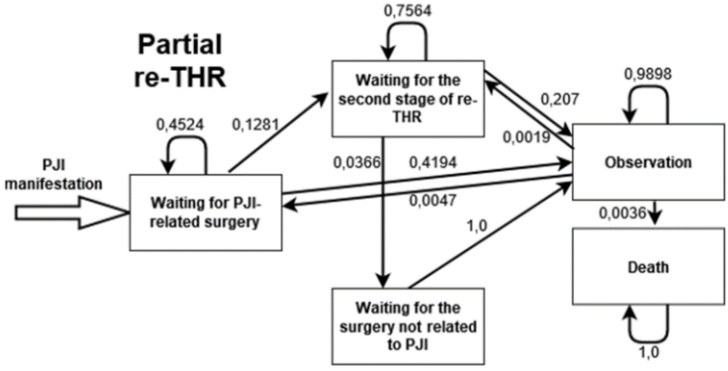


Figure 2. Markov model calibration results for partial re-THR (confidence intervals for the probabilities are omitted)

The treatment impact for a fixed individual patient trajectory is calculated according to the formula $C(t) = \sum_i t_i c_i$, where c_i are the monthly costs or QALY units associated with the patient state i in the model, and t_i is the patient's time of staying in the state i (the number of months). Unlike in case of the decision tree, accurate numbers of c_i (in both rubles and QALY units) are not available due to generalized states used in the model. They could be taken by averaging the costs for the patient state based on the set of particular interventions related to that state and on the experts' opinion. For example, it might be assumed that the ideal state of the patient in terms of QALY units relates to 'Observation' status. The lowest QALY values correspond to 'PJI'. The quality of life of a patient waiting for the second stage or additional surgeries is higher than in case of PJI, but lower than in the 'Observation' state due to corresponding health issues (particularly, the patients waiting for the second stage of the treatment have limited mobility due to spacer installation which badly affects their QALY count). Average QALY values calculated for different methods are shown in Tables 3 and 4. Following expert assumptions,we set the QALY for 'PJI' state equal to 0.35, for 'Second stage' equal to 0.7, for'Non-PJI operation' equal to 0.5, for 'Observation' equal to 0.85, and for 'Death' equal to 0.

Table 3. Average QALY for treatment methods in the retrospective group according to the Markov simulations

Treatment method	Re-THR-PE	Two-stage > 2 months	RA
QALY	1.88	1.92	1.79

Table 4. Average QALY for treatment methods in the prospective group according to the Markov simulations

Treatment method	Two-stage 2-3 weeks	Two-stage 6-8 weeks	One stage	Partial
QALY	1.925	2.055	1.99	1.952

4. Discussion

In this paper, a modeling framework is presented, which aims at facilitating the decision making for healthcare professionals in the area of periprosthetic joint infection treatment. By using two different approaches within one framework, which is the decision tree approach and the Markov modeling approach, we can obtain a detailed static analysis of

the prospected patient treatment trajectories, depending on the selected strategy, or alternatively perform a dynamic simulation of a patient trajectory of transitions between the generalized states in an imitational model. The former helps to calculate detailed total operational costs and obtained quality of life, their average values and their distributions, whereas the latter gives an opportunity to monitor and forecast the dynamics of costs and QALY units. After an additional verification of the models, specialists of Russian Scientific Research Institute of Traumatology and Orthopedics "R.R. Vreden" will test the proposed framework in clinical practice. Based on the results obtained, we plan to develop this tool further and to apply it for other problems related to PJI treatment. One of the possible framework applications consists in forecasting disease trajectories for patients using their individual characteristics (age, gender, body mass index) as parameters affecting the transition probabilities. Another one is to consider the prediction of PJI cases at the city level, using synthesized populationsas a model input [9, 10]. The framework coupled with the statistical model of PJI occurrence probability and the synthetic demographic data will make it possible to assess the middle- and long-term expenses of PJI treatment, the hospital occupancy and the prospected potential years of life lost in the urban population depending on the prevalent treatment methods.

Acknowledgement

This research is financially supported by The Russian Science Foundation, Agreement #19-11-00326.

References

[1] Shubnyakov II, Tikhilov RM, DenisovAO, Akhmedilov MA, Cherny AZ, et al. What Has Changed in the Structure of Revision Hip Arthroplasty?. Traumatologyand Orthopedics of Russia, 2019; 25(4):9-27.
[2] Masters EA, Trombetta RP, de Mesy Bentley KL, Boyce BF, Gill AL, et al. Evolving concepts in bone infection: redefining "bio-film","acute vs. chronic osteomyelitis","the immune proteome" and "local antibiotic therapy". Bone research, 2019; 7(1):1-18.
[3] Guyatt G, Cairns J, Churchill D, Cook D, Haynes B, et al. Evidence-based medicine: a new approach to teaching the practice of medicine. JAMA,(1992; 268(17):2420-2425.
[4] Li C, Renz N, Trampuz A. Management of periprosthetic joint infection. Hip & pelvis, 2018; 30(3):138-146.
[5] Hernández-Vaquero D, Fernández-Fairen M, Torres A, Menzie AM, Fernández-Carreira JM, et al. Treatment of periprosthetic infections: an economic analysis. The Scientific World Journal, 2013; 2013:821650. doi:10.1155/2013/821650.
[6] Kaliberda YE, Leonenko VN, Artyukh VA. Towards Cost-Effective Treatment of Periprosthetic Joint Infection: From Statistical Analysis to Markov Models. In International Conference on Computational Science (pp. 494-505). Springer, Cham, 2021, June.
[7] Srivastava K., Bozic KJ, Silverton C, Nelson AJ, Makhni EC, Davis JJ. Reconsidering strategies for managing chronic periprosthetic joint infection in total knee arthroplasty: using decision analytics to find the optimal strategy between one-stage and two-stage total knee revision. JBJS, 2019;101(1):14-24.
[8] Dolan P. Modeling valuations for EuroQol health states. Medical care, 1997; 35(11): 1095-1108.
[9] Bates S, Leonenko V, Rineer J, Bobashev G. Using synthetic populations to understand geospatial patterns in opioid related overdose and predicted opioid misuse. Computational and Mathematical Organization Theory, 2019; 25(1):36-47.
[10] Leonenko VN. Analyzing the spatial distribution of acute coronary syndrome cases using synthesizeddata on arterial hypertension prevalence. In International Conference on Computational Science (pp. 483-494). Springer, Cham, 2020, June.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210582

Predictive Modeling of COVID and non-COVID Pneumonia Trajectories

NIKITA MRAMOROV[1], ILYA DEREVITSKII, SERGEI KOVALCHUK
ITMO University, Saint Petersburg, Russia

Abstract. Today pneumonia is one of the main problems of all countries around the world. This disease can lead to early disability, serious complications, and severe cases of high probabilities of lethal outcomes. A big part of cases of pneumonia are complications of COVID-19 disease. This type of pneumonia differs from ordinary pneumonia in symptoms, clinical course, and severity of complications. For optimal treatment of disease, humans need to study specific features of providing 19 pneumonia in comparison with well-studied ordinary pneumonia. In this article, the authors propose a new approach to identifying these specific features. This method is based on creating dynamic disease models for COVID and non-COVID pneumonia based on Bayesian Network design and Hidden Markov Model architecture and their comparison. We build models using real hospital data. We created a model for automatically identifying the type of pneumonia (COVID-19 or ordinary pneumonia) without special COVID tests. And we created dynamic models for simulation future development of both types of pneumonia. All created models showed high quality. Therefore, they can be used as part of decision support systems for medical specialists who work with pneumonia patients.

Keywords. Pneumonia, disease simulation modeling, hidden markov models, bayesian network, machine learning

1. Introduction

With appearance of the COVID-19, our lifestyle has changed. It was uncommon practice to learn and work from home remotely, to wear a mask while going outside and generally to be in isolation for a large amount of time. The world manages to cope withCOVID-19 by creating vaccines and doing researches. However, curing and diagnosing one of the most infamous consequences of the COVID-19, pneumonia, is still a challenge for science. According to [1], 4,574,089 people around the world died to the consequences of COVID-19 by 2021 which includes pneumonia as on of the major factors. With 221,134,742 registered cases of COVID-19 infection, this numbers grows each week. There are some expectations of the COVID-19 third wave exists nowadays. Many researches are made in order to explore of COVID-19 pneumonia, to predict its behavior or mutations on different types of patients. The general approach for this task is the usage of mathematical models, such as Hidden Markov Model and Bayesian Network. Considering the facts above, pneumonia nowadays is considered as a strong threat towards elderly and humanity in general, so the pneumonia modeling task has a large significance these days.

[1] Corresponding Author: Nikita Mramorov, ITMO University, Saint Petersburg, Russia; Email: nmramorov98@gmail.com

2. Literature Review

For solving disease modeling tasks one of the best approaches are Hidden Markov Models and Markov Chains. For example, Kehinde Adiguna, Ayan Adeleke modeled the infections diseases trajectories with such methods. They established that the Markov chain could be used as a mathematical model for infectious diseases and revealed that the past history would affect the future only through the present state of infectious in disease which is a strong property of the Markov Chain [2]. Such approach could be applied to another modeling task. Tilahun Ferede Asena, Ayele Goshu, Mebratu Senbeta, Derbachew Asfaw Teni applied the semi-Markov modeling approach for HIV/AIDS progression. They claim that semi-Markov models are more agile in terms of parametric distribution choice than traditional Markov Chain [3].

Another option for such task includes usage of are Bayesian Networks. This class of models is famous for its graphical modeling approach that models the conditional probabilistic relationships of certain independent variables. For example, Hasan Aykut Karaboga, Aslihan Gunel and their colleges used Bayesian network as a decision tool for predicting ALS disease. They claim that Bayesian networks are widely used medicine or biology and extremely useful in terms of ease of use of posterior probabilities especially on risk assessment studies [4]. Another study group from National University of Singapore states that the underlying causal assumptions and interpretability make Bayesian networks a decent tool for medical applications and in CAD risk prediction [5].

Scientific research of disease using Computer Science and Data Analysis approaches, such as Deep Learning, is a significant research fields these days. Currently, COVID-19 pneumonia might be determined by generalized convolutional neural network model from chest X-ray images [7]. However, deep learning approach does not only used for tasks with COVID-19 or pneumonia. Authors [8] created recurrent disease progression networks for modeling risk trajectory of heart failure. Their study explored the ability of Recurrent Neural Networks in predicting long-term trajectories of recurrent events.

Each approach mentioned above has both advantages and disadvantages: Hidden Markov Models do not consider “past” states of variables and relies on the current state of presented variable; Bayesian Networks requires expert models with conditional probabilities for model evaluation and may not have all interconnections for complex systems. Our approach was developed in order to take into account advantages of both methods and decrease negative impact of both models’ disadvantages.

3. Data

Data for this research was taken from the Federal State Budgetary Institution “V.A. Almazov National Medical Research Center” of the Ministry of Health of the Russian Federation. It consists of 234 patients with different pneumonia symptoms and other diagnosis. Among patients there are 63 % male and 37 % female correspondingly. In terms of pneumonia types the most common is nosocomial type.

Data consists of patients with COVID and non-COVID pneumonia. Average age for non-COVID pneumonia is 66, for COVID pneumonia is 71. P-value is more than 0.05 so our hypothesis is considered valid.

The suggested method in this article implies to figure out dependencies and connections for COVID and non-COVID pneumonia. Patients were divided into groups wherever any of them have COVID pneumonia or not. Such patients were explicitly

marked by literals “U07.1” in their anamnesis. We propose a thesis that COVID pneumonia disease flow lies a greater pressure on patient health than non-COVID one. A comparison of pneumonia features that are considered sever according to clinical recommendations for COVID and non-COVID patients had been made. To properly estimate the thesis, we used Chi-Squared statistical test to compare two samples from separated COVID pneumonia and non-COVID pneumonia patient’s data. Finally, the statistical patterns have been obtained and used for building disease simulation models.

4. Methods

The aim of this research is to explore COVID and non-COVID pneumonia trajectories and make its comparison based on statistical models, such as Bayesian Networks and Hidden Markov models. Figure 1 describes the algorithm for finding COVID and non-COVID pneumonia disease trajectories.

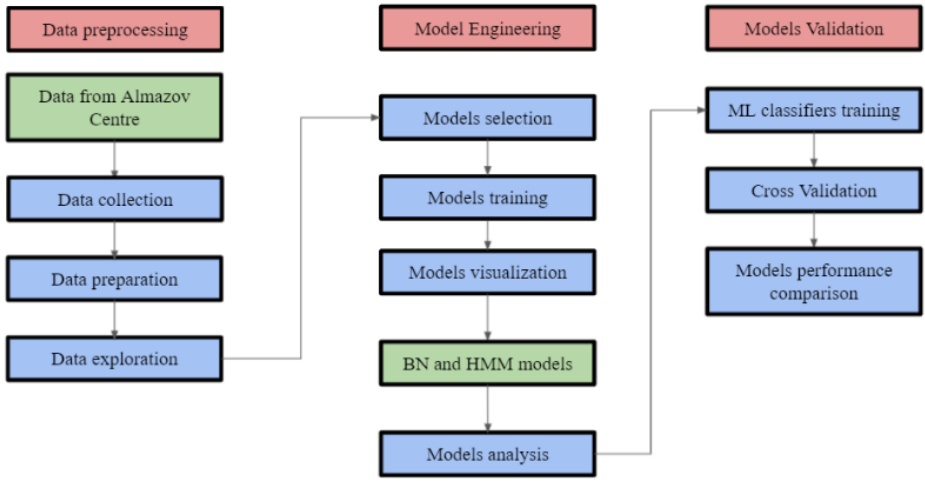


Figure 1. The diagram of the search algorithm for the best models.

There are three main steps in our research: data preprocessing, model engineering and models validation. Data preprocessing mostly consists of extracting data from clinical anamneses. Each patient has records of its every health manipulation.

5. Modeling and Results

This section refers to models analysis, interpretation and validation. Each model has produced unique set of trajectories for both COVID and non-COVID pneumonia. Authors applied modularity maximization techniques in order to locate the nodes groups and search for patterns and common trajectories of the disease.

First step involves building Hidden Markov Models for non-COVID and COVID pneumonia using Viterbi algorithm [6] (see Figure 2).

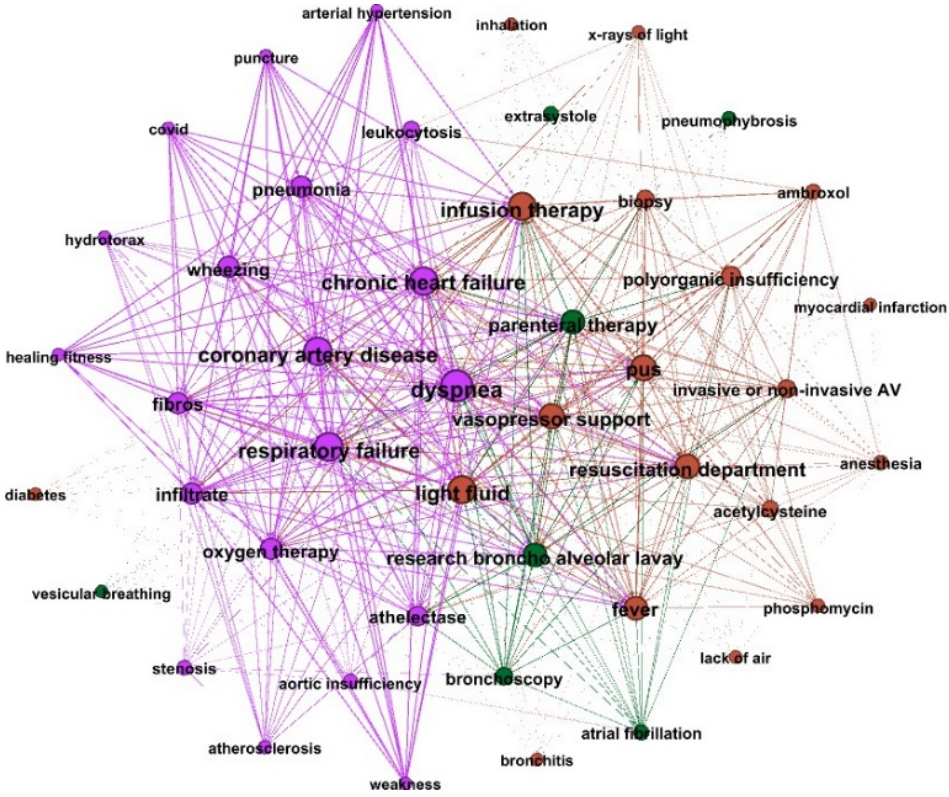


Figure 2. COVID Hidden Markov Model.

For both graphs cluster analysis had been applied. Each graph has three clusters with different features. Every cluster describes the trajectory of COVID pneumonia. Purple one describes COVID pneumonia directly as it has both COVID and pneumonia features within it. Largest nodes are coronary artery disease, chronic heart failure, dyspnea, infiltrate. Most of them describes one of the most common features of non-COVID pneumonia. Second cluster, colored in red, marks nodes, such as light fluid, infusion therapy, pus, invasive or non-invasive AV, fever, vasopressor support and resuscitation department as nodes with highest degree, hence, the most frequent one in COVID pneumonia trajectory. Last cluster consists of mostly features connected with breathing process, such as atrial fibrillation, bronchoscopy, vesicular breathing etc.

Next step involved creation of Bayesian Network models for both COVID and non-COVID pneumonia. For learning graph structure from the dataset we used Hillclimb algorithm due to its simplicity and fast convergence speed, and for parameter learning we chose to use BDeu score (see Figure 3).

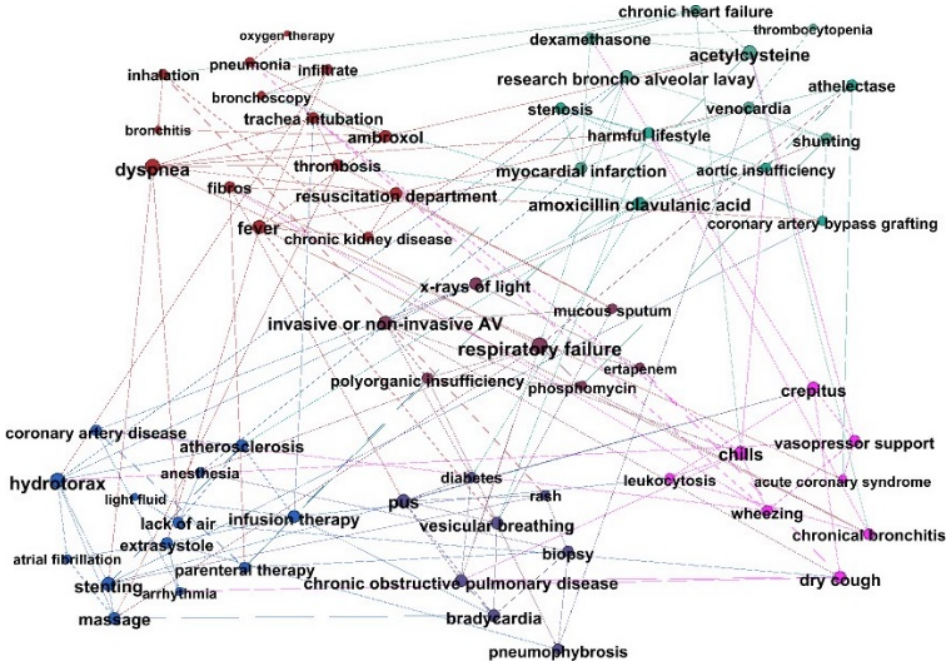


Figure 3. COVID Bayesian Network.

COVID pneumonia model has 5 clusters, contains most of the features of non-COVID pneumonia. The key difference from the non-COVID model here is the existence of severe COVID pneumonia cluster. It contains respiratory failure, polyorganic insufficiency, invasive or non-invasive AV, mucous sputum. Those features are considered severe according to clinical recommendations.

5.1. Models Validation

This subsection involves model validation through task of making binary classification for a given set of patients whatever type of pneumonia (COVID or non-COVID) they have and evaluating its precision. For better comparison we offer four another classifiers that would challenge two build models: XGBoost linear model, Random Forest Classifier, Logistic Regression model.

Results of predictions are shown in Table 1.

Table 1. Predictions

Method	Precision, %	Recall, %	Accuracy, %	F1-score, %
Hidden Markov Model	91	82	92	95
Bayesian Network	92	54.5	53.1	68.6
XGBoost linear model	88	51.2	53.4	66.6
Random Forest Classifier	91.5	90.3	91.4	95.5
Logistic Regression	92	32.5	38.2	49

Bayesian Network and Logistic Regression models reach the highest score among the models in terms of precision. However, in general, Hidden Markov Model shows better metrics on par with Random Forest Classifier.

6. Conclusions

In this paper authors proposed two dynamical models of simulating COVID and non-COVID pneumonia based on Hidden Markov Model and Bayesian Network architectures. Every model has shown COVID and non-COVID pneumonia unique trajectories, however Hidden Markov model turned out to be more preferable than the Bayesian Network due to its performance on models validation. Such process also provided that both models are able to detect most significant features of COVID and non-COVID pneumonia and might be useful for future research and medical analysis. Further work might include comparison of the built models with graphs made by experts for final validation and future tuning.

Acknowledgements

The reported study was funded by RFBR according to the research project № 20-31-90135.

References

[1] WHO Coronavirus (COVID-19) Dashboard, https://covid19.who.int/, 2021
[2] Kehinde Adiguna, Ayan Adelekeb, Oluwasesan Adewusic, Oluwafunmilola Olubiyid, Omobolaji Halide, Bayowa Babalolaf. Modelling Infectious Diseases Using Markov Chain, American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS) **61(1):280**, August 2020
[3] Tilahun Asena, Ayele Goshu, Mebratu Senbeta, Derbachew Teni, Semi-Markov Modelling of HIV/AIDS Disease Progression, December 2020
[4] Hasan Aykut Karaboga, Aslihan Gunel, Senay Vural Korkut, Ibrahim Demir and Resit Celik. Bayesian Network as a Decision Tool for Predicting ALS Disease, Brain Sciences **11(2):11**, January 2021
[5] Qiongyu Chen1, Guoliang Li, Bin Han, Chew Kiat Heng, and Tze-Yun Leong. Coronary Artery Disease Prediction with Bayesian Networks and Constraint Elicitation, September 2006
[6] G.D. Forney. The viterbi algorithm, Proceedings of the IEEE (Volume: 61, Issue: 3, March 1973), Codex Corporation, Newton, MA, USA, 1973
[7] Adi Alhudhaif, Kemal Polat, Onur Karaman, Determination of COVID-19 pneumonia based on generalized convolutional neural network model from chest X-ray images, 2021
[8] Xing Han Lu, Aihua Liu, Recurrent disease progression networks for modelling risk trajectory of heart failure, 2021

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210583

Data-Driven Modeling of Complex Business Process in Heterogeneous Environment of Healthcare Organization with Health Information Systems

Alexander KSHENIN[a 1] and Sergey KOVALCHUK[a,b]
[a]*ITMO University, St. Petersburg, Russia*
[b]*Almazov National Medical Research Centre, St. Petersburg, Russia*

Abstract. Business process modeling aims to construct digital representations of processes being executed in the company. However, models derived from the event logs of their execution tend to overcomplicate the desired representation, making them difficult to apply. The most accurate recovery of the business process model requires a comprehensive study of the various artifacts stored in the company's information system. This paper, however, aims to explore the possibility to automatically obtain the most accurate model of business process, using mutual optimization of models recovered from a set of event logs. Further, the obtained models are executed in multi-agent simulation model of company, and the resulting event logs are examined to determine patterns that are specific to distinct employees and those that generally characterize business process.

Keywords. Business process modelling, model recovery, agent-based simulation modeling, pattern identification, health information system

1. Introduction

Business Process (BP) modeling aims to construct digital representations of processes being executed in the company, that are required for the implementation of Business Process Management (BPM) in the organization and its sustainable development in constantly changing demands of the competitive environment (1).

An extensive research and optimization of processes in a medical organization make it possible to efficiently allocate and use available resources and provide treatment with no delays and at the proper volume (2). Moreover, the construction, research, and simulation of the BPs is performed when implementing the digital twin of the organization used in decision support systems (DSS) for scenario analysis (3). The health information system (HIS) of a medical organization stores large volumes of heterogeneous information about its functioning, which can be used to build and explore the BP of the healthcare system (4).

In this paper, authors consider the possibility of obtaining the most accurate model of BP in application to healthcare, using mutual optimization of models recovered from

[1] Corresponding Author, ITMO University, 197101, St. Petersburg, Russia; E-mail: Ksh.Al.Dm@gmail.com.

a set of event logs, as well as information about complex BP by examining structural and temporal patterns of BP simulation logs. The complexity of BP in this research means its diversity of execution, observed in uneven limited event logs.

Heterogeneous information about the BPs of a large medical center is used to show the proposed method. Obtainment and investigation of a set of BP implementations are performed using multi-agent simulation model developed by the authors.

2. Related Works

When the BP model is defined by an expert, the resulting representation may not reflect the actual execution of BP by employees. Models derived from the execution event logs may show excessive detail of actions, as well as they may vary from one employee executing a BP to another (5).

Different Process Mining algorithms aim to solve this problem for different types of event logs, assumptions about the investigated BPs (6). For example authors of the article (7) implement filtering of logs from rare actions that noise up the event log using their relative frequencies. Furthermore, heuristics are used to identify noise in logs. In (8) the researchers propose an approach that develops a systematic methodology for clearing event logs. Based on several datasets, 11 log imperfections patterns were identified and appropriate ways to detect and eliminate them were suggested. However, the goal of current research lies in automatic filtering of BP logs, with main assumption contained in splitting event logs on specific feature, and construction of an accurate BP model.

Recurring patterns can be found in the BP model, obtained from good quality event log without noise, they make the understanding of the processes easier or can be used to build new processes. In paper (9) the authors propose their approach for automatic pattern discovery using specified ontology from the event log generated during the execution of BPs. The sequence of events and semantics are used to determine if a pattern is present in a process. Authors of the article (10) describe an approach to identify process execution priority patterns, using the event log, to discover the impact on the overall process. Also patterns can be used to compare the performance of BP as shown in (11), where authors compare the elements of the graph built on event logs to evaluate the divergence of two logs.

3. Data

Data used in this research was provided by an Almazov National Medical Research Centre (Almazov Centre). It includes information about the movement of medical staff, derived from the Access Control System (ACS), and various actions with Electronic Health Records (EHR) from the HIS. ACS and HIS data were combined and used to extract the BP event logs for a set of employees. This study uses event logs of 24-hour nurses' shifts (log cases), generally described in table 1.

Table 1. Description of nurses' event logs

Employee	Shifts	Activities	Unique activities
№0	137	6118	239
№1	187	13644	286
№2	127	3595	239
№3	113	6427	246

Event logs belong to four nurses working in one department, with similar or close responsibilities, implementing a similar set of actions, allowing to identify and explore the model of BP. Thus, the event log of shifts contains nurse's activities – actions with HIS and movement through a set of locations in the center, with timestamps.

4. Methods

The ProFIT library (https://github.com/Siella/ProFIT), already applied to mining BPs in healthcare (4), was used in this research to obtain a set of BP models from the logs. Each BP model is obtained in form of a weighted graph, described by a weighted transition matrix. Graph nodes represent activities from the log, edges – sequential transitions between them. The weight of an edge is the frequency of transition occurrence in logs.

Authors propose the identification of the most accurate model of BP which includes:

- The construction of executable BP models $\{M_i\}_{i=1}^{n}$ for a set of n event logs.
- Mutual optimization of models – obtainment of a set of filtering parameters $\{f_i\}_{i=1}^{n}$ for activities and transitions (with preassigned bounds). Let $M_i(f_i)$ be a BP model, obtained from event logs i with activities' filtering f_i. $N_{i,f_i} = N\left(M_i(f_i)\right)$ – is the corresponding set of activities, then

$$\{f_i\}_{i=1}^{n} : \frac{1}{n}\sum_i \frac{N_{i,f_i} \setminus \cup_{j \neq i} N_{j,f_j}}{N_{i,f_i}} \to \min \quad (1)$$

The set of event logs can be derived by splitting general event log on specific feature, for example, employee – as presented in this research. Thus, proposed frequency filtering automatically remove nodes and edges, which occur very rarely in event logs and turn the result into "spaghetti-like" models, and consider information about other models in every BP model, making them more accurate.

To construct an executable BP model, graph is supplemented with transition probabilities and distributions of execution times, and the process of BP occurrence. Models are executed in multi-agent BP simulation model, developed by the authors to produce simulation logs used in pattern identification. Each pattern is a sequence of activities, that appear in several event logs with specified significance threshold.

5. Results

A set of four executable BP models of nurses' shifts were selected, obtained, optimized. For each nurse, BP simulation modeling was performed, resulting in a set of event logs of BP model implementation. Further, patterns of length 2 – 5 were identified for each simulation log. The analysis of patterns showed that nurses №0 and №3 have the closest BP models, followed by nurse №1 and then №2. The filtering parameters obtained with the optimization of BP model on the subset of original event logs of two employees (№0 and №3) are shown in table 2.

Table 2. Filtering parameters for actions and transitions of two BP models optimization

Employee	f_i of activities	f_i of transitions
№0	82	5
№3	77	

The visualization of optimized BP model, that corresponds to employee №0 with filtering parameters from table 2, is shown on Figure 1.

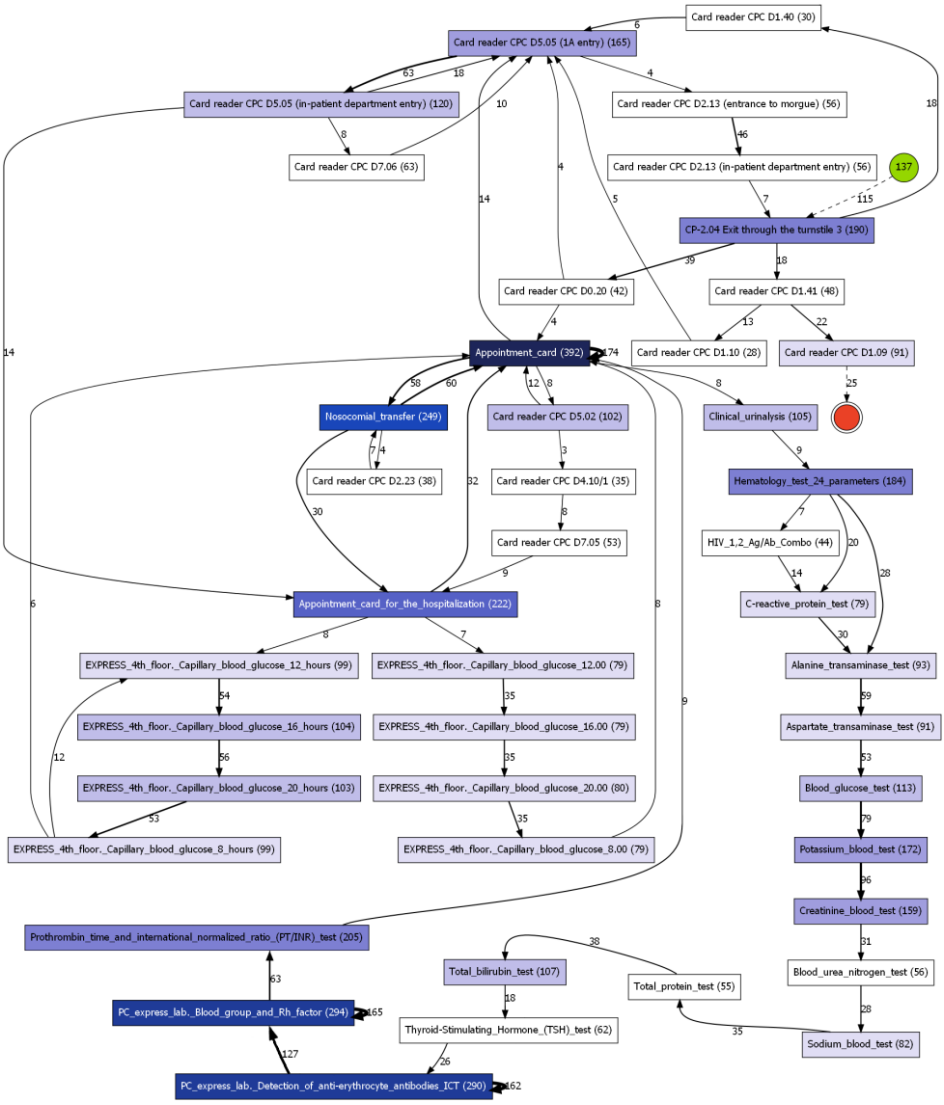


Figure 1. Visualization of optimized graph BP model corresponding to nurse №0.

Figure 2 provides the box plots of four movement patterns, that are mutual for all BP models, and consequently, employees. These patterns are situated at the very beginning of BP and may show small time loss in employee's №1 execution (left), and higher dispersion of longer patterns' execution times (right) that is due to untracked sub-pathways and different arriving time. The example of movement patterns, №7, is the sequence: 'CP-2.04 Exit through the turnstile 3', 'Card reader CPC D0.20', 'Card reader CPC D5.05 (1A entry)', 'Card reader CPC D5.05 (in-patient department entry)'.

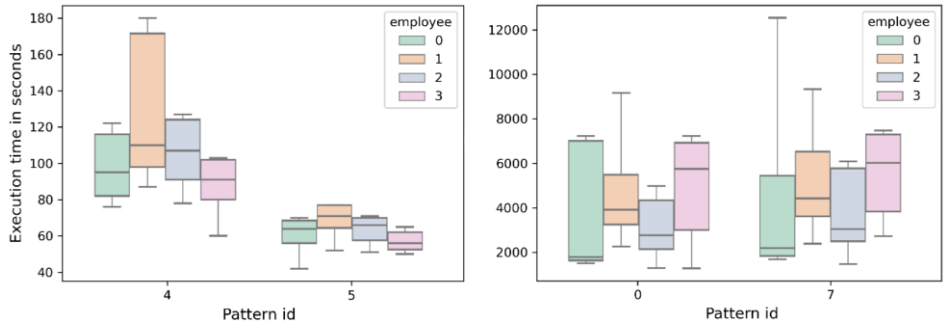


Figure 2. (left) length of two movement patterns; (right) length of four movement patterns.

Figure 3 provides the box plots of patterns related to various blood analysis tests. Though all three employees possess such competencies, employee №1 performs them in a different scenario not during BP closest to other employees' models. Employees №0 and №3 show similar results in executing these patterns, as they are subsequences of clinical blood analysis that is highly automated. The example of tests patterns, №13, is the sequence: 'Thyroid-Stimulating Hormone (TSH) test', 'PC express lab. Detection of anti-erythrocyte antibodies ICT', 'PC express lab. Blood group and Rh factor'.

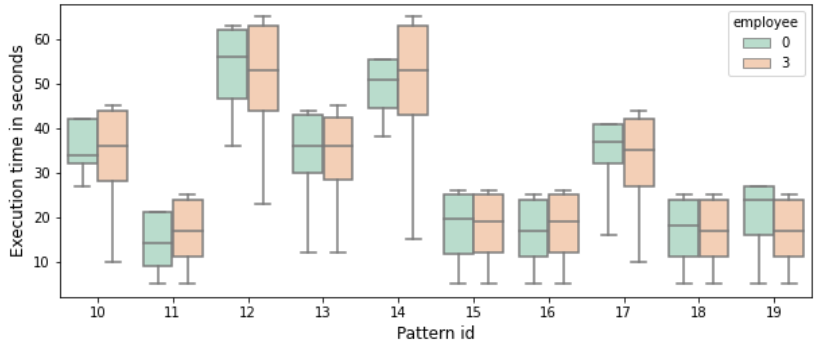


Figure 3. Blood analysis patterns.

6. Discussion

BP, explored in this research does not possess parallel execution of several activities, and moreover, is entirely executed by a single employee. The opposite case of the second feature, nonetheless, can be considered within the proposed method. In this case, the resulting activity patterns will correspond to all possible subsets of employees, that are involved in executing them. One can analyze the time and structure diversity in such patterns and identify activities and/or employees where the loss of efficiency occurs.

Execution patterns in simulation logs are rather simple, however, they reflect the process of execution – as a sequence of activities, so they are easy to interpret and analyze. Box plots comparison can show both the personalized loss of execution time and the efficiency of automation. When changing subsets of models one can identify structural differences between BP executed by different employees.

The results can be improved with a more representative variable for graph model optimization. The procedure, that can handle possible changes in models' subset, will allow to dynamically explore more diverse models and save time, not starting from the very beginning. The most resource-demanding part of the research method is simulation modelling; however, all BP models can be executed in parallel, vastly reducing time.

7. Conclusion

In this research, authors propose method to obtain the most accurate model of BP, using mutual optimization of models, derived from a set of event logs. The identification of BP model is followed by the examination of structural and temporal patterns in BP simulation logs. The study includes the identification and research of nurses' shifts BP, using the data of Almazov Centre, that show an interpretable result.

Acknowledgments

This research is financially supported by The Russian Scientific Foundation, Agreement #17-15-01177.

References

[1] Shaw DR, Holland CP, Kawalek P, Snowdon B, Warboys B. Elements of a business process management system: Theory and practice. Bus Process Manag J. 2007;13(1).

[2] Rojas E, Munoz-Gama J, Sepúlveda M, Capurro D. Process mining in healthcare: A literature review. J Biomed Inform [Internet]. 2016 Jun;61:224–36. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1532046416300296

[3] Sulis E, Terna P, Di Leva A, Boella G, Boccuzzi A. Agent-oriented Decision Support System for Business Processes Management with Genetic Algorithm Optimization: an Application in Healthcare. J Med Syst. 2020 Aug 2;44(9):157. doi: 10.1007/s10916-020-01608-4.

[4] Elkhovskaya L, Kabyshev M, Funkner A, Balakhontceva M, Fonin V, Kovalchuk S. Personalized assistance for patients with chronic diseases through multi-level distributed healthcare process assessment. Stud Health Technol Inform. 2019;261:309-312.

[5] van der Aalst W. Process mining: discovering and improving Spaghetti and Lasagna processes. 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM) 2012. DOI: 10.1109/CIDM.2011.6129461

[6] Van Dongen BF, Alves De Medeiros AK, Wen L. Process mining: Overview and outlook of Petri net discovery algorithms. Lecture Notes in Computer Science 2009; 5460: 225-242.

[7] Conforti R, La Rosa M, Ter Hofstede AHM. Filtering Out Infrequent Behavior from Business Process Event Logs. IEEE Trans Knowl Data Eng. 2017;29(2).

[8] Suriadi S, Andrews R, ter Hofstede AHM, Wynn MT. Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. Inf Syst. 2017;64:132-150.

[9] Ferreira DR, Thom LH. A semantic approach to the discovery of workflow activity patterns in event logs. In: International Journal of Business Process Integration and Management. 2012; 6(1). https://doi.org/10.1504/IJBPIM.2012.047909

[10] Suriadi S, Wynn MT, Xu J, van der Aalst WMP, ter Hofstede AHM. Discovering work prioritisation patterns from event logs. Decis Support Syst. 2017;100.

[11] van Beest NRTP, Dumas M, García-Bañuelos L, La Rosa M. Log delta analysis: Interpretable differencing of business process event logs. Lecture Notes in Computer Science 2015;9253:386-405. DOI: 10.1007/978-3-319-23063-4_26

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210584

Reinforcing Health Data Sharing Through Data Democratization

Yuhang WANG[1] and Bian YANG
Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Gjøvik, Norway

Abstract. In this paper, we propose a health data sharing infrastructure which aims to empower a democratic health data sharing ecosystem. Our project, named Health Democratization (HD), aims to enable seamless data mobility of health data across trust boundaries, through addressing structural and functional challenges of its underlying infrastructure with a throughout core concept of data democratization. A programmatic design of HD platform was elaborated, followed by an introduction about one of our exploratory designs —an "reverse onus" mechanism that aims to incentivize creditable data accessing behaviors. This scheme shows a promising prospect of enabling a democratic health data sharing platform.

Keywords. eHealth, data democratization, health data infrastructure, privacy enhancing

1. Introduction

Sharing health data creates value for clinical care, trials, and case studies, as well as improved knowledge base[1][2] for healthcare researchers and healthcare organizations. Health data has also immense commercial value [3] for other parties such as pharmaceutical industry, data analytics providers, insurers, data markets, business intelligence.

The huge value associated with health data can lead to data misuse, for example, targeted use of ransomware, participation in black market[4], and other cybercrimes. The conventional health data infrastructure was not designed for anticipating value-driven data mobility and the associated cyber threats. There is a structural deficiency in the conventional infrastructure on which patch-like remedies only add to the complexity of the challenge.

Related works such like the national eHealth infrastructure (e.g., Norsk Helsenett) [5] in Norway has been built since middle 1990s which emphasized localized data retention and confidentiality. The "one citizen – one journal" plan was proposed in 2012 with the laws regarding medical records and health registers updated in 2015 in order to facilitate data mobility. The national pilots Helseplattformen and Helseanalyseplattformen [6] were launched in recent years to technically implement the connectivity and coordination in data sharing. On the EU level, the effort has so far

[1] Corresponding Author: Yuhang Wang, Norwegian University of Science and Technology, Gjøvik, Norway; E-mail: yuhang.wang@ntnu.no

mainly focused on the technical (e.g., the epSOS project) and legal [7] interoperability towards the EU eHealth strategy 2020 [8].

Rather than the above-mentioned works which reinforcing the infrastructure from a traditional view of vulnerability identification, protection, detection, and response, our work aims to define, architect, implement, and evaluate a democratic health data infrastructure which is expected to incentivize all parties, including individuals, to prove, negotiate, and configure their rights associated with health data. The conflicts of interest among different parties can be reconciled through a set of automated mechanisms so that data can be mobilized across trust boundaries.

We dedicated to architecting and constructing a data transaction model by strikingly practice the concept of **Data Democratization** (or say, democratic data sharing). Formally, this indicates two kernel idea, which will be followed throughout the design of our HD platform: 1) All stakeholders are treated identically without discrimination, and 2) when facing the inequivalence reality among each party, to promote fairness as a complementary.

State-of-the-art research and ethical & legal efforts have pay extensive attention on the first, however, we argue that the fairness promotion is also critical with regard to the data democratization, due to the extremely unequal actuality exists between the individuals and the colossus entity. We gazed deep into the platform in a hierarchical perspective and proposed our **Conceptual Layered Architecture** that is promised to achieve our goal of data democratization.

The following context are organized as listed: We present the terms of the stakeholders defined in this paper and the formal conceptual architectures in Section 2. Section 3 will illustrate how the concept of reverse onus could be applied into the democratic-promoting designs. We then conclude our work in Section 4.

2. Conceptual Layered Architecture of HD Platform

2.1. Stakeholders Description

The prior task for our work is to distinct discrepant stakeholders with significant behavior characteristics and interest relationship. We first classify our HD platform-relevant stakeholders into 7 types, then we present a sample of matching between these types of roles and the roles defined in GDPR. HD platform will “circulate” among diverse stakeholders, e.g., some roles are tent to get the health data for their point of interest, while some others have the right of disposal of the health data. Some stakeholders may also tend to provider the data processing/storage/analyzing fundamentality. We classify these stakeholders into 7 different types as shown in Table 1.

2.2. HD Architecture

To fulfill the principle of data democratization and the promised capabilities, we gazed deep into the platform in a hierarchical perspective. The HD platform is responsible for developing and managing the democratic negotiation procedures during the healthcare data business, for use in and exchange of clinical and individual healthcare information between the potential DS/DM and the potential data consumer.

Table 1. Stakeholders in HD platform

Stakeholder	Description
computing resource manager (CRM)	Supporter participant which assists each player in managing computing, storage, and communication resources in facilitating data sharing with other players.
data consumer (DC)	Player participant which can access data directly, query a database, or receive data from DS, DG, or DSP in order to exploit the value of the shared data.
data generator (DG)	Player participant which directly generates data from a data subject or converts sensed signal to formatted data.
data manager (DM)	Supporter participant which assists each player in processing, managing, and exchanging the data with other players.
dataset provider (DSP)	Player participant which creates and maintains, under the consent given by DS and possibly the agreement with DG, one or several structured datasets sourced from DS or / and DG.
data rights manager (DRM)	Supporter participant which assists each player in managing their rights with other players, i.e., proving, negotiating, and recording the terms and conditions describing the rights and obligations about the data for sharing.
data analysis service provider (DASP)	Participant which provides data analysis as a service to DS, DG, DC, or DSP.

For each principle and the potential promised scenario, the executive process could be considered as a correlation between the data sharing participants and an affair-relevant data sharing function in different executive level.

Guiding by the eHealth standardization in the Nordics countries[9] with respect to the interoperability [10][11], the data sharing function ranges from the incipient data provenance to the rights and obligation tracking after the agreement. We stratify our platform into four conceptual layers, named “Computing Infrastructure Layer”, “Data Sharing Operation Layer”, “Data Sharing Logic” and “Healthcare Business Layer”. Our Architecture also obtains references from the peer work on a diverse eHealth networking and healthcare data sharing solutions [12][13]. Figure 1 expounds our conceptual layered Architecture in detail, it describes the layers of the data sharing hierarchical structure, the data sharing participants, and the data sharing function.

The main systematic-level functions required in our platform are listed in Table 2.

Table 2. Main systematic-level functions

Function	Description
Data provenance	To provide a backward traceability of medical device, personal device in homecare environment, etc., and the health data sourced from these devices need to be audited in a trusted way of their rights and operation status.
Risk Assessment	Enabling each data subject has different risk acceptance tolerance and incentive degrees when they are entitled to rights and benefits from data.
Computational negotiation	Negotiating agents can operate and negotiate decisions. The requirements will be developed in compliance with the GDPR, healthcare regulations, and other relevant regulations. When processing and exchanging personal data between the agents, the design of the infrastructure will address such key requirements of the GDPR as data protection by design and by default, accountability, pseudonymization, right of access and right to erasure.
Multi-lateral security policing	Enabling individuals be able to share and control access to health data without having to place extensive trust in entities, and institutions must also be able to share data responsibly for research, innovation, and quality across institutional boundaries.

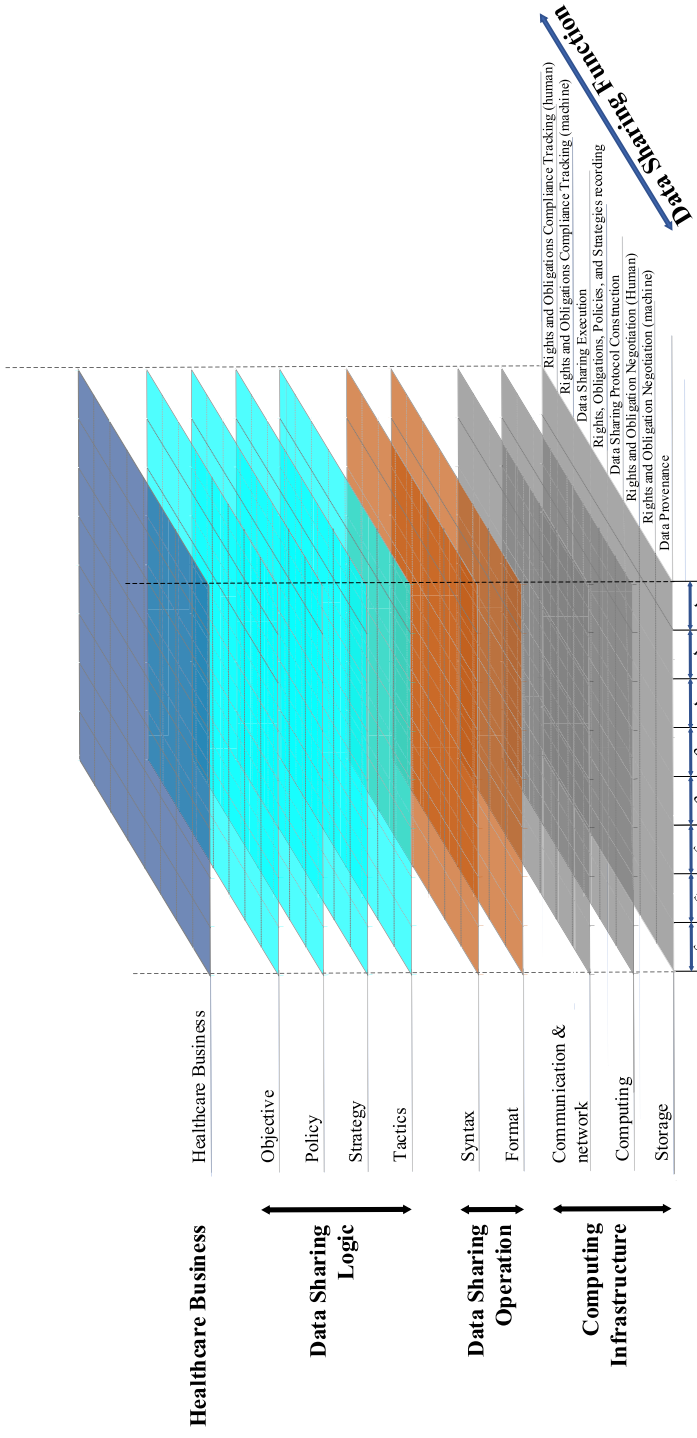


Figure 1. Conceptual Layered Architecture of HD platform.

3. "Reverse Onus" in Health data Negotiation

In most of the current ecosystem of digital market, an incommensurable inequality exits between the individuals and the so called "digital oligarchy" [14]. The fast growing of these consolidation power tries to gain monopoly in various of aspects, including the power of interpretation of the privacy data usage, and the health data sharing cannot be righteous alone.

In our HD platform, the DC could be played by such kind of roles. For example, DC is an influential giant company who wishes to constitute its global health big data warehouse, while the DM is just a small agent of DSs. In this design, we assume an inequitable situation between such two kinds of stakeholders and seek to resolve a potential unfair issue of knowledge asymmetry.

During the negotiation process in our platform, dominant DC may have much more right to argue 1) the (social or monetary)value of the health data, and 2) the scale and granularity of health data are demanded to perform a certain healthcare service. On the contrary, the DM may lack of knowledge to assess the opponent's proposal.

Our HD platform utilizes the concept which similar to "Reverse Onus" to mitigate this problem. Whenever a health data relevant negotiation happened between to stakeholders, say, a dominant DC and a regular DM, with great disparity, the platform shifts the burden of proof onto the DC specified to prove the necessity of the health data claim. When DM raises a discontent against the proposal with regard to data minimization (an essential privacy enhancing principle defined in GDPR), data value, ethical issue, etc., DC is in the position to provide convincible specification on his proposal.

Data Usage Approval:

The negotiation procedure is protected by requiring the DC to submit an application form (*appFm*) on the usage of the health data. Including: 1) usage purpose. 2) data precision upper limit in percentage. 3) data requesting schedule instant/time period/data manager triggered/etc. 4) requiring pattern in frequency distribution. 5) if necessary, reasonableness report.

The *appFm* will be assessed by the platform, based on the history usage log, will consider: 1) purpose to precision. 2) purpose to schedule. 3) purpose to pattern. 4) history comparison across entities. In this paper, we only assess the privacy-leakage risk and register the *appFm* in the following credit system proposed in the next subsection.

Credit Mechanism for Promoting Reverse Onus:

After reaching the mutual-agreement and the contract was built, the **credit mechanism** inherited from our previous work [15] will monitor the execution of the protocol to stimulate the DC to follow the terms. We applied a credit score mechanism upon the DC to encourage conformity and generate the virtual credit of DC based on his record, this credit will be further used to consult the future negotiations.

We set a credit score for each DC, denoted as α ($\alpha \in [0, 1]$) and with 1 means DC is with the highest credit score. In view of the HD platform, one observation of DC could lead to a downgrade of its credit score, which is an excessive access to the DM's health data, this is possible happening when DC misuses his interpretation clauses and collects health data exceeds the defined amount, granularity, etc. The DC with lower credit score will face a more arduous negotiating process than usual, and hence loss the potential health data application value. Guaranteed by this credit mechanism, a rational would DC tends to behave responsible and honest to the reverse onus scheme, and therefore the fairness of HD platform will be strengthened.

4. Conclusion

In this paper, we raised the concept of data democratization which will reinforce the health data sharing with respect to privacy enhancement and benefit insurance. An overall conceptual layered architecture was proposed which aims to enable such vision.

We further introduced an advanced concept of data democratization, which emphasized the fairness promotion in HD platform. A credit-mechanism-powered incentive scheme for promoting "reverse onus" on data usage was proposed. This mechanism rebalances the inequitable situation among all the stakeholders.

The future work will keep on implementing and integrating the proposed conceptual designs, several landing cases studies will be put into effort to improve the practicability of our work. Specially, the concept of reverse onus and the corresponding credit mechanism should be verified prudently by applying it onto the health data ecosystem.

References

[1] Olson S, Downey AS. Sharing clinical research data: Workshop summary. Washington, District of Columbia: The National Academies Press; 2013.
[2] Walport M, Brest P. Sharing research data to improve public health. The Lancet. 2011;377(9765):537-539.
[3] Czeschik C. Black Market Value of Patient Data. Digital Marketplaces Unleashed. 2017;:883-893.
[4] G. Hartvigsen. Lessons learned from 25 years with telemedicine in Northern Norway. Tromsø: NST-rapportserie; 2015.
[5] Helseanalyseplattformen Project: https://ehelse.no/helsedataprogrammet/helseanalyseplattformen
[6] Overview - Public Health - European Commission [Internet]. Public Health - European Commission. 2021 [cited 6 September 2021]. Available from: https://ec.europa.eu/health/cross_border_care/policy_en
[7] Topic 1: EU eHealth strategy towards 2020 – Public Health – European Commission [Internet]. Public Health - European Commission. 2016 [cited 4 November 2016]. Available from: https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20161121_co01_en.pdf
[8] Herlof N, Christine B. Nasjonal e-helsestrategi og mål 2017 – 2022 [Internet], Direktoratet for E-helse, 2017. Available from: https://ehelse.no/publikasjoner/nasjonal-e-helsestrategi-og-mal-2017-2022-oppdatert-2019
[9] Vincent van Pelt, Michiel Sprenger. Adoption and take up of standards and profiles for eHealth Interoperability [Internet]. Antilope. 2015 [cited 2018 Nov 19]. Available from: https://www.antilope-project.eu/wp-content/uploads/2013/05/D1.1-Refinement_of_Antilope_Use_Cases_v1.2.pdf
[10] Kouroubali A, Katehakis DG. The new European interoperability framework as a facilitator of digital transformation for citizen empowerment. Journal of biomedical informatics. 2019 Jun 1;94:103166.
[11] Cole AM, Stephens KA, Keppel GA, Lin CP, Baldwin LM. Implementation of a health data-sharing infrastructure across diverse primary care organizations. The Journal of ambulatory care management. 2014 Apr;37(2):164.
[12] Mandl KD, Simons WW, Crawford WC, Abbett JM. Indivo: a personally controlled health record for health information exchange and communication. BMC medical informatics and decision making. 2007 Dec;7(1):1-0.
[13] Shen M, Zhu L, Xu K. Layered Data Sharing Architecture with Blockchain. InBlockchain: Empowering Secure Data Sharing 2020 (pp. 29-37). Springer, Singapore.
[14] Digital oligarchy. Knowledge for policy [Internet]. Knowledge4policy.ec.europa.eu. 2021 [cited 6 September 2021]. Available from: https://knowledge4policy.ec.europa.eu/foresight/topic/diversifying-inequalities/new-digital-oligarchy_en
[15] Wang Y, Tian Z, Sun Y, Du X, Guizani N. LocJury: an IBN-based location privacy preserving scheme for IoCV. IEEE Transactions on Intelligent Transportation Systems. 2020 Feb 10.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210585

Dynamic Aortic Aneurism Risk Factors

Oleg METSKER[a], Georgy KOPANITSA[b,1], Olga IRTYUGA[a], Vladimir USPENSKIY[a]
[a] *Almazov National Medical Research Centre, Saint-Petersburg, Russia*
[b] *ITMO University, Saint-Petersburg, Russia*

Abstract. According to different systematic reviews incidence of thoracic aortic aneurysms (TAA) in the general population is increasing in frequency ranging from 5 to 10.4 per 100000 patients. However, only few studies have illustrated the role of different risk factors in the onset and progression of ascending aortic dilatation. Currently, noninvasive imaging techniques are used to assess the progression rate of aortic and aortic valve disease. Transthoracic (TT) Echocardiographic examination routinely includes evaluation of the aorta It is the most available screening method for diagnosis of proximal aortic dilatation. Since the predominant area of dilation is the proximal aorta, TT-echo is often sufficient for screening. We retrospectively analyzed the ECHO database with 78499 echocardiographic records in the Almazov National Medical Research Centre to identify patients with aneurysm. Detailed information including demographic characteristics, ECHO results and comorbidities were extracted from outpatient clinic and from hospital charts related to hospitalizations occurring within a year before index echocardiography was performed. Comorbid diseases were similarly extracted from outpatient clinic and/or hospital admissions. The classifier showed an AUC-ROC for predicting of aneurism detection after a repeated ECHO at 82%.

Keywords. Aneurism, Machine learning, risk factor, prognosis

1. Introduction

Non-coronary heart diseases are a large heterogeneous group of diseases [1]. Given their prevalence and mortality in the general population, they have an important medical and social significance. We have chosen aortic valve and ascending aorta diseases as a model of non-coronary heart disease. These diseases are diagnosed most late due to the absence of clinical symptoms before the onset of complications up to lethal outcomes.

According to different systematic reviews incidence of thoracic aortic aneurysms (TAA) in the general population is increasing in frequency ranging from 5 to 10.4 per 100000 patients [2,3]. However, only few studies have illustrated the role of different risk factors in the onset and progression of ascending aortic dilatation [4,5]

Currently, noninvasive imaging techniques are used to assess the progression rate of aortic and aortic valve disease [6].

Transthoracic (TT) Echocardiographic examination routinely includes evaluation of the aorta. It is the most available screening method for diagnosis of proximal aortic dilatation. Since the predominant area of dilation is the proximal aorta, TT-echo is often sufficient for screening [7,8].

[1] Corresponding Author, Georgy Kopanitsa, ITMO University, Saint-Petersburg, Russia, E-mail: georgy.kopanitsa@gmail.com

Criteria for the progression of these diseases have not yet been developed. Both conservative and surgical therapies are not personalized. At the same time, the quality of life of patients can change rapidly depending on diagnosis, treatment and lifestyle.

The aim of this work is to identify anamnestic predictors of aortic aneurysm and to develop a model for identifying the need for re-diagnosis with ECHO.

2. Methods

We retrospectively analyzed the ECHO database with 78499 echocardiographic records in the Almazov National Medical Research Centre to identify patients with aneurysm. Detailed information including demographic characteristics, ECHO results and comorbidities were extracted from outpatient clinic and from hospital charts related to hospitalizations occurring within a year before index echocardiography was performed. Comorbid diseases were similarly extracted from outpatient clinic and/or hospital admissions.

2.1. Inclusion criteria

Age 18-75 years

Widening of the thoracic aorta >40 mm and/ or hemodynamically significant impairment of aortic valve function (aortic valve velocity >2.0 and/ or degree 2 or greater aortic insufficiency)

Congenital aortic valve abnormality (bicuspid, monocuspid, quadricuspid aortic valve)

All patients underwent at least 2 comprehensive 2-dimensional and Doppler transthoracic echocardiography

2.2. Data preprocessing

We removed 1% of values having the highest z-score to filter out some obvious outliers. Furthermore, we applied the min and max normalization to the remaining values.

2.3. Classification model and feature importance

To train a model for a prediction of the need of a second ECHO the experiment ran in the setting of stratified 5-fold cross-validation (i.e. random 80% of patients were used for training and 20% for testing target class ratios in the folds were preserved). A random forest algorithm was applied to calculate the feature importance. A correlation heatmap was used to analyze correlations between top predictors. The algorithm was implemented using Python 3.6.3 and the scikit-learn 0.19.12 library.

As an additional performance assessment score, we used AUC of ROC, which represents the trade-off between sensitivity and specificity of the model. The AUC was calculated based on an average of 5 curves (one curve per fold in the setting of 5-fold cross-validation).

[2] https://scikit-learn.org/stable/

According to the results of the ECHO study, all patients can be divided into three groups:

Class 0. patients with normal aortic dimensions and low probability of aneurysm development, not requiring any intervention.

Class 1. patients with an existing or developing aortic aneurysm requiring treatment and dynamic monitoring according to current clinical guidelines.

Class 2. patients with normal-sized aorta, but at extremely high risk of aneurysm development in the coming years. In this group, strict control of cardiovascular risk factors and, possibly, certain drug interventions to prevent the disease seem to be rational;

3. Results

The classifier showed an AUC-ROC for predicting of aneurism detection after a repeated ECHO at 82%. Features importance analysis is presented in the Figure 1.

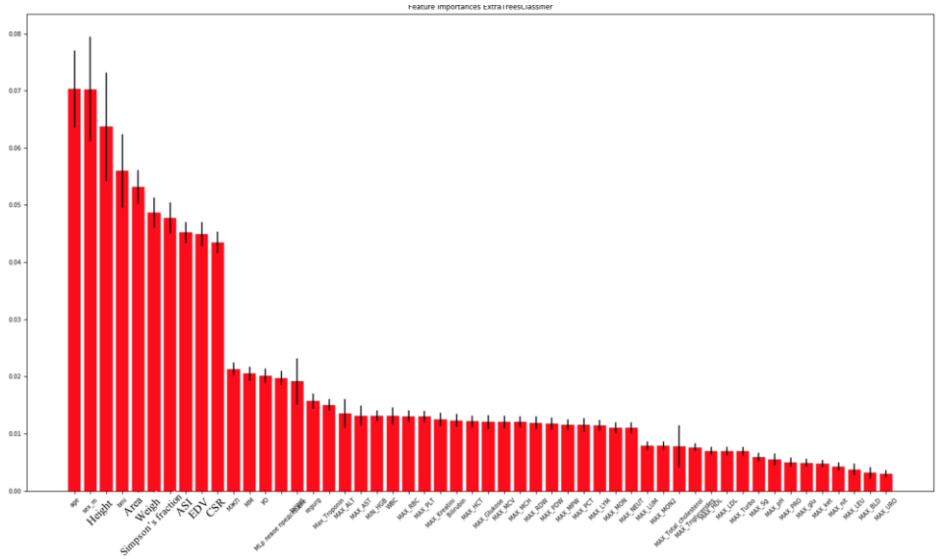


Figure 1. Feature's importance

The correlation matrix (Figure 2) demonstrates the fact that in the absence of structural changes in the ascending aorta, they are unlikely to develop in subsequent years in women, young adults, and low body weight. Interestingly, the likelihood of developing an aortic aneurysm is lower in patients with an existing aortic valve defect (stenosis or insufficiency). A Correlation matrix is presented in the Figure 2.

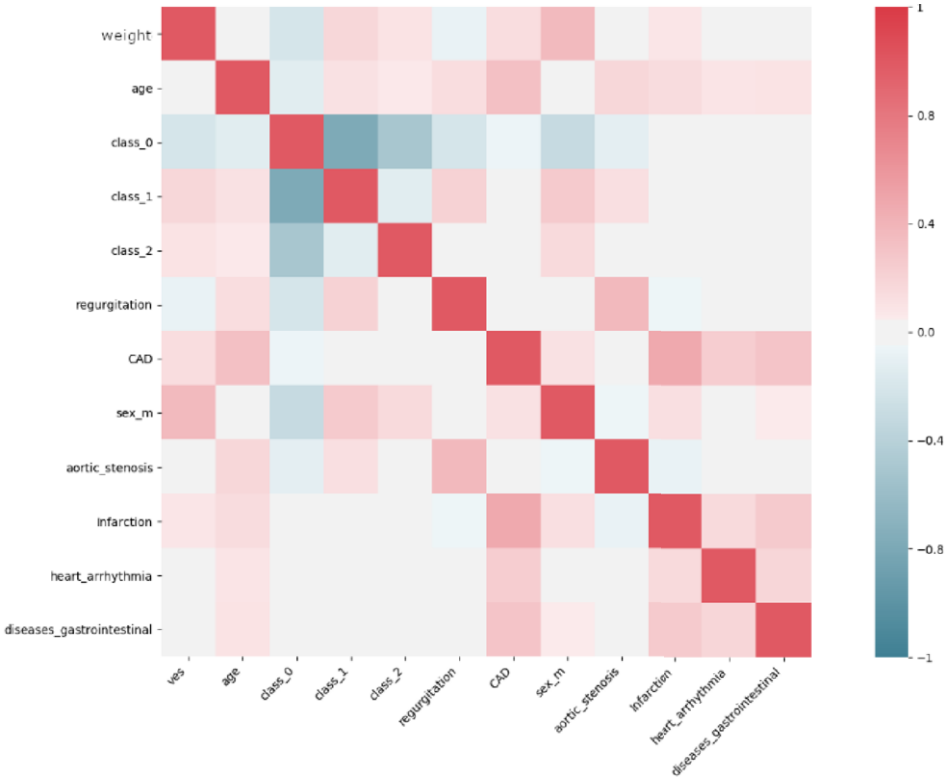


Figure 2. Heatmap

4. Discussion

Analyses of the findings confirm the significance of the known risk factors for aortic aneurysm development - age and male gender [3,5]. It should be noted that these risk factors are nonspecific, associated with the risk of a number of cardiovascular diseases, and, above all, atherosclerosis with lesions of various vascular basins and its acute complications, myocardial infarction and stroke [9]. Height is also a known risk factor for aortic aneurysm. In particular, it has been shown that height, rather than weight or body surface area, is most associated with the risk of aneurysm development [10]. It is possible that high height is more associated with connective tissue abnormalities associated with a high risk of aortic aneurysm formation, most pronounced in Marfan syndrome. The most significant appears to be the contribution of global left ventricular contractile function (Simpson ejection fraction). It has the leading position among all factors analyzed except for gender, age, height and body weight. Meanwhile, the risk of aneurysm development is associated with high contractile function of the left ventricle. It can be assumed that mechanical impact of blood flow on the aortic wall makes a significant contribution to the development of aortic aneurysm in predisposed patients with high cardiac contractile function. This is fundamentally different from the vast majority of patients with cardiovascular disease. For those an adverse prognosis is associated with decreased left ventricular contractile function, and low values of Simpson ejection fraction are considered as a universal marker of adverse prognosis [11].

Although in our study Simpson ejection fraction correlated with height, the independent association of preserved cardiac pump function with the risk of aortic aneurysm is supported by the association with other echocardiographic indices, including those reflecting valve function. We should also note the association of aortic aneurysm risk with several laboratory parameters, among which troponin level is the leading one. This index is a known predictor of adverse cardiovascular outcomes [12]. Troponin levels are also associated with abdominal aortic risk [13]. Since troponin, along with D-dimer, is also a marker of acute aortic rupture [14], it can be assumed that in patients with aortic aneurysm the increased troponin level is associated with subclinical damage of the aortic wall, but not of the myocardium.

5. Conclusion

We have identified the risk factors for the dynamic development of aortic aneurism. These factors can be obtained from the screening methods that allow early diagnostics of the aneurism development.

Acknowledgements

This work was financially supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement No.075-15-2020-901).

References

[1] Neil HAW, Hawkins MM, Durrington PN, Betteridge DJ, Capps NE, Humphries SE. Non-coronary heart disease mortality and risk of fatal cancer in patients with treated heterozygous familial hypercholesterolaemia: a prospective registry study. Atherosclerosis. 2005 Apr;179(2):293-7. doi: 10.1016/j.atherosclerosis.2004.10.011.

[2] Sampson UKA, Norman PE, Fowkes FGR, Aboyans V, Song Y, Harrell FE, et al. Global and regional burden of aortic dissection and aneurysms: mortality trends in 21 world regions, 1990 to 2010. Glob Heart. 2014 Mar;9(1):171-180.e10. doi: 10.1016/j.gheart.2013.12.010.

[3] Devereux RB, De Simone G, Arnett DK, Best LG, Boerwinkle E, et al. Normal limits in relation to age, body size and gender of two-dimensional echocardiographic aortic root dimensions in persons >15 years of age. Am J Cardiol. 2012 Oct 15; 110(8): 1189–1194. doi: 10.1016/j.amjcard.2012.05.063

[4] Kälsch H, Lehmann N, Möhlenkamp S, Becker A, Moebus S, et al. Body-surface adjusted aortic reference diameters for improved identification of patients with thoracic aortic aneurysms: Results from the population-based Heinz Nixdorf Recall study. Int J Cardiol. 2013 Feb 10;163(1):72-8. doi: 10.1016/j.ijcard.2011.05.039.

[5] Rogers IS, Massaro JM, Truong QA, Mahabadi AA, Kriegel MF, et al. Distribution, determinants, and normal reference values of thoracic and abdominal aortic diameters by computed tomography (from the framingham heart study). Am J Cardiol. 2013 May 15;111(10):1510-6. doi: 10.1016/j.amjcard.2013.01.306.

[6] Doris MK, Everett RJ, Shun-Shin M, Clavel A-M. Dweck MR. The Role of Imaging in Measuring Disease Progression and Assessing Novel Therapies in Aortic Stenosis. JACC Cardiovasc Imaging. 2019 Jan;12(1):185-197. doi: 10.1016/j.jcmg.2018.10.023.

[7] Evangelista A, Flachskampf FA, Erbel R, Antonini-Canterin F, Vlachopoulos C, et al. Echocardiography in aortic diseases: EAE recommendations for clinical practice. Eur J Echocardiogr. 2010 Sep;11(8):645-58. doi: 10.1093/ejechocard/jeq056.

[8] Mądry W, Zacharska-Kokot E, Karolczak MA. Methodology of echocardiographic analysis of morphological variations of the aortic arch and its branches in children – own experience. J Ultrason. 2019;19(76):24-42. doi: 10.15557/JoU.2019.0004.

[9] Libby P, Buring JE, Badimon L, Hansson GK, Deanfield J, et al. Atherosclerosis, Nature Reviews Disease Primers 2019; 5(56): 1–18. doi:10.1038/s41572-019-0106-z.
[10] Zafar MA, Li Y, Rizzo JA, Charilaou P, Saeyeldin A, et al. Height alone, rather than body surface area, suffices for risk estimation in ascending aortic aneurysm. J Thorac Cardiovasc Surg. 2018 May;155(5):1938-1950. doi: 10.1016/j.jtcvs.2017.10.140.
[11] Kutyrev K, Yakovlev A, Metsker O. Mortality Prediction Based on Echocardiographic Data and Machine Learning: CHF, CHD, Aneurism, Procedia Computer Science. 2019; 156: 97–104. doi:10.1016/J.PROCS.2019.08.184.
[12] Willeit P, Welch P, Evans JDW, Tschiderer L, Boachie C, et al. High-Sensitivity Cardiac Troponin Concentration and Risk of First-Ever Cardiovascular Outcomes in 154,052 Participants. J Am Coll Cardiol. 2017 Aug 1;70(5):558-568. doi: 10.1016/j.jacc.2017.05.062.
[13] Molacek J, Treska V, Zeithami J, Hollan I, Topolcan O, et al. Blood biomarker panel recommended for personalized prediction, prognosis, and prevention of complications associated with abdominal aortic aneurysm. EPMA J. 2019 Jun 3;10(2):125-135. doi: 10.1007/s13167-019-00173-2.
[14] Li Z-D, Liu Y, Zhu J, Wang J, Lu F-L, Han L, Xu Z-Y. Risk factors of pre-operational aortic rupture in acute and subacute Stanford type A aortic dissection patients. J Thorac Dis. 2017 Dec;9(12):4979-4987. doi: 10.21037/jtd.2017.11.59.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210586

A New Paradigm for Ensuring Digital Drug Care Quality Monitoring Based on Ontology Tools

Tatiana LUGOVKINA[a,1], Sergey GORSHKOV[b], Evgeniy SVALOV [b]
[a] *Ural Research Institute of Phthisiopulmonology, Ekaterinburg, Russia*
[b] *TriniData LLC, Ekaterinburg, Russia*

Abstract. Systematizing and conceptualizing of the components of the patient's diagnosis at definite Clinical Situations are the important steps in constructing the medical electronic platforms to monitor the drug treatment quality and to realize the process of risk-management of drug care. The risk-management is a hard and expensive process for the non-profit hospitals. The Information technologies have a high potential for solving these problems. The conceptual schemes to construct the multimodal medical electronic platform were discussed. The Information Space of Clinical Practice, Information Field for formulating the detailed patient's diagnosis at the definite Clinical Situations, Information Environment for the components of the detailed patient's diagnosis were described.

Keywords. Conceptual schemes, drug quality care, medical electronic platform

1. Introduction

The basic principle of the drug treatment is to choose the right drug for a patient with a correct diagnosis, to administer the drug at the right time and in the right dose. The principles of "patient-centricity" are included into the law of the Russian Health Care system [1]. The process of safety drug treatment is one of the most important, complicated and expensive tasks for the non-profit hospitals. The Information technologies have a high potential for solving these problems. The methodological and theoretical approaches to improve the medication treatment process by using the tools of information technologies are very actual.

Any subject area of knowledge operates with specific terminology. The accuracy of the terms and concepts is especially important in medicine. Contemporary development of the knowledge about the accumulated clinical practice skills is in the process of transformation into a digital format [2]. *The* precise definitions of the terms and connections with their concepts are strongly needed for all professional areas in medicine. Conceptual apparatus is highly important and actual for constructing the medical electronic platforms.

1.1. Objective of the Work

The objective of the work is the deployment of the conceptual scheme of the Clinical Information Space based on the methodology of systematization of the Clinical

Situations to support clinicians at the moments of drug choice decisions, to predict the drug adverse events and to monitor the real Clinical Situations in the databases of medical electronic platform.

Object: conceptualization of the components of the *Information Field of the patient's diagnosis* for coding of the individual features (facets) of the patient's diagnosis at the definite Clinical Situations.

Subject: conceptualization of the Information Environment of the components in the information vectors of the individual patient's diagnosis at the definite Clinical Situations.

2. Material and Methods

The concepts of the Clinical Information Space were formulated on the base of the systemic principles: hierarchy, uniformity, sufficient diversity, symmetry of reflection, irreversibility of the time etc. [2,3,4]. The methodology of constructing schemes was based on the principles of the philosophical foundations of the science: logical, ethical and ontological [3,5,6,7,8]. The ontologies, the electronic intelligent system based on the knowledge (expert decision support system) and the system for digital coding of the individual patient's diagnosis at the definite Clinical Situations were simultaneously in demand.

The conceptual scheme of the multi-module medical electronic platform included 5 program modules: 1) data bases, 2) knowledge bases, 3) program for digital coding of the individual components of patient's diagnosis, 4) the module for identification: a) the drug prescriptions deviations from the recommended models, b) the triggers of the adverse events marked with the special key words controlled in the different parts of the patients' data, c) the signals of the adverse events in the laboratory and clinical parts of database; 5) module for monitoring of the Clinical Situations dynamic and clinical results in digital codes.

The prototype of the first module for decision support and drug prescriptions monitoring was made by the authors and registered. The algorithm for coding the of the patient's diagnosis components in the definite Clinical Situation was tested and showered the opportunity to solve these tasks [9]. The tools of ontology allowed to personalize the choice of the drugs, when the Clinical Situations had the additional factors and conditions affecting the decisions. The inference rules for decision support were carried out by the ontology tools from the knowledge-base. Other path for decision support by the ontologies took the start in the databases of real Clinical Situations. The new inference rules were constructed. In order to use these rules according to the regulatory demands the agreement with the medical commission was made.

3. Results

Conceptual scheme of the Information Space of Clinical Practice was constructed as the hierarchical tree: General Information Space of Clinical Practice > Information Spaces of the specialized (subject) areas of Clinical Practice > Information space of the diseases' names based on ICD-10 > Information Space of Clinical Situations at the moments of decisions and etc.

The Information Space is the result of human semantic activity [8]. *Information Space of Clinical Practice* was systematized. The first conceptual scheme of the Information Space was made for the subject area "Phthisiology".

Information Field - this concept is used in a great number of the scientific studies for description of the properties of the real space and real world [2,3,4,8].

Information Field of *Clinical Situation* was constructed with two interconnected parts:

1. Information Field of the Patient's diagnosis at definite Clinical Situation;
2. Information Field of the medications corresponding with the patient's diagnosis at definite Clinical Situation.

Information Field of *Patients' diagnosis at definite Clinical Situation* was constructed with two types of information vectors:

1. *The discrete information vectors* - stage, phase, severity of the patient's Clinical Status and the level of complexity of the definite Clinical Situation. These discrete vectors were constructed in the format of the constant dimensionality to monitor the process of treatment in dynamic by the program. All discrete information vectors had 5 levels of gradation. The corresponded definite criteria for each component of the information vectors were described.
2. The vector of information continuum - the ICD codes as the names of the diseases, the numbered entries of a linear array of the factors and conditions that affected the drugs' choice and medical interventions.

Information Field of medications for treatment – included the appropriate models of the medications for treatment according to the National Clinical Recommendations, Standards of medical care at the definite Clinical Situations, the collection of the declarative rules and the regulatory requirements for the quality drug care, background information on drugs interactions, risks of adverse events, etc.

The Information Environment for the components of the patients' diagnosis was described by using of the criteria for their boundaries. The digital code of the patient's diagnosis of Clinical Situation from database was used as the key to find the concordant Standard model of medication treatment in knowledge-base.

4. Discussion

The transformation of the scientific knowledge forms affects the process of cognition. The range of applications of the information technologies in the Clinical Practice expands very fast. But the medical information is still designated by numerous synonyms in the medical texts. For example, very important concept "stage" may be named as: stage=period=phase [10,11,12]. This requires to accelerate the improvement and harmonization of the scientific terms and concepts in medicine. The vocabulary of the terms and concepts for the components of patients' diagnosis at the Clinical Situations - stage, phase, severity of Clinical status of the patient and complexity of the Clinical Situation were formulated by the authors, published and used for coding the detailed patients' diagnosis [9,13,14]. The medical electronic platforms for monitoring of medication care, decision support, predicting the adverse events are in emergence of demand for unambiguous of terms and definitions in order to collect the clinical diagnosis with exact details in codes in databases without disclosure of the private information about the patient.

The new paradigm for decision support and monitoring of drug treatment quality with the multi-module medical electronic platform using simultaneously the ontology tools, expert decision support systems and the detailed Clinical Situations coding in the conditions of the systematization of the clinical Information Space and conceptualizing the medical terms has the great opportunity to be useful in the improving of medication care.

5. Conclusions

1. The conceptual schemes of Clinical Practice Information Space gave the opportunity to use the digital codes for the personal patient's diagnosis at the definite Clinical Situations without disclosure of the private information about the patient.
2. The conceptual apparatus for the components of patient's information vectors allowed to use the digital codes for searching the standard models in the knowledge base.
3. The tools of ontologies were successful for decision support in Clinical situations with the additional factors and conditions affecting the clinical results.
4. The new inference rules were constructed by the ontologies, when the standard models were absent.
5. The tools of ontologies showed to be useful to identify the triggers and signals by the key words and laboratory data in the process of monitoring treatment.

References

[1] Ob osnovah ohrany zdorov'ya grazhdan v Rossijskoj Federacii [On the basement of the health protection for citizens of the Russian Federation. Federal Law]: federal'nyj zakon ot 21.11.2011 № 323-FZ (red. ot 07.03.2018) Gl.5 st.37 p.4.; red. ot 25.12.2018 st.37 p.1.3: FZ №489 [Online]. [Loaded 08.03.2018] URL: http://base.garant.ru/12123352/. (Cite visit: 20.12.2018). (In Russian).

[2] Pankov AV, Shevchenko SV. Obosnovanie roli i formirovanie konceptual`noj modeli sistemy` intellektual`noj obrabotki informacii v edinom informacionnom prostranstve VS RF. [Justification of the role and conceptual model formation for the system of intellectual information processing in the unified information space of the Russian Federation armed forces]. *Informatika i komp`yuterny`e texnologii.* Izvestiya SPbGE`TU «LE`TI». 2018;1:38-43. (In Russian).

[3] Ozherel`eva TA. Ob otnoshenii ponyatij informacionnoe prostranstvo, informacionnoe pole, informacionnaya sreda i semanticheskoe okruzhenie. [About the relation of the concepts of information space, information field, information environment and semantic environment]. *Mezhdunarodny`j zhurnal prikladny`x i fundamental`ny`x issledovanij.* 2014;10(2): 21-24. [Online]. [Loaded 05.10.2014]URL: https://applied-research.ru/ru/article/view?id=5989 (Site visit: 25.05.2020). (In Russian).

[4] Bezdushny`j A.N., Gavrilova E`.A., Serebryakov V.A., Shkotin A.V. Mesto ontologij v edinoj integrirovannoj sisteme RAN. [The place of ontologies in the unified integrated system of the Russian Academy of Sciences.]. [Online]. [Loaded15.09.2019].URL: http://www.olap.ru/home.asp?artId=2362 (Site visit:20.05.2020). (In Russian).

[5] Neobratimy`j process.Fizicheskaya e`nciklopediya.[Irreversible process. Physical encyclopedia]. [Online]. [Loaded 05.09.2017] URL:https://rus-physical-enc.slovaronline.com/2488 (Site visit 06.05.2020). (In Russian).

[6] Filosofiya mediciny` [pod red. Shevchenko Yu.L.].[Philosophy of medicine] M: GE`OTAR – MED;2004.480 s. (In Russian).

[7] Filosofskie osnovaniya nauki [Philosophical foundations of the science]. [Online]. [Loaded15.09.2019]URL: https://helpiks.org/8-2257.html(Site visit: 25.05.2020). (In Russian).

[8] Filosofskie osnovaniya nauki. E`lektronnaya biblioteka IF RAN. Novaya filosofskaya e`nciklopediya. [Philosophical foundations of science. Electronic library of the IPh RAS. New philosophical

encyclopedia. [Online]. [Loaded 05.09.2019]. URL:https://iphlib.ru/library/collection/newphilenc/document/HASHfa8b4ead682875851d14fa (Site visit 06.05.2020). (In Russian).

[9] Sertifikat gosudarstvennoj registracii komp`yuternoj programmy` OOO "Trinidata" (No. 2018617324 ot 21 iyunya 2018).[Certificate of state registration of the computer program of LLC "Trinidata"].(In Russian).

[10] Svetun`kov S.G., Khan T.V. Logiko-gnoseologicheskaya terminologiya v e`konomike (kratkij slovar`).[Logico-epistemological terminology in Economics (short dictionary)]. SPb.: SPbGUE`F; 2004:115 - 117. (In Russian).

[11] Aleksandrova Z. E. Slovar` sinonimov russkogo yazy`ka. Prakticheskij spravochnik.[Dictionary of synonyms of the Russian language. A practical guide] M.: Russkij yazy`k. 2001;568. [Online]. [Loaded 15.09.2016]. URL: https://azbyka.ru/otechnik/Spravochniki/slovar-sinonimov-russkogo-jazyka (Site visit: 25.05.2020). (In Russian).

[12] Slovar` medicinskix terminov. [Vocabulary of medical terms].[Online]. [Loaded 05.09.2011] URL:http://deus1.com/vocab-17.(Site visit: 06.05.2020). (In Russian).

[13] Lugovkina T., Skornyakov S.,Egorov E., Medvinskiy I. Methodology for Electronic Monitoring of Quality Drug Prescriptions in Phthisiopulmonology. *Clinical Therapeutics*. 2017;39, (8),Supplement: e43–e44[Online]. [Loaded:15.10.2017] URL: http://www.clinicaltherapeutics.com/(Site visit: 25.11.2017). (Eng).

[14] Lugovkina T.K. Konceptualizaciya komponentov diagnoza klinicheskix situacij dlya e`lektronnogo monitoringa kachestva lekarstvennoj pomoshhi.[Conceptualization of the diagnosis components of the clinical situations for electronic monitoring of the drug care] *Social'nye aspekty zdorov'a naselenia /Social aspects of the population health* [on line]. 2021; 67(1):10. Available from: http://vestnik.mednet.ru/content/view/1236/30/lang,ru/(In Russian).

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210587

Digital Phenotypes for Personalized Medicine

Carlos MOLINA[a,1] and Belén PRADOS-SUAREZ[a]
[a] *Software Engineering Department, University of Granada, Spain*

Abstract. In this paper we propose a new definition of *digital phenotype* to enrich the formulation with information stored in the Electronic Health Records (EHR) plus data obtained using wearables. On this basis, we describe how to use this formalism to represent the health state of a patient in a given moment (retrospective, present, or future) and how can it be applied for personalized medicine to find out the mutations that should be introduced at present to reach a better health status in the future.

Keywords. Personalized medicine, Digital phenotype, Wearables, Artificial Intelligence, Electronic Health Record

1. Introduction

Personalized medicine ([10]) arises from the differences between the results of treatments on distinct patients, depending on their performance at molecular level, which is conditioned by the patient's genotype. In this sense, the presence or not of a concrete gen can make a patient more or less receptive to one treatment or another ([11]). The application of this principle is giving promising results, but so far it is limited to a small number of samples and very concrete diseases, because of the high cost of genetic sequencing ([3]).

In parallel, there is an increasing interest on storing phenotype related data ([15]), but usually considering the classic focus of the observation based-medicine for diagnosis, so only symptoms are stored. As an example, in the literature we can find some recent researches using mobile devices with sensor (*wearables*) ([2]) for medical purposes. Most of them design especial devices to detect concrete pathologies (e.g. heart attack, [12], stress [19][23]). Other approaches look for patterns in patient behavior (e.g. in the care of elder people [21], scoliosis patients [8]). All these proposals are for concrete cases providing ad-hoc solutions, but do not consider high level concepts (like life habits).

Recently, the data records from the interaction of patients with mobile devices has been used in Psychology ([9]) to obtain three patients' characteristics: the behaviour, the conscience and the mood. This approach has proven to be useful for identifying some mental disorders ([20]). In this proposal, only a very limited piece of information is considered and it could be quite enriched by incorporating the great data source that is the patient's Electronic Health Record (EHR). In the previous presented Psychology approach ([9]) the authors use these high level concepts (e.g the patient mood) but do not integrate with EHR data.

On the other hand, the data stored in the EHR has been used in combination with data mining methods, mainly oriented to diagnosis purposes ([22]). Among these methods, there are some interesting proposals considering simple temporal relations

[1] Corresponding Author: Carlos Molina, University of Granada; E-mail: carlosmo@ugr.es. ORCID: Carlos Molina (0000-0002-7281-3065), Belén Prados-Suarez (0000-0002-3980-102X)

([4][7]). These approaches find temporal relations by means of rules (looking for temporal relations between the antecedent and consequent) or by performing predictions with a fixed future window (e.g. one month). To our best knowledge, there are no proposals that take into account the complex temporal relations between elements in the EHR like sequences of facts with non lineal relations, different intervals depending on the patients, parallel evolutions of pathologies, interactions between drugs in multi-pathologies patients, etc.

The integration of both health information sources (wearables and EHR) is currently under study ([6]), mainly from the point of view of data integration than with the focus on patient care solutions; since to achieve this goal the first step is the integration of data from different sources. Solutions as the *Electronic Health Records Aggregators* (EHR_{agg},[17]) and the *European Health Data Space* ([5]) are the perfect starting points to solve this issue.

In this paper we redefine the concept of *digital phenotype* to incorporate information from EHR and wearables, to find out complex relations that opens another way to personalize treatments, parallel to the current genetic via. Our proposal thus is framed into the Smart Data area ([18]) because it aims to deal not only with the volume of the Big Data but also with the correct representation of the information to be able to model and learn more complex relations.

2. Methods

In this section we present our proposal to define the enriched basic concepts related to digital phenotypes.

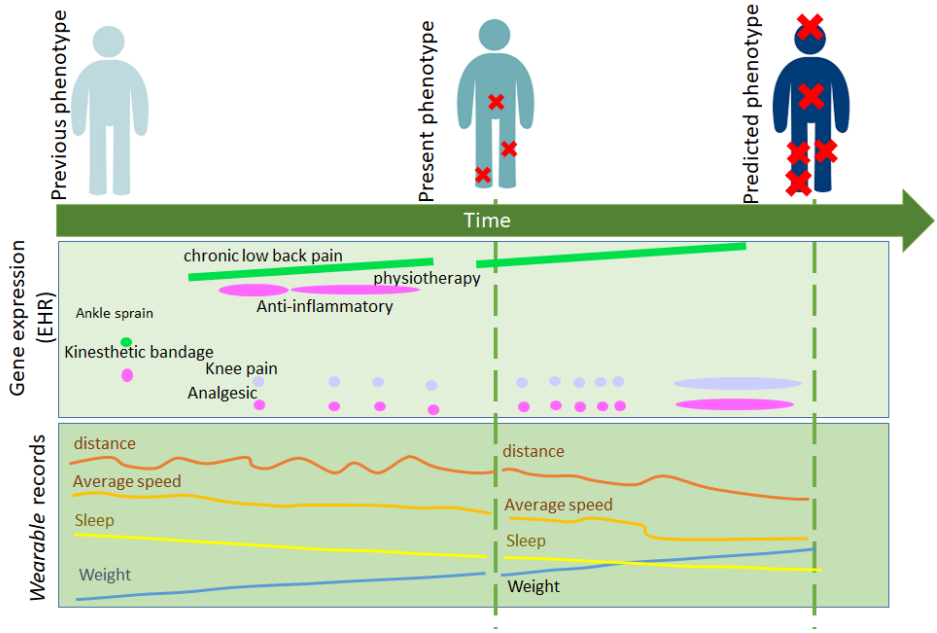


Figure 1. Structure of *Digital Phenotype* for patient *Mr. Smith*.

Definition 1. A *Digital Nucleotide* (*dn*) is each one of the elements stored in the EHR (a diagnosis, a treatment, an analysis result, a surgical intervention, a symptom, etc.) or from wearables devices (activity, sleep habits, mood, diet data, etc.).

In Figure 1, we have some examples of *dn*. Some are from the patient's EHR (e.g. a diagnosis like *ankle sprain*, symptoms like *chronic low back pain* or *knee pain*, and different treatments). Other *dn* are high level concepts defined over the data from wearables like *decreasing activity*, *increasing of weight* or *bad sleep routines*.

Definition 2. A *Digital Fen* (*df*) is a set of $j \in [1, +\infty)$ nucleotides that includes *dn*'s from the EHR, wearables, or both; and a relation $<_t$ that establishes the temporal ordering between them.

$$df\_i = (\{dn_1, \dots, dn_j\}, <_t) \tag{1}$$

To exemplify, as shown in Figure 1, we may define a *Digital Fen df* modelling *lumbar facet syndrome* as the *digital nucleotides* dn_1={*chronic low back pain*}, dn_2={*Anti-inflammatory*}, and dn_3={*physiotherapy*}, with the relation $<_t$ defined as:

$$\begin{aligned} dn_1 &<_t dn_2 \\ dn_1 &<_t dn_3 \end{aligned} \tag{2}$$

meaning that the dn_1 appears before dn_2 and dn_3, but there is no order between these two later nucleotides; so they can be applied before, after or concurrently but always after dn_1 appears.

These *digital fens* are the result of applying specific data mining methods over the data from the EHR databases and the wearable records, so these hidden relations could be discovered.

With these definitions, we are able to present the concept of *Digital Phenotype*.

Definition 3. The *Digital Phenotype* (DP_t^X) of a patient *X* at a specific time *t* is the set of *digital fens* that the patient presents:

$$DP_t^X = \{df_1, \dots, df_n\} \tag{3}$$

As an example, the *DF* at present time of the patient *Mr. Smith* shown in Figure 1 would be:

$DP_{present}^{Smith}$= {*lumbar facet syndrome, decreasing activity, increasing weight, bad sleep routines, knee pain*}

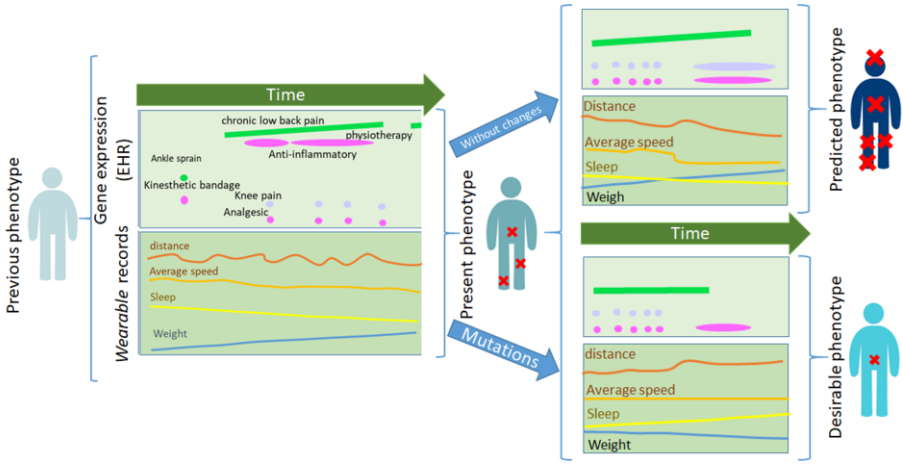


Figure 2. Evolution of the *Digital Phenotype* for *Mr. Smith*.

3. Results

According to this definition above, a patient has different *DPs* evolving from one to another over time. For example, Figure 1, applying prediction methods we can estimate the patient's future digital phenotype, if the patient does not change his habits. Then his *DP* for the future (what we can call *Predicted Digital Phenotype*) would be

$$DP^{Smith}_{predicted} = \{lumbar\ facet\ syndrome,\ obesity,\ depression\} \quad (4)$$

However the idea behind the *DP*'s is not only to be able to model in a better way the patient's health and predict his/her evolution, but to be able to identify what we should be done to redirect the patient's health towards a *Desirable Digital Phenotype* ($DP_{desirable}$)} (Figure 2). To follow the *genome simile* we want to find which *mutations* (changes in the *Digital Phenotype*) must be applied to achieve the $DP_{desirable}$.

Our proposal, to find the mutations is to identify which *df*'s should be modified (added or removed) in the patient's $DP_{present}$. To be able to do it, we need to learn how the *DP* changes depending on the *digital fens* considered:

$$DP_t + \{df_1, \dots, df_m\} \rightarrow DP_{t+1} \quad (5)$$

This task can be achieved by means of machine learning methods capable to work with the presented structure. By learning these *transformation functions* (→), we can predict how a patient's health status can change and identify the *df*'s to induce a better health evolution.

4. Discussion

Once we have the basic concepts, we define the stages of the process that complete the framework, indicating the methods that should be developed at each step:

- *Identification stage.* This step identifies the *dn*'s from the EHR and wearable records. We are currently developing data mining methods capable of working over multi-source data. As shown in [13], it is essential that these methods take into account the reliability and quality of the data.
- *Evolution stage.* This step learns the mechanisms though which the DP evolves; i.e. which mutations results in which health state. Techniques in this step must be explainable, so medical staff and patients can understand the evolution process and results. This is the case of methods such as temporal association rules [1] or gradual dependencies [14] able to deal with imprecision and interpretable to the user (e.g [16]).
- *Prediction stage.* At this phase, the knowledge discovered in previous steps is used to estimate the future health state. Here estimation and prediction methods should be applied to figure out the $DP_{Predicted}$.
- *Mutation stage.* Methods in this final step will be able not only to predict the evolution but also to identify the changes to introduce in patient's treatments or life habits, so the $DP_{Desriable}$ can be achieved. This process will involve data mining methods to learn these changes and their effects in the DP's.

5. Conclusions

We have proposed and enriched concept of *Digital phenotype* as a data model able to represent not only the data from the patient's wearable records but also the data from his/her EHR and their complex temporal relations. This framework, represents a step into the real SmartData modelling to give an holistic view of the patient. With the *digital phenotypes* we can model the patients evolution, learn hidden patterns and relations.

We are working on the definition of methods to automatically identify the relations between *df*, not only frequent, but also those existing in rare diseases (less frequent). They can help in early diagnosis and in the improvement of the expected evolution of patients.

The integration of the *digital phenotypes* with patients' genetic information will be another step to improve the *personalized medicine* as a global concept.

Acknowledgements

This research is partially supported by PGC2018-096156-B-I00 Recuperación y Descripción de Imágenes mediante Lenguaje Natural usando técnicas de Aprendizaje Profundo y Computación Flexible of the Ministerio de Ciencia, Innovación.

References

[1] Ale JM, Rossi GH. An approach to discovering temporal association rules. In Proceedings of the 2000 ACM Symposium on Applied computing. 2000; Volume 1: 294–300.

[2] Banaee H, Ahmed MU, Loutfi A. Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. Sensors 2013; 13 (12): 17472–17500.

[3] Borisov N, Tkachev V, Muchnik I, Buzdin A. Individual drug treatment prediction in oncology based on machine learning using cell culture gene expression data. In Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics (2017), pp. 1–6.

[4] Concaro S, Sacchi L, Cerra C, Fratino P., Bellazzi R. Mining healthcare data with temporal association rules: Improvements and assessment for a practical use. In Conference on Artificial Intelligence in Medicine in Europe (2009), Springer, pp. 16–25.

[5] European Commission. European health data space, https://ec.europa.eu/health/ehealth/dataspace_en. Project, July 2021.

[6] Giordanengo A, Bradway M, Muzny M, Woldaregay A, Hartvigsen G, Arsand E. Systems integrating self-collected health data by patients into ehrs: a state-of-the-art review.

[7] Hanauer DA, Ramakrishnan N, Seyfried LS. Describing the relationship between cat bites and human depression using data from an electronic health record. PLoS One 2013; 8(8): e70585.

[8] Iftikhar O. Designing an Automated System using Wearable Devices for Compliance Monitoring and Activity Detection in Scoliosis Patients. PhD thesis, University of Michigan-Dearborn, 2018.

[9] Insel TR. Digital phenotyping: technology for a new science of behavior. Jama 2017; 318(13): 1215–1216.

[10] Jain KK. Personalized medicine. Current opinion in molecular therapeutics 2002; 4(6): 548–558.

[11] Katsios C, Roukos DH. Individual genomes and personalized medicine: life diversity and complexity. Personalized Medicine 2010; 7(4): 347–350.

[12] Manisha M, Neeraja K, Sindhura V, Ramaya P. Iot on heart attack detection and heart rate monitoring. International Journal of Innovation in Engineering and Technology (IJIET) (2016).

[13] Molina C, Prados-Suarez B. Measuring the quality of data in electronic health records aggregators. In 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (2020), pp. 1–6.

[14] Molina C, Serrano J-M, Sanchez D, Vila MA. Measuring variation strength in gradual dependencies. In EUSFLAT Conf. (1) 2007; 7: 337–344.

[15] Oellrich A, Collier N, Groza T, Rebholz-Schumann D, Shah N, et. al. The digital revolution in phenotyping. Briefings in bioinformatics 2016; 17(5): 819–830.

[16] Prados de Reyes M, Molina C, Prados-Suarez B, et al. Interpretable associations over datacubes: application to hospital managerial decision making. Stud Health Technol Inform. 2014;205:131-5.

[17] Prados-Suarez B, Molina C, Peña Yañez C. Electronic health records aggregators (EHRagg). Methods of Information in Medicine 2020; 59 (2,3): 96–103.

[18] Sheth A. Smart data-how you and i will exploit big data for personalized digital health and many other activities. In Proc. IEEE Int. Conf. Big Data (2014).

[19] Simões L, Gonçalves J, Silva J. Mobile application for stress assessment. In 2017 12th Iberian Conference on Information Systems and Technologies (CISTI) (2017), IEEE, pp. 1–7.

[20] Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela J-P. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. NPJ digital medicine 2018; 1(1): 1–9.

[21] Tsai C.-H, Chu C-H, Liu S-W, Hsieh S-Y, Tseng VS. Mining life patterns from wearable sensors data for elderly anomaly detection. In 2017 Conference on Technologies and Applications of Artificial Intelligence (TAAI) (2017), IEEE, pp. 66–71.

[22] Yadav P, Steinbach M, Kumar V, Simon G. Mining electronic health records (EHRs) a survey. ACM Computing Surveys (CSUR) 2018; 50(6): 1–40.

[23] Zubair M, Yoon C, Kim H, Kim J, Kim J. Smart wearable band for stress detection. In: 5th International Conference on IT Convergence and Security (ICITCS) (2015), IEEE, pp. 1–4.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210588

A Multilayer LSTM Auto-Encoder for Fetal ECG Anomaly Detection

Inna SKARGA-BANDUROVA[a,1], Tetiana BILOBORODOVA[b], Illia SKARHA-BANDUROV[c], Yehor BOLTOV[b], Maryna DERKACH[d]
[a] *School of Engineering, Computing and Mathematics, Oxford Brookes University*
[b] *G.E. Pukhov Institute for Modelling in Energy Engineering*
[c] *Luhansk State Medical University*
[d] *Volodymyr Dahl East Ukrainian National University*

Abstract. The paper introduces a multilayer long short-term memory (LSTM) based auto-encoder network to spot abnormalities in fetal ECG. The LSTM network was used to detect patterns in the time series, reconstruct errors and classify a given segment as an anomaly or not. The proposed anomaly detection method provides a filtering procedure able to reproduce ECG variability based on the semi-supervised paradigm. Experiments show that the proposed method can learn better features than the traditional approach without any prior knowledge and subject to proper signal identification can facilitate the analysis of fetal ECG signals in daily life.

Keywords. Auto-encoder, long short-term memory (LSTM), anomaly detection, fetal electrocardiography (ECG)

1. Introduction

The modern wearable ECG recorders can store a large amount of information, which allows recording throughout the day, tracking changes with high resolution and data transmission. In addition, the analysis of this data based on machine learning models using computational algorithms makes it possible to detect changes in a timely manner and personalize treatment. At the same time, the use of wearable and intelligent devices for telemedicine and outpatient care requires accurate and reliable analysis of biomedical signals in real-time, the study of their characteristics and the corresponding physiological events.

In this context, we consider the problem of detecting abnormalities in the fetal ECG obtained from a portable device attached to the mother's abdomen, which allows long-term monitoring during pregnancy. Fetal ECG reflects the electrophysiological activity of the fetal heart. Due to the fact that the size of the maternal ECG detected in the abdominal cavity is approximately 2-10 times the size of the fetal ECG [1], the analysis of abnormalities in the fetal ECG signal is still challenging. Moreover, it needs to detect fetal QRS complex in a very noisy environment [2] and analyze the huge amount of collected data. The process is therefore expensive and time-consuming. It is

[1] Corresponding Author: Inna Skarga-Bandurova, School of Engineering, Computing and Mathematics, Oxford Brookes University, Wheatley Campus, Oxford, UK; E-mail: iskarga-bandurova@brookes.ac.uk.

advantageous to characterize data in advance on ECG states and thereby use consequent deep analysis only if needed.

Any change in the standard ECG signal leads to heart disease (abnormal). Based on this statement, differentiating the ECG into normal and abnormal will help reduce the time it takes experts to interpret the ECG. Expert analysis will be necessary only if there are deviations in the signal from the normal ECG.

Different approaches are used to detect deviations in the ECG signals. In [3], the novel S-transform and frequency division technique was proposed to analyze the time-frequency sub-band matrices stemming from the ECG signals. The autoregressive time-frequency analysis was applied in [4] for finding the coefficients and time-frequency description of the ECG signal. The study [5] proposed a segmented-beat modulation method as a new template-based filtering procedure able to reproduce ECG variability. A model that utilizes long short-term memory (LSTM) prediction errors as input features for the second stage of anomaly class prediction was proposed and described in [6]. The results show that the error vectors represented by their summary features carry useful predictive information about the type of the ECG anomaly. In [7], a deep learning model based on residual network (ResNet) that adopts the 1-D octave convolution was proposed to extract fetal ECG. The study [8] proposed a cardiac abnormality detection algorithm for optical photoplethysmography (PPG) based on an autoencoder. The autoencoder was trained to recognize typical normal PPG morphology and rhythm.

This article aims to extend these studies and describe a multilayer LSTM auto-encoder network to obtain information about changes or abnormalities in the ECG during fetal monitoring via a non-invasive wearable device. Following [9], the LSTM-based encoder is used to detect deviations and assess the presence of abnormalities in the fetal ECG rhythm as well as to help determine the need to visit a doctor between regular schedules for a more detailed and comprehensive examination.

2. Methods

In this study, we tested a semi-supervised learning approach to assess ECG abnormalities using the LSTM autoencoder. The key idea is to learn an LSTM autoencoder on a multivariate ECG signal to perform binary classification. This can be achieved in the following way. The LSTM autoencoder is trained on ECG data with normal rhythm and on new reconstructed data. Then the specific threshold is applied, and the loss function is calculated. If the loss is higher than the threshold, it is defined as an anomaly. The defined anomaly is recorded and counted, and if the recording is higher than the threshold, it is triggered to warn the patient to visit a doctor for a more thorough examination.

Further, the description of the proposed methodology is presented. A semi-supervised learning pipeline to detect abnormalities in the ECG using the LSTM autoencoder includes the following main stages: data preprocessing, model training, anomaly detection.

2.1. Methodology

Traditionally, preprocessing includes ECG signals normalization and segmentation. Since we used normalized data, the normalization stage was skipped. The sliding window is used for signal segmentation. The waveform divides into overlapping segments. This approach is used to get instances of waveform shape. The necessary primary condition

in this model is the size of the window. A window is represented as a shape of consecutive data elements per time unit. The result of signal segmentation is the data set of ECG shapes.

In situations when some labels are unavailable for one of the classes, for example, when negative class in unlabeled, it is possible to utilize these unlabeled data drawn from marginal density [10]. In case of binary classification, having one labeled class and set of unlabeled data the classifier can be trained using only one class (for example, positive) and unlabeled data.

The classifier should discriminate between positive class and unlabeled data. However, a naive attempt to classify positive class and unlabeled data in many cases lead to bias. To address this challenge, we utilize unlabeled data to evaluate the risk for negative samples [11] and introduce the indicator function defined as binary values (0,1), which takes the value of 0 if the predicted class is positive and 1 if the predicted class is negative.

The anomaly detection is performed using the L1 Loss function. The L1 Loss or Mean Absolute Error (MAE) is used for model evaluation. MAE is the sum of absolute differences between the true value and predicted values that measures the average magnitude of errors in a set of predictions, without considering their directions.

$$L1LossFunction = \sum_{i=1}^{n} \left| y_{true} - y_{predicted} \right| \tag{1}$$

The ECG abnormalities can be assessed based on the following assumptions. The anomaly shapes that represent the segment S follow the Boolean function {0, 1}

$$f(S) = \begin{cases} 1, \text{ if the anomaly is detected} \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

where $f(S)$ is a function of segment anomaly estimation.

Abnormalities assessment of ECG signal α is defined as a function $g(\alpha)$ that calculated as a percentage of detected anomaly segments in ECG signal

$$g(\alpha) = \frac{\sum_{j=1}^{w_i} (f(S_i^j))}{w_i} \tag{3}$$

where α is an ECG signal, w_i is a number of all segments in ECG signal α, $f(S_i^j)$ is a Boolean function defined in (2).

Further, if a number of anomaly segments is higher than the defined threshold, ECG signal α will be assessed as abnormal.

2.2. Experimental Data

For the experiment stage, we used open-source data [12] from PhysioNet [13]. In this dataset, non-invasive fetal ECG recordings were collected from pregnant women during routine medical visits. The data include recordings with diagnosed fetal arrhythmias by echocardiography, and equal control recordings diagnosed with a normal rhythm. All data were acquired at a sampling frequency of 500 Hz (ARRs 6-9, Table S1) or 1000 Hz

(ARRs 1-5 and 10-11 and all NRs) and with a 16-bit resolution and a range of ± 8 mV. Data were recorded continuously.

Each entry consists of four or five abdominal canals that correspond to the fetus ECG: Abdomen_1, Abdomen_2, Abdomen_3, Abdomen_4, Abdomen_5, and one channel that corresponds to the highlighted mother's ECG: ECG. The entries are annotated as follows: ARR - arrhythmia fetus, NR - normal rhythm fetus. Records NR_01 and ARR_01 were used in the experiment.

A fragment of the ARR_01 includes six records (see Fig. 1). The upper signal corresponds to maternal ECG, five lower signals correspond to fetal ECG.

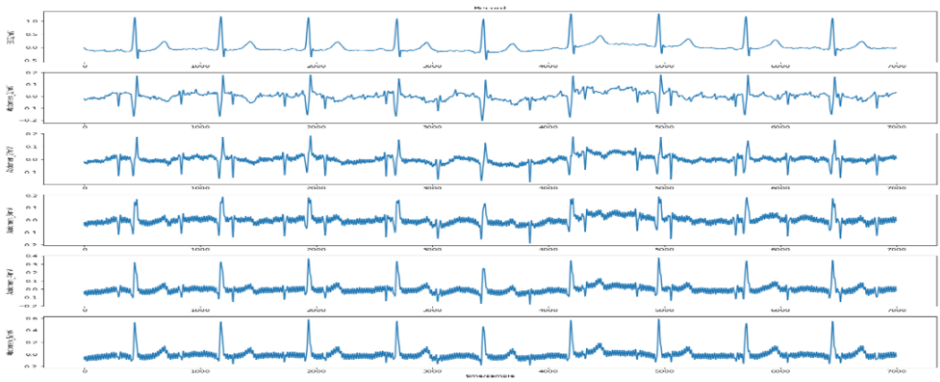


Figure 1. The fragment of data.

For the experiment, we previously removed the maternal ECG from the studied ECG signals [14].

3. Results

Segmentation was carried out using a sliding window with a size of 550 epochs. The offset is 1%, i.e. 6 epochs. The model was trained using two configurations of LSTM auto-encoders. The model was trained on segmented fetal data with a normal rhythm. The number of epochs is 350 for the experiment setup. The architecture of used classifiers is shown in Table 1.

Table 1. The parameters of the LSTM classifiers.

Parameters	LSTM cells	Network layers	Optimizer	Dropout /Dense	Epoch	Batch size	Cost function
Default settings	64	4	NAdam	0	350	128	Focal Loss
Our settings	80	5	Adam	1	250	128	L1Loss

For the second classifier, we increased the number of the LSTM layers to 5 resulting in that the network learned features at different time scales.

Anomalies were identified by classifying abdominal lead data that correspond to segmented fetal ECG data, which are annotated as ARR. The control data used 15% of the segmented fetal ECG dataset with a normal rhythm, which was not used in training. The threshold value for the loss function, which indicates a prediction error. All values

of loss function above the threshold are flagged as an anomaly. If the loss function is below the threshold, it was classified as a normal heartbeat.

The visualization of original and reconstructed normal and anomalous ECG segments is shown in Fig. 2.

Figure 2. The examples of simulated data with detected anomalies.

The resulting threshold value for the loss function, which indicates a prediction error, was defined to be 52. All values of loss function above this value are flagged as an anomaly. If the loss function is below the threshold, we classify it as a normal heartbeat. As noted earlier, exceeding the threshold could be a signal to see a doctor for a more specialized examination.

4. Conclusion

In this work, we used LSTM auto-encoders to assess the presence or absence of features that determine the normalcy of the signal. Auto-encoders have made it possible to reduce the multidimensional problem of whether an ECG segment matches the patterns that define normality in just two dimensions: reconstruction errors in the time and frequency domains. A detailed study of the obtained results of detecting ECG signal anomalies showed that the recordings alternate segments of low and high quality, and the presence of an anomaly affects the likelihood that the following segments will be abnormal. The neural networks make it possible to analyze the ECG signal without additional data processing; however, more accurate detection is needed for additional signal processing in order to identify the QRS-complex elements used for further analysis [15]. There are at least three benefits of the proposed methodology: (i) feature extraction, and selection techniques are not needed, (ii) the proposed approach can help to inform about significant abnormalities, (iii) it is possible to predict ECG anomaly using only one class of labelled data. One more positive effect of this approach is that the final anomaly estimate has a certain range at (0, 1) and is clearly interpreted as the probability that the segment is an anomaly, which would not be so unambiguous if we were left with the reconstruction error as an anomaly estimate.

References

[1] Ungureanu GM, Bergmans JWM, Oei SG, Ungureanu A, Wolf W. The event synchronous canceller algorithm removes maternal ECG from abdominal signals without affecting the fetal ECG, Computers in Biology and Medicine 2009; 39(6): 562-567.

[2] Yuen B, Dong X, Lu T. Detecting Noisy ECG QRS Complexes Using Wavelet CNN Autoencoder and ConvLSTM, in IEEE Access, 2020; 8: 143802-143817, doi: 10.1109/ACCESS.2020.3012904.

[3] Tripathy RK, Paternina MR, Arrieta JG, Zamora-Méndez A, Naik GR. Automated detection of congestive heart failure from electrocardiogram signal using Stockwell transform and hybrid classification scheme. Computer methods and programs in biomedicine, 2019; 173: 53-65.

[4] Gupta V, Kanungo A, Kumar P, Sharma AK, Gupta A. Auto-regressive Time Frequency Analysis (ARTFA) of Electrocardiogram (ECG) signal. International Journal of Applied Engineering Research, 2018; 13(6): 133-138.

[5] Agostinelli A, Sbrollini A, Giuliani C, Fioretti S, Nardo FD, Burattini L. Segmented beat modulation method for electrocardiogram estimation from noisy recordings. Medical Engineering & Physics, 2016; 38(6): 560-568. doi:10.1016/j.medengphy.2016.03.011

[6] Chauhan S, Vig L, Ahmad S. ECG anomaly class identification using LSTM and error profile modeling. Comput Biol Med. 2019 Jun;109:14-21. doi: 10.1016/j.compbiomed.2019.04.009.

[7] Vo K, Le T, Rahmani AM, Dutt N, Cao H. An Efficient and Robust Deep Learning Method with 1-D Octave Convolution to Extract Fetal Electrocardiogram. Sensors (Basel). 2020 Jul 4; 20(13), 3757. doi: 10.3390/s20133757.

[8] Whiting S, Moreland S, Costello J, Colopy G, McCann C. Recognising Cardiac Abnormalities in Wearable Device Photoplethysmography (PPG) with Deep Learning. *arXiv preprint*, 2018, arXiv:1807.04077.

[9] Hou B, Yang J, Wang P, Yan R. LSTM-Based Auto-Encoder Model for ECG Arrhythmias Classification, IEEE Transactions on Instrumentation and Measurement, 2020; 69(4): 1232-1240, doi: 10.1109/TIM.2019.2910342.

[10] Sakai T, du Plessis MC, Niu G, Sugiyama M. Semi-Superised Classification based on classification from positive and unlabelled data, https://arxiv.org/abs/1605.06955.

[11] du Plessis MC, Niu G, Sugiyama M. Analysis of learning from positive and unlabeled data. In NIPS, pp. 703–711, 2014., du Plessis, M. C., Niu, G., and Sugiyama, M. Convex formulation for learning from positive and unlabeled data, In ICML, **37**, (2015), 1386-1394.

[12] Behar JA, Bonnemains L, Shulgin V, Oster J, Ostras O, Lakhno I. Noninvasive fetal electrocardiography for the detection of fetal arrhythmias. Prenat Diagn. 2019 Feb;39(3):178-187. doi: 10.1002/pd.5412.

[13] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation. 2000 Jun 13;101(23):E215-20. doi: 10.1161/01.cir.101.23.e215.

[14] Biloborodova T, Scislo L, Skarga-Bandurova I, Sachenko A, Molga A, et al. Fetal ECG signal processing and identification of hypoxic pregnancy conditions in-utero, [J]. Mathematical Biosciences and Engineering, 2021; 18(4): 4919-4942. doi: 10.3934/mbe.2021250.

[15] Filatova A, Skarga-Bandurova I, Brezhniev E, Fahs M. Evaluating the Effectiveness of Electrocardiological Study Using Cardiological Decision Support Systems, CEUR Workshop Proceedings, 2020; 2711: 294–308.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210589

A NLP Pipeline for the Automatic Extraction of Microorganisms Names from Microbiological Notes

Sara MORA[a,1], Jacopo ATTENE[a], Roberta GAZZARATA[b], Giustino PARRUTI[c] and Mauro GIACOMINI[a]
[a] *Department of Informatics, Bioengineering, Robotics and System Engineering (DIBRIS), University of Genoa, Italy*
[b] *Healthropy, Corso Italia 15/6, 17100 Savona, Italy*
[c] *Department of Infectious Diseases, AUSL Pescara, Pescara, Italy*

Abstract. According to the "Istituto Superiore di Sanita`" (ISS), hospital infections are the most frequent and serious complication of health care. This constitutes a real health emergency which requires incisive and joint action at all levels of the local and national health organization. Most of the valuable information related to the presence of a specific microorganism in the blood are written into the notes field of the laboratory exams results. The main objective of this work is to build a Natural Language Processing (NLP) pipeline for the automatic extraction of the names of microorganisms present in the clinical texts. A sample of 499 microbiological notes have been analysed with the developed system and all the microorganisms names have been extracted correctly, according to the labels given by the expert.

Keywords. Microbiological infections, natural language processing, automatic extraction, standard coding system, laboratory information systems

1. Introduction

After the beginning of the well-known worldwide pandemic of COVID-19, it has become even more evident that hospitals, assisted residences and shelters for the elderly nowadays represent areas where the circulation of pathogenic microorganisms is increasingly worrying and widespread. According to the *"Istituto Superiore di Sanità"* (ISS) [1], hospital infections are the most frequent and serious complication of health care. They can be defined as infections that arose during hospitalization or after the patient's discharge, which at the time of admission were not clinically manifest, nor were they incubating. This constitutes a real health emergency which requires incisive and joint action at all levels of local and national health organization. The goal is to uniformly activate stable and automatic systems of reporting and epidemiological surveillance capable of promptly identifying pathogenic microorganisms, multi-resistant or not, responsible for infections and to allow the immediate adoption of specific control measures.

[1] Corresponding Author: Sara Mora, Department of Informatics, Bioengineering, Robotics and System Engineering (DIBRIS), University of Genoa, Italy; E-mail: sara.mora@edu.unige.it

This kind of information can be found into the results of laboratory exams, in particular when microbiological cultures are executed on the blood sample. It is more than 10 years that laboratory analyses have been managed in a computerized manner through the use of *Laboratory Information Systems* (LISs). The non-contemporary and highly localized development of this type of systems led each center to create its own vocabulary for the coding of individual laboratory analyses. In order to make the results coming out of the single laboratories comparable, appropriate standard systems have been devised for the management of terminology, for example the *Common Terminology Services re- lease 2* (CTS2) [2], whose specifications derive from the synergy between the *Object Management Group* (OMG) [3] and *Health Level 7* (HL7) [4].

The main problem that computerized systems have faced is precisely the management of microbiology. This is because it is a highly variable discipline and linked to the habits of individual laboratories (for example which coding system is used for the nomenclature of bacteria, how sensitivity analyses are performed, ...). Therefore, the management of this type of laboratory analysis should be more varied. On the other hand, national laws imposed the mandatory use of LISs, whose structure in some specific cases may be too stiff. This problem has been overcome during the years with the simple trick of writing lots of natural text in the clinical notes, probably because clinicians could not find more appropriate fields for that kind of information. So, clinical notes became a great source of valuable information both for patient care and biomedical research, but they require manual inspection which is very expensive from an effort and timing point of view.

To tackle the problem, artificial intelligence tools can be used such as *Natural Language Processing* (NLP), a branch of computer science that deals with the interactions between computers and natural human language; it studies the problems connected with the automatic generation and understanding of human language, written or spoken [5,6,7].

The goal of this work is to create a pipeline for the automatic extraction of specific information from microbiological reports through the use of NLP techniques. This work is a first step for the development of a hospital and territorial antibiotic prescription monitoring system in the Abruzzo Region [8].

2. Materials & Methods

2.1. Characteristics of the Sample

The clinical notes used in this study are extracted from 1 month of anonymized laboratory referral from the main hospital of Pescara in Abruzzo Region. The sample is composed by 499 texts, 276 of them containing the name of a microorganism. The presence of the specific microorganism was confirmed by an expert from the hospital.

2.2. Environment & Libraries

The pipeline was completely developed in Python and the environment used is Jupyter Notebook. The libraries used for the specific tasks of this project are:

1. *Pandas*: it is a Python library containing open source data analysis and manipulation tools [9].

2. *Natural Language Toolkit* (NLTK): it is the most used library to perform text analysis in multiple languages, in fact it is very popular in academia and for research [10]. Some examples of the supported operations are: tokenization, stemming, part of speech tagging and disambiguation.
3. *SpaCy*: is an open source library for NLP in Python. It supports different languages and it is particularly suitable for the creation of software applications in- tended for production.
4. *FuzzyWuzzy*: is a Python library that supports the comparison of strings with each other. In particular, its main functions compute the distance in different cases: strings with the same length or not, taking into account the order in which the words are arranged and how many times a string can be repeated. The comparison between strings is based on the Levenshtein distance:

$$lev_{a,b}(i,j) = \begin{cases} max(i,j), & \text{if } \min(i,j) = 0 \\ \min \begin{cases} lev_{a,b}(i-1,j)+1, \\ lev_{a,b}(i,j-1)+1, \\ lev_{a,b}(i-1,j-1)+1_{(a_i \neq a_j)} \end{cases} & \text{otherwise} \end{cases}$$

where i and j are the indexes of the last character of the substring.

2.3. Steps of the Pipeline

1. **Vocabulary building:** a vocabulary was created containing the names of microorganisms (bacteria, fungi, yeasts, viruses) respecting the current taxonomic subdivision proposed by Carl Woase in 1990. Together with the name of the family, genus and species, the microorganism has been mapped into 3 standard coding systems, at national or international level: *Italian Clinical Microbiologists Association* (AMCLI), *Systematized Nomenclature of Medicine – Clinical Terms* (SNOMEDCT) and *National Healthcare Safety Network* (NHSN).
2. **Data acquisition:** the data were imported together with the vocabulary through the use of pandas library.
3. **Tokenization, stopwords removal and genus extension:** in this phase, clinical notes have been divided into tokens. Then stopwords longer than 1 character have been deleted in order not to compare strings that are prepositions, articles and adverbs.
   Starting from the Linnaeus classification [11] the binomial nomenclature is used. It is formed by the name of the genus with the first capital letter and the name of the species in lowercase. Often, after a species name is introduced in a text, the genus is abbreviated to the first letter in subsequent mentions (followed by a fullstop). Unfortunately, however, since the notes are very short, it is a shared agreement to always use the abbreviated form, even without having first specified the entire genus once. In this binomial nomenclature, the use of a two-letter abbreviation for the genus has not been introduced. So, words composed by only one character haven't been deleted.

This check could be done also through the use of regular expressions. However, this choice was made in order to consider that abbreviations could be spelled incorrectly. For example abbreviations not followed by a fullstop or letters followed by a fullstop but lowercase. The last step of this phase was the extension of the microorganism genus, in particular the "n+1" token have been compared with each species of the vocabulary. If the two tokens had a very high similarity wuzzy index (greater than or equal to 98) the token "n" was checked. In the event that the token "n" began with the same letter of the genus belonging to the species found in position "n+1", the extension of the genus was made.

4. **Extraction of the desired microorganism from the clinical notes:** initially, an attempt was made to carry out a morphological and lexical analysis. However this approach did not produce any good results due to the lack of morphological structure of the reports. The extraction of the microorganism was done by comparing the tokens present within the report and the vocabulary using the library FuzzyWuzzy. As regards the extraction of the genus, the threshold on the similarity index is set at 75, while the threshold for the species is set at 85 (they were more frequently written correctly).

3. Results

3.1. Genus Extension

There were 107 abbreviated genera followed by species in the notes. Once all the notes have been elaborated by the system the abbreviation extension from all 107 genera have been extended. The extended genera completely matched with the indications of the expert.

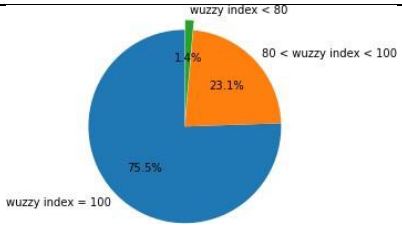
3.2. Microorganism Detection

The total number of available clinical notes was 499, 276 of them actually contained the name of a microorganism while 223 did not contain any microorganisms name. Two tests were carried out:

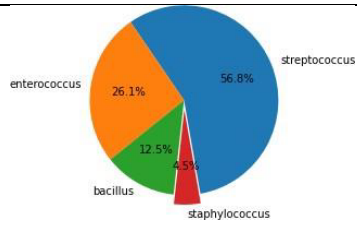
1. First all 499 notes were introduced into the microorganism extraction module.
2. Then only the notes that actually contained the microorganisms were introduced.

In both cases the system extracted all the microorganisms names, this suggests that it is not necessary a pre-processing phase that filters the data in some way. In particular 416 genera of microorganisms were found, most of them (321) with a wuzzy index of 100, also thanks to the process genus extension.

As can be seen in figure 1b below, the microorganism with the lowest score is Staphylococcus. If a species is not specified, Staphylococcus tends to have a very low similarity index, between 76 and 80. This is because Staphylococcus and 'stafilococco/stafilococchi' (term referable to the Staphylococcus microorganism that is tran- scribed in the notes) have respectively 14 and 12 letters. In addition, only 9 letters coincide, so they have a Levesthein distance of 5, as it takes 5 changes to transform the first word into the other. This is due to the fact that the Staphylococcus is one of the most widespread bacteria, therefore it is common to mention it in the natural discourse and therefore it is frequent to find the Italian term and not only the strictly scientific term in the notes.



a) Wuzzy indexes percentage distribution of genera.



b) Percentage distribution of genera with lower index.

Figure 1. System performance on genera extraction.

Species had a wuzzy index always greater than 88.

Finally, a weight parameter was introduced. It was a decimal value, between 0 and 1, associated with each couple composed by genus and species, or only with the genus if present alone, in order to highlight the maximum wuzzy indexes. This because the same word (genus and/or species) could be associated with more than one genus/species. For example, if the genera *Acetobacter* and *Acinetobacter* were compared, these would have a similarity index of 92, therefore quite high. In order to confirm the selected genus the following token was compared to the species of that specific genus in the vocabulary. If a matching was found (wuzzy index over 98) than that genus assumed weight equal to 1 and the other 0.

Gram negativi profilo proteomico riferibile ad A baumannii. Propensione di A baumanni alla pan-resistenza eccetto colistin a(Microorganismo alert)

Gram	gram	100.0	positive	positivi	88.0	[95, 'gram', 'positivi', 'streptococchi', 'ent...	1
Streptococcus	streptococcus	100.0	pneumoniae	pneumoniae	100.0	[95, 'gram', 'positivi', 'streptococchi', 'ent...	1
Streptococcus	streptococchi	85.0	NaN	NaN	NaN	[95, 'gram', 'positivi', 'streptococchi', 'ent...	0.85
Peptostreptococcus	streptococcus	84.0	NaN	NaN	NaN	[95, 'gram', 'positivi', 'streptococchi', 'ent...	0
Enterococcus	enterococchi	83.0	NaN	NaN	NaN	[95, 'gram', 'positivi', 'streptococchi', 'ent...	0.83
Peptococcus	streptococcus	83.0	NaN	NaN	NaN	[95, 'gram', 'positivi', 'streptococchi', 'ent...	0
Cryptococcus	streptococcus	80.0	NaN	NaN	NaN	[95, 'gram', 'positivi', 'streptococchi', 'ent...	0
Gram	Gram	100.0	negative	negativi	88.0	[94, 'Gram', 'negativi', 'profilo', 'proteomic...	1
Acinetobacter	A	100.0	baumannii	baumannii	100.0	[94, 'Gram', 'negativi', 'profilo', 'proteomic...	1
Acinetobacter	A	100.0	baumannii	baumannii	100.0	[94, 'Gram', 'negativi', 'profilo', 'proteomic...	1
Acetobacter	A	92.0	NaN	NaN	NaN	[94, 'Gram', 'negativi', 'profilo', 'proteomic...	0

Figure 2. Columns are: genus, word in text matching the genus, genus wuzzy index, species, word in text matching the species, species wuzzy index, tokenized clinical note, weight. In the upper part of the figure the full text of the highlighted clinical note is available.

Otherwise, if there was no species present and the two words had the same wuzzy index, for example due to a spelling error, then the algorithm would return both genera but with a weight of 0.5.

4. Discussion

In general the genus extension led to good results in the phase of microorganisms names extraction compared to the labels given by the expert. Anyway some ambiguities can be

found during this phase. In fact, there are several microorganisms that have identical species and the genus begins with the same letter. Among these it is worth mentioning the *intermedius* species as it is the most probable among the ambiguous ones. It has both *Streptococcus* and *Staphylococcus* as genus. *Staphylococcus intermedius* is a very rare human pathogen. There are very few cases in the literature describing *S. intermedius* as a cause of infection in humans. Most of these cases have been described in association with exposure to animals, mainly dogs. While *Streptococcus intermedius* is the major cause of brain abscesses (70%). In Italy, however, there are very few cases of brain abscesses per year, in fact the incidence is less than 0.1% per year. In the event that an extreme case similar to the one mentioned above appears, the clinical note will be duplicated and both the microorganisms will be extracted, but the weight of the single couple is 0.5.

5. Conclusions

The main objective of this work was building a NLP pipeline that supported the automatic extraction of the names of microorganisms contained in microbiological notes. All the microorganisms present were extracted correctly, so the main goal was achieved. The next step of the aforementioned project will be to proceed with the automatic extraction of the antibiotic prescription from the same clinical notes. In particular, a key information will be the specific sensitivity of the microorganism to each single antibiotic tested.

Finally, being based on international nomenclature standards, this pipeline can be applied with microbiology notes in Italian from hospitals of all over the national territory.

References

[1] https://www.epicentro.iss.it/
[2] Gazzarata R, et al. *A terminology service compliant to CTS2 to manage semantics within the regional HIE.*, European Journal of Biomedical Informatics 13.1 (2017).
[3] https://www.omg.org/
[4] https://www.hl7.org/
[5] Matheny, Michael E., et al. *Detection of blood culture bacterial contamination using natural language processing.*, AMIA Annual Symposium Proceedings. Vol. 2009. American Medical Informatics Association, 2009.
[6] Maganti N, et al. *Natural language processing to quantify microbial keratitis measurements.*, Ophthalmology 126.12 (2019): 1722-1724.
[7] Fu S, et al. *Automated detection of periprosthetic joint infections and data elements using natural language processing.*, The Journal of Arthroplasty 36.2 (2021): 688-692.
[8] Gazzarata R, et al. *A SOA based solution for MDRO surveillance and improved antibiotic pre- scription in the Abruzzo region.*, pHealth 2019. IOS Press, 2019. 49-54.
[9] https://pandas.pydata.org/
[10] Bird S. *NLTK: the natural language toolkit.*, Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions. 2006.
[11] Linné Cv. *Systema naturae.* Ed 10 (1758): 551.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210590

Understanding the Gap Between Information Models and Realism-Based Ontologies Using the Generic Component Model

Mathias BROCHHAUSEN[a,1], Sarah J. BOST[b], Nitya SINGH[c], Christoph BROCHHAUSEN[d], and Bernd BLOBEL[e,f,g]
[a] *Dept. of Biomedical Informatics, University of Arkansas for Medical Sciences* [a], *USA*
[b] *Dept. of Health Outcomes and Biomedical Informatics, University of Florida, USA*
[c] *Emerging Pathogens Institute & Dept. of Animal Sciences, University of Florida, USA*
[d] *Institute of Pathology & Central Biobank, University and University Clinic of Regensburg, Regensburg, Germany*
[e] *Medical Faculty, University of Regensburg, Germany*
[f] *eHealth Competence Center Bavaria, Deggendorf Institute of Technology, Germany*
[g] *First Medical Faculty, Charles University of Prague, Czech Republic*

Abstract. The wide-spread use of Common Data Models and information models in biomedical informatics encourages assumptions that those models could provide the entirety of what is needed for knowledge representation purposes. Based on the lack of computable semantics in frequently used Common Data Models, there appears to be a gap between knowledge representation requirements and these models. In this use-case oriented approach, we explore how a system-theoretic, architecture-centric, ontology-based methodology can help to better understand this gap. We show how using the Generic Component Model helps to analyze the data management system in a way that allows accounting for data management procedures inside the system and knowledge representation of the real world at the same time.

Keywords. Information Models, Biomedical Ontologies, Knowledge Representation, Systems Theory, eHealth

1. Introduction

As Biomedical Informatics (BMI) is increasingly moving towards using Artificial Intelligence (AI), including Knowledge Representation and Reasoning approaches on datasets created by integrating multiple databases and datasets, BMI practice and research continues encountering problems created by the difference in requirements for information models and data integration/harmonization. Since information models are

[1] Corresponding Author, Mathias Brochhausen, Department of Biomedical Informatics, University of Arkansas for Medical Sciences, 4301 W. Markham St., Slot 782, Little Rock, AR 72205, USA; E-mail: mbrochhausen@uams.edu.

specifically designed to inform database schema development, the obstacles to data integration created in the information models are inherited in database schemata. Data about one person is frequently stored in more than one database system, with the different systems partially holding different types of data. The integration of multiple databases is crucial in order to draw meaningful inferences, e.g., about causative infectious agents, their temporal patterns, and outbreak detection, which is prime to infection prevention and public health best practices. A unique key identifier is typically associated with each entry in each database and can be used to retrieve the related information from either system. However, the same patient can be infected or sampled multiple times and at multiple locations in the course of a disease or infection, generating multiple associated entries in each system and multiple unique key identifiers, one for each database system. Thus, lack of strategic decisions on knowledge representation considerations can hinder the integration of the multiple databases due to data discrepancies. This is a major hurdle in drawing meaningful inferences from integrated data sets, especially in real-time, for making informed medical decisions. They negatively affect our ability to use and integrate data, for example when linking the database, managing the patient's demographic and follow-up patient-reported data about travel, food intake, exposure to potential carriers, etc. with the biobank samples, and their integrated database, and managing different molecular analyses related to infectiological results. Other limitations of information models are described and overcome, e.g., in Blobel et al. and Oemig and Blobel [1, 2].

Likewise, in many biobanks the entry number of a specimen is used as the reference number. Over the course of time, a tumor patient will have several entries, each receiving a new entry number. This does not only apply to specimens derived from the primary tumor, but also to those derived from secondary tumors or metastases. In the biobank, these samples must be identified and documented in the so-called sample history as follow-up samples for the first tumor. Typically, the sample history assignment is done using the patient identifier. Many systems enforce that there be only one patient identifier for each patient. If the patient has all samples taken at the same healthcare institution, this does not present a problem. However, if some of the sampling is done at another healthcare provider, tracing disease progression via the patient identifier is no longer possible because the entry number cannot be matched with the patient identifier at the location of the tissue bank. This is a common challenge especially in regional Comprehensive Cancer Centers in Germany, where patients are treated by different health care providers, and the central biobank collects samples from all the health care providers within the Cancer Center intake area. The fact that the information model enforces the existence of one and only one patient identifier hinders the integration of data about specimens curated outside of the biobank system. Thus, the requirement for a unique patient identifier, while understandable from the perspective of intra-system development, negates the fact that human beings are patients with multiple healthcare providers and, consequently, may have multiple patient identifiers.

There is no doubt that information models are useful and that using information models to inform database schemata is an important strategy. However, the example of integrating biobank data presented above raises the question of how the interplay between information model and domain representation can be orchestrated to ensure that both are compatible and do not lead to errors and false inferences.

In a recent paper, Brochhausen et al. [3] discussed the lack of computable semantics in medical Common Data Models and problems arising from falsely ascribing semantic capabilities to those important tools. In this paper, we explore a representational issue

that frequently occurs when using IT-oriented information models as sole representational resources. IT-oriented information models fulfill a crucial role in planning, defining, and describing the operational behavior of an IT system, such as an EHR system or a biobank information system. Due to the operational focus, the resulting representation, especially when relationships are represented, is frequently focused on data in that system alone and not on what the data represents, in our case the medical world, e.g., the patients, encounters, and prescriptions described in an EHR or the specimens, donors, and storage properties in a biobank. This problem also creates difficulties for approaches that seek to develop a knowledge representation resource (e.g. ontology-driven data management) based on an information model [4]. We stress the importance of bringing in biomedical expertise when representing the biomedical domain and the relations therein. In this paper, we discuss a use-case showing how the Generic Component Model (GCM) (Figure 1) in interplay with realism-based ontologies can be used to bolster information representation with domain specific knowledge, in this case based on clinical care and research.

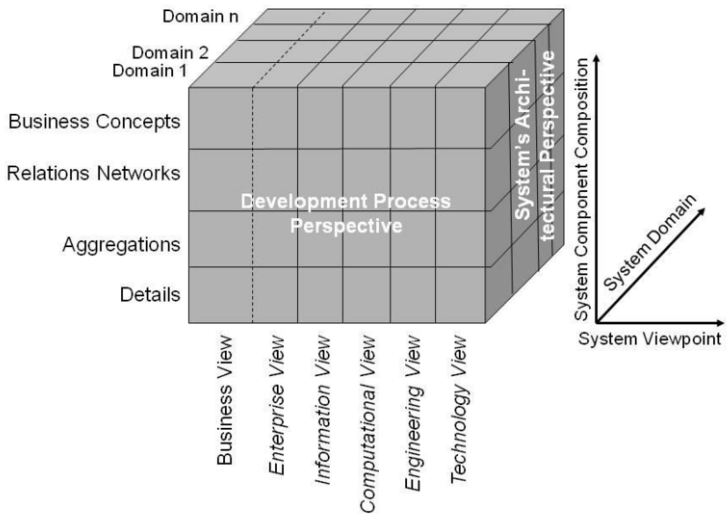


Figure 1. The Generic Component Model

The GCM is a top-level architectural model for multi-domain systems, describing the system components, their functions and interrelations structurally and behaviorally, thereby representing specific aspects (domains) by related subsystems. It is described and specified in ISO 23903 [5]. For each business case, the subsystem components and their functions and interrelations are instantiated by naming and representing them using the specific terminologies and ontologies of the domains involved in that business case. For enabling this representation of the real world system by its information technology-independent domain ontologies, the GCM specifies a Business View in addition to the five views defined by the ISO 10746 Open Distributed Processing Reference Model (RM-ODP) [6]. Furthermore, ISO 23903 introduces generic granularity levels for correctly representing and interrelating compositions/decompositions of elements. The views prescribed by the RM-ODP are Enterprise View (purpose, scope, and policies of

the system), Information View (information processing, semantics of information), Computational View (functionality of the system, functional decomposition), Engineering View (distribution of processing performed by the system), and Technology View (choice of technology for the system) [6]. Notably, a descriptive representation of the domain and its composition/decomposition is missing, which the GCM compensates by adding the Business View and granularity levels, that way correctly representing multi-domain real world systems.

The general problem of unique identifiers is not the main focus of this paper. We use longitudinal identification of tissue samples in a biobank as the leading example of how restrictions from data models are hindering integration approaches for consumption in AI methods. This is of crucial importance for the further development of precision medicine.

2. Methods

Applying the methods to the use-case at hand first requires considering the role of the information models and the database schemata according to the views of the GCM. Both belong to the Information View, since their primary focus is to guide information processing within the system. Hence, they do not provide real world knowledge, such as the fact that one person can have more than one patient ID and that sampling of tumor progression is frequently done by different healthcare providers. Notably, those aspects lie outside each individual EHR system and are ill-fitted to be represented within the Information View. However, from our database example, it is clear that not accounting for those aspects of the Business View may lead to errors in the system.

Our approach is to use representation of data in the Resource Description Framework (RDF) along with realism-based ontologies. Smith & Ceusters [7] propose a realism-based ontology development as the general methodology to inform knowledge representation. Their approach is based on experience with multiple implementations in biomedical informatics, which have influenced the evolution of their methodology from the beginning [8-11]. Following this approach, we use realism-based ontologies to represent the Business View. Brochhausen and Blobel [12] have outlined a strategy to use the Generical Component Model (GCM) to assist the representation of relations in realism-based biomedical ontologies (Figure 2).

RDF [13] is a Semantic Web Standard that allows the representation of information in a machine-interpretable way. For each entity, RDF provides a unique identifier. RDF data can be annotated and used along domain descriptions provided in the Web Ontology Language [14]. We are using RDF representations and OWL ontologies following the principles described by Smith and Ceusters [7].

3. Results

Figure 2 shows the result of using an RDF representation of the data in our leading example. This representation allows for the representation of multiple entry numbers referring to specimens from a single donor. The different classes are modeled based on best practice of realism-based ontology developed as described in [7].

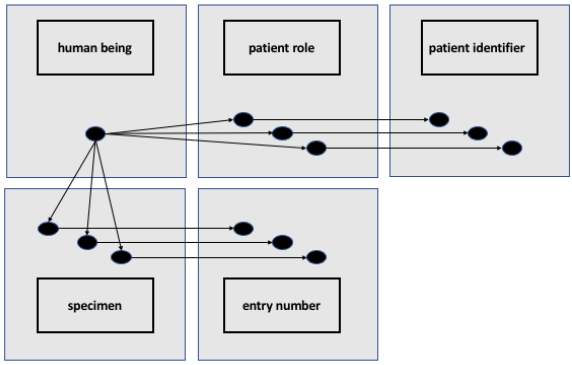


Figure 2. Representation of RDF individuals and OWL classes representing a human being having multiple patient roles, corresponding to multiple patient IDs and multiple specimens derived from that human being corresponding to multiple entry numbers.

4. Discussion

Our example shows that an information model restricting each patient to one and only one patient identifier is not sufficient to fulfill the GCM Information View, which provides the semantics. From a real world perspective, we do know that one person may be identified by multiple patient IDs throughout their life. The use of those patient IDs might overlap temporally, if the person sees multiple healthcare providers over the same period of time. In addition to the information model, a realism-based representation is needed that specifies that a person can be a patient at multiple healthcare providers and may have multiple patient IDs associated with them.

In the current stage, we have applied the GCM and the basic tenets of realism-based ontology development to one use-case. The use-case presented is an extremely common use-case in managing specimens from multiple organizations or biobanks. In addition, the successful deployment of ISO 23903 for integrating different domains and knowledge spaces including their specific models has been demonstrated for the integration of HL7 privacy and security specifications [15] in ISO 13606 EHR communication [16], the harmonization of concepts from ISO 12967 [17] and ISO 13940 [18] or the mapping of open EHR (ISO 13606) archetypes [19], ISO 13972 clinical models [20] and HL7 FHIR resources [21].

The next step is to develop computational ways to leverage the GCM system analysis and its results immediately in developing information models, database schemata, and ontologies. In addition, we plan to explore using the GCM to transition from information models to ontologies and *vice versa*.

Traditionally, technical standards have been focused on a specific domain or technology, while healthcare is multi-disciplinary by nature and therefore in need of the GCM multi-domain architecture, impacting many standards from HL7 [22] or ISO/TC215 [23]. With the turn to more complex approaches such as IoT or Smart Cities, etc., also other domains and their standards developing organizations such as OMG [24], IEEE 70xx projects [25] or ISO/IEC JTC1 [26] adopt ISO 23903.

5. Conclusion

Based on these preliminary results, we conclude that the GMC can inform development of information models and knowledge representation resources, thus, filling a relevant gap in AI usage for biomedical data.

References

[1] Blobel B, Ruotsalainen P, Oemig F. Why Interoperability at Data Level Is Not Sufficient for Enabling pHealth? Stud Health Technol Inform. 2020;273:3–19.
[2] Oemig F, Blobel B. (2010) Harmonizing the Semantics of Technical Terms by the Generic Component Model. Stud Health Technol Inform. 2010;155:115–121.
[3] Brochhausen M, Bona J, Blobel B. The role of axiomatically-rich ontologies in transforming medical data to knowledge. Stud Health Technol Inform. 2018;249:38–49.
[4] Zhang Y-F, Tian Y, Zhou T-S, Araki K, Li J-S. Integrating HL7 RIM and ontology for unified knowledge and data representation in clinical decision support systems. Comput Methods Programs Biomed. 2016 Jan 1;123:94–108.
[5] International Organization for Standardization. ISO 23903:2021 Health informatics - Interoperability and integration reference architecture - Model and framework (retrieved from: https://www.iso.org/standard/77337.html)
[6] International Organization for Standardization. ISO/IEC 10746 Information technology - Open distributed processing.
[7] Smith B, Ceusters W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. Appl Ontol. 2010 Nov 15;5(3–4):139–88.
[8] Hogan WR, Hanna J, Joseph E, Brochhausen M. Towards a Consistent and Scientifically Accurate Drug Ontology. CEUR Workshop Proc. 2013;1060:68–73.
[9] Ceusters WM, Spackman KA, Smith B. Would SNOMED CT benefit from Realism-Based Ontology Evolution? AMIA Annu Symp Proc. 2007;2007:105–9.
[10] Lin Y, Xiang Z, He Y. Brucellosis Ontology (IDOBRU) as an extension of the Infectious Disease Ontology. J Biomed Semant. 2011 Oct 31;2:9.
[11] Grenon P, Smith B, Goldberg L. Biodynamic ontology: applying BFO in the biomedical domain. Stud Health Technol Inform. 2004;102:20–38.
[12] Brochhausen M, Blobel B. Architectural approach for providing relations in biomedical terminologies and ontologies. Stud Health Technol Inform. 2011;169:739–43.
[13] Resource Description Framework (RDF) Primer, https://www.w3.org/TR/rdf11-concepts/ (retrieved 09/05/2021).
[14] Web Ontology Language (OWL) Primer, https://www.w3.org/TR/owl2-primer/ (retrieved 09/05/2021).
[15] HL7 International. HL7 Version 3 Domain Analysis Model: Composite Security and Privacy, Release 1
[16] International Organisation for Standardisation. ISO 13606:2019 Health informatics – Electronic health record communication. ISO: Geneva; 1998.
[17] International Organisation for Standardisation. ISO 12967:2020 Health informatics – Service architecture (HISA). ISO: Geneva; 2020.
[18] International Organisation for Standardisation. ISO 13940:2015 Health informatics – System of concepts to support continuity of care. ISO: Geneva; 2015.
[19] openEHR Foundation. www.openEHR.org (last accessed on 10 December 2015).
[20] International Organisation for Standardisation. ISO 13972:2015 Health informatics - Detailed clinical models – Characteristics and processes. ISO: Geneva; 2015. (Currently under revision to ISO 13972:2020 Health informatics - Clinical information models - Characteristics, structure and requirements.)
[21] HL7 International. HL7 Fast Healthcare Interoperability Resources (FHIR). http://www.hl7.org/fhir/
[22] HL7 International. http://www.hl7.org
[23] International Organization for Standardization. TC 215 Health informatics. https://www.iso.org/committee/54960.html
[24] Object Management Group. *https://www.omg.org*
[25] IEEE 70xx Standards on Ethics in Autonomous and Intelligent systems Standards. https://ethicsinaction.ieee.org/p7000/
[26] ISO/IEC JTC1 Information Technology. https://www.iso.org/isoiec-jtc-1.html

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210591

Machine Learning Based Metagenomic Prediction of Inflammatory Bowel Disease

Andrea MIHAJLOVIĆ [a,1] Katarina MLADENOVIĆ[b],
Tatjana LONČAR-TURUKALO[b] and Sanja BRDAR[a]
[a] *BioSense Institute, University of Novi Sad, Novi Sad, Serbia*
[b] *Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia*

Abstract. In this study, we investigate faecal microbiota composition, in an attempt to evaluate performance of classification algorithms in identifying Inflammatory Bowel Disease (IBD) and its two types: Crohn's disease (CD) and ulcerative colitis (UC). From many investigated algorithms, a random forest (RF) classifier was selected for detailed evaluation in three-class (CD versus UC versus nonIBD) classification task and two binary (nonIBD versus IBD and CD versus UC) classification tasks. We dealt with class imbalance, performed extensive parameter search, dimensionality reduction and two-level classification. In three-class classification, our best model reaches F1 score of 91% in average, which confirms the strong connection of IBD and gastrointestinal microbiome. Among most important features in three-class classification are species Staphylococcus hominis, Porphyromonas endodontalis, Slackia piriformis and genus Bacteroidetes.

Keywords. microbiome, imbalance, machine learning, feature selection

1. Introduction

Inflammatory bowel disease (IBD) is an umbrella term used to describe chronic inflammation of digestive tract. It includes two types of disease: Crohn's disease (CD) and ulcerative colitis (UC). They share many common features – diarrhea, bloody stools, weight loss, abdominal pain, fever, and fatigue, even though they affect different part of digestive tract. The exact cause of IBD is unknown, but some risk factors are known. It is a genetic disease, manifested under certain external influences. Research findings imply that microbiome has a fundamental role in patients with IBD, in all aspects: the development, progression, and treatment [1,2]. Machine learning (ML) algorithms applied on microbiome data have huge potential in uncovering patterns and aiding diagnosis of diseases including IBD. Early diagnosis is crucial in helping patients particularly in cases of diseases which are caused by microbial and environmental factors, since prevention in that case could be more efficient.

In this study we investigate faecal microbiota composition, in an attempt to evaluate performance of classification algorithms in identifying IBD state. The aim is to predict patients state based on their metagenomic taxonomic profile in different time points. Since the IBD is a genetic disease, it is considered that the state remains unchanged during time. There are three possible states: CD, UC and nonIBD. Two of them (CD and

[1] Corresponding Author: Andrea Mihajlović, BioSense Institute, Dr Zorana Đinđića 16, 21000 Novi Sad, Serbia; E-mail: andrea.mihajlovic@biosense.rs.

UC) represent unhealthy individuals or IBD. We perform three-class classification (CD versus UC versus nonIBD) to distinguish between the states. In addition, two binary classification tasks (nonIBD versus IBD and CD versus UC) are examined. Our approach differs in using only metagenomic data for prediction unlike other studies on the same database [3]. More on related work can be found in [4].

2. Data

Resource for IBD data used in this paper is the Inflammatory Bowel Disease Multi'omics Database (IBDMDB, https://ibdmdb.org), part of the Integrative Human Microbiome Project [5]. For the classification task we use table of metagenomic taxonomic profiles and metadata, specifically sample and subject identifiers and true labels indicating clinical diagnosis. Dataset contains 429 samples from 27 healthy subjects, 750 samples from 65 CD subjects and 459 samples from 38 UC subjects. In total 1209 samples from 103 IBD subjects.

Microbes are grouped into 1479 operational taxonomic units (OTUs). Grouping of microorganisms is done in the following taxonomic levels: kingdom, phylum, class, order, family, genus, species and strains. Detailed description of the workflow used for producing the OTU table from the raw DNA sequences is available on the bioBakery 2.0 GitHub repository https://github.com/biobakery/hmp2_workflows. The OTU table is transposed such that each row represents one sample and each column one feature, i.e. OTU. Values in the OTU table are given as relative abundances of particular OTU in some sample, relative to all reads from that sample assigned at the same taxonomic level. Sum by level in each sample is approximately 1 (due to the rounding errors).

3. Methods

3.1. Workflow

The adopted approach consists of three main steps: (i) choosing of suitable ML model and forming of model pipeline; (ii) hyperparameters search; and (iii) training and evaluation of the best model. Further, mentioned steps are explained in details. The proposed workflow is presented in the Figure 1. It is worth noting that in the subsequent steps the care was taken with respect to which sample belongs to which subject along with its diagnosis. The model used is described in Section 3.2. Parameter searching was conducted using a group cross-validation approach. In the given problem, group cross-validation is used in order to ensure that all samples from one subject are either in the training set or in the validation set. Initial set of parameters was created at random, in order to narrow parameters searching space. Best performing parameters from the initial random search were used as a guidance for the more thorough grid search and thus further fine-tuned. Training was performed for each model in 100 iterations with 10-fold group cross-validation. This ensures insight into stability of performance and more comprehensive evaluation.

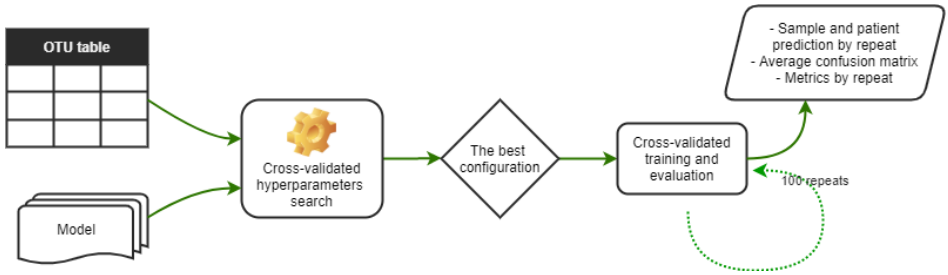


Figure 1. The workflow overview.

Upon sample-wise binary classification, subjects are labelled as positive (in one case IBD, in other UC) if average decision probability of their samples that model outputted is above the certain threshold. The threshold was as well treated as a parameter and varied. In three-class classification, the decision becomes somewhat complicated. For that purpose, we employ two thresholds and tune them. Firstly, the average decision probability for each class is estimated by the model used. By summing two probabilities for IBD classes (CD and UC) and comparing with average probability for nonIBD class, we decide if the subject is an IBD (above *th1*) or not. Furthermore, if it is an IBD, we decide which type of disease subject might have by normalizing average probabilities for CD and UC on the IBD event and decide if the subject is UC or not (above *th2*).

3.2. Model

Data imbalance have an impact on the classification since some events becomes so rare that it is impossible for classifier to learn useful patterns about them. One possibility is to use class weighting if applicable with the learning algorithm. Another solution are widely adopted resampling techniques, which assume either down- and over- sampling. Downsampling reduces the number of samples in the majority class to balance the classes, while over-sampling increases the number of samples in the minority class. Apart from the random sampling with replacement, there is a popular method to over-sample minority class(es) *the Synthetic Minority Oversampling Technique* (SMOTE) [6].

From many investigated algorithms, a Random Forest (RF) classifier [7] is selected for detailed evaluation in our classification tasks. RF randomizes decision tree through the features/samples sub-sampling (bootstrap) and groups trees to make a final decision on the basis of majority voting / averaging. In learning extremely imbalanced data, there is a significant probability that a bootstrap sample contains few or even none of the minority class, resulting in a tree with poor performance for predicting the minority class [8]. To overcome this issue, Chen et al. [8] proposed algorithm for balancing downsampled data in a bootstrap process called *Balanced Random Forest* (BRF). Using a random selection of features to split each node (tree growing), in both RF and BRF, each tree, gives an internal estimate which leads to feature importance [9]. This property of (B)RF will be very useful for microbiome analysis after classification since it can tell us which features most helped in making decision.

In the dataset used classes are highly imbalanced, with IBD samples making up 74% of the dataset and almost two times more CD samples than UC. To approach class imbalance problem, three model pipelines were evaluated and compared: (1) class weighting of RF; (2) concatenation of SMOTE and RF; and (3) BRF. The feature selection was used in all scenarios to reduce data dimensionality. Experiments included

hand-picked taxa and/or selecting k best scored features (SKB), an (*Univariate Feature Selection* method based on *ANOVA* F-test [10]).

The workflow is implemented in Python 3.0. For the ML algorithms *scikit-learn* 0.24.2 library [11] is used in addition with *imbalanced-learn* 0.8.0 library [12] for the resampling and BRF.

4. Results

In order to find best suitable hyperparameters, the parameters search was repeated several times. It was noticed that the best parameters' vicinity is similar for the same model type between the classification tasks. Also, change in classification performance as a function of the employed thresholds was noted. Table 1 contains best parameters configuration for each of the pipelines. Results of three-class classification is shown in the Figure 2.

Table 1. Parameters configuration for the three pipelines in three-class classification

Parameters	SKB_BRF	SMOTE_SKB_RF	SKB_RF
RF max depth	15	3	3
RF # of estimators	150	200	200
RF class weight	None	None	Balanced subsample
SKB *k*	500	300	300
*th*1	0.45	0.6	0.55
*th*2	0.5	0.53	0.5

a) b)

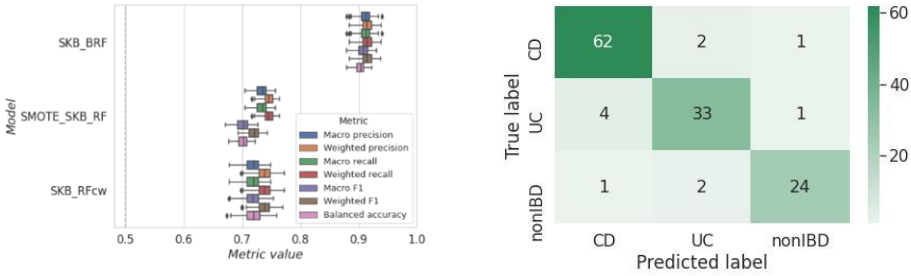


Figure 2. Three-class classification results: (a) metric values in 100 repeats and (b) confusion matrix on patient level for the best model (SKB_BRF).

Each model performance metric can be seen on the left (Figure 2a) and average number (in repeats) of well classified and misclassified patients for the pipeline containing BRF classifier on the right (Figure 2b). For the best evaluated model (SKB_BRF) all metrics are more or less similar (balanced accuracy is slightly worse in average) with an average score 91%. Other two model pipelines achieved significantly worse results with the weighted metrics slightly better than unweighted (macro).

Additionally, importance of each feature for correct classification of the best performing model was calculated. The *Bacteroidetes* genus has shown to be the most important. The rest of the features are mainly on species taxonomic level, but also some strains show and two orders *Coriobacteriales* and *Lactobacillales*. Top 20 selected

features and their importance are presented in the Figure 3. Feature order is constant in repeats. Since OTU name is too large, we show only last taxonomic level of particular OTU. The first letter in the feature name indicates taxonomic level: o - order; g - genus; s - species and t - strain.

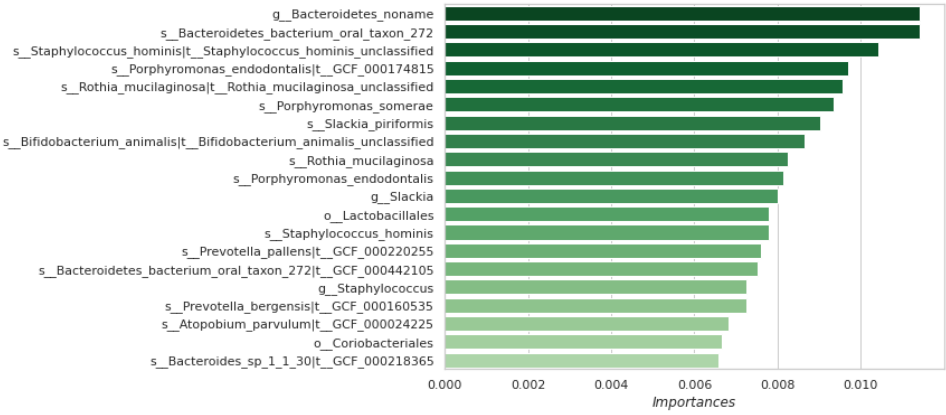


Figure 3. Feature importance for the best model in three-class classification.

In binary classification tasks, SKB_BRF model pipeline again significantly outperformed the rest. In CD versus UC case after 100 repeats, both average AUC and balanced accuracy reached 90% and average F1 score (unweighted and weighted) was 91%. In nonIBD versus IBD case, these values were 92%, 91% (unweighted) and 94% (weighted), respectively. Confusion matrices are in the Figure 4.

a) b)

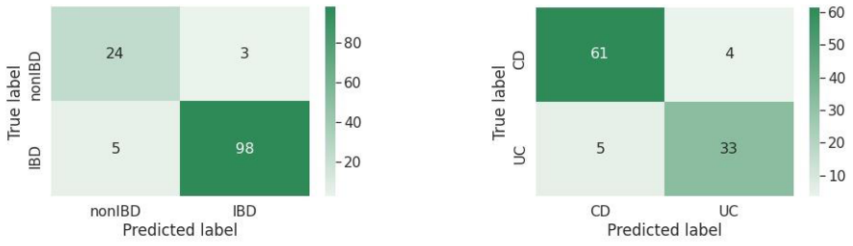


Figure 4. Confusion matrices for the best model (SKB_BRF) on patient level: (a) nonIBD versus IBD case and (b) CD versus UC case.

Based on feature importance, the top 20 features differ slightly in these two cases. In nonIBD versus IBD case, the top features comprise mainly *Alistipes* and *Bacteroides* genus, with *A. putredinisi* as most important, and *A. shahii* and *B. ovatus* among most common species. In CD versus UC case, the highlighted features are Clostridium, *Odoribacter*, *Dorea* and *Alistipes* genus, with species *C. clostridioforme*, *D. formicigenerans*, and strain *O. splanchnicus* GCF_000190535.

5. Conclusion and discussion

We addressed the classification problem in IBD dataset, taking in consideration the associated problems: class imbalance, number of samples per subject imbalance and an overall lack of data, in an attempt to avoid overfitting, a usual pitfall in applying ML algorithms. Moreover, we managed to reduce considerably large hypeparameters searching space emerged in dealing with all these problems. Among the evaluated classifiers, Balanced Random Forest showed the best performance and achieved the balance among different metrics. An additional advantage of the model used is the information on feature importance. However, to investigate the level of the OTUs (features) presence in different subject groups further analysis on a dataset is needed including the domain experts. The development of standardized ML pipeline would benefit from more data and an enhanced model explainability.

6. Acknowledgement

The financial support of the Ministry of Education, Science and Technological Development of the Republic of Serbia is acknowledged (Grant No. 451-03-9/2021-14/200358).

References

[1] Glassner K.L, Abraham B.P, Quigley E.M.M. The microbiome and inflammatory bowel disease. *J Allergy Clin Immunol*. 2020;145,1:16-27.
[2] Azimi T, Nasiri M.J, Chirani A.S, Pouriran R, Dabiri H. The role of bacteria in the inflammatory bowel disease development: a narrative review. *APMIS*. 2018;126,4:275-283.
[3] Hassouneh S.A.D, Loftus M, Yooseph S. Linking Inflammatory Bowel Disease Symptoms to Changes in the Gut Microbiome Structure and Function. *Frontiers in Microbiology*. 2021;12.
[4] Marcos-Zambrano L. J, Karaduzovic-Hadziabdic K, Loncar Turukalo T, Przymus P, Trajkovik V, Aasmets O, Magali B, Gruca A, Hasic J, Hron K, Klammsteiner T, Kolev M, Lahti L, Lopes M.B, Moreno V, Naskinova I, Org E, Pacienca I, Papoutsoglou G, Shigdel R, Stres B, Vilne B, Yousef M, Zdravevski E, Tsamardinos I, de Santa Pau E.C, Claesson M, Moreno-Indias I, Truu J. Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment. *Frontiers in Microbiology*. 2021;12,313.
[5] Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. *Cell Host Microbe*. 2014;16,3:276-89.
[6] Chawla, N, Bowyer, K, Hall, L, Kegelmeyer, W.P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res*. 2002;16: 321-357.
[7] Breiman L. Random Forests. *Machine Learning*. 2001;45:5–32.
[8] Chen C, Breiman L. Using Random Forest to Learn Imbalanced Data. Berkeley: *University of California*. 2004.
[9] Huynh-Thu V.A, Saeys Y, Wehenkel L, Geurts P. Statistical interpretation of machine learning-based feature importance scores for biomarker discovery, Bioinformatics. 2012;28,13:1766–1774.
[10] Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics* (Oxford, England). 2007;23:2507-2517.
[11] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thriton B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau, D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12,85:2825-2830
[12] Lemaitre, G, Nogueira F, Aridas C.K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*. 2017;18,17:1-5.

Personal Health Records, Patient-Centered Data Management, and Cloud Services

This page intentionally left blank

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210593

Toward an Agile System: Iranian Information System for Covid-19-Affected Patients Data Collection from Iranian Hospitals

Somayeh ABEDIAN[a,b,1], Pirhossein KOLIVAND[a], Hamid Reza LORNEJAD[a]
[a] *National Emergency Medical Services Organization, Ministry of Health and Medical Education, Tehran, I.R. Iran*
b Department of Information Technology Management, Islamic Azad University, Qazvin, Iran

Abstract. COVID-19's rapid spreads has caused a global pandemic. On 19th February 2020, Iran reported its first confirmed cases of infections in Qom City and the number of diagnosed cases and the death toll rose exponentially in March [1-3]. Managing the disease, which is considered a pandemic according to the World Health Organization (WHO) [4], requires definite approaches differing according to various factors in each country, which may also lead to (in)effective dealing with the disease. In addition, using international data and information, and WHO advice, especially in the crisis and therapeutic procedures, is one of the best crisis management strategies [5]. For every plan by governances, the first step is collecting information on epidemic distribution for the purpose of isolating provinces and cities at a national scale. Thus, Ministry of Health and Medical Education of Iran (MOHME) attempted to collect the minimum required data on the infection-affected patients based on medical records and epidemiological factors, such as demographic data (gender, age and national code), exposure history (close contact with the infected, suspect patients or even having traveled) and signs and symptoms (fever, cough, shortness or difficulties in breathing, fatigue, anorexia, hemoptysis, sputum production, dyspnea, Myalgia, Pharyngalgia, nausea, vomiting, Diarrhea, Headache, Abdominal pain, Dizziness, etc.). Therefore, to ensure accuracy and validity, and to speed up data collection in an area, Information Technology (IT) tools were required [6]. In this regard, developing an information system with a simple format and user-friendly interface in the shortest possible time was the aim. This study presents the local information system developed in March 2020, which has been registering hospitalized Covide-19-affected patients in Iranian hospitals up till now. In other words, this paper introduces features and procedures of one of the national systems as a health registry that includes clinical information on admitted Covid-19 patients in Iranian hospitals from admission to discharge or death. This system is supported by MOHME, and along with outpatient Point of Care Information Systems (POCS), feeds the national and international pandemic reports and decisions.

Keywords. COVID-19; crisis management; information technology, Health Information System (HIS), Case Report Form (CRF)

[1]Corresponding Author: Somayeh Abedian, Ministry of Health and Medical Education, Tehran, Iran, E-mail: s.abedian@behdasht.gov.ir, so.abedian@gmail.com

1. Introduction

In February 2020, some countries were already fighting COVID-19, and after China, the Islamic Republic of Iran was the second country to declare two deaths due to coronavirus, within 50 days after China on February 18th, 2020. It is surmised that the virus may have been transmitted to the country by a merchant from Qom who had travelled to China [1].

During the COVID-19 outbreak, more than in any other public health conditions, Health Information Systems (HIS) perform a critical role in information management [7]. The policies are made according to crucial evidence by health professionals in taking the post appropriate action, making the most informed decisions to allow for better insights on measures with the aim of improving citizens' healthcare. Emerging automation and information have the potential to improve public health. HISs arrange for immediate, expeditious, and coordinated data access and sharing, and they facilitate the care priority and access, especially for the people in vulnerable conditions. With properly disaggregated data, it is possible to plan activities that reduce probable health inequities at all levels of care, and facilitate the implementation of strategies to address such cases of injustice. In addition, they provide reliable statistical information for planning and healthcare with respect to human resources allocation, and medicine and medical equipment assignment [8-10].

After the first case of the outbreak was detected in Iran, COVID-19 Disease Management Policy Committee (CDMPC) was informed at the MOHME. The committee approved to collect information of the patients suffering pneumonia with acute respiratory distress from hospitals as the main healthcare providing centers with cases of severe illness. According to the experts' opinion and solutions analysis, in seeking an Agile HIS, the use of a national web-based system was considered.

This paper presents the experience of Iran in applying a live and dynamic system for data gathering about the infection-affected (or suspicious of Covide-19) patients from hospitals. The process would include reports from the data entry at the hospitals to monitoring dashboard development for medical science universities and MOHME.

2. Methods

With COVID-19 outbreak, countries need to adopt a broad and long-term vision to improve data systems to help manage information about the pandemic. A whole-government, multi-sectoral approach is required to use the outbreak data to manage consumer supplies, hospital beds, sharing of resources among healthcare provider centers and deployment of healthcare workers [11]. This article aims to introduce the live system that is utilized in Iran for data gathering from hospitals. It focuses on the patients referred to a hospital, suspected of an acute respiratory condition. The system's role is similar to a health registry for decision making, while facing the effects of COVID-19, in terms of prevention and control, based on the collected and processed data, as well as the received and generated information. This national project benefited from the professional advice of public health professionals, IT experts and CDMPC members. Based on committee statements in the limited time for crisis management, use or development of a national integrated system require some phases and steps throughout the process as following:

2.1. Data Requirements and Modeling

Information systems have played a crucial role in reacting to the outbreak of COVID-19, not only they have served as platforms for efficient and effective communications but they have also essentially changed the dynamics of interactions among healthcare providers and patients, and the healthcare services delivery methods [12-13]. This should be based on commonly agreed-upon standardized data and fit-for-purpose indicators in the use of HISs infrastructures. In this project, the primary phase was developing optimal clinical and non-clinical data elements identification and Case Report Form (CRF) design since there was not any template announced by WHO at the start of the pandemic. Thus, CDMPC members discussed the main goals of CFR count and related items, which led to three finalized CFR categories.

1. Hospitalized pneumonia report form containing admission and demographics (date-time of admission, date of initial symptoms prior to diagnosis, full name, nationality, national code, age, gender, weight, travel history, level of education, occupation status, etc.), Past Medical History (PMH: pre-existing health conditions such as diseases or pregnancy, smoking, addiction or substance abuse, medications, etc.), clinical services, examinations (signs and symptoms), any type of Covid-19 test results (such as PCR test, lg test and CT scan), re-infection information, etc.
2. Discharge report form containing date of discharge, conditions on discharge, prescribed medical orders, etc.
3. Death report form containing time of death, infection intensity (Mild, Moderate, Severe or Critical), Covide-19 virus identification or suspicion, and cause of death based on ICD10.

Indeed, after CRF's first version was published to be confirmed by WHO, it was revised [14].

2.2. Software Development Process

In the first step, using hospital information systems were recommended, but this idea was rejected because of the variety in the types of products (more than 30 software programs with different technology) and challenges in their data integration. The second reachable solution was discussed around the utilization of national Electronic Health Record (EHR). Iranian EHR receives data from all POCSs based on WSDL standards and essential components of SOA architecture [15], but the challenge of time of data transferring from hospitals still persisted since the data were provided after patients' discharge (inpatient clinical items) and MOHME needs them as early as the pandemic is picking up. Therefore, this solution was rejected too (for this goal, development of new online web services was both necessary and time-consuming and saving time was very important to us at the beginning of the pandemic).

Finally, according to experts' experience, using a national web-based system developed earlier in order to manage patient transferring process among hospitals was selected due to the following reasons:

1. The system's technical infrastructure was ready and development and launch time would be saved.
2. The system was form-based and designing new forms was possible.
3. Hospital users were healthcare providers already familiar with the system (more than 1200 hospitals). They are trained to work with the system.

4. Usernames and passwords are defined for users.
5. There are supervisory units in the universities of medical sciences to monitor the accuracy of data entry in electronic forms, which assures MOHME of evidence accuracy and validity.
6. The system was integrated on a national scale and did not require software or new web-service development.

According to above-mentioned points, CRFs were developed in the system and MOHME required the hospitals to fill them twice a day (12 a.m. and 2 p.m.).

2.3. Supervision and Revision

After one week, the results of CFRs fillings were evaluated. Based on the results of supervisory report on more than 1200 hospitals (governmental, private, charity, military and other), CDMPC members provided some recommendations, and consequently the system was updated for the purpose of better data analysis.

2.4. Regional/National Dashboard Development

MOHME has developed one multi-resource monitoring dashboard. It shows current and reliable data on COVID-19 cases submitted directly from healthcare provider centers and is used by decision makers at regional and national access level. In addition, the dashboard has a clean and modern interface with several data visualization tools to better grasp the current status of COVID-19 as the situation unfolds [16]. The dashboard resources were supplied by HISs (such as this project for hospital cases, laboratory information systems, primary care information systems, mobile-based or home-care systems and clinical information systems) including information on the hospitalized/outpatient definite or suspected patients based on WHO guidelines. To this aim, MOHME released a number of substantial indicators to visualize and monitor COVID-19 outbreak dispersion seeking the following issues:

1. New and confirmed cases, recovered and deaths nationally with daily statistics.
2. City and province-specific information by clicking on any city/province/medical science university on the interactive map with intelligent alarms such as color.
3. Interactive charts showing the reported cases by regions including daily and cumulative statistics and weekly, monthly and annual trends.
4. Confirmed cases, deaths, and changes over time in a specific province, region, or territory, on interactive charts.
5. A new explorer tab designed to provide complex datasets for easy access and use. It allows health policy-makers to select variables which help spot correlations and relationships that can provide insights into COVID-19 and how communities are responding to it.

It is critically important that all healthcare centers specially hospitals be able to register their COVID-19 situations. The powered dashboard will now provide more comprehensive insights into the epidemiology and response to COVID-19 at regional and national scales.

3. Results

Health information systems are one of the six WHO health system building blocks: "A well-functioning health information system is one that ensures the production, analysis, dissemination and use of reliable and timely information on health determinants, health system performance and health status" [6]. With the COVID-19 outbreak, the most basic issue was the spreading virus volume in the country which was manageable just by means of an electronic information system to meet reliable and timely evidence.
In this paper, an agile information system development for COVID-19 outbreak monitoring in Iranian hospitals is introduced. It depicts a multi-step process from planning, CRF design and case registration to monitoring dashboard development which is shown in Figure 1.

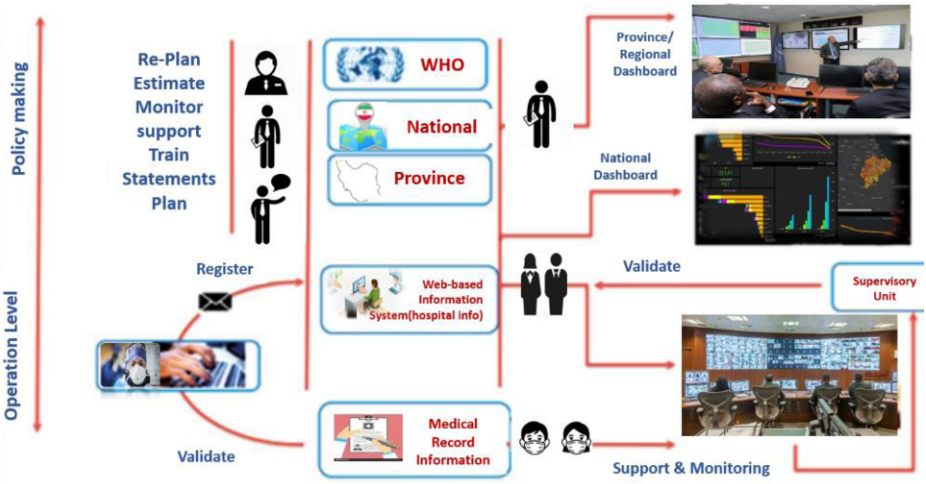


Figure 1. The COVID-19 data collection process from Iranian hospitals

4. Discussion

A national information system is designed based on professional advice and emerging requirements. In this regard, a committee consisting of all relevant experts was formed as early as the first coronavirus case declaration in Iran. In this process, supported by the Iranian Ministry of Health and Medical Education, attempts were made to improve the reliability and validity of COVID-19-affected patients' information and statistics.

This paper seeks to illustrate a nation-wide effort to provide and manage valid information on the COVID-19 pandemic for public health planning in the country. The government is using this data repository beside other clinical data resources in order to improve the quality of health services acceptability, continuity/sustainability, clinical effectiveness in a minimum time and prevent further spreading of the virus. Moreover, national reports shown in the dashboard feed the national and international reports.

Acknowledgment

The authors express their gratitude to Dr. Saeid Namaki (Minister of Health and Medical Education) for his supports and all efforts throughout COVID-19 pandemic in Iran. We would also like to appreciate our experts for their advices especially Dr. Mohammad Hussein Norouzi.

References

[1] York Times. "How Iran Became a New Epicenter of the Coronavirus Outbreak". New Archived from the original on 27 June 2021. Available from: https://www.newyorker.com/news/our-columnists/how-iran-became-a-new-epicenter-of-the-coronavirus-outbreak

[2] Iranian Ministry of Health and Medical Education. Coronavirus epidemic in Iran. Available from: http://ird.behdasht.gov.ir/. Accessed August 3, 2020.

[3] Venkatesan P. COVID-19 in Iran: round 2. Lancet Infect Dis. 2020 Jul;20(7):784. doi: 10.1016/S1473-3099(20)30500-4. PMID: 32592672; PMCID: PMC7314444.

[4] Prevention UCfDCa. *Coronavirus Disease 2019 (Covid-19)*. (2020). Available online at: https://www.cdc.gov/coronavirus/2019-nCoV/summary.html

[5] Rassouli Maryam, Ashrafizadeh Hadis, Shirinabadi Farahani Azam, Akbari Mohammad Esmaeil, "COVID-19 Management in Iran as One of the Most Affected Countries in the World: Advantages and Weaknesses", JOURNAL: "Frontiers in Public Health", VOLUME:8, 2020, Available from: https://www.frontiersin.org/article/10.3389/fpubh.2020.00510, DOI: 10.3389/fpubh.2020.00510, ISSN:2296-2565

[6] Al-Shorbaji N. E-health in the Eastern Mediterranean Region: a decade of challenges and achievements. East Mediterr Health J. 2008;14 Suppl: S157-73. PMID: 19205616.

[7] Pan American Health Organization," COVID-19 and the importance of strengthening Information Systems, PAHO/EIH/IS/COVID-19/20-0021, 2020-05-12, Available at: https://iris.paho.org/handle/10665.2/52127

[8] WHO Regional Office for Africa 2020, Monitoring and evaluation framework for the Covid-19 response in the WHO African Region. Available from: https://apps.who.int/iris/bitstream/handle/10665/334344/9789290234432-eng.pdf

[9] World Health Organization. Everybody's business: strengthening health systems to improve health outcomes: WHO's framework for action. Production [Internet]. 2007;1–56. Available from: http://www.who.int/healthsystems/strategy/everybodys_business.pdf

[10] Saxena N, Gupta P, Raman R, Rathore AS. Role of data science in managing COVID-19 pandemic. Indian Chemical Engineer, 2020; 62(4): 385-395. DOI: 10.1080/00194506.2020.1855085 To link to this article: https://doi.org/10.1080/00194506.2020.1855085

[11] Ros F, Kush R, Friedman C, Gil Zorzo E, Rivero Corte P, Rubin JC, et al. Addressing the Covid-19 pandemic and future public health challenges through global collaboration and a data-driven systems approach. Learn Health Syst. 2020 Dec 6;5(1):e10253. doi: 10.1002/lrh2.10253

[12] Mykytyn PP. COVID-19 and Its Impacts on Managing Information Systems, Information Systems Management. 2020;37(4):267-271. DOI: 10.1080/10580530.2020.1818900

[13] Duckett S. What should primary care look like after the COVID-19 pandemic? Australian Journal of Primary Health 2020;26(3):207-211. https://doi.org/10.1071/PY20095

[14] World Health Organization. Revised case report form for confirmed novel coronavirus COVID-19 (report to WHO within 48 hours of case identification), 27 February 2020. Available from: https://apps.who.int/iris/handle/10665/331234?show=full

[15] Abedian S, Kazemi H, Riazi H, Bitaraf E. Cross hospital bed management system. Stud Health Technol Inform. 2014; 205:126-130. PMID: 25160159.

[16] Irwansyah E, Budiharto W, Widhyatmoko D, Istamar A, Putra Panghurian F. Monitoring Coronavirus COVID-19/SARS-CoV-2 Pandemic using GIS Dashboard: International and Indonesia Context. Preprints 2020, 2020080415 (doi: 10.20944/preprints202008. 0415.v1).

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210594

Prerequisites of Personal Health Record for Chronic Kidney Disease: A Scoping Review and Evaluation of the Content Validity

Fatemeh SALEHI[a], Peivand BASTANI[a], Leila AHMADIAN[b], Katayoon SAMADI[c], Azita YAZDANI[a], Roxana SHARIFIAN[a1]
[a]*Health Human Resources Research Center, School of Management & Information Sciences, Shiraz University of Medical Sciences, Shiraz, Iran*
[b]*Department of Health Information Sciences, Faculty of Management and Medical Information Sciences, Kerman University of Medical Sciences, Kerman, Iran*
[c]*Kidney Transplantation Complication Research Center, Mashhad University of Medical Sciences, Mashhad, Iran*

Abstract. Background: It is obvious that the Personal Health Record (PHR) is a major cornerstone for "improving the self-management of patient". However, lack of an effective and comprehensive personal health record system prohibits the widespread use of PHRs. The aim of this study was to identify the core data sets and required functionalities for designing a PHRs for chronic kidney disease (CKD) management and assess their validity. **Methods:** It was a study including two phases. In the initial phase, a scoping review was conducted with the aim of determination the core data sets and required functionalities for designing PHRs. Then in the second phase, the validity of data items and functionalities was determined by 25 multidisciplinary experts. **Results:** 22 studies were eligible after screening 1335 titles and abstracts and reviewing 88 full texts. We determined 20 core data set and 8 required functionalities of PHRs. From the perspective of experts, 'health maintenance' and 'advance directives' were most often marked as useful but not essential, while 'test and examination', 'medication list' and 'diagnosis and comorbid conditions" were predominantly considered as essential by all experts (n= 25,100%). **Conclusion:** This research is a step that we have taken to identify pre-requisites that could be used for the design, development, and implementation of an effective and comprehensive electronic personal health record.

Keywords. Chronic Kidney Disease, Personal Health Record, PHR, CKD, Core data sets

1. Introduction

Significant social and economic burdens of chronic diseases have led to a shift in the health policy, involving a focus on health promotion, chronic disease prevention, and self-management [1]. Chronic kidney disease (CKD) is a major public health concern [2-

[1] Corresponding Author: Roxana Sharifian, Associate Professor, School of Management & Information Sciences, Shiraz University of Medical Sciences, Shiraz, Iran. E-mail: sharifianr@sums.ac.ir

4]. More than 70 million individuals worldwide have CKD, and according to estimates, the prevalence will further increase as will the already enormous impact CKD has on health system resources related to its care [5,6]. Health information technologies (ITs) have the potential to significantly increase the engagement of patients by using personal health records (PHRs) to electronically connect them to their health information and clinical team and continuity of care [7]. PHR is a tool that has the potential to change and possibly to improve patient–provider relationship and enable the healthcare system to evolve a more personalized medical model and promising results to address some of these challenges [8-10]. Unfortunately, designing and developing programs that improve patient care and obtaining complete and high-quality data in nephrology have remained a challenge [11]. The literature does not yet adequately describe the potential functions and prerequisites of PHRs design [12-15]. So, the objectives of these study were to derive core data sets and functionalities specifically for PHRs for patients diagnosed with CKD and determine the validity of these core data sets and functionalities.

2. Method

This review was guided by Arksey and O'Malley's 6-stage scoping review framework [16]. We searched for relevant articles written in English between 1990 and Jan 2021 using PubMed, Science Direct, Web of Science and Embase databases, and the related websites such as guideline.gov, IEEE, and WHO. A combination of keywords and Medical Subject Headings (MeSH) were used as follows: group A included PHR-related terms and group B included terms related to "kidney failure chronic." We considered all the full text papers with quantitative, qualitative, or mix method designs and full reports that studied PHRs and determined the data elements and functionalities of chronic kidney disease PHR. However, the papers in the formats of letter to editor, short communication and commentary, and articles in non-English language or those with English abstracts published in languages other than English were excluded. Additionally, if the study was about a personal electronic record but had not the state data items and functionalities, it was excluded from the study. After searching the studies from all databases and eliminating duplicates, the studies were independently reviewed and screened by two members of the research team (FS and RSH) in three phases by title, abstract, and then the full text of the articles. Studies meeting the inclusion criteria were critically reviewed using Arksey and O'Malley's summative analysis method [16] according to the frequency of the items in the included studies. To validate the core data sets and functionalities, we formed an expert panel. The panel consisted of 25 multidisciplinary experts that were recruited base purposive sampling. In this way, the experts are requested to specify whether the core data sets and functionalities is necessary for designing a PHR for CKD or not via email. To this end, they are requested to score each item from 1 to 3 with a three-degree range of "not necessary, useful but not essential, essential" respectively. The formula of content validity ratio is CVR= (Ne - N/2)/ (N/2), in which the Ne is the number of panelists indicating "essential" and N is the total number of panelists. The numeric value of content validity ratio is determined by Lawshe table [17]. Ethical approval was received from the Shiraz University of Medical Sciences by Dr Abbas Rezaeianzadeh, (Ethical number: IR.SUMS.REC.AC.IR.1399.1310).

3. Results

In total, 1335 studies were selected after searching the databases. After removing the duplicates, screening, and applying inclusion and exclusion criteria, 88 studies were eligible for further full-text review. Thereafter, 16 articles, 4 reports, and 2 guidelines were selected for the final analysis.

Most studies were journal article (n=16, 76%), published in the USA (n=10, 45%), and published between 2012 and 2018 (n=15, 71%) (Table1).

Table1. Description of included study

First author name & [Ref]	Resource type	Publication Date	Country/Institution
Venuthurupalli [18]	Article (Cross sectional)	2017	Australia
Navaneethan [19]	Article (Cross sectional)	2012	USA
Nakashima [10]	Article (Cross sectional)	2019	Japan
Mendu[6]	Article (Prospective study)	2014	USA
Drawz [20]	Article (RCT)	2012	USA
Mendu [21]	Article (Cross sectional)	2019	USA
Drawz [22]	Article (Review)	2015	USA
Khan [23]	Article (Cross sectional)	2013	USA
Venuthurupalli [24]	Article (Cohort)	2018	Australia
Bruland [25]	Article (Case report)	2016	Germany
Do [26]	Article (project report)	2011	USA
Archer [9]	Article (Review)	2011	Canada
Kaelber [11]	Article (Cross sectional)	2008	USA
Gearon [27]	Report	2007	California
Tran [28]	Guideline	2012	USA
Burke-Bebee [29]	Report	2010	USA
Johnston [30]	Report	2007	AMIA
Roehrs [31]	Article (Review)	2017	Brazil
Tang [32]	Article	2006	USA
Dickinson [33]	Guideline	2014	USA
Unknown [34]	Report	2012	European
Katehakis [35]	Article (Review)	2017	Greece

Table2. Core data sets for designing PHRs for chronic kidney disease based on evidence and expert panel.

Core Data Sets	Frequency (%)	Expert Panel			CVR*	Interpretation
		Essential	Useful not essential	Unnecessary		
Problem list	7(31.81%)	24(96%)	1(4%)	0	**0.92**	Remained
Surgical procedures	2(9.00%)	23(92%)	1(%4)	1(%4)	**0.84**	Remained
Diagnosis/comorbid conditions	9(40.90%)	25(100%)	0	0	**1**	Remained
Medications list	12(54.54%)	25(100%)	0	0	**1**	Remained
Risk factors & allergies	8(36.36%)	23(92%)	2(%8)	0	**0.84**	Remained
Demographics data	9 (40.90%)	24(96%)	1(%4)	0	**0.92**	Remained
Health maintenance	1(4.54%)	16(64%)	7(28%)	2(8%)	**0.28**	Eliminated
Disease characteristic	2(9.00%)	24(96%)	1(4%)	0	**0.92**	Remained
Advance directives	2(9.00%)	17(%68)	2(8%)	6(22%)	**0.36**	Eliminated
Physical examination	2(9.00%)	22(88%)	2(8%)	1(%4)	**0.76**	Remained
Wellness management	3(13.62%)	22(88%)	3(12%)	0	**0.76**	Remained
Care plan	2(9.00%)	23(92%)	1(%4)	1(%4)	**0.84**	Remained
Health summary	6 (27.27%)	23(92%)	2(%8)	0	**0.84**	Remained
Family record& history	7 (31.81%)	24(96%)	1(%4)	0	**0.92**	Remained
Genetic data	2(9.00%)	20(80%)	3(12%)	2(%8)	**0.60**	Remained
Health patterns	7(31.81%)	20(74%)	4(14%)	1(%4)	**0.60**	Remained
Test and examination	17(77.27%)	27(100%)	0	0	**1**	Remained
Functional status	2(9.00%)	21(84%)	3(12%)	1(%4)	**0.68**	Remained

Finally, 124 data items were identified from the literature that classified in 20 core data set. “Test and examination” was the most common core data set examined (n = 17) in the included studies. Other common core data item examined included medication list (n=12), “diagnosis and comorbid conditions”, “preventive care & immunization” and “demographics data” (n=9). About data sets, ‘health maintenance’, and ‘advance directives’ were most often marked as useful but not essential or unnecessary, while ‘test

and examination', 'medication list' and 'diagnosis and comorbid conditions", were predominantly considered as essential by all experts (n= 25, 100%) (Table2).

Table 3. Required functionality in designing PHRs for CKD based on evidence and expert panel.

Required functionality	Sub items	Frequency	Expert plan			CVR*
			Essential	Useful, not essential	Unnecessary	
Historical data	Manage historical clinical data	4(18.18%)	24(96%)	1(4%)	0	0.92
Management observations	Manage clinical observations	3(13.62%)	18(72%)	7(28%)	0	0.44
	Manage test results	3(13.62%)	25(100%)	0	0	1
	Manage provider care plans	2(9.00%)	23(92%)	1(4%)	1(4%)	0.84
	Manage health calendar	2(9.00%)	23(92%)	1(4%)	1(4%)	0.84
	Manage medication	9(40.90%)	25(100%)	0	0	1
Management of decision support	Manage drug interaction checking	3(13.62%)	18(72%)	7(28%)	0	0.44
	Manage guidelines and protocols	2(9.00%)	21(84%)	2(8%)	2(8%)	0.68
	Manage health alerts	3(13.62%)	23(92%)	2(8%)	0	0.84
	Manage health reminders	6(27.27%)	23(92%)	0	2(8%)	0.84
Management of patient support	Manage custom patient education	6(27.27%)	23(92%)	2(8%)	0	0.84
	Manage family education	2(9.00%)	23(92%)	2(8%)	0	0.84
	Manage data input errors	3(13.62%)	24(96%)	1(4%)	0	0.92
	Manage trading patterns	2(9.00%)	21(84%)	2(8%)	2(8%)	0.68
	Manage shared patient experience	4(18.18%)	15(60%)	1(4%)	9(36%)	0.20
	Manage results notification	2(9.00%)	23(92%)	2(8%)	0	0.84
Management of security	Manage secure the access to PHR	2(9.00%)	24(96%)	1(4%)	0	0.92
	Manage entity authentication	2(9.00%)	24(96%)	1(4%)	0	0.92
	Manage entity authorization	2(9.00%)	24(96%)	1(4%)	0	0.92
	Manage secure data exchange	2(9.00%)	24(96%)	1(4%)	0	0.92
	Manage patient privacy	4(18.18%)	24(96%)	1(4%)	0	0.92
	Manage secure messaging	3(13.62%)	24(96%)	1(4%)	0	0.92
	Manage consents and authorizations	2(9.00%)	23(92%)	1 (4%)	1(4%)	0.84
	Manage data masking for sensitive	2(9.00%)	22(88%)	3(12%)	0	0.76
	Manage a registry of actors	2(9.00%)	24(96%)	1(4%)	0	0.92
Management of administrative issues	Manage demographics information	3(13.62%)	24(96%)	1(4%)	0	0.92
	Management scheduling	7(31.81%)	24(96%)	1(4%)	0	0.92
	Manage advance care directives	2(9.00%)	17(%68)	2(8%)	6(24%)	0.36
	Manage insurance eligibility	3(13.62%)	21(84%)	1(4%)	3(12%)	0.68
	Manage clinical trial recruitment	2(9.00%)	22(88%)	0	3(12%)	0.76
	Manage multiple views of data	2(9.00%)	22 (88%)	3(11%)	0	0.76
	Manage donor information	2(9.00%)	16 (64%)	5(20%)	4(16%)	0.28
	Manage access to public health	2(9.00%)	20(80%)	4(16%)	1(4%)	0.60
	Manage clinical research	2(9.00%)	24(96%)	0	1(4%)	0.92
	Manage clinical dashboard	3(13.62%)	22(88%)	2(8%)	1(4%)	0.76
Management of electronic communication	Manage team coordination	2(9.00%)	23(92%)	2(%8)	0	0.84
	Manage of communication	9(40.90%)	25(100%)	0	0	1
	Manage contact information	5(22.72%)	23(92%)	2(%8)	0	0.84
	Manage referral authorizations	3(13.62%)	16 (64%)	6(24%)	3(12%)	0.28
Management health monitoring	Manage Home monitoring	8(36.36%)	23(92%)	2(%8)	0	0.84
	Manage wellness, preventive, life style	4(18.18%)	23(%92)	1(%4)	1(%4)	0.84

NOTE: *CVR or Content Validity Ratio = (N_e-N/2)/ (N/2) with 25 persons at the expert panel (N=25), the items with the CVR bigger than 0.37 remained at the instrument and the rest eliminated.

In terms of required functionalities, 'manage of communication, 'manage medications' and 'manage test and examination' were considered as essential by all experts (n= 25, 100%). Management of clinical research information and clinical trial recruitment were other functionalities recommended by experts. According to result of expert panel, 2 items out of core data sets items (health maintenance & advance directives) and 4 items out of functionalities (manage shared patient experience, manage advance directives, manage donor information & manage referral authorizations) were eliminated (Table 2&3).

4. Discussion

Based on the results of our study, 20 core data sets were determined. Core data items proposed by this study covered all 11 data components essential for PHRs that were prepared by consensus set of standards of CCD, CCA, CCR, AHIMA, AMIA, DICOM, immunizations, medications, allergies, family history, lab/test results, and procedures/surgeries [28,32]. The corresponding PHRs for CKD, 'advanced directives', was checked as unnecessary in most of responses by experts. Considering that advanced directives are not popular in Iran, the number was expected. These findings contrast with other countries which "advanced directives" is very important and, indeed, the legal right of the patient [28,36]. These differ may be due to cultural differences between Iran and other countries might influence the choice of key data sets/functionalities. Essential functionalities that recommended by experts in designing PHRs for CKD were consistent with the results of other on literature [6,10,11,18,19,26,28,29]. An innovative function under strong focused of experts is the "custom patient education". Health care delivery moves towards a more consumer focused, personalized care, patients and individuals' roles grow, and many potential advantages of the PHR have been portrayed [37].

5. Conclusions

We propose pre-requisites of personal health record consisting of 20 core data sets and 8 main functionalities for CKD patient. These pre-requisites could be used for designing and implementing effective and comprehensive PHRs for chronic kidney disease management.

Acknowledgment

This article is adopted from a PhD dissertation approved by Shiraz University of medical sciences, Shiraz, Iran

References

[1] Jordan JE, Osborne RH. Chronic disease self-management education programs: challenges ahead. Medical Journal of Australia 2007; 186: 84 – 87.
[2] Zhang L, Wang F, Wang L, Wang W, Liu B, Liu J, et al, Prevalence of chronic kidney disease in China: a cross-sectional survey. Lancet 2012;379:815-822.
[3] GBD 2013 Mortality and Causes of Death Collaborators: a systematic analysis for the global burden of disease study. Lancet 2015;385:117-171.
[4] Braun L, Sood V, Hogue S, Lieberman B, Copley-Merriman C, High burden and unmet patient needs in chronic kidney disease. Int J Nephrol Renovasc Dis 2012;5:151-163.
[5] Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY, Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. N Engl J Med 2004;351:1296–1305.
[6] Mendu ML, Schneider LI, Aizer AA, Singh K, Leaf DE, Lee TH, Waikar SS, Implementation of a CKD checklist for primary care providers. Clin J Am Soc Nephrol 2014;9**:**1526–1535.
[7] Wells, S et al. Personal health records for patients with chronic disease: a major opportunity. Applied clinical informatics 2014;5:416-29.
[8] Roehrs A, da Costa CA, da Rosa Righi R, de Oliveira KSF. Personal Health Records: A Systematic Literature Review. J Med Internet Res 2017;19.

[9] Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE (2011) Personal health records: a scoping review. J Am Med Inform Assoc 2011;18:515-522.
[10] Nakashima N et al. Recommended configuration for personal health records by standardized data item sets for diabetes mellitus and associated chronic diseases. J Diabetes Investig 2019;10:868-875.
[11] Kaelber DC, Jha AK, Johnston D, et al. A research agenda for personal health records (PHRs). J Am Med Inform Assoc 2008;15:729.
[12] Public Health England -Trust Development Authority. NHS England Care Quality Commission Health Education England Monitor NHS England Care Quality Commission Health Education England Monitor (2014).
[13] National Information Board. Personalised Health and Care 2020: Using Data and Technology to Transform Outcomes for Patients and Citizens.
[14] Ford EW, Hesse BW, Huerta TR, Personal Health Record Use in the United States: Forecasting Future Adoption Levels. J Med Internet Res 2016;18:e73.
[15] Tang, Paul C et al. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. Journal of the American Medical Informatics Association 2006;13:121-6.
[16] Sea1-Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8:19-32.
[17] Lawshe CH. A quantitative approach to content validity. Personnel psychology 1975;28:563–75.
[18] Venuthurupalli SK, Hoy WE, Healy HG, Cameron A, Fassett RG, CKD.QLD: establishment of a chronic kidney disease [CKD] registry in LQueensland, Australia. BMC Nephrol 2017;18:189.
[19] Navaneethan SD, Jolly SE, Sharp J, et al. Electronic health records: a new tool to combat chronic kidney disease? Clin Nephrol 2013;79(3):175-183.
[20] Drawz et al. Impact of a chronic kidney disease registry and provider education on guideline adherence – a cluster randomized controlled trial. BMC Medical Informatics and Decision Making 2012;12:62.
[21] Mendu ML, Ahmed S, Maron JK, Rao SK, et al. Development of an electronic health record-based chronic kidney disease registry to promote population health management. BMC Nephrol 2019;20:72.
[22] Drawz PE, Archdeacon P, McDonald CJ, et al. CKD as a Model for Improving Chronic Disease Care through Electronic Health Records, Clin J Am Soc Nephrol 2015;10:1488-1499.
[23] Khan IA. Personalized Electronic Health Record System for Monitoring Patients with Chronic Disease. Systems and Information Engineering Design Symposium (SIEDS) (Conference paper) IEEE (2013).
[24] Venuthurupalli SK, Rolfe A, Fanning J, Cameron A, Hoy WE, NHMRC CKD.CRE and the CKD.QLD Collaborative. Chronic Kidney Disease, Queensland (CKD.QLD) Registry: Management of CKD with Telenephrology. Kidney Int Rep 2018;3:1336-1343.
[25] Bruland et al. Common data elements for secondary use of electronic health record data for clinical trial execution and serious adverse event reporting. BMC Medical Research Methodology 2016;16:159.
[26] Do NV, Barnhill R, Heermann-Do KA, Salzman KL, Gimbel RW. The military health system's personal health record pilot with Microsoft HealthVault and Google Health. J Am Med Inform Assoc 2011;18:118-124.
[27] Georghen C, et al. Perspective of future of personal health record. California health care foundation (2007).
[28] Tran BQ, Gonzales P. Standards and Guidelines for Personal Health Records in the United States: Finding Consensus in a Rapidly Evolving and Divided Environment. J Health Med Information (2012).
[29] Evaluation of Personal Health Records Pilots for Fee-for-Service Medicare Enrollees from South Carolina. NORC at the University of Chicago (2010).
[30] The value of Value of Personal Health Records, a Joint Position Statement for Consumers of Health Care (2007) AHIMA/AMIA https://library.ahima.org/doc?oid=100380. Accessed 01 June 2021
[31] Roehrs A, da Costa CA, Righi RD, de Oliveira KS (2017) Personal Health Records: A Systematic Literature Review. J Med Internet Res 2017;19:e13.
[32] Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13:121-126.
[33] Personal Health Record System Functional Model (PHR-S FM), R1. HL7/ANSI (2014)
[34] PHR Design and Architecture. Project # FP7-ICT-StrokeBack-288692 (2012) Version 1.
[35] Katehakis DG, Kondylakis H, Koumakis L, Kouroubali A, and Marias K, Integrated Care Solutions for the Citizen: Personal Health Record Functional Models to Support Interoperability. EJBI 2017;13.
[36] Fong KC, Chiong W. Understanding Advance Directives as a Component of Advance Care Planning. The American Journal of Bioethics 2020;20:67-69.
[37] Lee M, Delaney C, Moorhead S. Building a personal health record from a nursing perspective. Int J Med Inform. 2007;76:S308-S316.

Solutions for Personalized Care

This page intentionally left blank

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210596

A User-Centered, Integrated Model to Improve Medication Prescription, Administration and Adherence in Switzerland

Katherine BLONDON[a,1] and Frederic EHRLER[b]
[a] *Medical Directorate, University Hospitals of Geneva, Switzerland*
[b] *Direction of Information System, University Hospitals of Geneva, Switzerland*

Abstract. Medication adherence remains an obstacle for the ideal medical care. Communication issues arise between care-providers, and the patient is left to deal with potentially conflicting information. The new electronic patient record (EPR) that will soon be implemented nationally opens new perspectives to improve patient medication management. In this context, we propose an integrated model that could help further empower the patient with better communication about medications and considerations for reconciliation processes. We discuss important considerations for our proposed solution.

Keywords. Care Network, Medication Management, EPR, Medication Adherence

1. Introduction

Despite the many improvements in healthcare over the past years, the patient's road from being diagnosed to receiving medical treatment and subsequent improvement remains bumpy [1]. Patients often have more than one healthcare provider, who may each prescribe medications without necessarily knowing about the patient's other prescriptions. Patients can go to any pharmacy to receive their medicine, so the pharmacist may not know about other on-going treatments [2]. Patients may even receive contradictory information about their medication from their multiple healthcare providers, leading to confusion and yet lower adherence to treatment [3]. The family doctor often receives reports from the specialists about their shared patients, allowing them to play an important role in the medication reconciliation process. Many gaps during transitions and difficulties in sharing information to collaboratively manage a patient remain present, resulting in fragmented care [4,5]. The current reconciliation process largely relies on the patient to communicate about their treatments to other healthcare providers.

Despite the implementation of electronic medical records in hospitals, clinics and many private practices in Switzerland, there is no secure, systematic communication channel between the various healthcare networks [6]. Patient portals also show potential

[1] Corresponding Author, Katherine Blondon, University Hospitals of Geneva, Gabrielle Perret Gentil, 4 1205 Geneva, Switzerland E-mail: Katherine.Blondon@hcuge.ch.

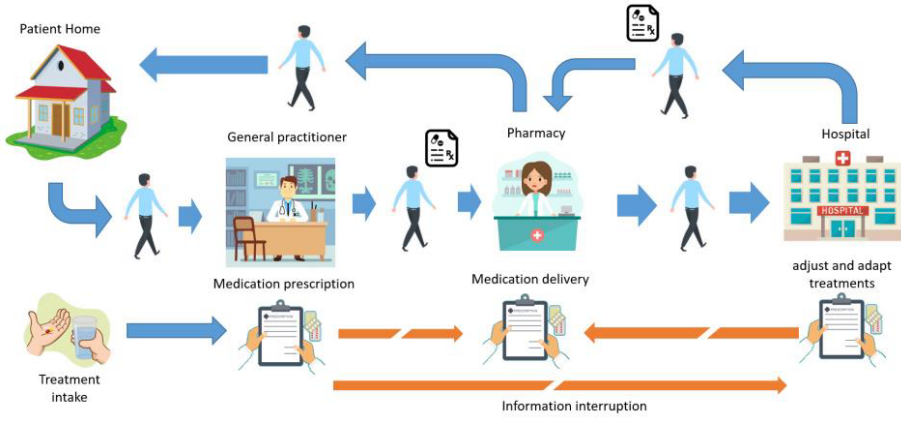
in helping patients management their medications [7]. The long-awaited electronic patient record (EPR) may provide new opportunities for patient empowerment, particularly for medication management.

In this paper, we propose a user-centered model to help improve the flow of information about a patient's treatments, including the patient's self-management at home. This model is proposed based on the experience collected through several projects around medication such as the redesign of our hospital computerized order entry system, the development of an application for the daily management of patients' treatment [8], the participation to the design of the shared medication plan of the Swiss EPR [9] and a working group around medication reconciliation. All these inputs processed by the authors helped us to model the current patient journey around their medication and to identify the existing gaps. Based on the evolution of the digital landscape, especially with the introduction of the EPR, we propose an integrated model aiming to fill these gaps. We then discuss important considerations for the implementation of this model.

2. Current Considerations and Existing Tools

As mentioned above, there are several gaps in the communication process between the doctor's prescription, the patient's self-management and long-term adherence to treatment. In a simplified, single doctor-patient model, the doctor diagnoses a condition, then prescribes an appropriate medication: after explaining why and how to take the medicine, the doctor hands the paper with the prescription to the patient. The patient heads to the pharmacy, and receives the box of medicine with instructions and precautions on how to take the medicine. Back at home, the patient is then in charge of taking the medication regularly as prescribed. The patient needs to understand what to do, and why it should be done, which in turn should help drive motivation to follow through with the recommendations until the end of the prescribed treatment.

In a multiple-provider model with a family doctor, the family doctor could ideally help centralize the patient's medication information, and adjust and adapt treatments as needed. Practically, however, the patient is the one who goes from one provider to another: as there is yet little systematic communication between healthcare professionals, the flow of information about medications often depends on what the patient understands, recalls, and can report. It also depends on what the provider discusses with the patient.

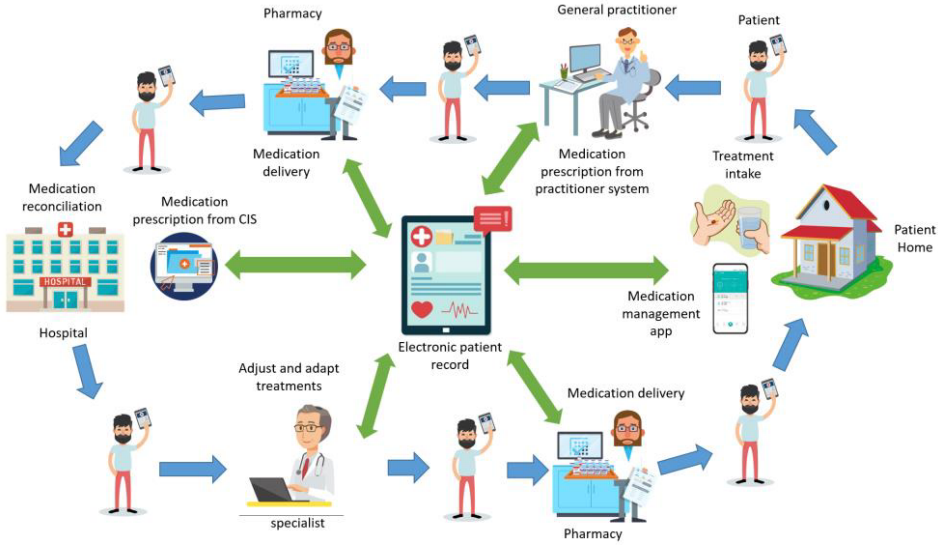


Conflicting health information is a growing concern, with increasing sources of information (e.g. digital sources [10]), in particular for medication [11]. In a recent study of 405 pharmacy customers, 47% of individuals declared that they had received some discordant information in some way about their medication (from their various healthcare providers, from TV programs or commercials, from social media or from family and friends). About two thirds subsequently changed their medication behavior.

The future national EPR record will soon be implemented, and will include a module called "shared medication plan"[12]. This electronic record aims to help allow the patient to have access to her medical information, and to share it with whoever she wishes. This medication plan module will allow doctors in various settings (private practice or institutions) to export their prescriptions to the patient's EPR. The pharmacist's system will import the prescriptions and will link them with the given medication (exact medication name, dose and amount in the package). The patient will be able to access ant interact with the medication list (such as over the counter medications). The patient can add comments to all medications, but can only edit those entered by the patient.

3. Proposed Model

We propose an integrated model based on the future national electronic patient record and its medication plan module, which will also be integrated with our local institution's electronic medical record.



Our aim with the integrated model is to address the patient's daily needs for medication self-management and to discuss some practical considerations in the integration process. In this model, the EPR centralizes the information from the hospital and from the various care-providers: when the patient sees the family doctor, the report and medication are synchronized with the EPR. The patient goes to the pharmacy: the pharmacist can access the prescribed medications, and adds the exact package that was delivered to the patient (brand or generic, number of pills, etc.). Once the patient arrives

at home, she can double check the medications she has (those just received, as well as any prior medications) by scanning the barcodes in the app: this creates a reconciled list that can also include any over-the-counter medications. If the patient is admitted to the hospital, the medication list from the app can be synchronized with the hospital's computerized prescription tool, with additional information from the tracking and monitoring tools if the patient has used them. At discharge from the hospital, the prescribed medication list is again updated in the EPR, and available for the pharmacist to deliver the medication.

The patient is the only person that is present at each step of his own care: he needs to have access to an updated medication list at all times, and to be able to share it with whomever he wishes at any time. One approach to empower the patient is to provide support for the medication plan, and to help them manage their medications daily with reminders and tracking tools [13,14]. Based on focus groups and patient interviews, with iterative testing and improvements, a medication app called Swiss-meds was developed to provide these functionalities[15]. Patients discharged from the hospital often receive a list of new medications, and one difficulty that the project participants reported was not knowing which medicines to take, between the new prescriptions and the medication that their family doctor had prescribed before the hospital stay. This app therefore also included an alert for when similar medications are entered in the medication plan, suggesting that the patient ask the pharmacist or doctor about these potential duplicates. Another request for those who track their medicine was to be alerted in time to fetch a new box of medication before running out. Furthermore, low-literacy information sheets and links to the official patient information website in the app allow the patients to learn about their medication.

4. Implications of an App for an Integrated Solution

The medication app for the patient addresses self-management needs, but can also help provide links between the patient and the other actors in the model: besides empowering the patient with information about the medications, it allows the patient to share the collected data with the care-providers. In an integrated model between the patient and the healthcare system, the medication list from the electronic medication plan, enriched with the actual medication packages received at the pharmacy, should be synchronized with the app information with reconciliation functionalities. First, as in the current app, when the patient scans new medicines to add to the list, the app needs to identify similar drugs and suggest potential duplicates. Scanning the packages should be a routine use of the app, to help keep correct counts of the countdown until refills are needed, and second to check for duplicates and perhaps also to detect drug interactions. Second, the patient can track the medications that are taken, which is a key information for the healthcare team, when it is available.

Although the medication list in the EPR is the list of prescribed (and delivered) medications, there may be differences with the patient's actual medication list: some doctors may not be connected to the patient's EPR, the patient may take over the counter drugs, and may actually have stopped taking some of the medications. This is why the patient needs to be able to edit his medication list without any restrictions (possible in the app). For example, during a phone consultation, the doctor may recommend a dose adjustment (e.g. "take half the pill instead of a full pill during the next 5 days") without changing the prescription, since the patient already has the medication.

5. Important Considerations for an Integrated Medication Management System

Several existing barriers and anticipated obstacles need to be considered when implementing an integrated medication management system. An integrated system requires all stakeholders to have access to the system, and to use it regularly to have up-to-date data. The first step is therefore to provide a secure access to the EPR for patients and providers, and to facilitate access to a free app for patients.

Patients will need an ID for the new EPR system to ensure a unique identity in Switzerland. The first challenge is that users of existing systems, such as our institutional patient portals and the medication app, do not yet rely on such an identifier. Therefore the existing identity and the new EPR ID must be reconciled.

Usability decreases accordingly with the efforts that are required for its use. The current EPR integrates high security standards that require health professionals to log into a system through a two-factor identification process. Since their usual electronic patient charts are not integrated yet, they will need to go through a double identification process with an SMS challenge for each patient chart. This is an interruptive, cumbersome process. With high clinical and administrative workload, these access issues may become an important barrier for regular use. The same issues exist for pharmacists.

The implementation of an integrated system must take into consideration the low or partial adoption of each tool in each part of the process. Individuals may adopt the EPR, but only use it occasionally to view their reports and prescriptions. Individuals may choose to only use the medication app, with or without sharing their information with their care-providers or with the hospital clinicians. The medication app is available in the app store, and will not be limited to individuals with EPR identification. Therefore, an ID reconciliation process with the unique ID must be anticipated for those who want to connect their chart to the EPR system after creating a personal login in the app.

The medication reconciliation process has been recognized as a key moment for patient safety, particularly during care transitions [5]. It needs to be carefully considered for each context:

- When admitting a patient to the hospital, the patient may have a different medication lists in the app and in the EPR. It is only after clarification with the patient that the hospital doctors can establish the list of current medications, because both sources may be discordant.
- When discharging the patient, the EPR medication list should be overridden by the hospital medication list. This list may be edited by the patient (who may stop some medications by themselves) or by the family doctor at the next visit.
- When the patient uses the app after discharge, the EPR medication list, annotated with the delivered medication list, will be available in the app. When scanning the medications, the patient can then check if there are doubles or similar medications from the prior medications at home (from before the hospitalization) that may need further clarification by the pharmacist or doctor.

We see that the reconciliation process is complex, particularly when the patient deals with medication at home. Even if the medication list may simply transit from the initial prescriber to the pharmacist, there is no way to guarantee correct information throughout the process due to high risk of undocumented changes or decisions from the patient (by both patient and care-providers). We must in particular be wary to avoid believing that what the doctor prescribes is what the patient actually takes at home, and allow the patient to correct and change the data as needed (orally or electronically).

6. Conclusion

We conclude that an integrated medication system in Switzerland is feasible with the arrival of the new EPR system, and would further empower the patient for medication management. Besides having access to the EPR list, the availability of an app can help with daily adherence, medication safety (similar drugs and possible interactions) and allow the patient to edit the medication list without limitations. There are anticipated hurdles that need to be addressed, for access (identification and usability issues) and for the reconciliation process in each different situation.

References

[1] Peterson AM, Takiya LT, Finley R. Meta-analysis of trials of interventions to improve medication adherence. Am J Health Syst Pharm. 2003 Apr 1;60(7):657-65. doi: 10.1093/ajhp/60.7.657.
[2] Foppe van Mil JW, Westerlund T, Brown L, Chen TF, Henman M, et al. Medical care and drug-related problems: Do doctors and pharmacists speak the same language? Int J Clin Pharm. 2016;38(2):191–4.
[3] Viswanathan M, Golin CE, Jones CD, Ashok M, Blalock SJ, Wines RCM, Coker-Schwimmer EJL, Rosen DL, Sista P, Lohr KN. Interventions to Improve Adherence to Self-administered Medications for Chronic Diseases in the United States. Ann Intern Med [Internet]. 2012 Dec 4;157(11):785. Available from: http://annals.org/article.aspx?doi=10.7326/0003-4819-157-11-201212040-00538
[4] Davis MM, Devoe M, Kansagara D, Nicolaidis C, Englander H. "Did I do as best as the system would let me?" Healthcare professional views on hospital to home care transitions. J Gen Intern Med. 2012;27(12):1649–56.
[5] Redmond P, Grimes TC, Mcdonnell R, Boland F, Hughes C, Fahey T. Impact of medication reconciliation for improving transitions of care. Cochrane Database Syst Rev. 2018;2018(8).
[6] De Pietro C, Francetic I. E-health in Switzerland: The laborious adoption of the federal law on electronic health records (EHR) and health information exchange (HIE) networks. Health Policy (New York) [Internet]. 2018;122(2):69–74.
[7] Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: Implications for medication management. J Med Internet Res. 2013;15(7).
[8] Ehrler F, Gschwind L, Meyer P, Blondon K. Swiss-Meds: An app fostering medication adherence of Swiss patient. Stud Health Technol Inform. 2019;259:71–6.
[9] Blondon K, Skalafouris C, Simonet ML, Piguet V. Using a Medication Plan as a Quality Indicator: Feasibility and Satisfaction Results from an Observational Study. Elzohry AAM, editor. J Fam Med [Internet]. 2019 Aug 24;1(3):15–23.
[10] Carpenter DM, Geryk LL, Chen AT, Nagler RH, Dieckmann NF, Han PKJ. Conflicting health information: a critical research need. Heal Expect. 2016;19(6):1173–82.
[11] Carpenter DM, Elstad EA, Blalock SJ, DeVellis RF. Conflicting Medication Information: Prevalence, Sources, and Relationship to Medication Adherence. J Health Commun [Internet]. 2014 Jan 9;19(1):67–81.
[12] Bugnon B, Geissbuhler A, Bischoff T, Bonnabry P, von Plessen C. Improving primary care medication processes by using shared electronic medication plans in Switzerland: Lessons learned from a participatory action research study. JMIR Form Res. 2021;5(1):1–17.
[13] Park JYE, Li J, Howren A, Tsao NW, de Vera M. Mobile phone apps targeting medication adherence: Quality assessment and content analysis of user reviews. JMIR mHealth uHealth. 2019;7(1):1–12.
[14] Náfrádi L, Nakamoto K, Schulz PJ. Is patient empowerment the key to promote adherence? A systematic review of the relationship between self-efficacy, health locus of control and medication adherence. PLoS One. 2017;12(10):1–23.
[15] Ehrler F, Gschwind L, Meyer P, Blondon K. Swiss-Meds: An app fostering medication adherence of Swiss patient. Vol. 259, Studies in Health Technology and Informatics. 2019.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210597

Lifestyle Cancer Survival Predictors: Influence of Vegetarian Diet on the Relapse of Endometrial Cancer

Georgy KOPANITSA[a], Oleg METSKER[b,1], Ekaterina BOLGOVA[a], Sergey KOVALCHUK[a]
[a] *ITMO University, Saint-Petersburg, Russia*
[b] *Almazov National Medical Research Centre, Saint-Petersburg, Russia*

Abstract. Endometrial cancer (EC) is the most common gynecological tumor in high-income countries, and its incidence has increased over time. The most critical risk factor for EC is the long-term unopposed exposure to increased estrogens both exogenous and endogenous. Machine learning can be used as a promising tool to resolve longstanding challenges and support identification of the risk factors and their correlations before the clinical trials and make them more focused. In this paper we present the results of the research of the correlation analysis of Endometrial cancer risk factors. The study was performed with EC patients of the Almazov center in Saint-Petersburg, Russia. All women involved in the current study underwent radical surgical intervention due to EC. After initial cancer treatment, they were referred to the Almazov center outpatient specialists for follow-up visits. Many of them were readmitted of the inpatient clinic due to relapse. We extracted a variety of parameters related to lifestyle, dietary habits, socioeconomic, and reproductive features from the inpatient and outpatient databases of Almazov center. The medical records of the women with enough data were included in the study. Prediction of Progression-free survival (PFS) and overall survival (OS) were analyzed respectively. The AUC of ROC was calculated for PFS = 0.93 and for OS = 0.94.

Keywords. Oncogynecology, Machine learning, Prediction, Lifestyle

1. Introduction

Oncogynecology embraces a group of diseases that originate in the female reproductive organs. There are five main sites of gynecologic cancers: ovaries, uterus (corpus and cervix separately), vulva, and vagina. In the year 2019, more than 1.3 million new cancer cases, arising from those five sites occurred worldwide [1].

Although being grouped, each disease represents a distinct clinical entity with different incidence, risk factors, clinical presentation, and prognosis [2].

Endometrial cancer (EC) is the most common gynecological tumor in high-income countries, and its incidence has increased over time [3-4]. In 2018, there were 382,069 cases of EC worldwide [1]. The majority of patients suffering from EC are over 40 years, although rare cases also occur in younger women [5]. Overall, the median age of women diagnosed with EC peaks around 60 years.

[1] Corresponding Author: Oleg Metsker, Akmazov National Research Center, Saint-Petersburg, Russia, E-mail: olegmetsker@gmail.com

EC belongs to so-called hormone-responsive cancers, which implies that hormone exposure, more specifically, estrogen exposure, is involved in EC occurrence and progression. Unbalanced estrogen exposure may originate from drugs (e.g., hormone replacement therapy) or could be a consequence of early menarche, late menopause, and obesity [3]. Other factors that mimic estrogens, such as endocrine disruptors or cadmium, are also associated with increased risk of EC [6].

The most critical risk factor for EC is the long-term unopposed exposure to increased estrogens both exogenous and endogenous [7]. This can be potentially associated with dietary preferences of a woman. All of them directly or indirectly affect the balance of estrogen in the body of a woman. Early diagnosis of the EC and its relapses plays an essential role. The risk of EC relapse depends on patient age and tumor characteristics, such as type, stage, present of metastases and progesterone receptor expression [8].

Despite the known risk factors, several challenges and poor diagnostic and treatment outcomes have been reported in the recent systematic review, where failure to diagnose malignancies early affect treatment provision and prognosis [9]. Therefore, a deeper analysis ofthe risk factors and their combinations can provide better insights on the diagnostics andprevention of Endometrial cancer. As clinical trials take much time and effort, study thecombination of factors can become even more challenging task. Machine learning approach helps to process large amount of information within shorter time, which makesit useful in the search for factors affecting the progression of cancer. Although it is clearthat the disposition to cancer is due to unmodified factors, it may affect when the cancer"shows itself".

Machine learning can be used as a promising tool to resolve longstanding challenges and to support identification of the risk factors and their correlations prior to clinical trials and make them more focused.

In this paper, we present the results of the research of the correlation analysis of Endometrial cancer risk factors.

2. Methods

The study was performed with EC patients of the Almazov national research center.

2.1. Participation Criteria

Women, diagnosed with EC and admitted to the Almazov National research center hospital within the period 2011-2020.

2.2. Clinical Data Collection

All women involved in the current study underwent radical surgical intervention due to EC. After initial cancer treatment, they were referred to the outpatient specialists for follow-up visits. If the relapse was suspected, a patient was referred back to the inpatient hospital.

From the Almazov inpatient and outpatient databases we extracted parameters related to lifestyle, dietary habits, socioeconomic, and reproductive features. Corresponding medical records were retrieved from the existing database. All data was analyzed in unidentified fashion. In total, data from 3845 women were processed. We processed only data of women,who had enough information about their diet in the

anamnesis. Of these, 9.8% women had relapse or died due to EC progression. Patients' median age (IQR) at the time of diagnosis was 70.0 (65.3 - 77.0) and BMI = 26.3 (23.7 - 30.1).

2.3. Endpoint Events

The following indicators were calculated – progression-free survival (PFS) and overall survival (OS), respectively. PFS is the time which a patient lives with cancer without it worsening. OS is the length of time from the defined start point to death fromany cause.

2.4. Data Analysis

Each experiment ran in the setting of stratified 5-fold cross-validation (i.e., random80% of patients were used for training and 20% for testing, target class ratios in the foldswere preserved), which represents the trade-off between sensitivity and specificity of the model. TheAUC was calculated based on an average of 5 curves (one curve per fold in the setting of 5-fold cross-validation). All the measurements were performed separately per datasetand per model parameter value to determine the best parameters for classifiers as well asoptimal data preprocessing.

After determining the optimal dataset and model parameters, we performed a more thorough ROC analysis (100 x 5-fold cross-validation) with the given parameters to find the optimal probability threshold for further label assignment, and the final assessment that utilizes the threshold to calculate the remaining prediction quality characteristics.

We used a series of classification models available within scikit-learn as a pool for the selection of the best predictive methods to be applied within the proposed scheme. A random forest method was applied for the classification with the following parameters: rf = RandomForestRegressor(n_estimators = 100, n_jobs = -1, oob_score = True, bootstrap = True, random_state = 42). A decision tree was developed to partition the data space into cluster regions.

3. Results

The AUC of ROC was calculated for PFS = 0.93 and for OS = 0.94. The decision tree is presented in Figure 1.

The model makes different decisions about outcomes for different ages, in particular physical activity for patients over 78 years of age (patients with low physical activity are more at risk of recurrence. Patients between 78 and 83 years of age with low physical activity were more likely to die (OS event) if prunes were present in their diet. For patients with normal exercise, sunscreen was used as a protective cream. For patients between 60 and 78, depending on age and physical activity, the following factors affected cancer progression: type of diet (consumption of prunes increased the risk, consumption of liver pate increased the probability of survival).

For patients younger than 60, depending on age and physical activity, the following factors influenced the cancer progression: carbohydrate intake, exposure to sun. Both factors increased the risk of cancer progression.

Random forest regression showed 89 % AUC of ROC. The R^2 Training Score was 0.88, the OOB Score came up with 0.06, and the R^2 Validation Score with 0.10.

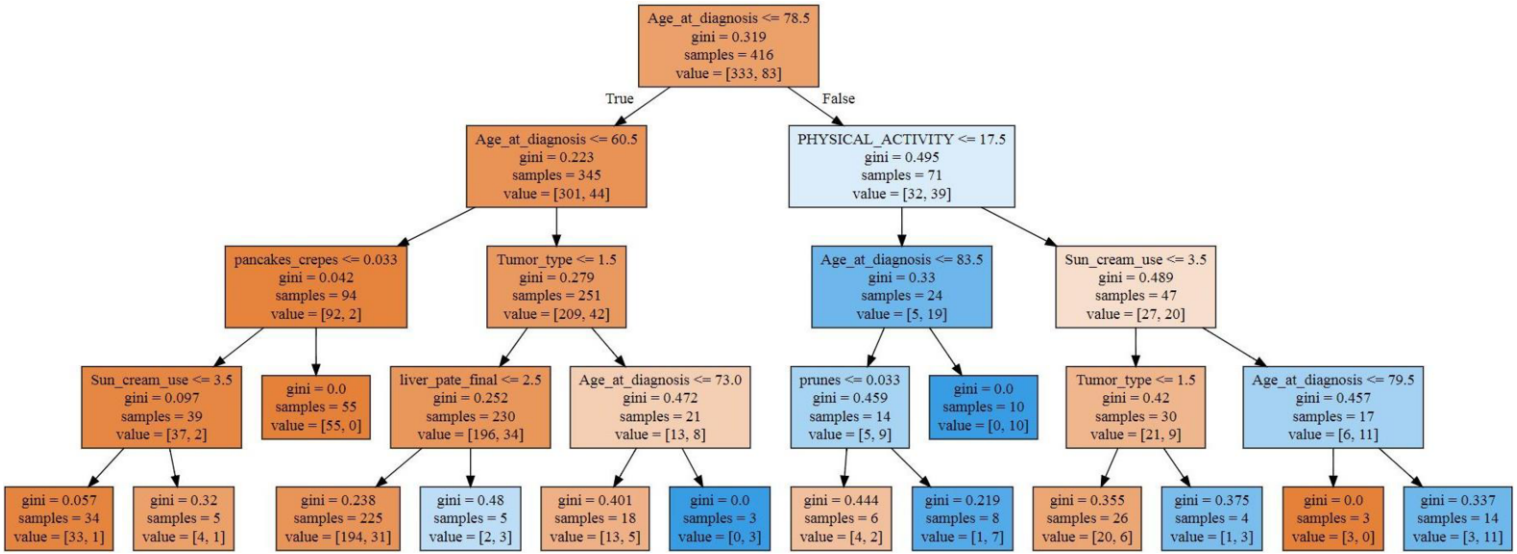


Figure 1. PFS Decision tree

The correlation analysis showed the existence of Relapse and a Diet Type (diet_type). It is an interesting conclusion on Diet Type, because the variable is encoded as follows. 0=mixed, 1=vegetarian, 2=vegan. It turns out that vegans have a higher risk of recurrence. This may be due to a higher consumption of soybeans and plant products that contain phytoestrogens, potential endometrial cancer agents.

Nut consumption can also play its role in the vegetarian diet due to cadmium levels in them. Cadmium in the body has an estrogen-like effect, which is a known EC risk factor.

4. Discussion

Based on our results, the risk of relapse and death due to EC correlated with stage, tumor type, its differentiation and patient's age. These findings are consistent with common knowledge, as it is well-known that prognosis worsens with advanced stage and age. Type 2 EC is also associated with lower survival rate.

Interestingly, relapse and death rate inversely correlated with physical activity. These data complement previous studies that have shown the benefits of regular physical activity in the prevention of endometrial cancer [8-10].

Soft drinks consumption was also correlated with risk of relapse and even stronger with risk of death due to EC. This is probably due to the fact, that high-sugar foods and beverages stimulate insulin production. This is related to increased BMI and type II diabetes and most probably directly to endometrial cancer progression [11,12]. However, further studies are needed due to contradictory results from other works.

Less explained fact relates to sun cream use, which is correlated with increased risk of both relapse and death due to EC. People have different skin phototypes, which affect their sun cream use. Since phototype is inherited the underlying mechanism behind our finding might be potentially hidden in various genetic features.

Other findings related to dietary habits are to be thoroughly studied in well-designed case-control studies.

The patients were followed-up for more than eight years, which is relatively long, while medical records were retrieved from the well-established national database.

The retrospective design of the study has inherent disadvantages. Besides, since patients completed the questionnaires on their own, bias cannot be excluded entirely.

5. Conclusion

The study demonstrates how dietary factors effect risk of EC development. Especially their influence on the relapse of the cancer.

To conclude, here we demonstrate the feasibility of using machine learning in the analysis of the outcome predictors in EC patients.

Aknowledgement

This work was financially supported by the Government of the Russian Federation through the ITMO fellowship and professorship program. This work was supported by the Russian Foundation for Basic Research, grant number 20-37-70047.

References

[1] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Global cancerObservatory: cancer today. Lyon, France: international agency for research on cancer. Cancer Today. 2018.
[2] The Centers for Disease Control and Prevention's Inside Knowledge: Get the Facts about Gynecologic Cancer Campaign. Series: Inside Knowledge: Get the Facts about Gynecologic Cancer. Centers for Disease Control and Prevention 2012, June 9.
[3] Colombo N, Creutzberg C, Amant F, Bosse T, Gonzalez-Martin A, LedermannJ, et al. ESMO-ESGO-ESTRO Consensus Conference on Endometrial Cancer: diagnosis, treatment and follow-up. Ann Oncol. 2016 Jan;27(1):16-41. doi: 10.1093/annonc/mdv484. Epub 2015 Dec 2.
[4] Lortet-Tieulent J, Ferlay J, Bray F, Jemal A. International Patterns and Trends in Endometrial Cancer Incidence, 1978-2013. Journal of the National Cancer Institute. 2018;110(4):354-61.
[5] Lee NK, Cheung MK, Shin JY, Husain A, Teng NN, Berek JS, et al. Prognostic factors for uterine cancer in reproductive-aged women. Obstetrics and gynecology. 2007;109(3):655-62.
[6] Safe S. Cadmium's disguise dupes the estrogen receptor. Nat Med. 2003;9(8):1000-1.
[7] Endometrial cancer treatment Physician Data Query (PDQ) 2018 [Available from: http://www.cancer.gov/cancertopics/pdq/treatment/endometrial/healthprofessional.
[8] Huijgens A, Mertens H. Factors predicting recurrent endometrial cancer. Facts,views & vision in ObGyn. 2013;5(3):179.
[9] Collaborators M, Beral V, Bull D, Reeves G. Endometrial cancer and hormone-replacement therapy in the Million Women Study. Lancet. 2005;365:1543-51.
[10] Furness S, Roberts H, Marjoribanks J, Lethaby A, Hickey M, Farquhar C. Hormone therapy in postmenopausal women and risk of endometrial hyperplasia. Cochrane Database of Systematic Reviews. 2004(3).
[11] Henderson BE. The cancer question: an overview of recent epidemiologic and retrospective data. Am J Obstet Gynecol. 1989;161(6 Pt 2):1859-64.
[12] Onstad MA, Schmandt RE, Lu KH. Addressing the Role of Obesity in Endometrial Cancer Risk, Prevention, and Treatment. J Clin Oncol. 2016 Dec 10;34(35):4225-4230. doi:10.1200/JCO.2016.69.4638.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210598

Gait Analysis Platform for Measuring Surgery Recovery

Marc CODINA[a,1], Manuel NAVARRETE[a], Ashkan REZAEE[a], David CASTELLS-RUFAS[a], Maria Jesús TORRELLES[a], Stefan BURKARD[b], Holger ARNDT[b], Sabine DREVET[c], Medhi BOUDISSA[c], Jerome TONETTI[c], Isabelle MARQUE[d], Alexandre MOREAU-GAUDRY[d], Armand CASTILLEJO[e], Jordi CARRABINA[a]
[a] *CEPHIS. MiSE Dpt. Universitat Autònoma de Barcelona, Bellaterra, Spain*
[b] *SpringTechnoGMBh, Bremen, Germany*
[c] *Dpt. of Orthopedic and Trauma Surgery. CHU Grenoble-Alpes, La Tronche, France*
[d] *INSERM. CHU Grenoble-Alpes, CIC1404, Grenoble, France*
[e] *STMicrolectronics, Grenoble, France*

Abstract. Gait analysis has evolved significantly during last years due to the great development of the Medical Internet of Things (MIoT) platforms that allow an easy integration of sensors (inertial, magnetic and pressure in our case) to the complex analytics required to compute, not only relevant parameters, but also meaningful indexes. In this paper, we extend a previous development based on a fully wireless pair of insoles by implementing an updated version with more reliable and user-friendly devices, smartphone app and web front-end and back-end. We also extend previous work focused on fall analysis (with the corresponding fall risk index or FRI) with the proposal of a new surgery recovery index (SRI) to account for the individual speed recovery speed that can be measured either at clinical facilities or at home in a telemedicine environment or while doing daily life activities. This new index can be personalized for different types of surgeries that affect gait such as hip, knee, etc. This paper presents the case of hip recovery and is built on top of the clinical standard SPPB test and allows obtaining quantitative parameters directly from the sensors.

Keywords. Gait Analysis, Insoles, Hip surgery recovery, Recovery Index, Medical Internet-of-Things

1. Introduction

Wearables have been significantly relevant in increasing the motivation for wellbeing all around the world. It should be agreed that this should improve the general health care situation for all age ranges and profiles. Wearable recordings started with counting steps and progressively added new measures (oxygen saturation, pulse, etc.) to even more complex functionalities (electrocardiographic analysis, fall detection, gait analysis (e. g. for runners). Wearables have the advantage that can take and store continuous measurements on health parameters. These parameters are not yet being intensively used and/or integrated in the health records of the users or patients nor in the health platforms

[1] Corresponding Author: Marc Codina; Universitat Autònoma de Barcelona, Centre de Prototips i Solucions Hardware-Software, CEPHIS. Escola d'Enginyeria, Campus UAB, 08193 Bellaterra, Spain; Email: Marc.Codina@uab.cat.

from public and private health providers what can be considered as a loss of useful information. It is commonly agreed that some of the reasons are the different precision of the different sensors used or the complexity of the integration of patients' data into the health records. Nonetheless, the evolution of the sensor technology and related calibration procedure plus the increasing access to platform providers through APIs easier those walls. Still, it remains one relevant issue: how to move or transform existing criteria and protocols based on measurements done in clinical installations to integrate this existing data in the health records of the patients. This platform will be used for validating relevant data (raw or derived) according to the golden standards. In this paper, we present a proposal for such framework devoted to measure the degree of recovery after a surgery that affects gait.

Gait analysis has become a widely used clinical tool, enabling clinicians to provide better patient care. The analysis of the gait of patients offers new opportunities for diagnosis and treatment of walking problems and to find patterns in gait (1,2).

Our hypothesis if that with semi-automated gait analysis it is possible to evaluate the recovery status of a patient from surgery and provide a new Surgery Recovery Index (SRI) to quantify the individual rate of recovery. This analysis can be carried out in both ways, in the clinical facility or at home while performing activities of daily life. In the former WIISEL project, we worked on the analysis of falls and the elaboration of a Fall Risk Index (FRI), since falls are a major cause of injuries for the elderly people (3). This project was also validated by several clinical studies (4). Nonetheless, the insole prototype was far from a final product concerning reliability of its components and integration.

In the scope of the SERENE project (5), we present an updated MIoT platform to collect real-time gait data from users wearing a custom pair of wireless insoles, oriented to the analysis of the recovery process for patients after a surgery that affects their gait that is currently measured by the score of a given test and we propose to measure by defining a numerical surgical recovery index.

2. The SERENE Platform

The SERENE project is a collaborative project with a consortium of 20 partners from France, Germany and Spain with the goal of providing better remote care and diagnostic tools using Medical IoT technologies. Our contribution is centered on providing a platform for gait analysis composed of a pair of smart-insoles and a smartphone per patient, a cloud data-base, a software platform for the analytics and a web access for 3 types of users (patients, caregivers / relatives and clinicians).

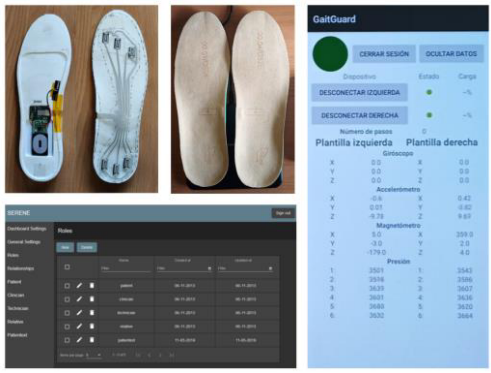


Figure 1. Platform components: Insoles, Android APP and Web interface

The wearable insoles were designed to look and feel like regular insoles in terms of comfort with the support of a commercial insole maker: Flexor. The internal electronics were separated into two flexible parts for power and computation/communication plus two antennas, allowing them to be adjusted to the desired shape or to flex during use. The communication with the mobile phone is via Bluetooth Low Energy (BLE). The insoles have no charging connectors, as they are equipped with Qi wireless charging. An Inertial Measurement Unit (IMU) plus a pressure sensor layer are used as the source of data. The IMU outputs acceleration and gyroscope data, each 3-axis, together with self-calibrated orientation in real time. The pressure sensor layer includes 6 force-sensitive resistors (FSRs) positioned under the heel and the metatarsal. Sensor data acquisition system is 33 samples per second for all sensors while the IMU has 14 bits to16 bits resolution while pressure sensors have 10 bits. The IMU includes a Cortex-M0 processor able to compute quaternions that will also be used for gait analysis. Captured data is formatted in 20 bytes packets and sent to the smartphone through Bluetooth Low Energy (BT5.1)

An application in the smartphone is used to monitor and display the data acquired from the pressure sensors and the information generated by the IMU. The bidirectional communication between the smartphone and the smart insoles is based on Bluetooth Low Energy using a custom GATT profile. The data collected from both smart insoles is processed and appended into a data file that is stored for later analysis.

In addition, the mobile phone acts as a bridge between the insoles connected via Bluetooth and the cloud server which we access via LTE/5G or Wi-Fi. This communication with the server allows access to the patient's personal profile, to retrieve statistical data of the user as well as to send the files with the sensor data captured from the insoles.

The data analysis platform consists of a back-end in which a service responds to requests from the mobile application once the clinical trial recording session is finished. It stores collected data in a database from which 49 parameters related to the gait analysis are computed.

The front-end allows the access of the different users to the processed data through a dashboard designed according to specific profiles: clinician, patient and relative. The user interface graphically displays the gait analysis data and allows the selection of relevant parameters by the clinician what allows personalizing the medical treatment according to the status of the patient.

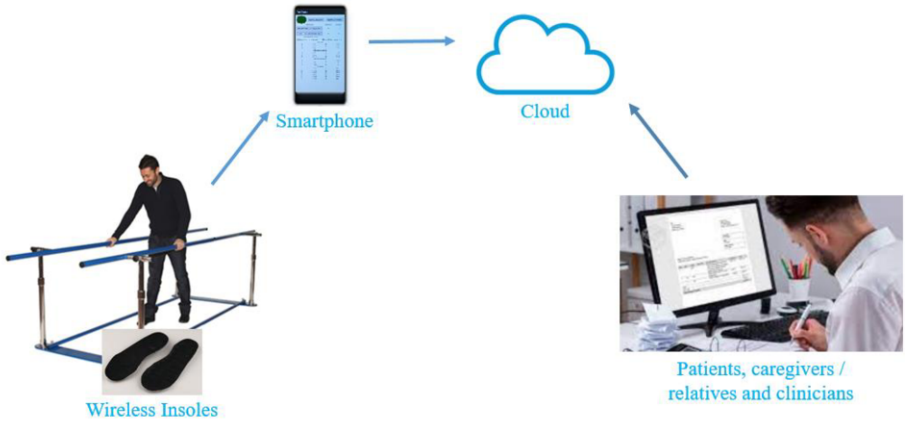


Figure 2. Diagram of the SERENE platform

3. Gait analysis.

Using the new sensors implemented in this revision of the technological part, the number of gait features extracted after processing the sensor data have been extended from 26 to 49. The gait analysis framework has been designed and organized into layers, each one having a specific functionality. The framework consists of five main layers: Raw data processing; Gait parameter definition; Pattern recognition and feature extraction; Gait parameter processing and Surgery Recovery Index (SRI) and exploitation of gait analysis results.

From IMU sensors data, we estimate the position and orientation of the foot with respect to the world. Madgwick's gait tracking algorithms were implemented to determine the position of the patient as they move either through a controlled environment as a clinical facility (6) or at home. Among the parameters that are extracted with the data obtained from the IMU (3D accelerometer, 3D gyroscope and 3D magnetometer) and pressure sensors, we can find the following ones: Cadence, Double Support, Single Support, Swing Phase, Stance Phase, Gait Symmetry, Heel Strike Force Slope, Average Acceleration Amplitude Mediolateral. Some similar approaches can be found in the literature (7).

4. Surgery Recovery

Both surgery recovery and fall risk assessment are based on the analysis of gait parameters. The main differences among them are: the temporal scale of the patients' evolution (usually faster and more deterministic in recovery processes and slower and unpredictable in the evolution of the elderly) and also the relevant gait parameters and their combination to obtain the index that is much more related to the specific surgery rather than to the individual health state as in the case of the elderly).

We focused this new version of the platform in hip surgery recovery. According to the literature, several aspects related to gait can be used to evaluate recovery improvement (8): walking velocity, stride length, range of hip motion and hip abduction

moments. Other authors consider further measurements that can be considered such as: Muscle strength (voluntary isometric force of the hip and knee muscles related to walking ability) (9–11) or balance (postural stability) (12). Some of these parameters must be directly obtained or estimated by the observation of the clinician while others can be provided by our gait analysis platform. Our idea of SRI is to indicate whether the patient is progressing faster, equal or slower than the average for a given recovery time after surgery.

After evaluating the different tests proposed in the literature (13–15), we decided to use SPPB (16) as the basis for our tests. This test evaluates balance, gait speed and the ability to get up and down a chair (as indicator of strength) by measuring the time of each activity. It currently provides a combination of scores, giving more precision in the differences in stride and balance (17). The values obtained from IMU and pressure sensors for each one of the three stages of the SPPB test will be captured with higher confidence. In the case of balance, we use a posturometer as a gold standard to fine tune our algorithms.

We consider two different scenarios. The first one is related to rehabilitation tracking in clinical facilities, where multiple patients will be able to use the insoles (according to their feet size) in their rehabilitation routines and clinicians can analyze the data obtained through our platform. This leads to a sporadic use of the system. The second model focuses on domestic use, where the patient wears a pair of insoles on their daily life activities. Then, clinicians track that patient at any time on their dashboards. When any anomaly is detected, the clinician will contact the patient for a more detailed follow-up. This model can also be used for remote monitoring of specific exercises in telemedicine scenarios that are becoming popular during the COVID-19 pandemic.

The hip SRI complements the existing SPPB test with the parameters obtained by the insoles (pressure and motion) for each stage of the test. Studies about the weight of the parameters in the index and the way to personalize them according to the patient will be carried out at later research.

5. Conclusion

This paper presents a solution of a platform for surgical recovery analysis composed of a pair of fully wireless insoles, a smartphone app and a cloud solution that includes both database and web personalized dashboards. We proposed a new Surgical Recovery Index (SRI) to account for the evolution of the patient after a surgery that affects its gait.

We apply this procedure to hip surgery recovery upon the existing SPPB test, whose score will be complemented by the data obtained and computed from sensor parameters, thus becoming easy to interpret in terms of recovery rate, that depends on the health state and specific surgery done to every person.

This SRI model can be extended to other surgeries that affect gait such as hip, knee, achilles, etc. (8,18,19). In some cases, when data is available, that index will consider the status before the surgery or even before the lesion (e.g. in sports).

This platform can be deployed for remote surgery recovery in several scenarios, all of them under the premise that secure servers must hold medical records. We initially plan to use the devices in both clinical facilities and at patients' home. In this second case, two models are available: a telemedicine scenario to proceed with tests supervised remotely by clinicians and an unsupervised scenario based on daily living activities with the automatic identification of relevant situations to compute SRI. It is planned to start

soon the clinical trials to analyze the recovery status of several patients at Grenoble hospital.

References

[1] Pagano G, D'Addio G, De Campi M, Donisi L, Biancardi A, Cesarelli M. Rehabilitation Outcome in Patients undergone Hip or Knee Replacement Surgery using Inertial Technology for Gait Analysis. IEEE Med Meas Appl MeMeA 2020 - Conf Proc. 2020;

[2] Ghent F, Mobbs RJ, Mobbs RR, Sy L, Betteridge C, Choy WJ. Assessment and Post-Intervention Recovery After Surgery for Lumbar Disk Herniation Based on Objective Gait Metrics from Wearable Devices Using the Gait Posture Index. World Neurosurg [Internet]. 2020;142:e111–6. Available from: https://doi.org/10.1016/j.wneu.2020.06.104

[3] Arndt H, Burkard S, Talavera G, Garcia J, Castells D, Codina M, et al. Real-time constant monitoring of fall risk index by means of fully-wireless insoles. Stud Health Technol Inform. 2017;237:193–7.

[4] Bravo J, Cook D, Riva G. Ambient intelligence for health environments. Vol. 64, Journal of Biomedical Informatics. 2016. 15–25 p.

[5] SERENE-IoT site [Internet]. Available from: https://penta-eureka.eu/project-overview/penta-call-1/serene-iot/

[6] Madgwick SOH, Harrison AJL, Vaidyanathan R. Estimation of IMU and MARG orientation using a gradient descent algorithm. IEEE Int Conf Rehabil Robot. 2011;

[7] Nouredanesh M, Godfrey A, Howcroft J, Lemaire ED, Tung J. Fall risk assessment in the wild: A critical examination of wearable sensor use in free-living conditions. Gait Posture [Internet]. 2021;85(December 2019):178–90. Available from: https://doi.org/10.1016/j.gaitpost.2020.04.010

[8] Ewen AM, Stewart S, St Clair Gibson A, Kashyap SN, Caplan N. Post-operative gait analysis in total hip replacement patients-A review of current literature and meta-analysis. Gait Posture [Internet]. 2012;36(1):1–6. Available from: http://dx.doi.org/10.1016/j.gaitpost.2011.12.024

[9] Rantanen T, Era P, Heikkinen E. Maximal Isometric Strength and Mobility among 75-year-old Men and Women. Age Ageing [Internet]. 1994 Mar 1;23(2):132–7. Available from: https://doi.org/10.1093/ageing/23.2.132

[10] Rantanen T, Avlund K, Suominen H, Schroll M, Frändin K, Pertti E. Muscle strength as a predictor of onset of ADL dependence in people aged 75 years. Aging Clin Exp Res. 2002;14(3 Suppl):10–5.

[11] Manini TM, Clark BC, Nalls MA, Goodpaster BH, Ploutz-Snyder LL, Harris TB. Reduced physical activity increases intermuscular adipose tissue in healthy young adults. Am J Clin Nutr. 2007;85(2):377–84.

[12] Fiatarone MA, Marks EC, Ryan ND, Meredith CN, Lipsitz LA, Evans WJ. High-Intensity Strength Training in Nonagenarians: Effects on Skeletal Muscle. JAMA J Am Med Assoc. 1990;263(22):3029–34.

[13] Soares Menezes KVR, Auger C, de Souza Menezes WR, Guerra RO. Instruments to evaluate mobility capacity of older adults during hospitalization: A systematic review. Arch Gerontol Geriatr [Internet]. 2017;72(October 2016):67–79. Available from: https://doi.org/10.1016/j.archger.2017.05.009

[14] Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the timed up and go test. Phys Ther. 2000;80(9):896–903.

[15] Tinetti M. Performance-oriented assessment of mobility problems in elderly patients. J Am Geriatr Soc. 1986;34(2):119–26.

[16] Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: Association with self-reported disability and prediction of mortality and nursing home admission. Journals Gerontol. 1994;49(2).

[17] Lopes KJY, Pin ADS, Miranda RM da S, Ghelfi A. Comparison of the Tinetti and Short Physical Performance Battery test specificities for detecting neuromotor pathologies. ConScientiae Saúde. 2019;18(1):1–7.

[18] McClelland JA, Webster KE, Feller JA. Gait analysis of patients following total knee replacement: A systematic review. Knee. 2007;14(4):253–63.

[19] Don R, Ranavolo A, Cacchio A, Serrao M, Costabile F, Iachelli M, et al. Relationship between recovery of calf-muscle biomechanical properties and gait pattern following surgery for achilles tendon rupture. Clin Biomech. 2007;22(2):211–20.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210599

The Effectiveness of Telemedical Monitoring Program DiabCare Tirol for Patients with Gestational Diabetes Mellitus

Gihan EL MOAZEN[a], Bernhard PFEIFER[a,b], Agnes LOID [c], Peter KASTNER[a] and Christian CIARDI[d]

[a] *AIT Austrian Institute of Technology GmbH, Graz, Austria*
[b] *Landesinstitut für Integrierte Versorgung Tirol, Innsbruck, Austria*
[c] *Tirol Kliniken GmbH, Innsbruck, Austria*
[d] *Krankenhaus St. Vinzenz Zams, Landeck, Austria*

Abstract. The aim of this paper was to evaluate the effect of telemedical care of gestational diabetes mellitus (GDM) patients with the digital treatment pathway model DiabCare Tirol. Methods: 27 courses of patients with GDM, who were telemonitored through the integrated care program DiabCare Tirol in a diabetes outpatient clinic in Tyrol, Austria during the COVID-19 pandemic in 2020, were analyzed. In addition, randomized controlled trials (RCTs) on telemedicine interventions for GDM were researched, and their results were used for comparison with this disease management method. The patient outcome analysis was used to examine the effects of the integrated care program involving telemonitoring support and compared them to the results of RCTs in which participants were randomly assigned to one of two groups, either mobile monitored or standard treatment group. Results: The feasibility of the digital treatment pathway model was confirmed in practice, as the trend analysis of the 27 GDM patients involved showed significantly improved glycaemic control. Results of RCT studies tend to support the findings of DiabCare Tirol. Conclusion: Benefits of telemonitoring with integrated care to support conventional therapy cannot be dismissed, especially in times of the pandemic. Continuous outcome research with larger patient numbers will be necessary to confirm the effectiveness of telemonitoring in a regular care setting.

Keywords. gestational diabetes mellitus, telemonitoring, disease management program, integrated care program

1. Introduction

The prevalence of diabetes continues to rise across the world [1]. In terms of gestational diabetes mellitus (GDM), which is defined as a glucose intolerance occurring first during pregnancy, it affects 6% to 15% of all pregnancies, depending on screening criteria, and is usually diagnosed between 24 and 28 weeks of gestation [2,3]. The incidence thereof increases constantly, especially due to the prevalence of obesity, as with type 2 diabetes, and additionally due to the rising number of pregnancies at an advanced maternal age [2]. Adverse perinatal and postpartum complications associated with GDM for the mother include shoulder dystocia, clinical neonatal hypoglycemia, preterm delivery and

cesarean delivery, as well as the development of cardiovascular disease (e.g. coronary artery disease) and a sevenfold increased risk of developing type 2 diabetes mellitus due to GDM compared to women with normal-glycemic pregnancies. High blood glucose (BG) levels of the mother lead to increased insulin levels in the infant, which causes increased birth weight [2,3].

With timely diagnosis and treatment of GDM, a decrease of pregnancy- and birth-related complications can be achieved with an outcome of pregnancy similar to that of women without GDM. However, for this to happen, regular physical activity, controlled weight gain and nutritional counseling to restrict the amount of carbohydrates, must be incorporated into the patient's routine. For control purposes, BG measurements and documentation thereof should be performed four times a day. If BG limits are exceeded despite adherence to these measures, insulin therapy is required in approximately 20-30% of pregnant women suffering GDM.

In principle, clinical guidelines state that women with GDM should undergo regular follow-up examinations and track self-monitoring data, such as BG and ketonuria, in order to stabilize maternal BG after delivery [2]. With the advancement of telemedicine,which uses information and communication technologies (ICT) as a means of bridging distances to enable health-related information exchange between healthcare professionals and patients-, it is gaining popularity as it offers a practical and sustainable intervention method in GDM management [4-7]. Abundant research studies analyzing telemonitoring in diabetic care found significant improvement in self-care, as diabetes-related data could be transmitted electronically to professionals. As a result, the number and necessity for outpatient clinic visits could be reduced [2,3]. Based on the mentioned aspects, the AIT Austrian Institute of Technology GmbH developed a disease management program (DMP) for diabetes patients called DiabCare Tirol, using a telemonitoring system, which is incorporated in a comprehensive network of specialists. A methodically similar example of a successfully implemented telemonitoring DMP would be HerzMobil Tirol, which is a multi-dimensional post-discharge program for heart failure patients that has been used in regular care and was gradually extended across the region of Tyrol since 2018 [8,9].

After a pilot phase with type 1 and type 2 diabetes patients, the project concept was adapted to the needs of gestational diabetes in 2020. This paper shows first results of patients with GDM while participating in the DiabCare program in Tyrol and compares the outcomes with those of different RCT studies to check validity.

2. Method of Telemonitoring with DiabCare Tirol

DiabCare Tirol is a telemonitoring platform for remote treatment support of type 1-(DM1), type 2 diabetes (DM2) and GDM since it replaces paper diary records with online documentation and remote access for caretakers. The measured data are transferred from devices[1] to the DiabCare mobile application[2] via Bluetooth or through manual entries and is accessible online on a data management system[3]. Patients can communicate with their assigned clinical specialists via the app and receive feedback on their progress. A DMP with a specific treatment pathway was developed considering requirements for

[1] Blood glucose meter - LifeScan Verio Reflect, Blood Pressure Meter - BOSO Medicus System

[2] DiabCare Mobile Client Application, Version 4.3.2 (AIT)

[3] DiabCare Data Management System, https://diabcare.tirol-kliniken.at, Version 1.27 (AIT)

treatment of GDM (e.g. monitoring plan and intensity, tasks for and interaction between caretakers). The diabetes outpatient clinic at Krankenhaus St. Vinzenz Zams, Landeck served as the recruitment center for GDM patients willing to participate in the project. Due to the temporary overload of COVID-19 cases in hospitals of the Landeck region the participation helped to keep outpatient visits as low as possible and thus minimize the risk of infection for this particularly vulnerable group.

The care process (Figure 1) for GDM patients started with the visit of the diabetes outpatient clinic for an initial consultation. A specialized nurse for diabetes counseling and dieticians, in consultation with the attending physician, provided hypoglycemia training, individualized nutritional counseling, and offered technical advice regarding the telemonitoring equipment. This was followed by assigning the patients to the first phase for a duration of one month where close monitoring took place.

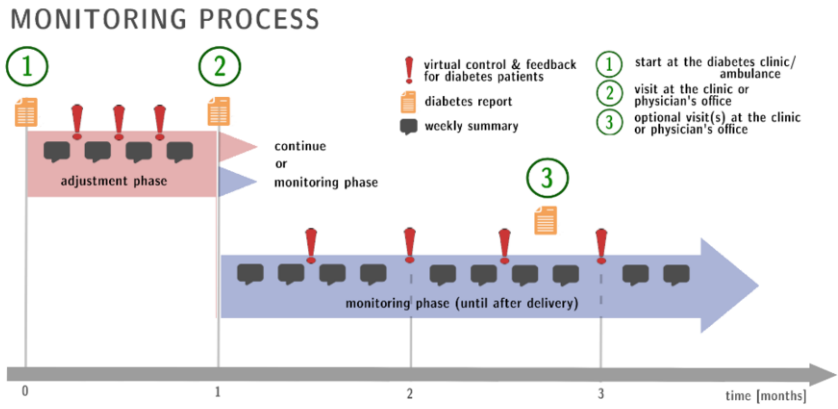


Figure 1: Monitoring process of GDM patients through DiabCare Tirol.

For each phase, a monitoring schedule (weekly or bi-weekly) for virtual checkups by diabetes counselling nurses and, if necessary, additional nutritional counselling was established. After one month, a physician visit was scheduled for interim control. For GDM patients, the participation ended after delivery. In case of an acute medical situation, patients were advised to contact their family or emergency doctor directly.

3. Data Collection and Parameter Analysis of DiabCare Tirol

Starting with the first wave of COVID-19 cases in the first quarter of 2020 in Austria, the DiabCare concept was increasingly used. From March 2020 to June 2020, 29 gestational diabetics were included in DiabCare of which 27 were incorporated in the analysis, as two patients dropped out early from the program. On average, GDM patients had a mean duration of follow-ups lasting 9.6 (sd = 6.1) weeks and were between the ages of 20 to 38 (32.6, sd = 5.6) years at the time of inclusion - as summarized in Table 1 - with various forms of therapy, ranging from oral antidiabetics to insulin.

The total number of BG values transmitted was 8.918, equivalent to 34,4 data transmission weekly, and that of blood pressure (BP) and pulse values amounted to 878 transmissions, equivalent to 3,4 weekly ones. The patients received feedback two times a week on average from a diabetes consultant. Communication between involved caregivers (physicians, specialized nurses, dieticians) took place via notes in the data management system with restricted access for authorized users.

Table 1. Descriptive data of all participants

Number of participants	27
Type of diabetes	gestational diabetes mellitus
Age	32.6 (sd = 5.6) years
Mean duration of participation	9.6 (sd = 6.1) weeks
Number of blood glucose measurements	8.918 (~34 per week)
Number of blood pressure measurements	878 (~3-4 per week)
Number of received feedback	2 per week on average

The mean heart rate of the included gestational diabetic women equaled 84 bpm (sd = 13 bpm) at the beginning. An increase in heart rate to an average of 93 bpm was detected, which is known through literature to be normal in pregnant women. The weekly mean systolic BP value was 123 mmHg (sd = 14 mmHg), and diastolic value was 80 mmHg (sd = 11 mmHg). Over the observation period, there was an increase in BP typical for pregnant women [10]. A total of 8.918 glucose self-measurements was obtained, with an overall mean of 112 mg/dl (sd = 32 mg/dl). As shown in the boxplot-analysis of weekly mean BG values in Figure 2, the readings are largely within an adequate range except for a few outliers (weeks 2, 4 and 13). The standard deviation became narrower over the duration of care (weeks 14-16), where i.e., fewer hyper- and hypoglycemic values occurred. Further analysis showed a significant reduction in BG levels (p-value < 0.001). The patients were managed well within the individual limits set by the physician and remained largely within the target range.

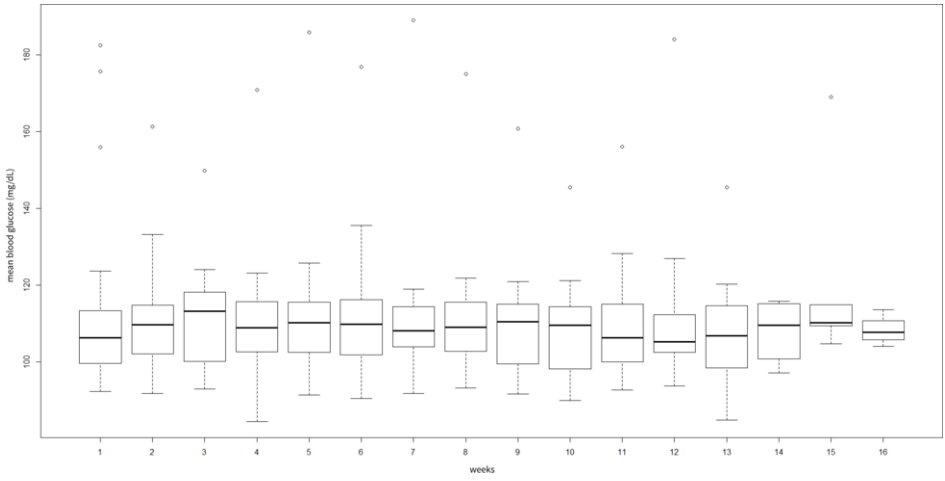


Figure 2: Boxplot - analysis of weekly mean blood glucose values of all gestational diabetic women during the observation period of 16 weeks.

4. Randomized Controlled Trials (RCTs)

Sung et al. [11] reported on a randomized controlled trial (RCT), conducted throughout four months in 2017, where 21 women diagnosed with GDM at 24-28 weeks of gestation were recruited and put into two groups at a ratio of 1:1. The mobile management (MM) group was trained on a telemonitoring mobile phone application (Huraypositive Inc. Korea) to record their BG values regularly. Patients received feedback and data were controlled with tailored medical and nutritional guidance. In the control group for conventional management (CM), patients were allowed to access the application and review their information but were not observed by specialists, as the MM group was. All participants were asked to adhere to regular prenatal visits for usual care. All patients had full-term deliveries. The results measured 4-12 weeks postpartum showed significantly lower values for the median maternal BMI (23.72 (CM) versus 20.22 (MM)), weight (62.58 (CM) versus 54.31 (MM)) and percentage of body fat (38.12% (CM) versus 29.20% (MM)) in the MM group. No difference was observed regarding the diagnoses of DM2, with two diagnosed patients per group post-delivery.

A multicentre RCT, where women with GDM - diagnosed between 23 and 30 gestational weeks - were randomized to a WeChat group chat-based BG management group or a routine clinical prenatal care group, was analyzed by Tian et al. [12]. A total of 309 women with GDM participated in the trial, with 162 women randomized to the control group and 147 to the intervention group. Results show that the glycemic qualification rate of the intervention group was better than that of the control group at nearly all times with gradually increased glycemic qualification rate as gestational weeks progressed in both groups, regardless of the intervention method.

In another, similar RCT (n = 120, 60 per group) with the same group allocation, Miremberg et al. [13] reported that early intervention for the management of GDM through a mobile feedback system from the time of diagnosis led to improvement in glycemic control (low mean BG: 105.1 mg/dL versus 112.6 mg/dL) and a low rate of insulin-dependent patients (13.3% versus 30.0%).

5. Discussion

Telemedicine has been transforming conventional diabetes care by increasing the use of technology, managing information in an electronic medium, and allowing patients and physicians to communicate despite different schedules. The use of telemedical support allows patients' conditions to be assessed and monitored more often than with conventional clinical procedures. The goal is to create a monitored autonomy for the patient, allowing them to make their own decisions with professional remote counselling and infrequent practice visits in case of need. This is exactly what the DiabCare program is designed to provide. The results and statistical evaluations confirm the validity and effectiveness of the methodology used and at the same time, motivate the continuation and expansion for the benefit of all diabetics in Tyrol. Results from RCTs investigating comparable methods reported by Sung et al. [11], Tian et al. [12] and Miremberg et al. [13] sustain the use of telemonitoring as treatment support for GDM patients.

In conclusion, telemonitoring with DiabCare showed benefits for the patients, due to the significantly positive progression in terms of therapy management for BG. Throughout the COVID-19 pandemic, digital DMPs gained even more importance since they encouraged active patient care and reduced ambulance and practice visits. DiabCare

Tirol was able support patients, especially in the early stages of the disease, by enabling diabetologists and physicians to react quickly to the vital signs and BG data transmitted by the patient and to adjust or optimize therapy if necessary. However continuous outcome research with larger patient numbers will be necessary to confirm the effectiveness of telemonitoring in a regular care setting.

References

[1] Ogurtsova K, da Rocha Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, Cavan D, Shaw JE, Makaroff LE. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040, Diabetes Research and Clinical Practice, Volume 128, 2017, Pages 40-50, ISSN 0168-8227, https://doi.org/10.1016/j.diabres.2017.03.024.

[2] Khalil C. Understanding the Adoption and Diffusion of a Telemonitoring Solution in Gestational Diabetes Mellitus: Qualitative Study. JMIR Diabetes. 2019 Nov 28;4(4):e13661. doi: 10.2196/13661. PMID: 31778118; PMCID: PMC6913512.

[3] García-Sáez G, Rigla M, Martínez-Sarriegui I, Shalom E, Peleg M, Broens T, Pons B, Caballero-Ruíz E, Gómez EJ, Hernando ME. Patient-oriented Computerized Clinical Guidelines for Mobile Decision Support in Gestational Diabetes. J Diabetes Sci Technol. 2014 Mar;8(2):238-246. doi: 10.1177/1932296814526492. Epub 2014 Mar 6. PMID: 24876573; PMCID: PMC4455422.

[4] Zahmatkeshan M, Zakerabasali S, Farjam M, Gholampour Y, Seraji M, Yazdani A. The use of mobile health interventions for gestational diabetes mellitus: a descriptive literature review. J Med Life. 2021 Mar-Apr;14(2):131-141. doi: 10.25122/jml-2020-0163. PMID: 34104235; PMCID: PMC8169150.

[5] Rasekaba TM, Furler J, Blackberry I, Tacey M, Gray K, Lim K. Telemedicine interventions for gestational diabetes mellitus: A systematic review and meta-analysis. Diabetes Research and Clinical Practice. 2015 Oct;110(1):1–9.

[6] Raman P, Shepherd E, Dowswell T, Middleton P, Crowther CA. Different methods and settings for glucose monitoring for gestational diabetes during pregnancy. Cochrane Database Syst Rev. 2017 Oct 29;10(10):CD011069. doi: 10.1002/14651858.CD011069.pub2. PMID: 29081069; PMCID: PMC6485695.

[7] Nouhjah S, Jahanfar S, Shahbazian H. Temporary changes in clinical guidelines of gestational diabetes screening and management during COVID-19 outbreak: A narrative review. Diabetes Metab Syndr. 2020 Sep-Oct;14(5):939-942. doi: 10.1016/j.dsx.2020.06.030. Epub 2020 Jun 18. PMID: 32593935; PMCID: PMC7299858.

[8] Ammenwerth E, Modre-Osprian R, Fetz B, Gstrein S, Krestan S, Dörler J, Kastner P, Welte S, Rissbacher C, Pölzl G. HerzMobil, an Integrated and Collaborative Telemonitoring-Based Disease Management Program for Patients With Heart Failure: A Feasibility Study Paving the Way to Routine Care. JMIR Cardio. 2018 Apr 30;2(1):e11. doi: 10.2196/cardio.9936. PMID: 31758765; PMCID: PMC6857958.

[9] Poelzl G, Egelseer-Bruendl T, Pfeifer B, Modre-Osprian R, Welte S, Fetz B, et al. Feasibility and effectiveness of a multidimensional post-discharge disease management programme for heart failure patients in clinical practice: the HerzMobil Tirol programme. Clin Res Cardiol [Internet]. 2021 Jul 16 [cited 2021 Sep 5]; https://link.springer.com/10.1007/s00392-021-01912-0

[10] Hunter S, Robson SC. Adaptation of the maternal heart in pregnancy. Br Heart J. 1992 Dec;68(6):540-3. doi: 10.1136/hrt.68.12.540. PMID: 1467047; PMCID: PMC1025680

[11] Sung J-H, Lee DY, Min KP, Park C-Y. Peripartum Management of Gestational Diabetes Using a Digital Health Care Service: A Pilot, Randomized Controlled Study. Clinical Therapeutics. 2019 Nov;41(11):2426–34.

[12] Tian Y, Zhang S, Huang F, Ma L. Comparing the Efficacies of Telemedicine and Standard Prenatal Care on Blood Glucose Control in Women With Gestational Diabetes Mellitus: Randomized Controlled Trial. JMIR Mhealth Uhealth. 2021 May 25;9(5):e22881. doi: 10.2196/22881. PMID: 33783365; PMCID: PMC8188321.

[13] Miremberg H, Ben-Ari T, Betzer T, Raphaeli H, Gasnier R, Barda G, et al. The impact of a daily smartphone-based feedback system among women with gestational diabetes on compliance, glycemic control, satisfaction, and pregnancy outcome: a randomized controlled trial. American Journal of Obstetrics and Gynecology. 2018 Apr;218(4):453.e1-453.e7.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210600

A Web Based Tool to Support a Personalized Therapeutic Path Through the Use of Psychological Tests

Elena LAZAROVA[a, 1], Sara MORA[a], Davide ARMANINO[b], Alessandro POIRE[b], Fabio FURLANI[c], and Mauro GIACOMINI[a]
[a] *Department of Informatics, Bioengineering, Robotics and System Engineering (DIBRIS), University of Genoa, Italy*
[b] *Genoa Constructivist Association, Genoa, Italy*
[c] *Como Cognitive Therapy Center, Via Rusconi 10, Como, Italy*

Abstract. Psychology tests are tools used to evaluate specific aspects of individual psyche and behavior, in clinical practice and for research purposes. They are validated and standardized both in the administration procedures and interpretation of results. Nowadays, most of these tests are questionnaires administered to patients on paper support. The patient's answers to the questionnaires constitute a basis for self-presentation and self-awareness at the beginning of the therapeutic path. The computer is a valuable aid allowing a quickly consultation of all the answers and highlighting the most salient ones. The main aim of this work was to design, develop and test a computerized support tool for the interpretation of psychological tests that allow good interaction between groups of therapists sharing the same operating modes. The developed system allows: first the storing of the numerical values corresponding to the answers to the questionnaires of the patients; then it creates a complete 'Picture' of the patient and allows the automatic computation of the correlation between the indexes of the various scales. The graphical correlations between scales can be also a valuable aid in finding the outliers, so patients far away from the trend line.

Keywords. Psychological test interpretation, web based professional collaboration, correlation matrix, outliers identification, Z-score normalization

1. Introduction

Psychology tests are tools used to evaluate specific aspects of individual psyche and behavior, in clinical practice and for research purposes. They are validated and standardized both in the administration procedures and interpretation of results. They rely on statistical methods that allow the performance comparison between the tested group of people and the controls. Psychological tests can be divided into two main categories: personality tests, aimed at identifying the individual's personality traits, and efficiency tests, aimed at investigating the individual's knowledge and skills. Nowadays, most of these tests are questionnaires administered to patients on paper support. The patient can choose within a set of fixed answers. This choice, widely shared by virtually

[1] Corresponding Author: Elena Lazarova, Department of Informatics, Bioengineering, Robotics and System Engineering (DIBRIS), University of Genoa, Italy; E-mail: elena.lazarova@dibris.unige.it

the entire scientific community of psychologists, is intended to obtain the most truthful possible answer with- out being influenced by the eventual subjection to any technological tool used to replace paper. The patient's answers to the questionnaires constitute a basis for self-presentation and self-awareness at the beginning of the therapeutic path. As the process goes on, the patient and the therapist review the most significant answers together, interpreting their meaning in an even deeper way, with knowledge and self-awareness that is broader and more meaningful for the patient himself. The computer, in its simplicity of operation and immediate usability, is a valuable aid allowing a rapid consultation of all the answers and highlighting the most salient ones. In this way the patient re-appropriates the symptom and his initial emotive discomfort by translating and progressively declining it as an emotional, cognitive and behavioral style of functioning. From the strangeness of the symptom, the patient moves on to understanding his personal initial functioning and grasps its adaptive, albeit painful and not optimal, meaning. The diagnostic tool, when strategically shared, becomes the basis of the therapeutic alliance based on transparency, sharing and mutual trust in the reading and understanding of the patient's way of being, as well as the direct object of therapeutic work along the way. This modality of setting up the therapeutic work also appears to be fully consistent with a dimensional vision of the diagnosis, overcoming an exclusively categorical, descriptive and nosographic approach. It is not a question of understanding the presence of a psychopathological syndrome but of grasping the patient's level of functioning (mobile, variable and slowly modifiable) along dimensions that describe styles and methods of adaptation. The main aim of the work presented here was to design, develop and test a computerized support tool for the interpretation of psychological tests that allow good interaction between groups of therapists sharing the same operating modes. The produced system allows the entry and orderly filling of clinical questionnaires for the evaluation of personality traits and dimensions, early maladaptive patterns, salient emotional and cognitive styles, as well as the main symptoms of the nosographic spectrum and the quality of the functioning of consciousness with respect to the presence, type, and quantity of dissociative symptoms in the patient. The set of these data composes a multifaceted and in-depth clinical "picture" of the patient immediately usable by the clinician, who consults it not only in the diagnostic phase while setting the therapeutic strategy but also along the treatment path, even directly during the psychotherapeutic session.

2. Methods and Materials

2.1. Selected Psychological Tests

As part of this research project, the included psychological tests are:

1. *Young Schema Questionnaire version L3* (YSQ-L3): identifies maladaptive patterns and builds dynamic profile of the patient.
2. *Personality Inventory Domain* (PID-5): a personality inventory which supports diagnosis process according to the: *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5).
3. *Symptom Checklist 90* (SCL-90): is a checklist of diagnostic relevance symptoms.
4. 'Questionario per la valutazione delle organizzazioni di Significato Personale' (QSP): helps to find out which meaning organization belongs to a person.

For each of these tests we collected both raw data and we calculated the summary indexes according to the tests guidelines. These indexes are shown to help the psychologists in order to have the 'Picture' of the patient at a glance.

2.2. Data Sharing Platform

This project originates from the idea of a group of psychologists who would like to improve their daily therapeutic routine with an IT-supported data sharing, since web platforms are the easiest and most effective tools for data sharing in the scientific and medical field [1,2,3,4]. An interactive web site was developed to give telematics support to type shared patient evaluations.

It is composed of two main parts: a SQL Server Database (DB), that stores anamnestic information and patient test results, and a website, which guarantees an easy and re- mote interaction with the DB. The access to patient's data is restricted by a login page and it is reserved to the psychologists involved in the project. The architecture is similar to other solutions developed in the laboratory of some of the authors, that are also routinely used in clinical practice [5,6].

2.3. Characteristics of the Sample

In the period May 2017 - May 2021, a group of 248 patients were involved (130 women and 118 men) and the total number of compiled tests are 902. No discrimination was made regarding gender or occupation and the mean sample age was 41 years (range 30-52). Questionnaires were first completed by patients on paper and then imported into the web platform by psychologists thanks to the support of pre-arranged spreadsheets.

2.4. System of Anonymization

The data collected for this research project were completely anonymous.

The patient is uniquely identified within the platform through a progressive number. The only anagraphic information stored about the patient are the birth sex and the year of birth. Each clinician maintains a matching between the unique identifier of the patient given by the platform and his anagraphic information. It's responsibility of the single physician to guarantee that no duplicates are inserted in the platform. Given the close cooperation with the doctors that enter the data, the authors are sure that this task is done correctly. Since the data is completely anonymous and unidentifiable, the GDPR does not apply.

2.5. Missing Data Management

As psychological tests were administered on paper and not all the answers were expressively mandatory, the collected dataset contained some missing values. This problem could be addressed in two ways: listwise deletion (or complete case analysis) by removing all the rows with a missing value from the dataset, or imputation methods by estimating the value of missing data. In order to keep the sample as big as possible, the second option was selected.

According to literature, a generally effective algorithm in this scenario of data imputation is based on *K-Nearest Neighbor* (KNN). It imputes each sample's missing values using the mean value from n nearest neighbors found in the dataset [7].

2.6. Data Normalization

Considering that the numerical values associated to each answer belonged to different scales, the second mandatory step in data pre-processing was normalization. In particular the chosen algorithm for data normalization was based on the Z-score:

$$Z = \frac{x - \mu}{\sigma}$$

where μ is the mean and σ is the standard deviation.

The Z-score gives an idea of how far from the mean a data point is. It's a perfect strategy to deal with outliers because it measures how many standard deviations are below or above the population mean and raw score.

3. Results

3.1. Web Tool

After an authentication phase, the user can:

1. Decide if he/she wants to visualize data of a specific patient who has already been enrolled into the platform or add a new patient.
2. Store the scores of psychological tests through pre-formatted spreadsheets where the user has already manually inserted the numerical values corresponding to the patient's answers to the test.
3. Visualize a complete 'picture' containing all the relevant information about the patient and a summary of the scores obtained in each test ((Figure 1).

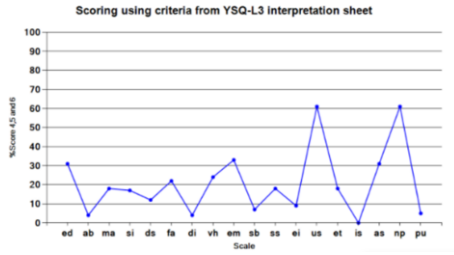
3.2. Correlation between scales

The developed system allows the automatic computation of the Pearson correlation coefficient between the indexes of the various scales. We considered 4 tests with a total number of 59 scales. In order to investigate the relations among this scales we calculated a complete correlation matrix. After a screening of this correlation matrix (1711 relations), we counted each correlation within the intervals defined by Akoglu in [8]. The obtained counts are summarized in Table 1.

Tratti di Personalità	Num_off_Quest	Totale	Average
Anedonia	8	8	1
Ansia	9	18	2
Ricerca di Attenzione	8	14	1,75
Insensibilità	14	1	0,07
Inganno	10	15	1,5
Depressività	14	7	0,5
Distraibilità	9	3	0,33
Eccentricità	13	1	0,08
Labilità Emotiva	7	9	1,29
Grandiosità	6	6	1
Ostilità	10	8	0,8
Impulsività	6	0	0
Evitamento dell'Intimità	6	1	0,17
Irresponsabilità	7	3	0,43
Manipolatorietà	5	9	1,8
Disregolazione Percettiva	12	3	0,25
Perseverazione	9	3	0,33
Affettività Ridotta	7	8	1,14
Perfezionismo Rigido	10	12	1,2
Tendenza a correre Rischi	14	8	0,57
Angoscia di Separazione	7	6	0,86
Sottomissione	4	2	0,5
Sospettosità	7	6	0,86
Convinzioni ed Esperienze Inusuali	8	11	1,38
Ritiro	10	7	0,7

(a) PID-5 results table

Disturbi QSP	Risultati
DAP	62
FOB	61
OSS	54
DEP	37



(b) The upper table refers to the personality traits extracted from the QSP questionnaire while the graph refers to the results of YSQ-L3 questionnaire.

Figure 1. Example of the personalized 'Picture' available for each patient and obtained collecting the summary results from the considered tests.

Table 1. Counts from correlation matrix.

Correlation	Interval	Count
Very strong positive correlation	0.90 to 1.00	0
Strong positive correlation	0.70 to 0.89	20
Moderate positive correlation	0.40 to 0.69	371
Weak positive correlation	0.20 to 0.39	760
Very weak positive correlation	0.00 to 0.19	409
Very weak negative correlation	-0.19 to 0.00	61
Weak negative correlation	-0.39 to -0.20	5

The minimum negative correlation found was -0.35.

4. Discussion

During these years of application with the working group of some of the authors, they found that the reading of the exceptions is very useful. Specifically, an exception can be defined as the presence of unexpected correlations or moreover, as the absence of expected correlations (that are present in other patients in the sample with the same characteristics). For example, a patient who has a borderline mode of operation and who, unlike the norm, does not have traits of impulsiveness or emotional lability will immediately attract the attention of the clinician.

This kind of information can be achieved in a simple and quick way looking at the 'picture' that fully characterizes the individual patient. This attention translates into an attitude of research together with the patient of his specific functioning, with an

attitude of respectful and real curiosity towards their personal way of being in addition to the diagnosis. These tools therefore allow the overcoming of the factors that unite the various patients, which becomes only a starting point for the therapeutic exploration of the specific characteristics of the patient. This in-depth work, in our opinion, is very valuable and innovative and would be impossible without the aid of the IT tool which, in addition to the immediate calculation of the correlations and their statistical significance, translates the correlation of the two variables taken into consideration with a graph giving the clinician an immediate vision of the patient with respect to others. This also makes it possible to detect similarities with other patients in the sample and therefore their comparison and to search for similarities and differences in the profile and style of functioning or in the anamnesis and life history. Examples of the previous are patients who have suffered a trauma or a precocious abandonment, significant emotional separations, significant medical pathologies, etc. Lastly, it should be noted that all patients in the sample were followed by therapists and they were clinical cases of which the therapist had deep knowledge and documentation.

5. Conclusions

It is our belief that the exploration of the potential of the IT tool as a clinical aid is only at the beginning. The graphic simplicity of consultation, the reliability of the tool and the ease and immediacy of consultation are its winning features. The calculated correlation indexes are under accurate study by a group of relevant psychologists (some of them are authors of this work), in order to define synergies and competitions between the used tests to improve their efficacy in the personalized diagnosis and treatment. The complexity and the amount of data that can be derived from it are the most promising characteristics for future research.

References

[1] Giacomini M, et al. Data modeling for tools and technologies for the analysis and synthesis of NANOstructures (TASNANO) project., Journal of Information Technology Research (JITR) 2.3 (2009): 49-70.
[2] Henrie A, et al. ClinVar Miner: Demonstrating utility of a Web-based tool for viewing and filtering ClinVar data., Human mutation 2018; 39(8): 1051-1060.
[3] Dyrlund TF, et al. MS Data Miner: A web-based software tool to analyze, compare, and share mass spectrometry protein identifications., Proteomics 2012; 12(18): 2792-2796.
[4] Hamosh A, et al. P heno DB: A New Web-Based Tool for the Collection, Storage, and Analysis of Phenotypic Features., Human mutation 2013; 34(4): 566-571.
[5] Giannini B, et al. From Liguria HIV Web to Liguria infectious diseases network: how a digital platform improved doctors' work and patients' care., AIDS research and human retroviruses 2018; 34(3): 239-240.
[6] Bonetto M, et al. I-Maculaweb: a tool to support data reuse in ophthalmology., IEEE journal of translational engineering in health and medicine 2015; 4: 1-10.
[7] Dong X, et al. TOBMI: trans-omics block missing data imputation using a k-nearest neighbor weighted approach. Bioinformatics 2019; 35(8): 1278-1283.
[8] Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med. 2018;18(3):91-93. Published 2018 Aug 7.

Challenges and Solutions in Health Education and Training

This page intentionally left blank

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210602

Development of a Didactic Online Course Concept for Heterogeneous Audience Groups in the Context of Healthcare IT

Matthias KATZENSTEINER[a,1], Stefan VOGEL[b], Jens HÜSERS[c], Jendrik RICHTER[b], Johannes HÖLKEN[d], Natalia LESNIEWSKA[e] and Oliver J. BOTT[a]
[a] *University of Applied Sciences and Arts Hanover, Germany*
[b] *Department for Medical Informatics, University Medical Center Göttingen*
[c] *Health Informatics Research Group, University AS Osnabrück, Germany*
[d] *University of Applied Sciences and Arts Hildesheim/Holzminden/Göttingen, Gesundheitscampus Göttingen*
[e] *Digital Learning and Teaching, University of Göttingen*

Abstract. Building a well-founded understanding of the concepts, tasks and limitations of IT in all areas of society is an essential prerequisite for future developments in business and research. This applies in particular to the healthcare sector and medical research, which are affected by the noticeable advances in digitization. In the transfer project "Zukunftslabor Gesundheit" (ZLG), a teaching framework was developed to support the development of further education online courses in order to teach heterogeneous groups of learners independent of location and prior knowledge. The study at hand describes the development and components of the framework.

Keywords. Didactic, Health IT, Citizens, eLearning, Digitalization, Digitization, Patient empowerment

1. Introduction

Digitalization is reaching all areas of society. [1] Not only since the Corona crisis has it become apparent that IT systems, as well as IT competencies, more and more unfold themselves as basic requirements for everyday work, personal comfort, and healthcare, e.g. in the guise of video conferences, kinds of tracking systems, or smart home and ambient assisted living infrastructures. A sensitive approach to these issues is required. [1, 2] Therefore, society must be empowered to understand IT, its concepts, and limits. [1, 2] For this, a learning program, and a continuous learning process for digital competencies are required for citizens throughout society. [3] Especially in the health sector, the IT empowerment of health care professionals and especially patients could contribute to new ways of care and supervision in the future. [4]

[1] Corresponding Author, Matthias Katzensteiner, University of Applied Sciences and Arts Hanover, Expo Plaza 12, 30539 Hanover, Germany; E-mail: Matthias.Katzensteiner@hs-hannover.de

In this research study, we address how a teaching and learning concept should be designed in order to provide such a heterogeneous group of learners with competences for digitalization.

2. Methods

The Zukunftslabor Gesundheit (ZLG) is a transfer project focusing at digitization in health care. Its educational subproject aims to spread the gathered knowledge from the other co-projects within medical and medical informatics professionals, patients and the general public alike based on online teaching. To understand the needs of this audience in terms of didactic concepts, we split them into coherent sub-groups using the following characteristics to describe them further: Sociographic factors as age, gender, place of residence, and family status. Psychographic factors as motivation, potentials, expectations, strengths, and weaknesses. Educational factors as educational background, an affinity for (digital) media, learning type, and discussion type. [5] These factors only partially describe the target groups. Therefore, we additionally used the sinus milieu model, a market research tool, to narrow down the groups even further. [6]

With this description of our targeted audience groups, we started to adapt common models of modern online didactic, esp. the five-stage model by Gilly Salmon [7], to their needs and to the planned content of our courses. Based on all these factors and models, we then selected the tools needed to build effective eLearning courses.

All courses are to be taught twice during the project period. With the evaluation concept, we aim at fine-tuning the didactic concept, understanding of the audience and course concepts with each iteration.

3. Results

3.1. Audience Groups

We divided the heterogeneous target group in four main audience groups:

Citizens and Interested Public to emend their digital and media literacy with respect to digitization in healthcare. Members of this group will participate voluntarily and therefore have a strong intrinsic motivation to learn about current developments in the digital health sector.

Patients, their carers and relatives are directly affected as potential end-users to emend their digital and media literacy regarding digitalization in healthcare. This group has a strong extrinsic motivation to learn about current specific (w.r.t their disease) developments in the digital health sector.

Healthcare Professionals as potential users of research results and prototypes of the ZLG will also benefit from the courses. Depending on the profession, members of this group will face new technologies and will have to implement new evidence in their daily job routines.

Medical informatics professionals to empower them to use the ZLG research results for further research and development.

The latter two groups allow for a relatively fine-grained description in terms of psychographic and educational factors. These descriptions are well known from experiences in using eLearning for university teaching. However, this is not the case for

the sub-groups of the general public. Here we had to develop an open didactic concept that is also suitable for addressing these heterogeneous groups.

3.2. The ZLG Didactic Core Concept (ZLG-DCC)

The development and implementation of an online course is time-consuming and characterized by many influencing factors. For our core didactic concept, we have limited our focus to four such factors: group size, group dynamics and communication, learning and working phases, as well as type and scope of supervision.

Based on these influencing factors, an adapted course structure was developed as a template (see Figure 1) according to the didactic concepts of Gilly Salmon [7] and the HiGHmed teaching project [8,9], which should support the development of online courses.

One main element of the template are so-called *E-tivities*. According to Gilly Salmon [7], E-tivities are standardized in the structure: introduction, objective, task, discussion. With the E-tivites, a constructivist approach to learning is followed step-by-step, starting with the pure exchange of information, to the construction of knowledge and finally to personal development.

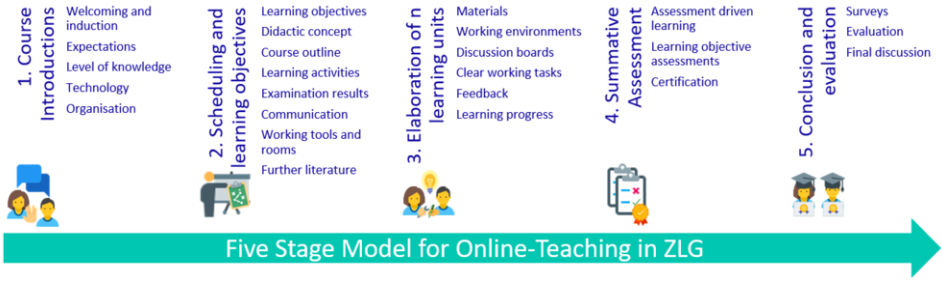


Figure 1. Five Stage Model for Course Template ZLG based on [7-9].

The template for course development contains five essential phases, each of which has to be prepared by the lecturer.

The first phase provides a shallow course introduction in which learners get to know the new learning environment, the other learners, and the lecturers. Organizational as well as technical framework conditions are clarified, but also expectations and the knowledge level of the participants play a role in this phase.

In **the second phase**, the course schedule and the learning objectives to be achieved by the participants are communicated. The contents of the course also concern the expected performance of the participants but also of the lecturers.

Phase three is the core of the didactic concept in which the participants go through *n* learning units, each of which is designed through topic-related tasks, i.e. in the form of E-tivities. The control of the achieved goals through the fulfillment of the tasks takes place in **phase four**.

The **last phase** is dedicated to evaluation, lessons learned and a wrap-up discussion.

Depending on the target group, the size of the learning group and the scope of the course, the ZLG-DCC is used in a reduced form.

3.3. Tools

The concept demands tools for content management and sharing as well as interacting with participants. A suitable learning management tool (LMS) is the key to deliver these requirements. For the presented concept, the LMS should support user management (to provide access to the content), a forum (enabling interaction between participants and lecturers), provide methods to present content (e.g., learning modules) and features for surveys and exams. Offering additional synchronous interactions to students through the LMS, e.g., video conferencing and chats, promises to engage and motivate participants in an asynchronous learning environment.

In ZLG we gathered 32 requirements in total to assess and evaluate different LMS. In the end, ILIAS[2] was selected. For synchronous interaction, BigBlueButton[3] (BBB) was designated as video conferencing tool for data protection reasons.

For the creation of online teaching material, even more tools (e.g. video cutting and editing) might be useful, but are out of scope of this presentation.

3.4. Evaluation

In general, evaluation of educational programs aims to monitor the quality by the participants [10]. The evaluation results then serve as the foundation for further course development. The ZLG courses are conducted at least twice during the project. Thus, the evaluation is an integral part of each course to edit the courses according to the participant's feedback to improve their quality.

We designed an evaluation framework, from which different evaluation strategies can be derived. The framework considered three dimensions: type, method, and objects of evaluation. The following section will briefly introduce them.

First, the **evaluation type** defines if the evaluation is formative, i.e., ongoing throughout the running course, or summative, i.e., final assessment when the course ends. Second, the **evaluation method** is either quantitative or qualitative. Both methods offer distinct advantages, e.g., to openly assess the subjective experiences of learners with more degrees of freedom, qualitative evaluation should be preferred [11]. However, if we want to assess general satisfaction in a standardised way, a quantitative survey may better meet the need of the evaluation goals. Third, the **evaluation object** focuses on the evaluated entity, e.g. course content, teachers and E-tivities.

From this framework, we derived a core evaluation that is part of each ZLG educational program. The core evaluation allows monitoring of the course quality across the ZLG. However, the standardized toolbox also supports course developers in extending the core evaluation and designing an evaluation plan to meet their specific needs.

4. Discussion

The didactic concept outlined poses various challenges, from creating of a learning course to its implementation and evaluating of the participants' performance and the course quality. The strength of the concept lies in the individual discussion and

[2] https://www.ilias.de/
[3] https://bigbluebutton.org/

interdisciplinary exchange in a learning group. In this context, the E-tivities form a common thread, which form a starting point for further targeted development. At the same time, however, the concept also has the weakness of being rolled out in groups of up to a maximum of 30 people per lecturer, leading to challenges when demand is high. Nevertheless, the e-learning format offers the possibility of location-independent learning. A special aspect of the concept is the creation of a common understanding of the teaching contents for the broad and heterogeneous target group of the didactic concept. For the first courses at the ZLG the particular challenge will be to prepare academic, technical but also political topics in a way that the topics can be conveyed in a captivating way and the participants are empowered and sensitized.

5. Conclusion

The first version of the ZLG educational framework is finalized. In the current project phase, the first courses are being developed by the lecturers in order to start the first course implementations in 2022. The course topics will be based on the other working groups of the ZLG and should thus become the bridge between research and society. The first courses will address the topics *"The learning healthcare system: How it learns! - Secondary use of Clinical Data for Medical Research"* and *"Sensor technology in patient support: application examples, perspectives and current research"*. Citizens in particular are to be addressed as audience members in the first phase. As a further subdivision of the target group, students in upper secondary education have been envisaged. All ZLG courses will be evaluated based on a common evaluation set to establish comparability between the course implementations and to further improve the courses.

Acknowledgement

The ZLG is funded by the Lower Saxony Ministry of Science and Culture under grant number ZN3491 within the Lower Saxony "Vorab" of the Volkswagen Foundation and supported by the Center for Digital Innovations (ZDIN).
We thank Ursula Hübner, Dagmar Krefting, Christoph Russmann and Marcus Wuttke for their support during the development of the ZLG-DCC.

References

[1] Vuorikari R, Punie Y, Carretero Gomez S and Van Den Brande G. DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: the Conceptual Reference Model. EUR 27948 EN. Luxembourg (Luxembourg): Publications Office of the European Union; 2016. JRC101254
[2] Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. *Mhealth*. 2017;3:38. Published 2017 Sep 14. doi:10.21037/mhealth.2017.08.07
[3] Pettersson, F. On the issues of digital competence in educational contexts – a review of literature. Educ Inf Technol 23, 1005–1021 (2018). https://doi.org/10.1007/s10639-017-9649-3
[4] Calvillo J, Román I, Roa LM. How technology is empowering patients? A literature review. Health Expect. 2015;18(5):643-652. doi:10.1111/hex.12089
[5] Köhn-Ladenburger C. Marketing für LOHAS. Wiesbaden: Springer Fachmedien Wiesbaden; 2013. S. 12-21

[6] Barth B, Flaig BB, Schäuble N, Tautscher M, Hrsg.: Praxis der Sinus-Milieus®: Gegenwart und Zukunft eines modernen Gesellschafts- und Zielgruppenmodells. Wiesbaden: Springer VS; 2018. available at : http://dx.doi.org/10.1007/978-3-658-19335-5.
[7] Gilly Salmon, E-Moderating: The Key to Online Teaching and Learning Paperback, Routledge; 2nd edition (19 July 2013)
[8] education.highmed.org [Internet]. HiGHmed Teaching and Training Programme for Medical Informatics [last checked: 07.09.2021]. available: https://education.highmed.org
[9] Witte ML, Behrends M, Benning NH, Hoffmann I; HiGHmeducation Consortium, Bott OJ. The HiGHmed Didactical Framework for Online Learning Modules on Medical Informatics: First Experiences. Stud Health Technol Inform. 2020 Jun 26;272:163-166. doi: 10.3233/SHTI200519. PMID: 32604626.
[10] Baldwin, Sally J. & Trespalacios, Jesús (2017). Evaluation instruments and good practices in online education. Online Learning 21 (2) doi: 10.24059/olj.v21i2.913
[11] Ghislandi P, Raffaghelli JE, Sangrà A, Ritella G. The Street Lamp Paradox: Analysing Students' Evaluation of Teaching through Qualitative and Quantitative Approaches. ECPS - Educational Cultural and Psychological Studies. 2020. DOI: 10.7358/ecps-2020-021-ghis

mHealth Applications

This page intentionally left blank

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210604

Social Media Chatbot for Increasing Physical Activity: Usability Study

Dillys LARBI[a,1], Elia GABARRON[a,b], Kerstin DENECKE[c]
[a] *Norwegian Centre for E-health Research, Tromsø, Norway*
[b] *Faculty of Education, Østfold University College, Halden, Norway*
[c] *Institute for Medical Informatics, Bern University of Applied Sciences, Bern, Switzerland*

Abstract. Fully automated self-help interventions integrated with social media chatbots could serve as highly cost-effective physical activity promotion tools for a large population. We have developed MYA, a Telegram-based chatbot for increasing physical activity. The objective of this study was to assess the usability of MYA. To identify usability issues, we recruited volunteers and asked them to interact with MYA and to answer the Chatbot Usability Questionnaire. Thirty volunteers participated in the study, 83.3% agreed MYA was welcoming during initial setup and 63.3% agreed MYA was very easy to use. MYA was perceived as realistic and engaging, easy to navigate, and its responses were useful, appropriate, and informative (all 53.3%). However, 63.3% of respondents agreed MYA failed to recognize most of their inputs, and 43.3% claimed it would be easy to get confused when using MYA. Although the results are encouraging, it remains unclear if a social media chatbot can motivate people to increase their physical activity. MYA has the potential to do that, with improvements in functionalities like challenge personalization. The efficacy of these approaches should be studied in a clinical trial.

Keywords. Chatbot, social media, physical activity, health, participatory health

1. Introduction

The growing burden of chronic diseases highlights the urgent need to increase physical activity [1] through low-cost preventive interventions for large populations. Regular physical activity is important for disease prevention [1,2], better disease management, and improvement in overall health and quality of life [1]. However, engaging large adult populations in public health interventions for increasing exercise behavior remains challenging.

Digital technology interventions can successfully increase physical activity among adults [3-6]. Higher levels of engagement and effectiveness are associated with interventions that incorporate social media [7]. Fully automated self-help interventions integrated with social media chatbots could serve as highly cost-effective physical activity promotion tools. Chatbots are easy to use and do not require familiarity with a specific user interface. The use of a chatbot is mainly through a text-based dialogue where the communication aims to reach a goal, such as an increase in physical activity.

[1] Corresponding Author: Dillys Larbi, Norwegian Centre for E-health Research, University Hospital of North Norway, Sykehusveien. 23 9019 Tromsø, Norway; E-mail: dillys.larbi@ehealthresearch.no.

Although the use of chatbots for health-related purposes is still an emerging field [8,9], its benefits related to increase in physical activity and user satisfaction have already been shown [10-12]. Thus, a chatbot on a social media platform (e.g., Telegram) could provide an inexpensive and widely available health intervention that appeals to a large population.

The objective of this study was to assess the usability of MYA, a social media-based chatbot prototype we designed to increase physical activity among adults.

2. Methods

2.1. Development of the social media chatbot

The chatbot was developed in 4 steps: strategy planning, design, implementation, and testing. As part of the strategy planning, we interviewed Psychology and Public Health experts about the intervention's requirements. The following functional requirements were specified: the chatbot should motivate a user to be physically active, communicate with a user, engage the user in an interesting conversation through a variety of conversation flows, have a rating feature, and offer exercise options. In addition, these non-functional requirements were specified: the chatbot should be rule-based, accessible through a social media platform, integrated with an activity tracker or step counter and retrieve the number of steps , and the chatbot/user dialogue should be stored and accessible. Furthermore, the chatbot should be designed as a friendly and empathic backslapper that informs a user of his/her current number of steps and encourages him/her to increase the daily step count.

We created a social media chatbot called MYA (Figure 1). To model the chatbot's input and its possible reactions, i.e., to create the dialogues, we used Business Process Model and Notation. Nine conversation flows were designed using Hillary Black's [13] conversation interface design template: first encounter, further encounter, menu, goals, challenges, steps today, facts, chatting, and help. The chatbot was developed using FlowXo [14], Google Sheets, and Telegram: a free cloud-based mobile and desktop social media app (Figure 2). FlowXo offers several functionalities to create, host, and maintain chatbots for social media platforms. The chatbot was integrated into Telegram using Botfather.

Flows, the "brain" of FlowXo chatbots, are activated by a specific trigger like a message sent by a user. This triggers MYA to send a message, picture, or video to the chat. The user can respond using either free text, predefined words, or button selection. The following information is gathered during the interaction and stored on Google Sheets with no encryption: username; current challenge and challenge number; current goal and goal number; daily step goal; and whether it is the first encounter. MYA currently has a simulated activity tracker, i.e., it has no integrated step counter. To make MYA more interesting, there is an integrated "small talk" functionality.

2.2. Usability study

We aimed to recruit at least 26 volunteers to test MYA, as suggested by Holmes et al. [15]. Volunteers aged 18 years or older were recruited from Bern University of Applied Sciences in Switzerland and Norwegian Centre for E-health Research and their affiliates.

Potential participants were contacted via e-mail or Telegram and invited to freely interact with MYA. They were asked to complete one or more of the following tasks: set

a goal; request number of steps; a challenge; and to hear a fact. Participants then answered an anonymous online Microsoft Office Forms survey available in English, German, and Spanish. The survey included the Chatbot Usability Questionnaire (CUQ) [15], which includes questions about the chatbot's personality, onboarding, navigation, understanding, responses, error handling, and intelligence rated on a scale of 1 (strongly disagree) to 5 (strongly agree). We distinguished between the positive and negative items of the CUQ and analyzed the number and its corresponding percentage of respondents answering "Agree" and "Strongly agree" to these.

We summarized the survey and CUQ responses using descriptive statistics. Quantitative data were analyzed with SPSS (version 25; IBM Corp) whereas NVivo 12 Plus was used for the qualitative data (participant open feedback). All collected data were treated confidentially and only used for this study. The study protocol was assessed by the Cantonal Ethics Committee in Bern (BASEC-No: Req-2021-00244).

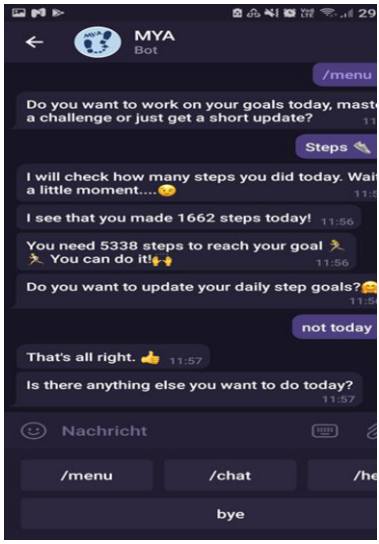


Figure 1. Screenshot of conversation between MYA (dark purple) and user (bright purple)

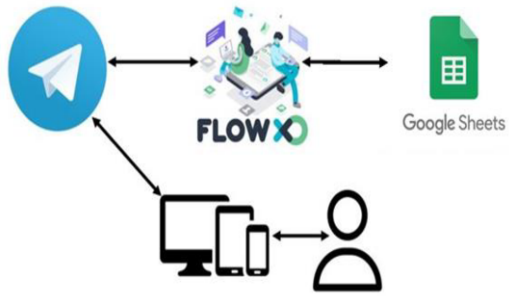


Figure 2. Technology stack of MYA

3. Results

Thirty volunteers, 15 males aged 18 to 69 years and 15 females aged 18 to 49 years, tested the chatbot and answered the survey. Nineteen participants (19/30; 63.3%) reported a chatbot interaction time of 5-15 minutes; 5 participants (5/30; 16.7%) used it for 15-30 minutes; 4 participants (4/30; 13.3%) less than 5 minutes, and 2 participants (2/30; 6.7%) used it for more than 60 minutes. To the question "Do you think MYA could help you in increasing your physical activity/change your activity behavior?", 16 of the respondents chose the option "maybe" (16/30; 53.3%), while 7 answered "yes" and 7 answered "no" (both 7/30; 23.3%).

The best quality attributed to MYA was that it was welcoming during the initial setup (83.3% of respondents agreed or strongly agreed). The next best quality attributed to MYA was that it was very easy to use (63.3%); followed by its personality was realistic

and engaging, easy to navigate, and its responses were useful, appropriate, and informative (all 53.3%). On the other hand, 63.3% of the respondents thought MYA failed to recognize a lot of their inputs, and 43.3% thought it would be easy to get confused when using it (See Table 1).

Table 1. Participants who agreed or strongly agreed with the different items of the CUQ

"Positive" features of the chatbot	**Agree/Strongly agree**
Q3-The chatbot was welcoming during initial setup	25 (83.3%)
Q15-The chatbot was very easy to use	19 (63.3%)
Q1-The chatbot's personality was realistic and engaging	16 (53.3%)
Q7-The chatbot was easy to navigate	16 (53.3%)
Q11-Chatbot responses were useful, appropriate, and informative	16 (53.3%)
Q9-The chatbot understood me well	7 (23.3%)
Q13-The chatbot coped well with any errors or mistakes	6 (20%)
Q5-The chatbot explained its scope and purpose well	2 (6.7%)
"Negative" features of the chatbot	**Agree/Strongly agree**
Q10-The chatbot failed to recognise a lot of my inputs	19 (63.3%)
Q8-It would be easy to get confused when using the chatbot	13 (43.3%)
Q2-The chatbot seemed too robotic	11 (36.7%)
Q14-The chatbot seemed unable to handle any errors	9 (30%)
Q12-Chatbot responses were not relevant	7 (23.3%)
Q6-The chatbot gave no indication as to its purpose	6 (20%)
Q16-The chatbot was very complex	4 (13.3%)
Q4-The chatbot seemed very unfriendly	3 (10%)

The last open-ended question allowed the respondents to write their general comments, including suggestions for improvement. In total, 73.3% (22/30) of the respondents wrote comments which can be grouped into 3 main themes: identified issues, preferred chatbot features, and suggestions for improvement. The identified issues with the chatbot, MYA, included: interaction difficulties; incomplete app design; spelling errors; and unresponsive/frozen app. The preferred features of MYA included the challenge feature and the goal feature. The participants suggested among others, a weekly challenge and wide subject variety for the 'small talk' feature.

4. Discussion

This study assessed the usability of MYA, a social media-based chatbot designed to increase physical activity using the Chatbot Usability Questionnaire (CUQ) [15]. More than 50% of the participants perceived MYA as realistic, engaging, and user-friendly. Compared with the negative features, the chatbot's positive features received higher ratings which suggest a positive participant attitude towards the social media chatbot. The positive attitudes could be a predictor of the chatbot's acceptability, as suggested by Nadarzynski et al. [16]. Like other chatbots aimed at improving physical activity [15, 22], MYA is integrated into a social media platform and does not need any installation, which eases access. MYA's ability to interact with users as a peer and remind them of their physical activity goals could contribute to continued chatbot use [17].

Further developed chatbot conversation flows, and its better interpretation of users' input might increase MYA's usability [15], which in turn could improve its acceptability and the users' confidence in its proper function. Fadhil et al. [18] associate a better chatbot experience that promotes user engagement with a well-designed and implemented chatbot. Since MYA was a prototype with a simulated activity tracker, the identification and reporting of issues by the participants was expected. These identified issues, together with the suggestions for improvement will help improve the features, functions, and design to provide a better user experience.

Approximately half of the participants responded that MYA could maybe help them to increase their physical activity. Among the remaining, half thought the chatbot could not help them while the other half thought it could. Most of the participants who responded that MYA could not help them, identified technical issues with the social media chatbot. Meanwhile those who responded that MYA could help increase their physical activity stated their preferred chatbot features and/or made suggestions for improvement. Previous studies on chatbots showed that they were effective in changing physical activity behavior [19-23]. However, there was a lack of sustained use and the chatbots failed to engage users [19]. Since an increased and sustained engagement has been reported among users of physical activity apps integrated with social media platforms [7], the integration of MYA with Telegram has the potential to engage users for continued use.

Limitations: Answers from the participants might not be comparable to the general population given the selection of participants, the majority of whom may be tech-savvy. Planned improvements include activity tracker integration and extension of MYA's rule-based communication knowledge to enable flexible conversations with users. Additional security issues (storage location, encryption, etc.) will be assessed and addressed. Further research on physical activity chatbots could explore the effect of additional functionalities, behavior change techniques, or its integration with different social media platforms.

5. Conclusion

Social media chatbots have the potential to increase physical activity among their users and through usability studies, all relevant features that increase user experience can be integrated. This paper presented usability test results of the initial prototype of MYA, a social media-based chatbot for increasing physical activity. Although the results are encouraging, it remains unclear if the chatbot can motivate people to increase their physical activity. MYA has the potential to do that with improvements in functionalities like challenge personalization. The efficacy of these approaches should be studied in future randomized controlled trials involving sedentary individuals.

Acknowledgments

The following Medical Informatics students at Bern University of Applied Sciences developed the chatbot as part of a student project: Patricia Romao, Stefanie Neuenschwander, and Apiwat-David Gaupp. We would also like to thank all the volunteers who participated in the study and Dr. Elisa Salvi for her feedback on the final draft.

References

[1] World Health Organization. Global action plan on physical activity 2018-2030. More active people for a healthier world. 2018; Available from: https://apps.who.int/iris/bitstream/handle/10665/272722/9789241514187-eng.pdf.

[2] Ding D, Lawson KD, Kolbe-Alexander TL, et al. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet. 2016;388(10051):1311-24.

[3] Kongstad MB, Valentiner LS, Ried-Larsen M, et al. Effectiveness of remote feedback on physical activity in persons with type 2 diabetes: A systematic review and meta-analysis of randomized controlled trials. J Telemed Telecare. 2019;25(1):26-34.

[4] Kramer LL, ter Stal S, Mulder BC, et al. Developing Embodied Conversational Agents for Coaching People in a Healthy Lifestyle: Scoping Review. J Med Internet Res. 2020;22(2):e14058.

[5] Laranjo L, Ding D, Heleno B, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med. 2021;55(8):422-32.

[6] Tsoli S, Sutton S, Kassavou A. Interactive voice response interventions targeting behaviour change: a systematic literature review with meta-analysis and meta-regression. BMJ Open. 2018;8(2):e018974.

[7] Petersen JM, Prichard I, Kemps E. A Comparison of Physical Activity Mobile Apps With and Without Existing Web-Based Social Networking Platforms: Systematic Review. J Med Internet Res. 2019;21(8):e12687.

[8] Denecke K, Tschanz M, Dorner TL, et al. Intelligent Conversational Agents in Healthcare: Hype or Hope? Stud Health Technol Inform. 2019;259:77-84.

[9] Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc. 2018;25(9):1248-58.

[10] Bickmore TW, Silliman RA, Nelson K, et al. A randomized controlled trial of an automated exercise coach for older adults. J Am Geriatr Soc. 2013;61(10):1676-83.

[11] Gabarron E, Larbi D, Denecke K, et al. What Do We Know About the Use of Chatbots for Public Health? Stud Health Technol Inform. 2020;270:796-800.

[12] Kramer JN, Tinschert P, Scholz U, et al. A Cluster-Randomized Trial on Small Incentives to promote Physical Activity. Am J Prev Med. 2019;56(2):e45-e54.

[13] Black H. How to Design a Chatbot Script from Scratch. 2019; Available from: https://gumroad.com/l/conversationdesign

[14] Flow XO LLC. Create a chatbot with zero coding skills required. 2020; Available from: https://flowxo.com.

[15] Holmes S, Moorhead A, Bond R, et al. Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? Proceedings of the 31st European Conference on Cognitive Ergonomics; Association for Computing Machinery; 2019. p. 207-214.

[16] Nadarzynski T, Miles O, Cowie A, et al. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digit Health. 2019;5:2055207619871808.

[17] Piao M, Kim J, Ryu H, et al. Development and Usability Evaluation of a Healthy Lifestyle Coaching Chatbot Using a Habit Formation Model. Healthc Inform Res. 2020;26(4):255-264.

[18] Fadhil A, Gabrielli S. Addressing Challenges in Promoting Healthy Lifestyles: The AI-Chatbot Approach. Pervasive Health. 2017.

[19] Zhang J, Oh YJ, Lange P, et al. Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to promote Physical Activity and a Healthy Diet: Viewpoint. J Med Internet Res. 2020;22(9):e22845.

[20] Kramer JN, Künzler F, Mishra V, et al. Which Components of a Smartphone Walking App Help Users to Reach Personalized Step Goals? Results From an Optimization Trial. Ann Behav Med. 2020;54(7):518-528.

[21] Kocielnik R, Xiao L, Avrahami D, et al. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2018;2:1-26.

[22] Piao M, Ryu H, Lee H, et al. Use of the Healthy Lifestyle Coaching Chatbot App to Promote Stair-Climbing Habits Among Office Workers: Exploratory Randomized Controlled Trial. JMIR Mhealth Uhealth. 2020;8(5):e15085.

[23] Maher CA, Davis CR, Curtis RG, et al. A Physical Activity and Diet Program Delivered by Artificially Intelligent Virtual Health Coach: Proof-of-Concept Study. JMIR Mhealth Uhealth. 2020;8(7):e17558.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210605

A Mobile App to Improve Patient Management in Emergency Departments: Caregiver Needs Analysis, Design and Early Technology Acceptance Assessment

Frédéric EHRLER[a], Carlotta TUOR[b], Robin REY[b], and Johan N. SIEBERT[c,1]
[a] *Division of Medical Information Sciences, Geneva University Hospitals, Geneva, Switzerland*
[b] *University of Geneva, Faculty of Medicine, Geneva, Switzerland*
[c] *Department of Pediatric Emergency Medicine, Children's Hospital, Geneva University Hospitals, Geneva, Switzerland*

Abstract. Emergency care is very complex in that it requires patient-centered care in a coordinated manner among multiple providers in a highly distractible, unpredictable and stressful environment. Sharing information efficiently between providers in this context is difficult. Connecting emergency providers with each other through a digital communication channel could improve the efficiency of information sharing and emergency care. This study describes the development process of PIMPmyHospital, a mobile app dedicated to emergency department physicians and nurses to collaboratively manage their patients. We relied on a user-centered design process involving caregivers from a pediatric emergency department. The process started with semi-structured interviews that informed the specifications of the app, followed by an iterative design and development approach. The resulting prototype was evaluated by end-users using the perceived usefulness dimension of the technology acceptance model questionnaire. Early user engagement during the design and development of a dedicated mobile app must be taken into account to improve its perceived usefulness and future adoption.

Keywords. Emergency Medicine, Information Technology, Interprofessional Relations, Mobile Applications, Patient Care Management

1. Introduction

An emergency department (ED) is a place that requires high functional operability. This involves dealing with an erratic patient flow, including life-threatening emergencies requiring immediate attention, while simultaneously providing safe and optimal care to all patients in a timely manner in an endemically overcrowded and resource-constrained environment. In this context, time wasted by caregivers on indirect patient and non-patient activities wastes resources to the detriment of patient workflow, staff satisfaction and, ultimately, patients. It has been observed that ED physicians and nurses spend nearly

[1] Corresponding Author, Johan N. Siebert, Department of Pediatric Emergency Medicine, Children's Hospital, Geneva University Hospitals, 47 Avenue de la Roseraie, 1205 Geneva, Switzerland; e-mail: Johan.Siebert@hcuge.ch

half of their time on indirect patient and non-patient activities, including traveling within the unit, getting test results, and locating colleagues to share medical information [1; 2]. It is also challenging for ED caregivers to maintain a high level of awareness of each patient's individual situation while multitasking in a highly distracting and rapidly changing environment. For example, clinicians may not be notified instantly that a perfusion has ended or nurses that laboratory tests were prescribed. As a result, obtaining real-time patient information and sharing it between caregivers can be suboptimal in terms of communication, efficiency and patient flow. Patients have to wait unnecessarily in the emergency room when they could have been sent home or hospitalized earlier, which would improve ED patient turnover and potentially reduce upstream waiting times. Evidence to demonstrate the benefits of computerized tools specifically dedicated to streamlining shared patient management to overcome these drawbacks in ED is scarce.

To address this issue, we developed a mobile application (app) dedicated to ED caregivers, PIMPmyHospital (Patients In My Pocket in my Hospital). This app aims to provide relevant information in real-time about the patients they care for, as well as a secure chat and messaging platform to virtually connect physicians and nurses caring for the same patients. We present here the caregivers' needs analysis, design and early technology acceptance of this app by ED physicians and nurses.

2. Methods

The design and development of the app entailed a user-centered and multidisciplinary design approach. A four-step development cycle was used.

2.1. Semi-structured interview

In the first step, interviews were conducted with pediatric ED physicians and nurses of a tertiary hospital with more than 33,000 visits per year. After obtaining consent from each participant, individual semi-structured, face-to-face interviews with 20 open-ended questions were carried out to explore ED caregivers' perceptions of everyday challenges. Interviews were conducted by two qualified investigators (CT and RR) that were not involved in the aspects of emergency care with the participants. All questions were asked to participants in the same order and each participant was blinded to all other participants' answers. Interviews were audio-recorded, anonymized, and transcribed verbatim. An in-depth thematic analysis was then conducted according to the six-step iterative framework [3, 4] to identify patterns of emerging themes, including: 1) familiarization with data by reading the transcripts; 2) assigning preliminary codes to interesting features of the data in a systematic fashion across the entire data set; 3) searching for themes among the codes across the interviews; 4) reviewing themes by organizing the data that can best fit together into sub-themes; 5) defining and naming final themes; and 6) producing the report. The first five steps were blinded between the two interviewers. Findings were reported according to the Consolidated Criteria for Reporting Qualitative Research recommendations (available upon request) [5].

2.2. App's functional specifications identification

The app's specifications were defined based on the thematic analysis described above. Iterative focus group sessions involving three computer scientists and two pediatric

emergency physicians, including those from the project team, were conducted to analyze the feasibility of transcribing these specifications into functionality in the core app.

2.3. Design and prototyping

Relevant specifications gathered from the previous stages informed the design and development of the app prototype. Static mock-ups were initially produced with Adobe XD (Adobe Inc) to visualize the information displayed in a format similar to that of the proposed mobile app. The design sessions were an iterative process in which the project team provided feedback until the prototype was considered as covering all functions identified as relevant.

2.4. Evaluation

In the final step, we conducted a preliminary evaluation of the prototype through interviews using the Technology Acceptance Model (TAM) as a guiding theoretical framework to provide insight into one key determinant that influences user acceptance of technology, i.e., perceived usefulness [6]. Seven-point Likert-type scales ranging from "strongly disagree" (1) to "strongly agree" (7) were used to assess the six statements related to participants' perceived usefulness of the app. We reported average scores on each statement and the internal consistency of the questionnaire was measured using Cronbach's coefficient alpha. Finally, participants were asked to provide oral feedback on the app, which was transcribed to identify expected benefits and potential drawbacks.

2.5. Ethical considerations

As this study was considered as falling outside of the scope of Swiss legislation regulating research on human subjects, the need for local ethics committee approval was waived (No. req-2021-00740).

3. Results

3.1. Interviews

Between June 8 and June 29, 2021, a convenience sample of participants was recruited and interviewed. Data saturation [7] was reached by 14 interviews. However, interviews were continued until 20 participants to ensure a maximum variation in interviewees. This included eight ED fellows, two postgraduate year-1 residents, and 10 ED nurses. Mean (standard deviation) age, years since certification, and years spent in the ED were 37.0 (8.1), 12.0 (8.8), and 6.0 (5.4), respectively. Interviews were completed by all participants without any missing items (average duration, 19.6 [range 9.6–31.7] minutes per participant [total 392 minutes]). Thematic analysis of the transcripts enabled the identification of four main themes (Table 1) as follows:

Theme 1: Mobility. Caregivers reported that they have to travel extensively. Places that bring caregivers together seemed conducive to fostering communication.

Design principle. Reducing back and forth flow would 1) streamline caregiver's activities, 2) avoid staff stress and fatigue, and 3) reduce discontinuity of care. The app was tailored to caregivers by connecting them virtually and providing them with targeted information in real time. In other words, the app was designed so that information reaches the caregiver rather than the caregiver having to reach out for the information.

Theme 2: Time optimization. Almost all participants (19/20; 95%), irrespective of their functions as physicians or nurses, noted the need to reduce the waste of time spent to find the other caregivers involved in the care of the patients. They also pointed out the need to quickly know who is caring for a specific patient.

Design principle. The design principle for time optimization was to 1) focus on patients managed by a given caregiver, 2) facilitate the identification of other caregivers caring for the same patients, and 3) provide the ability to reach them quickly. To address these needs, we have integrated a patient list tailored to each caregiver. Only the patients the caregiver is in charge of are displayed in real time on the app, with other caregivers sharing the same patients displayed to the right, together with the time since ED admission (Figure 1). The entry and exit of patients on the list are automatically controlled by the institution's patient management software linked to the app, thus requiring no user intervention. Other caregivers in charge of the same patient are represented by their initials in a color-coded box indicating their functions (i.e., physician/red, nurse/blue, etc.). A click on the box opens an instant chat system to reach them remotely.

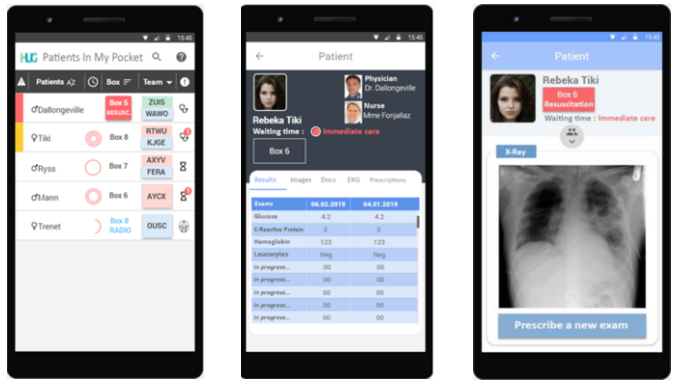


Figure 1. Prototype of the PIMPmyHospital app. Left screenshot: color-coded bars represent the five-level Canadian Triage and Acuity Scale [8]); patient gender and identity; time since admission as radial circular timers; patient allocation per room; color-coded box of individuals in charge of the patient; patient status (seen by a physician, waiting for results, CT-scan in progress, etc.). Push notifications prompt the user to be aware of the situations. Middle and right screenshots: selecting a patient opens a new page with scrollable contextual tab menus containing information related to laboratory results, imaging, patient files, electrocardiograms, and prescriptions entered in the computerized institutional prescription software.

Theme 3: Medical information. Almost all participants agreed that it was more convenient to access medical information on-the-go via their mobiles compared to static computers at disposal. The consequences of missing or delayed laboratory test results and, to a lesser extent, radiological results, was a common concern as no alert is generated directly to the caregiver. This was considered by many as a hindrance to patient

flow and a source of distraction and cognitive overload as they had to remember to check the results several times until they were made available.

Design principle. Simple, real-time display of laboratory and radiology results with prompts that integrate into the workflow can 1) help caregivers focus on their tasks while waiting for their availability, and 2) access them in a timely manner, wherever they are. To achieve this goal, we made these results available to caregivers by synchronizing and displaying them directly on the app, with prompts (Figure 1).

Theme 4: Communication. The caregivers recognized the practicality of the electronic ED patient status board already in place in the ED as a core medium for communicating via text messages and providing situational awareness. However, they also pointed out constraints linked to this multi-channel and disparate display of all ED patients at a glance, making reading complicated and at risk of fragmented or omitted information. They noted both the lack of push notifications to inform caregivers of incoming messages and validation of their receipt. Caregivers also expressed their desire to be able to communicate directly with others in charge of their patients, thus justifying the need for a complementary medium.

Design principle. Enabling caregivers to use a single and responsive communication channel could make care more efficient by promoting instantaneous and shared information exchange, while reducing fragmentation and information loss. To address this design principle, we created a chat area (Figure 2). To ensure responsiveness, push notifications were designed to inform caregivers of incoming messages.

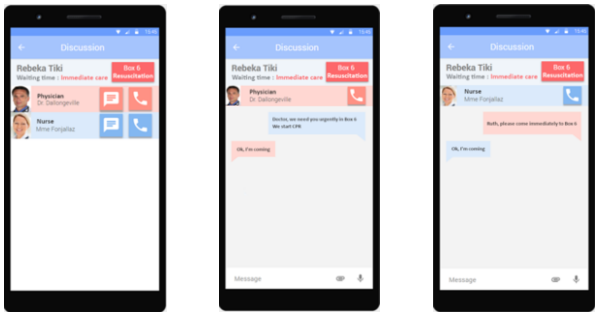


Figure 2. Prototype of the instant messaging system. The logos at the bottom of the screen represent the possibility to also import and link documents to the conversion, as well as to send voice memos.

3.2. Preliminary evaluation

The reliability of the questionnaire was excellent with a Cronbach's alpha of 0.98 for perceived usefulness. The scores for perceived usefulness suggested that the participants strongly agreed that the app is a useful tool (Table 1). Based on user feedback, the shared aspect of the information assisted by a mobile app was seen by many as beneficial in facilitating communication in a coordinated and collaborative manner. Although participants were likely to use the app, they also anticipated some potential drawbacks, such as being overwhelmed by messages occurring at inopportune times or that communication via the app would come at the expense of oral transmission.

Table 1. Items and scores for the perceived usefulness of the mobile app PIMPmyHospital

Perceived usefulness	Mean	Standard Deviation
1. Using this app in my job would enable me to accomplish tasks more quickly	5.75	1.02
2. Using this app would improve the quality of my work in providing better patient care	5.50	1.10
3. Using this app in my job would increase my productivity	5.40	1.57
4. Using this app would enhance my effectiveness on the job	5.45	1.50
5. Using this app would make my job easier to perform	6.0	1.26
6. I would find this app useful in my job	5.7	1.46
Overall	**5.63**	**1.32**

4. Discussion

EDs are highly dynamic and stressful care environments subject to overcrowding and acute situations requiring rapid and coordinated care among several stakeholders. In this context, wasted time, discontinuity of care, suboptimal communication, and omission or delay in seeking information can compromise patient outcomes and safety. This study sought to develop a user-centered mobile app connecting ED caregivers to each other. Preliminary results are encouraging, with users rating the usefulness of the app as high for a newly-developed digital tool. This outcome should however be validated through a formal evaluation in simulated setting before being used in real-life situations.

5. Conclusion

Early user engagement during the design and development of a dedicated mobile app is a crucial step to improve its perceived usefulness and future adoption. Providing ED caregivers with a communication and information sharing mobile app can be a light and affordable strategy to promote ED efficiency and improve patient care.

References

[1] Hollingsworth JC, Chisholm CD, Giles BK, Cordell WH, Nelson DR. How do physicians and nurses spend their time in the emergency department?, *Ann Emerg Med* **31** (1998), 87-91.
[2] Weigl M, Handl T, Wehler M, Schneider A. Time-allocation study of nurse and physician activities in the emergency department. *Med Klin Intensivmed Notfmed* **116** (2021), 229-237.
[3] Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131, *Med Teach* **42** (2020), 846-854.
[4] Maguire M, Delahunt B. Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars, All Ireland Journal of Higher Education **9** (2017).
[5] Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups, *Int J Qual Health Care* **19** (2007), 349-357.
[6] Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly* **13** (1989), 319-340.
[7] Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, Burroughs H, Jinks C. Saturation in qualitative research: exploring its conceptualization and operationalization, *Qual Quant* **52** (2018), 1893-1907.
[8] Beveridge R. CAEP issues. The Canadian Triage and Acuity Scale: a new and critical element in health care reform. Canadian Association of Emergency Physicians, *J Emerg Med* **16** (1998), 507-511.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210606

Assessment of mHealth Solutions Applied to Fall Detection for the Elderly

Fellipe Soares de OLIVEIRA[a], Camila Carvalho da SILVA[a], Talita Santos PINHEIRO[a], Larissa Mayumi YOKOI[a], Pablo Deoclecia dos SANTOS[a], Harki TANAKA[a], Priscyla Waleska SIMÕES[a1]
[a] *Engineering, Modeling and Applied Social Sciences Center (CECS), Biomedical Engineering Graduate Program (PPGEBM), Universidade Federal do ABC (UFABC), Alameda da Universidade s/n, 09606-045 São Bernardo do Campo, SP, Brazil*

Abstract. Mobile Health has been increasingly present in healthcare due to the wide availability of applications for smartphones, however, robust assessment methods must be considered, seeking to provide evidence for clinical practice and mHealth solutions. This research presents the assessment of applications aimed at detecting and preventing falls for the elderly, available for Android and IOS, through the Mobile App Rating Scale. Based on the results presented, it can be concluded that the fall detection and prevention applications for the elderly available for Android and IOS showed good quality after rigorous evaluation.

Keywords. Digital Health. Mobile Health. Prevention. IoT. Falls. Assessment.

1. Introduction

According to the World Health Organization (WHO), falls represent the second leading cause of accidental or unintentional deaths in the world (traffic accidents are the main cause in this group), resulting in approximately 650,000 deaths, and, of these, 80% in low-and middle-income countries. In this context, individuals over 60 years of age are the most susceptible to fatal falls (424,000 deaths), suggesting that falls for the elderly have become a public health problem[16; 17].

While many falls are not fatal, approximately 38 million of those that occur annually are severe enough to result in hospitalization. The costs of injuries related to falls are substantial, as, among the population of 65 years old or over, the average cost per injury can range from US$1,049.00 to US$3,611.00 in countries such as Australia and Finland, respectively[16].

Due to the increased use of the internet through smartphones and wearable devices (such as smart watches and bracelets), monitoring strategies are becoming increasingly popular through alerts, for example, in addition to contributing to training actions related to the treatment and prevention of diseases, and improving access to health services, clinical diagnosis and treatment adherence[6].

[1] Corresponding Author: Prof. Dr. Priscyla Waleska Simões, Universidade Federal do ABC, Engineering, Modeling and Applied Social Sciences Center (CECS), Biomedical Engineering Graduate Program (PPGEBM); Alameda da Universidade s/n, 09606-045 São Bernardo do Campo, SP, Brazil; E-mail: pritsimoes@gmail.com.

In this context, mHealth solutions have been developed for the detection and prevention of falls, through research using cell phone sensors (gyroscope and accelerometer), Machine Learning, among other technologies[10; 13].

Despite mHealth's advances[11] and the development of applications for the detection and prevention of falls aimed at the elderly population in recent years, such solutions rarely present a scientific mHeatlh assessment, and when assessment is performed, the instruments used are not specific to mHealth[7]. Therefore, there are gaps to be clarified for a broader adoption in clinical practice and based on scientific evidence[1].

Thus, the aim of this paper is to answer the question: Are the free applications currently available for the Android and IOS operating systems and aimed at detecting and preventing falls for the elderly, having good quality after rigorous scientific mHealth evaluation? Thus, the objective is to present an evaluation of mHealth solutions for the detection and prevention of falls for the elderly.

2. Methods

Study on the assessment of applications for the detection and prevention of falls for the elderly, which started with the search strategy carried out by 2 researchers independently, being followed by the description of the applications, categorization and analysis of features. The last step presented the assessment of applications by MARS, ending with the statistical analysis.

The search strategy was carried out through a search in the digital distribution service of Google (Google Play) and Apple (App Store), as they are the platforms associated with the most used operating systems in the world for mobile applications (Android and iOS, respectively).

The assessment was conducted in the 2nd semester of 2020 and 1st semester of 2021 and used the following keywords and synonyms: “Fall” and “Elderly”. Applications made available free of charge on the digital distribution service offered by Google (Google Play) and by Apple (App Store), and that presented the description provided by the developer, to identify the purpose of the study and the target population, were included in the study.

Applications with functionality presented in Portuguese, English and Spanish were also included, and applications with inoperable features, paid items (and not provided by the developer), and those that do not directly presented information on the detection and prevention of falls were excluded. The application screening and eligibility process is shown in figure 1.

In evaluating the applications, a scale focused on mHealth was used, the Mobile Application Rating Scale (MARS)[5], which is a questionnaire that measures the quality of health applications based on 23 questions organized into 4 objective categories (engagement, functionality, aesthetics and information quality) and 1 subjective category. Each question is rated on a 5-point scale (1-inadequate, 2-poor, 3-acceptable, 4-good, 5-excellent)[14].

The assessment of the apps was presented by the average, used as a measure of central tendency, and the standard deviation to quantify its variability. To identify the correlation between the categories (A, B, C, D, and E subjective) and the Final Score, the Pearson's correlation coefficient test was used. For all statistical analysis, a

significance level of 5% was considered. All statistical analysis were performed using Jamovi software (The Jamovi Project, Sydney, Australia) version 1.6.3.

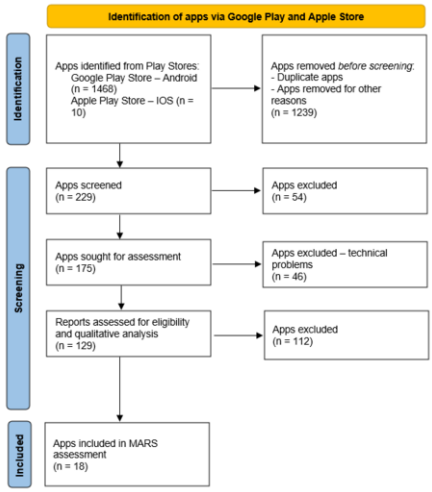


Figure 1. Flowchart of app selection

3. Results and Discussion

The assessment resulted in 18 apps that were rated by the MARS scale. The results indicated a positive correlation between the Final Score and Engagement ($r=0.872$; $R^2=76.0\%$; $p<0.001$); Functionality ($r=0.498$; $R^2=24.8\%$; $p=0.042$); Aesthetics ($r=0.909$; $R^2=82.6\%$; $p<0.001$); Information ($r=0.838$; $R^2=70.2\%$; $p<0.001$); Subjective ($r=0.953$; $R^2=90.8\%$; $p<0.001$).

Table 1 presents the main features of the apps. It is important to note that 4 applications were compatible with wearables (smartwatch), that is, the development of applications for wearable devices for monitoring falls is still rarely used. In the analysis of the use of sensors, applications aimed at health education actions for patients did not present sensors, however, all applications that detect falls use the accelerometer sensor, which is the main resource used to detect the vibrations associated with falls.

In addition to these features, some applications make use of the gyroscope sensor that together with the accelerometer sensor can improve the accuracy of the fall detection. The use of the GPS sensor by some applications contributes to the user's physical monitoring, thus, regardless of geographic location, GPS makes it possible to record the place where the fall occurs, which can be used as an alert to emergency contacts and immediate help.

The assessment was performed using the MARS scale shown in Table 2, which is organized in descending order of the Final Score, with general average ranging from 4.75 (± 0.25) to 1.86 (±0.69).

MARS has been widely used since 2015[9], and to present, the literature has more than 100 related publications (www.pubmed.com). This instrument has become widespread as can be seen in a review of urinary incontinence management applications[2]. Both studies[2; 9] suggest the importance of using scales such as MARS in the assessment stage, seeking to improve mHealth solutions.

Table 1. Sample characteristics

APP	Link	Language	OS	COMPATIBILITY	FREEWARE	CATEGORY	SUBCATEGORY	ACCESSORIES	SENSORS	ML
1	https://apps.apple.com/br/app/fallsafety-pro-safety-alerts/id870864283	English	Android / IOS	Requires Android 5.0 or higher / iOS 12.0 and watchOS 5.0 or higher.	No	Patient Centered	Health Tracking	Smartphone / Smartwatch	Accelerometer/ Gyroscope / GPS	Yes
2	https://apps.apple.com/us/app/fallsafety-home-personal-alert/id1097177984	English	IOS	iOS 9.0 and watchOS 2.0 or higher	No	Patient Centered	Health Tracking	Smartphone / Smartwatch	Accelerometer/ Gyroscope / GPS	Yes
3	https://play.google.com/store/apps/details?id=com.chkincam.fall&hl=pt_BR&gl=US	English	Android	Android 5.0 or higher	No	Patient Centered	Health Tracking	Smartphone / Smartwatch	Accelerometer/ Gyroscope / GPS	-
4	https://play.google.com/store/apps/details?id=com.caringcompany.mira&hl=pt&gl=US	English	Android / IOS	Android 5.0 or higher / iOS 13.0 or higher	No	Patient Centered	Health Tracking	Smartphone	Accelerometer/ GPS	Yes
5	https://play.google.com/store/apps/details?id=com.tidyware.workersafety.pro&hl=en_US&gl=US	English	Android / IOS	Android 5.0 or higher / iOS 13.0 or higher	No	Patient Centered	Health Tracking	Smartphone	Accelerometer/ Gyroscope / GPS	-
6	https://play.google.com/store/apps/details?id=com.GazGames.SalveVovo&hl=fr_CH	Portuguese	Android	Android 2.3 or higher	Yes	Educational	Patient Education	Smartphone	No need	No
7	https://play.google.com/store/apps/details?id=com.hcpastopfalls&hl=pt_BR&gl=US	English	Android / IOS	Android 5.0 or higher / iOS 10.3 or higher	Yes	Educational	Patient Education	Smartphone	No need	No
8	https://play.google.com/store/apps/details?id=com.healthappy.seizario2&hl=pt_BR&gl=US	English	Android	Android 5.0 or higher	Yes	Patient Centered	Health Tracking	Smartphone	Accelerometer/ GPS	-
9	https://play.google.com/store/apps/details?id=com.Sades.PrevencaoQuedasIdosos	Portuguese	Android	Android 4.4 or higher	Yes	Educational	Patient Education	Smartphone	No need	No
10	https://play.google.com/store/apps/details?id=com.Jump.FallsPrevention	English	Android / IOS	Android 4.1 or higher / iOS 10.0 or higher	Yes	Educational	Patient Education	Smartphone	Accelerometer/ Gyroscope	No
11	https://play.google.com/store/apps/details?id=com.fall.detection	English	Android	Android 4.1 or higher	Yes	Patient Centered	Health Tracking	Smartphone	Accelerometer	Yes
12	https://play.google.com/store/apps/details?id=com.HRscape.grandzangel	English	Android	Android 4.1 or higher	Yes	Patient Centered	Health Tracking	Smartphone	Accelerometer	--
13	https://apps.apple.com/us/app/alarm-free/id284728433	English	IOS	iOS 12.0 or higher	Yes	Patient Centered	Health Tracking	Smartphone	Accelerometer/ Gyroscope / GPS	No
14	https://play.google.com/store/apps/details?id=appinventor.ai_salekhan13.freefall	English	Android	Android 1.5 or higher	Yes	Patient Centered	Health Tracking	Smartphone/ Smartwatch	Accelerometer/ Gyroscope	-
15	https://apps.apple.com/pt/app/care24/id1152041483	English	IOS	iOS 12.0 or higher	Yes	Patient Centered	Health Tracking	Smartphone	Accelerometer/ GPS	No
16	https://play.google.com/store/apps/details?id=com.aniruth.falldetect	English	Android	Android 7.0 or higher	Yes	Patient Centered	Health Tracking	Smartphone	Accelerometer/ Gyroscope / GPS	Yes
17	https://play.google.com/store/apps/details?id=fall.detectionver2.acc	English	Android	Android 5.1 or higher	Yes	Patient Centered	Health Tracking	Smartphone	Accelerometer/ Gyroscope	--
18	https://play.google.com/store/apps/details?id=com.sjinnovation.gameoffalls.live	English	Android	Android 5.0 or higher	Yes	Patient Centered	Patient Education	Smartphone	-	No

OS – Operating System / ML – Machine Learning

Considering the MARS domains, category A, which covers issues related to engagement, was the category with the lowest overall average in our study (3.43±0.93). Engagement strategies such as gamification were not relevant to the context of falling for the elderly, however, such techniques have shown good results in mHealth solutions related to Nutrition in children and adolescents[12].

In category B related to features, there was a higher overall average (4.24±0.45), due to the easy use of the applications, simple and objective design, and few gaps related to navigation.

In category C (3.58±0.36), which evaluates the proportion of components in the interface, graphic quality of icons and aesthetic appeal, for example, the average was possibly influenced by applications that did not present a visually attractive interface (such as buttons in irregular positions).

In category D (3.60±0.92), the relevance of the information and the presence of illustrations were observed to facilitate navigation and understanding of the application. No apps were scientifically evaluated, and the patient-centric apps did not contain information relevant to fall prevention. Educational applications had a higher Final Average because their assessment took this category into consideration.

Some applications had manual pushbuttons and alerts to external contacts after the fall was detected, however, as the fall can immobilize the patient, this feature can influence the alert, help and assistance. In this context, in some studies[3; 4; 18], the accelerometer sensor on a wearable/smartphone is the main feature used in fall detection as it easily identifies the vibrations associated with the fall.

Table 2. MARS scores

App	Link	A	B	C	D	Average	E
1	https://apps.apple.com/br/app/fallsafety-pro-safety-alerts/id870864283	5.00±0.00	4.50±0.50	5.00±0.00	4.50±0.76	**4.75±0.25**	5.00±0.00
2	https://apps.apple.com/us/app/fallsafety-home-personal-alert/id1097177984	5.00±0.00	4.50±0.50	5.00±0.00	4.33±0.75	**4.71±0.30**	5.00±0.00
3	https://play.google.com/store/apps/details?id=com.chkincam.fall&hl=pt_BR&gl=US	4.40±0.49	5.00±0.00	5.00±0.00	4.33±1.11	**4.68±0.32**	5.00±0.00
4	https://play.google.com/store/apps/details?id=com.caringcompany.main&hl=pt&gl=US	4.40±0.80	5.00±0.00	4.67±0.47	4.60±0.80	**4.67±0.22**	5.00±0.00
5	https://play.google.com/store/apps/details?id=com.tidyware.workersafety.pro&hl=en_US&gl=US	4.80±0.45	4.25±0.83	5.00±0.00	4.00±0.82	**4.51±0.40**	3.25±0.43
6	https://play.google.com/store/apps/details?id=com.GazGames.SalveVovo&hl=fr_CH	4.00±1.26	4.75±0.43	5.67±0.47	4.17±0.69	**4.40±0.32**	3.50±1.12
7	https://play.google.com/store/apps/details?id=com.hcpastopfalls&hl=pt_BR&gl=US	3.40±1.96	5.00±0.00	4.33±0.47	4.67±0.75	**4.35±0.60**	4.50±0.87
8	https://play.google.com/store/apps/details?id=com.healthappy.seizario2&hl=pt_BR&gl=US	4.60±0.49	4.75±0.43	3.67±0.94	4.33±0.75	**4.34±0.42**	4.50±0.50
9	https://play.google.com/store/apps/details?id=com.Sades.PrevencaoQuedasIdosos	3.20±1.17	5.00±0.00	4.00±0.00	4.00±1.00	**4.05±0.64**	2.00±1.00
10	https://play.google.com/store/apps/details?id=com.Jump.FallsPrevention	4.20±1.17	3.50±0.50	3.67±0.47	4.50±0.76	**3.97±0.40**	2.50±0.50
11	https://play.google.com/store/apps/details?id=com.fall.detection	2.80±1.33	5.00±0.00	3.67±0.47	3.83±0.90	**3.83±0.78**	2.50±1.66
12	https://play.google.com/store/apps/details?id=com.HRscape.grandzangel	4.20±0.75	3.75±0.43	2.00±0.00	4.20±0.75	**3.54±0.91**	2.00±0.75
13	https://apps.apple.com/us/app/alarm-free/id284728433	1.80±0.75	4.75±0.43	2.67±0.94	2.50±1.26	**2.93±1.10**	1.25±0.43
14	https://play.google.com/store/apps/details?id=appinventor.ai_salekhan13.freefall	2.00±1.26	5.00±0.00	1.67±0.94	2.75±1.79	**2.85±1.30**	1.25±0.43
15	https://apps.apple.com/pt/app/care24/id1152041483	2.20±0.98	3.75±0.83	3.00±0.00	1.80±0.98	**2.69±0.75**	1.00±0.00
16	https://play.google.com/store/apps/details?id=com.anirudh.falldetect	2.00±0.89	4.00±1.22	1.33±0.47	2.25±1.09	**2.40±0.98**	1.25±0.43
17	https://play.google.com/store/apps/details?id=fall.detectionver2.acc	2.40±0.80	2.50±1.50	2.00±0.82	2.17±1.21	**2.27±0.20**	1.25±0.43
18	https://play.google.com/store/apps/details?id=com.sjinnovation.gameoffalls.live	1.40±0.49	1.25±0.43	3.00±0.00	1.80±0.40	**1.86±0.69**	1.00±0.00
	Final average	**3.43±0.93**	**4.24±0.45**	**3.58±0.36**	**3.60±0.92**	**3.71±0.59**	**2.88±0.47**

A – engagement; B – functionality; C – aesthetics; D – information; Final – final average, E – subjective

Also, optimization methods for fall detection algorithms have been researched in recent years, through a previously defined threshold[15; 19], and Machine Learning[8; 10], despite the fact that our sample has considered this last method little.

4. Conclusion

Based on the results presented, it can be concluded that the fall detection and prevention applications available for Android and IOS have good quality after a rigorous assessment carried out via MARS scale. A possible limitation was that many elderly people do not use the smartphone as often as young people, and this can impact the number of applications developed for this purpose.

It is recommended that future research address the development of assessment instruments which are adapted to a context where mHealth solutions are involved. Thus, we would suggest a new method of assessment or the continuation of studies for the development of an app that detect and prevents falls, aimed at the elderly population stemming from the gaps shown in the article.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001, and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Brazil.

References

[1] Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018; 1: 12.
[2] Dantas LO, Carvalho C, B.L.d.J. Santos, C.H.J. Ferreira, K. Bø, and P. Driusso, Mobile health technologies for the management of urinary incontinence: A systematic review of online stores in Brazil. Braz J Phys Ther. Jul-Aug 2021;25(4):387-395. doi: 10.1016/j.bjpt.2021.01.001.
[3] Er PV, Tan KK. 4 - Wearable solution for robust fall detection. In: Assistive Technology for the Elderly, N.K. Suryadevara and S.C. Mukhopadhyay, eds., Academic Press, 2020, pp. 81-105.
[4] Fakhrulddin SS, Gharghan KS. An Autonomous Wireless Health Monitoring System Based on Heartbeat and Accelerometer Sensors. J. Sens. Actuator Netw. 2019; 8(3): 39. https://doi.org/10.3390/jsan8030039.
[5] Gong E, Zhang Z, Jin X, Liu Y, Zhong L, et al. Quality, Functionality, and Features of Chinese Mobile Apps for Diabetes Self-Management: Systematic Search and Evaluation of Mobile Apps JMIR Mhealth Uhealth 2020; 8: e14836.
[6] Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. J Health Commun 2012; 17 Suppl 1: 82-104.
[7] Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A Fall Risk mHealth App for Older Adults: Development and Usability Study. JMIR aging 2018; 1: e11569-e11569.
[8] Jahanjoo A, Tahan MN, Rashti MJ. Accurate fall detection using 3-axis accelerometer sensor and MLF algorithm. In: 2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA), 2017, pp. 90-95.
[9] Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and Evaluation of Mindfulness-Based iPhone Apps. JMIR Mhealth Uhealth 2015; 3: e82.
[10] Nooruddin S, Milon Islam M, Sharna FA. An IoT based device-type invariant fall detection system. Internet of Things 2020; 9: 100130.
[11] Sahlab N, Jazdi N. AI-Based Elderly Assistance Systems. Stud Health Technol Inform 2020; 273: 163-169.
[12] Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. International Journal of Behavioral Nutrition and Physical Activity 2017; 14: 83.
[13] Silva J, Gomes D, Sousa I. Technological Solution for Pervasive Fall Risk Assessment. Stud Health Technol Inform 2020; 273: 176-181.
[14] Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. JMIR Mhealth Uhealth 2015; 3: e27.
[15] Wang F-T, Chan H.L, Hsu M-H, Lin C-K, Chao P-K, Chang Y-J. Threshold-based fall detection using a hybrid of tri-axial accelerometer and gyroscope. Physiological Measurement 2018; 39: 105002.
[16] World Health Organization. WHO, Falls, in, 2021.
[17] Wu H, Ouyang P. Fall prevalence, time trend and its related risk factors among elderly people in China. Arch Gerontol Geriatr 2017; 73: 294-299.
[18] Yacchirema D, de Puga JS, Palau C, Esteve M. Fall detection system for elderly people using IoT and ensemble machine learning algorithm. Personal and Ubiquitous Computing 2019; 23: 801-817.
[19] Yang L, Ren Y, Zhang W. 3D depth image analysis for indoor fall detection of elderly people. Digital Communications and Networks 2016; 2: 24-34.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210607

Continuous Stress Detection of Hospital Staff Using Smartwatch Sensors and Classifier Ensemble

Muhammad Ali FAUZI[1] and Bian YANG
Norwegian University of Science and Technology, Gjøvik, Norway

Abstract. High stress levels among hospital workers could be harmful to both workers and the institution. Enabling the workers to monitor their stress level has many advantages. Knowing their own stress level can help them to stay aware and feel more in control of their response to situations and know when it is time to relax or take some actions to treat it properly. This monitoring task can be enabled by using wearable devices to measure physiological responses related to stress. In this work, we propose a smartwatch sensors based continuous stress detection method using some individual classifiers and classifier ensembles. The experiment results show that all of the classifiers work quite well to detect stress with an accuracy of more than 70%. The results also show that the ensemble method obtained higher accuracy and F1-measure compared to all of the individual classifiers. The best accuracy was obtained by the ensemble with soft voting strategy (ES) with 87.10% while the hard voting strategy (EH) achieved the best F1-measure with 77.45%.

Keywords. Stress detection, Hospital, Machine Learning, Ensemble, Smartwatch

1. Introduction

Over recent years, stress has become an interesting topic in today's hectic world. There has been increasing awareness in many countries about the rise of work-related stress. Hospital is possibly one of the most important workplaces to be alarmed about this issue. Many studies reported that many hospital workers suffer from work-related stress [1,2]. Although stress at some level is normal, chronic stress can harm our physical, mental, and emotional wellbeing [3,4]. Specifically for hospital, many studies suggested that higher stress level has a relationship with low patient safety [5,6]. Another study also reported that higher stress level is significantly correlated with riskier cybersecurity practices [7].

Monitoring hospital workers' stress level has many advantages. Knowing their own stress level can help them stay aware and feel more in control of their response to situations and know when it is time to relax or take some actions to treat it properly [8]. Besides, this monitoring can help for early diagnosis of mental illness and disorders. The most common way to assess stress level is by using questionnaires (e.g., Perceived Stress Scale [9], Perceived Stress Questionnaire [10], etc.). However, this method takes time so that it is not convenient to be performed every day for continuous monitoring.

[1] Corresponding Author: Muhammad Ali Fauzi, Norwegian University of Science and Technology, Teknologivegen 22, 2815 Gjøvik, Norway; E-mail: muhammad.a.fauzi@ntnu.no.

The other stress level assessment method is by measuring the physiological responses related to stress such as heart rate, blood pressure, skin conductance, respiration activity, etc. Some sensors can be used to conduct the measurement task (e.g., electrocardiogram (ECG) to measure the heart rate, galvanic skin response (GSR) for skin conductance, etc.). The recent advance in wearable devices with sophisticated built-in sensors makes it feasible to passively collect multimodal data from people's daily lives for automatic continuous stress detection purposes. However, some wearable devices have a very low usability and not convenient to wear during work (e.g., chest-worn devices, finger placed GSR sensors, etc.) [11].

Smartwatch has recently emerged as a new platform that provides many successful applications. These devices have several built-in sensors that are useful for stress monitoring including Blood Volume Pulse (BVP), Electrodermal Activity (EDA), temperature, accelerometer, etc. Besides, the use of watches is well known and has a high degree of social acceptance by their ubiquity in everyday life [12]. Therefore, it has a high potential to be applied for multi-modal-based continuous stress detection.

Many previous works have been successfully leveraging multi-modal sensors data and machine learning methods to build automatic stress detection. The popular machine learning methods used are Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression [13,14,15,16]. In this work, we propose a multi-modal based continuous stress detection method using classifier ensemble and give comparative analysis between individual classifiers. Classifiers ensemble is a set of base classifiers whose individual classification outputs are combined in some way in order to enhance classification accuracy [17]. The individual classifiers used for this works include Naive Bayes (NB), Support Vector Machine (SVM), Neural Network (NN), K-Nearest Neighbors (KNN), Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT).

2. Proposed Work

2.1. Dataset

This research is based on the WESAD [13] dataset, which is available to the public. It includes data from 15 people who were measured with the Empatica E4 wrist-worn device and chest-worn RespiBAN device. However, because the focus of this work is on smartwatch sensors, only E4 data is used in this analysis. The E4 gadget incorporates skin temperature (ST), accelerometers (ACC), electrodermal activity (EDA), and blood volume pulse sensors (BVP) sensors. Data from three separate affective states (stress, amusement, and relaxation) were obtained during the data collection process. The stress situation lasted about 10 minutes, the amused situation 6.5 minutes, and the relaxed situation 20 minutes. For the stress detection task in this study, the amusement and relaxation classes were merged into one class: non-stress. As a result, the problem under investigation was binary (stress and non-stress).

2.2. Features

In this study, we used the data from all of the sensors available in the smartwatch including ACC, EDA, ST, and BVP. To extract the features, a sliding window with a window shift of 0.25 seconds was used to segment the data. Furthermore, the ACC

features were computed with a five-second window size, as this is a common window length for acceleration-based context detection [18]. Meanwhile, all other physiological features were calculated with a window size of 60 seconds following the suggestion by Kreibig et al. [19]. The AC, EDA, and ST features were extracted based on prior work by [20]. The features extracted including some statistical features (mean, standard deviation, maximum, and minimum). Besides, some derivatives and Discrete Wavelet Transform (DWT) were also applied to the data to extract other statistical features. Meanwhile, for BVP, statistical features (mean, standard deviation) were also computed. Moreover, some features based on energy in different frequency bands were also calculated.

2.3. Classifier

Seven machine learning methods were used as classifiers for stress detection tasks including Naive Bayes (NB), Support Vector Machine (SVM), Neural Network (NN), K-Nearest Neighbors (KNN), Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT). In addition, we also used two ensemble methods. In order to do stress detection, the ensemble technique trains numerous classification methods and then combines them using particular approach [21]. It is important to take note that the performance of the ensemble methods cannot be guaranteed to be higher than the best individual method in the ensemble. However, it would significantly minimize the chances of picking a poor-performing classifier [17].

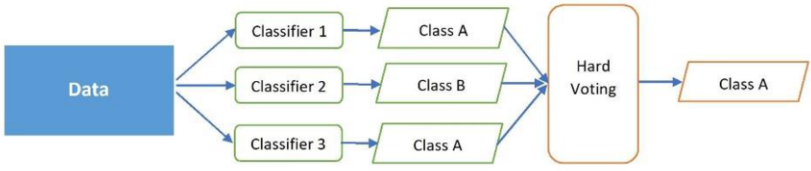


Figure 1. The hard voting strategy.

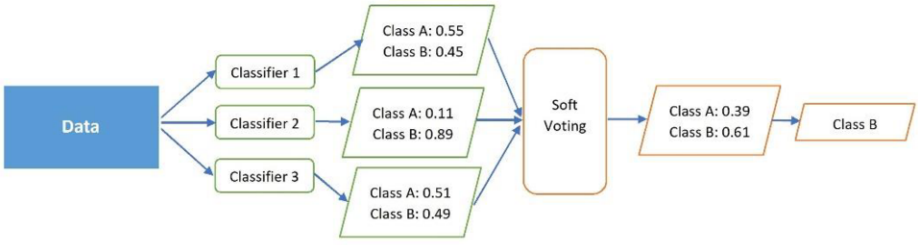


Figure 2. The soft voting strategy.

In this study, we employed three classification methods to build the ensemble learning method. Three individual classifiers with the highest accuracy were selected for the ensemble. Two ensemble strategies were used in this work as follows:

1. Hard voting (hard): As depicted in Figure 1, each classifier had one vote, and the class of the data was determined by the majority vote.

2. Soft voting (soft): As depicted in Figure 2, each classifier calculated the probability of each class in the first step. Then, the probabilities of each class from all classifiers were averaged, and the final class of the data was the one with the greatest average probability value.

2.4. Performance Evaluation

All classifiers were tested using the leave-one-subject-out (LOSO) cross-validation (CV) approach, which shows how a model will generalize and perform on previously unseen data. Several measurements including Accuracy, Precision, Recall, and F1-measure were employed for classifier performance evaluation.

3. Result

The stress detection result using individual classifiers is shown in Table 1 while the result using classifiers ensemble is displayed in Table 2. Table 1 depicts that all classifiers work quite well to conduct a stress detection task. All of them show adequate performance with an accuracy of more than 70%. RF obtained the best accuracy with 86.61% following by LR with a slight difference (85.46%). At third place was NN that has a slight margin to the first and second place (84.76%). These three top classifiers were then used for the ensemble methods. Meanwhile, the lowest accuracy was achieved by KNN with a value of only 73%.

In terms of precision, LR has the best precision among other individual classifiers with a value of 77.53%. Furthermore, in terms of recall, RF has the highest value with 89.87%. However, the precision of RF is quite low (69.17%) so that it could not obtain the highest F1-measure. It means that RF tends to successfully detect almost all of the stress data available but many non-stress data are incorrectly labeled as stress. Meanwhile, LR has a more balance precision and recall so that it could achieve the best F1-measure with 76.25%. Similar to the accuracy result, KNN also has the lowest F1-measure (52.43%).

The ensemble methods were build using the three best individual classifiers from the previous results (RF, LR, and NN). Table 2 shows that both ensemble methods obtained higher performance compared to all of the individual classifiers. Generally, most individual classifiers have their own inherent defects [22] and their performance is also domain-dependent [23]. By combining some classifiers, the advantage of one classifier is expected to cover the shortcomings of other classifiers so that the performance can be improved. Soft and hard voting have different strategies to combine the result from the individual classifiers so that they can lead to different decisions.

The result displayed in Table 2 shows that ES (soft voting) has a higher accuracy than EH (hard voting). In contrast, EH has a better performance in terms of F1-measure. Generally, the soft voting strategy tends to get better performance than the hard voting strategy as it takes into account more information. Soft voting is smoother as it uses probability information to get the final decision. However, the additional information could also lead to a worse decision. In this study, the best accuracy is obtained by ES with 87.10%. Meanwhile, the best F1-measure was achieved by EH with 77.45%.

Table 1. Stress Detection Result Using Individual Classifiers (%)

Method	Accuracy	Precision	Recall	F1-measure
NB	79.26	57.58	73.31	60.67
SVM	84.60	76.29	80.51	75.01
NN	**84.76**	76.53	80.12	74.97
KNN	73.71	52.31	63.74	52.43
LR	**85.46**	**77.53**	82.16	**76.25**
RF	**86.61**	69.17	**89.87**	73.05
DT	79.25	66.90	73.59	66.81

Table 2. Stress Detection Result Using Classifiers Ensemble (%)

Method	Accuracy	Precision	Recall	F1-measure
EH	86.99	76.00	88.02	**77.45**
ES	**87.10**	76.11	86.75	75.91

4. Conclusion

Enabling the workers to monitor their own stress level has many advantages. Knowing their own stress level can help them stay aware and feel more in control of their response to situations and know when it is time to relax or take some actions to treat it properly. This monitoring task can be enabled by using wearable devices to measure related physiological responses. Smartwatch is one of the devices that can be used for this task due to its usability for the working environment and its built-in sensors. In this work, we propose a multi-modal based continuous stress detection method using some individual classifiers and classifier ensembles.

The experiment results show that all classifiers work quite well to detect stress with an accuracy of more than 70%. RF obtained the best accuracy with 86.61% while KNN has the lowest accuracy with 73%. In terms of F1-measure, LR achieved the best F1-measure with 76.25%. Similar to the accuracy result, KNN also has the lowest F1measure (52.43%). The results also show that the ensemble method obtained higher performance compared to all individual classifiers. In this study, the advantage of one classifier can cover the shortcomings of other classifiers so that the accuracy can be improved. Furthermore, the results also show that ES (soft voting) has higher accuracy than EH (hard voting) in this study but EH has a better F1-measure than ES. In this study, the best accuracy is obtained by ES with 87.10%. Meanwhile, the best F1-measure was achieved by EH with 77.45%.

Our experimental study for the effect classifier ensemble is limited by the WESAD dataset that uses only two classes: stress and non-stress. In future work, the effect of the use of the ensemble method can be tested on a dataset that provides different stress levels (e.g., low stress, moderate stress, and high stress). Besides, a new dataset with more subjects could be created in the future in order to test the reliability of the proposed methods. The future dataset could also include not only label based on the intervention like in the WESAD dataset, but also the label from user-filled questionnaires (e.g., PSS).

References

[1] Marine A, Ruotsalainen JH, Serra C, Verbeek JH. Preventing occupational stress in healthcare workers. Cochrane Database of Systematic Reviews. 2006;(4).

[2] Weinberg A, Creed F. Stress and psychiatric disorder in healthcare professionals and hospital staff. the Lancet. 2000;355(9203):533–537.
[3] Pickering TG. Mental stress as a causal factor in the development of hypertension and cardiovascular disease. Current hypertension reports. 2001;3(3):249–254.
[4] Wang Y, Chen R, Zhang L. Reliability and validity of generalized anxiety scale-7 in inpatients in Chinese general hospital. J Clin Psychiatr. 2018;28:168–71.
[5] Welp A, Meier LL, Manser T. Emotional exhaustion and workload predict clinician-rated and objective patient safety. Frontiers in psychology. 2015;5:1573.
[6] Tsiga E, Panagopoulou E, Montgomery A. Examining the link between burnout and medical error: A checklist approach. Burnout Research. 2017;6:1–8.
[7] Fauzi MA, Yeng P, Yang B, Rachmayani D. Examining the Link Between Stress Level and Cybersecurity Practices of Hospital Staff in Indonesia. In: The 16th International Conference on Availability, Reliability and Security; 2021. p. 1–8.
[8] Liao W, Zhang W, Zhu Z, Ji Q. A real-time human stress monitoring system using dynamic Bayesian network. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)-workshops. IEEE; 2005. p. 70–70.
[9] Cohen S, Kamarck T, Mermelstein R. Perceived stress scale (PSS). J Health Soc Beh. 1983;24:285.
[10] Levenstein S, Prantera C, Varvo V, Scribano ML, Berto E, Luzi C, et al. Development of the Perceived Stress Questionnaire: a new tool for psychosomatic research. Journal of psychosomatic research. 1993;37(1):19–32.
[11] Spagnolli A, Guardigli E, Orso V, Varotto A, Gamberini L. Measuring user acceptance of wearable symbiotic devices: validation study across application scenarios. In: International Workshop on Symbiotic Interaction. Springer; 2015. p. 87–98.
[12] Lazaro MJS, Lim J, Kim SH, Yun MH. Wearable technologies: acceptance model for smartwatch adoption among older adults. In: International Conference on Human-Computer Interaction. Springer; 2020. p. 303–315.
[13] Schmidt P, Reiss A, Duerichen R, Marberger C, Van Laerhoven K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM international conference on multimodal interaction; 2018. p. 400–408.
[14] Garg P, Santhosh J, Dengel A, Ishimaru S. Stress Detection by Machine Learning and Wearable Sensors. In: 26th International Conference on Intelligent User Interfaces; 2021. p. 43–45.
[15] Indikawati FI, Winiarti S. Stress detection from multimodal wearable sensor data. In: IOP Conference Series: Materials Science and Engineering. vol. 771. IOP Publishing; 2020. p. 012028.
[16] Siirtola P. Continuous stress detection using the sensors of commercial smartwatch. In: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers; 2019. p. 1198–1201.
[17] Fauzi MA, Yuniarti A. Ensemble method for indonesian twitter hate speech detection. Indonesian Journal of Electrical Engineering and Computer Science. 2018;11(1):294–299.
[18] Reiss A, Stricker D. Introducing a new benchmarked dataset for activity monitoring. In: 2012 16th international symposium on wearable computers. IEEE; 2012. p. 108–109.
[19] Kreibig SD. Autonomic nervous system activity in emotion: A review. Biological psychology. 2010;84(3):394–421.
[20] Zhang Y, Haghdan M, Xu KS. Unsupervised motion artifact detection in wrist-measured electrodermal activity data. In: Proceedings of the 2017 ACM International Symposium on Wearable Computers; 2017. p. 54–57.
[21] Fauzi MA, Bours P. Ensemble Method for Sexual Predators Identification in Online Chats. In: 2020 8th International Workshop on Biometrics and Forensics (IWBF). IEEE; 2020. p. 1–6.
[22] Li L, Zhang Y, Zou L, Li C, Yu B, Zheng X, et al. An ensemble classifier for eukaryotic protein subcellular location prediction using gene ontology categories and amino acid hydrophobicity. PLoS One. 2012;7(1):e31057.
[23] Kumar BS, Ravi V. Text document classification with pca and one-class svm. In: Proceedings of the 5th international conference on frontiers in intelligent computing: theory and applications. Springer; 2017. p. 107–115.

Security, Privacy, Safety, and Ethics Issues in pHealth

This page intentionally left blank

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210609

Implementation of Privacy and Security for a Genomic Information System

Jaime DELGADO[1] [a], Silvia LLORENTE [a] and Guillem REIG [b]
[a] *Information Modeling and Processing (IMP) group – DMAG, Computer Architecture Dept. (DAC),*
[b] *inLab, Facultat d'Informàtica de Barcelona (FIB), Universitat Politècnica de Catalunya (UPC - BarcelonaTECH)*

Abstract. Genomic information is key for the implementation of real personalized medicine. Nevertheless, access to this kind of information must be controlled because of its high privacy and security requirements. Several genomic information formats exist, although we have started from MPEG-G as it includes metadata and protection mechanisms since its inception and provides a hierarchical structure to organize the information contained. The proposed GIPAMS modular architecture provides a secure and controlled access to genomic information, which may help on improving personalized medicine as described in this paper.

Keywords. Genomics, privacy, modular architecture, GIPAMS

1. Introduction

Personalized medicine is one of multiple examples of use of genomic information, which is a currently relevant research topic. In this context, genomic information sequencing and processing is gaining momentum. Sequencing price has decreased in the last years and now it is possible to sequence complete human genome for a thousand euros [1]. This is leading to an increase of information, making difficult its storage and management by different organizations.

On the other hand, genomic information should be protected from unauthorized access due to its specific characteristics, as it includes sensible information not only from a person but also her relatives. This means that, once the information is leaked, it cannot be revoked, like a certificate, and it is public "forever".

Nevertheless, protection mechanisms currently exist, but they have different limitations, such as being restricted for closed environments or being regulated by Data Access Committees, implying in this case a possible long and tedious process that may slow down reaching research results.

In order to provide a possible solution to privacy protection of genomic information, we present in this paper our proposal of a secure modular system, called GIPAMS (Genomic Information Protection And Management System), which defines mechanisms for privacy, protection, storage, search and access to genomic information. A first description of GIPAMS is given in [2]. We have implemented a first version of the

[1] Corresponding Author: Jaime Delgado, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain; e-mail: jaime.delgado@upc.edu

system. As information representation, we have chosen the file structure provided by ISO/IEC 23092, Genomic Information Representation (MPEG-G) [3][4]. The next subsection briefly introduces this structure.

In the rest of the paper we describe the proposed GIPAMS architecture, some details of its implementation with MPEG-G and some discussions on how it may help in the implementation of personalized medicine.

1.1. MPEG-G file structure

MPEG-G [3], currently developed and maintained by ISO/IEC JTC1 SC29/WG8, is a standard devoted to the representation of genomic information in a compressed and secure way, including metadata and protection features into the same file structure.

It is structured in hierarchical boxes, as represented in Figure 1. The boxes may contain data, like the File header (flhd) or subboxes, like Dataset Group (dgcn), which in turn may contain information boxes or container boxes, until the last level, which may be organized in access units (aucn) or in descriptor streams (dscn), depending on how the genomic information has to be accessed. In the end, genomic information is stored in blocks, regardless using access units or descriptor streams.

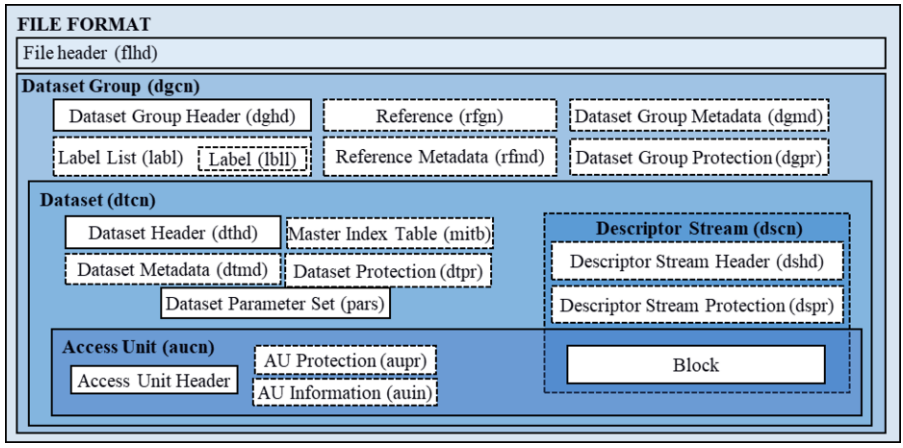


Figure 1. MPEG-G file format (adapted from MPEG-G standard).

2. Methods – GIPAMS Description

The architecture of the developed platform is an evolution of our original Multimedia Information Protection And Management System, MIPAMS [5]. As it now deals with genomic information, we have called it Genomic Information Protection And Management System, GIPAMS. Its structure is depicted in Figure 2. Not all the modules are currently implemented, but this is the expected final complete picture. The functionality of the different modules is as follows:

- User Application: Web application that sends requests to the Workflow Manager based on user actions. Communication between this application and the rest of the architecture is through a secure channel.

- Workflow Manager: Intermediate module that acts as a unique entry point to the system. It checks operation authorization before interacting with other modules.
- Authentication Service: User identification server, which uses OAuth 2.0 [6] and JSON Web Tokens [7]. Its implementation is currently based on Keycloak [8].
- Genomic Content Service: Module in charge of genomic archive management, both in reading and writing operations. It uses scripts as presented in 3.2.
- Authorization Service: Module for authorization rules validation.
- Search Service: Performs searches over genomic information. It provides extra filtering features with the use of a relational database.
- Policy Service: Module in charge of the creation of the authorization rules, which are organized into policies.
- Protection Service: Module which creates protection information metadata as well as applies the mechanisms defined (i.e. encryption, signature, etc.).
- Report / Track Service: Module in charge of reporting the operations done in the system, especially those not authorized.
- Certification Authority: Provides digital certificates to secure communications.

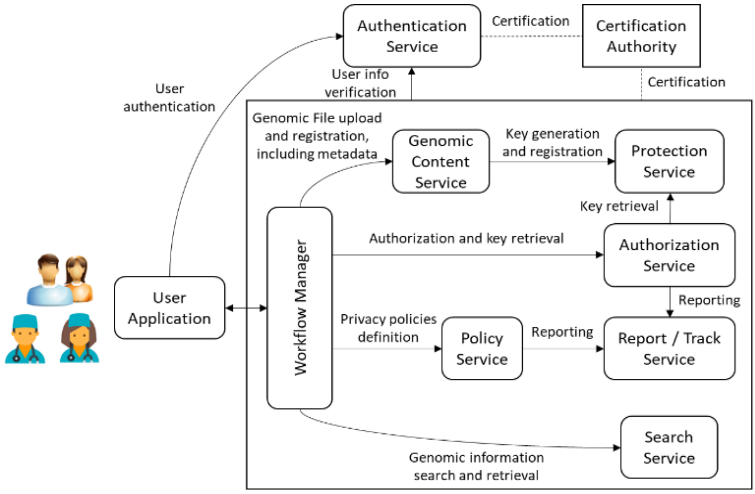


Figure 2. GIPAMS Architecture.

3. Results

When implementing GIPAMS with MPEG-G information, we need to deal with two relevant problems, first, integrating real metadata and, second, simplifying the file structure to facilitate development. We briefly describe how we have solved these problems in the next subsections.

3.1. Metadata Mapping

MPEG-G stores metadata in the information boxes Dataset Group Metadata (dgmd) and Dataset Metadata (dtmd) [9] using Extensible Markup Language (XML) [10].

To test the compatibility of the GIPAMS architecture with the MPEG-G metadata we have done a mapping of the metadata coming from two public organizations, the European Nucleotide Archive (ENA) [11] and the National Center for Biotechnology Information (NCBI) [12]. The structuring of metadata used in ENA is similar to the one defined in MPEG-G so the mapping is direct, as represented in Table 1. The structuring used at NCBI is more troublesome for us, as the Abstract field does not exist in the NCBI metadata and the Type field is not exactly the same as the one used in MPEG-G. The applied mapping is represented in Table 2.

MPEG-G also provides a mechanism to store additional information in the metadata fields. To test this mechanism we have defined three extensions for the NCBI metadata: one to store an additional identifier to the MPEG-G Sample field, another to store information about the organisms investigated in the project and store multiple investigation centers in a metadata file, and the last one to deal with an additional field in the NCBI BioSample, which contains extra information about the samples that cannot be stored in MPEG-G metadata fields. They are represented in Table 3.

MPEG-G Field	ENA Field
Title	Study - STUDY_TITLE
Type	Study - STUDY_TYPE
Abstract	Study - STUDY_ABSTRACT
ProjectCentre	Study - CENTER_PROJECT_NAME
Description	Study - STUDY_DESCRIPTION
Sample - TaxonId	Assembly - TAXON_ID
Sample - Title	Assembly – TITLE

Table 1. ENA metadata mapping.

MPEG-G Field	NCBI Field
Title	BioProject – Title
Type	BioProject - ProjectTypeSubmission
Abstract	Non existent
ProjectCentre	BioProject – Organization
Description	BioProject – Description
Sample - TaxonId	BioSample - TAXON_ID
Sample - Title	BioSample – TITLE

Table 2. NCBI metadata mapping.

AttributeExtension	NCBI Field
StudyDesign	BioSample - Attribute - study design
BodySite	BioSample - Attribute - body site
AnalyteType	BioSample - Attribute - analyte type
IsTumor	BioSample - Attribute - is tumor

Table 3. AttributeExtension fields.

3.2. File Structure Implementation

As stated in section 1.1, MPEG-G files are structured in hierarchical boxes forming a single file. To avoid having to deal with large files, we have used an alternative approach to simulate this structure, which is using the file system to represent the box structure.

In this approach, every box is represented by a directory, which contains multiple files like the headers, metadata or protection and some subdirectories representing the inner boxes.

The information files still need to be manipulated at bit level, so we have developed a Python script that can generate this whole directory structure and create the information files using valid data. The script can also integrate real metadata and protection policies into the files, using the data provided by the user. The complete MPEG-G file could be constructed from the directories and files stored in disk, if it is needed to share it with some other researcher or organization.

4. Discussion

Having a system like GIPAMS that can deal with genomic information in a modular and secure way may help in providing new services aimed at achieving personalized medicine.

An example of such a service is metadata search. By connecting metadata and genomic data in a more integrated, but flexible and efficient way, researchers may find subjects of interest for their research. The search may also include some information about the genomic data associated to metadata, especially if it can be accessed partially or completely. This point is controlled by means of the access rules defined for genomic metadata and data and of course supported by the encryption/decryption of the information. This could be for example very useful for rare diseases, where a few cases are available and the privacy and security standards should be the highest in order to not to reveal patient identity.

Furthermore, several GIPAMS' implementations could be established at different locations. The possibility of defining global access rules over the metadata stored in each location may facilitate the creation of a federated system providing federated search, still guaranteeing privacy and security. The advantage of having a GIPAMS federation is the fact that we may have several small modular systems, dealing only with their own genomic information (less storage and transmission required), but with the possibility of accessing to metadata describing other genomic studies that may be relevant for them. In a next step, request of the genomic information (or part of it), in addition to metadata search, could be implemented, also in a controlled and secure way.

Finally, the application of different security mechanisms provides privacy protection for the genomic information managed inside GIPAMS. First of all, the communication between the user application and the rest of the modules/services is done through a secure channel. Moreover, tracking of user actions and unique user identification are also implemented. For the protection of the genomic data and metadata, different encryption techniques can be applied. And to control which actions can users perform in the system, privacy protection rules can be defined with a high level of granularity. Although information could be leaked once at the user's application, we should be aware of two relevant facts: 1) users are identified and "trustable", 2) all actions on the information are tracked.

5. Conclusions and Future Work

This paper presents an initial implementation of GIPAMS, a modular architecture for the management of genomic information. Its first implementation is done based on the MPEG-G standard, which organizes genomic information, metadata and protection in a unique structure, facilitating access control, security and efficiency in storage.

The decision of using MPEG-G for the first implementation is twofold: It defines a clear hierarchy for representing and storing different kinds of genomic information and integrates security and protection mechanisms in it since its inception; i.e., by design. Based on this hierarchy, it seems feasible to integrate with other existing genomic information formats, facilitating its search and linkage through all the processes associated to the use of the information.

Moreover, the extension metadata mechanism offered by MPEG-G helps in the inclusion of new metadata, facilitating the implementation of more specific and accurate searches, giving access to more research results.

The implementation of such a system, which defines mechanisms for accessing genomic metadata and data in a secure and controlled way, may help in the implementation of more complex systems, like a federation of GIPAMS.

Apart from improving the federation facilities, next steps include to completely implement the modules conforming GIPAMS, and to extend supported genomic information formats, following a hybrid approach, which might consist on maintaining (at least part of) MPEG-G file structure including metadata and access rules, but supporting other genomic information representation formats in the lower levels.

Acknowledgement

This work is partly supported by the Spanish Government (GenClinLab-Sec, PID2020-114394RB-C31) and by the Generalitat de Catalunya (2017 SGR 1749).

References

[1] genome.gov, The Cost of Sequencing a Human Genome. 2020. https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost
[2] Delgado, J.; Llorente, S. FAIR aspects of a Genomic Information Protection and Management System. European Federation for Medical Informatics (EFMI) – Special Topic Conference (STC) 2021. 2021.
[3] ISO/IEC 23092-1:2020, Genomic information representation - Part 1: Transport and storage of genomic information. 2020. https://www.iso.org/standard/79882.html
[4] Alberti, C. et al. An introduction to MPEG-G, the new ISO standard for genomic information representation, bioRxiv, 2018. https://doi.org/10.1101/426353
[5] Llorente, S.; Rodriguez, E.; Delgado, J.; Torres, V. Standards-based architectures for content management, IEEE multimedia, 2012. https://doi.org/10.1109/MMUL.2012.58
[6] IETF. The OAuth 2.0 Authorization Framework, 2012. https://datatracker.ietf.org/doc/html/rfc6749
[7] IETF. JSON Web Token (JWT), 2015. https://datatracker.ietf.org/doc/html/rfc7519
[8] Keycloak, Open Source Identity and Access Management, 2021. https://www.keycloak.org/
[9] ISO/IEC 23092-3:2020, Genomic information representation - Part 3: Metadata and application programming interfaces (APIs). 2020. https://www.iso.org/standard/75625.html
[10] W3C, Extensible Markup Language (XML) 1.1 (Second Edition), 2006. https://www.w3.org/TR/xml11/
[11] European Nucleotide Archive (ENA). 2021. https://www.ebi.ac.uk/ena/browser/about
[12] National Center for Biotechnology Information (NCBI). 2021. https://www.ncbi.nlm.nih.gov/

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210610

Regulation Modelling and Analysis Using Machine Learning During the Covid-19 Pandemic in Russia

Egor TROFIMOV[a], Oleg METSKER[b], Georgy KOPANITSA[c], David PASHOSHEV[c,1]
[a]*The All-Russian State University of Justice, Moscow, Russia*
[b]*Almazov National Medical Research Centre, Saint-Petersburg, Russia*
[c]*ITMO University, Saint-Petersburg, Russia*

Abstract. Due to the specific circumstances related to the COVID-19 pandemic, many countries have enforced emergency measures such as self-isolation and restriction of movement and assembly, which are also directly affecting the functioning of their respective public health and judicial systems. The goal of this study is to identify the efficiency of the criminal sanctions in Russia that were introduced in the beginning of COVID-19 outbreak using machine learning methods. We have developed a regression model for the fine handed out, using random forest regression and XGBoost regression, and calculated the features importance parameters. We have developed classification models for the remission of the penalty and for setting a sentence using a gradient boosting classifier.

Keywords. COVID-19, machine learning, regulation, Russia

1. Introduction

Due to the specific circumstances related to the COVID-19 pandemic, many countries have enforced emergency measures such as self-isolation and restriction of movement and assembly, which are also directly affecting the functioning of their respective public health and judicial systems [1,2].

The breach of any disease containment measures during an emergency period constituted a criminal offense in many countries [3]. For example, Italy [4] was one of the first governments to declare the state of emergency on the 31st of January 2020. The Italian Government formally declared the state of emergency pursuant to Legislative Decree 1/2018 (Civil Protection Code). Later the states started to raise legal restrictions and change the regulations [5].

One of the main issues in many countries was the partition of authority between regional and national officials [6]. This had not only caused political tensions among the authorities themselves, but it resulted in different regulations, which did not allow legal certainty and consistency [7–9]. Various legal practices raised the discussion on the efficiency of the criminal sanctions to slow down the spread of the pandemic [10]. As the spread of pandemic is a complex multifactor, we don't think it is possible to

[1] Corresponding Author: David Paskoshev, ITMO University, Saint-Petersburg, Russia, E-mail: david.d.paskoshev@gmail.com

accurately identify the influence of only one legal measure [11]. At least it is worth and important to evaluate if the criminal restrictions themselves are efficient and clear to the people, public health and law enforcement services.

The goal of this study is to identify the efficiency of the criminal sanctions in Russia that were introduced in the beginning of COVID-19 outbreak using machine learning methods.

2. Methods

We believe that a well working administrative norm should have a low rate of s remission of a sentence, and an average amount of fine closed to the average amount of fine in the norm [12]. To measure the efficiency of the introduced regulation we have calculated features importance for the following outcomes:

- Amount of fine
- Remission of a penalty
- Setting a sentence

2.1. Data collection

The data for the study was extracted from the Russian national digital court decision database (http://www.cdep.ru) from the 1st January 2015 to the 30th of September 2020. In total we have extracted 58334 cases with 25682 cases from 2020. The dataset was randomly split into a training (80%) and a testing (20%) datasets.
We have extracted the narrative texts of every case and processed them using a data extraction software that was specially designed for Russian court decision processing [13].

2.2. Classification and Features importance

Features importance was analyzed using scikit learn library version 0.23 [14]. The following models were developed to calculate features importance.
We have developed a regression model for fine using a random forest regressor and a XGBoosting regressor, and calculated the features importance parameters.
We have developed classification models for the remission of the penalty and for setting a sentence using a gradient boosting classifier.
Each experiment ran in the setting of stratified 4-fold cross-validation i.e., random 80% of training dataset was used for training and random 20% of training dataset for testing. Target class ratios in the folds were preserved. After determining the optimal dataset and model parameters, we performed a validation with the testing dataset. As an additional performance assessment score, we used AUC of ROC, which represents the trade-off between sensitivity and specificity of the model.
Shapley values we calculated for the classification models on the remission of the penalty and setting a sentence.

3. Results

Feature's importance analysis for the prediction of amount of a fine is presented in Figure 1. The features "minimum" and "minimal" support the conclusion that the courts tended for a minimal possible sentence.

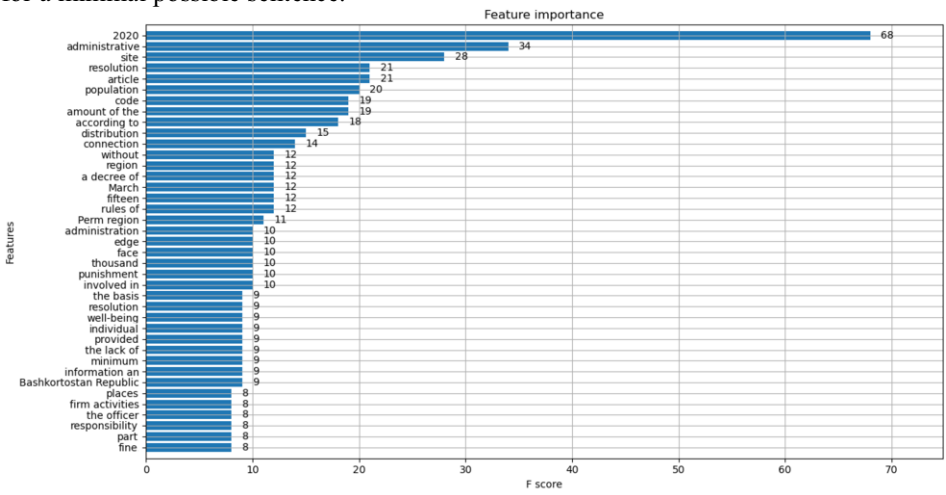


Figure 1. Features importance for the amount of fine

Figure 2 presents the Shapley values for the remission of penalty.

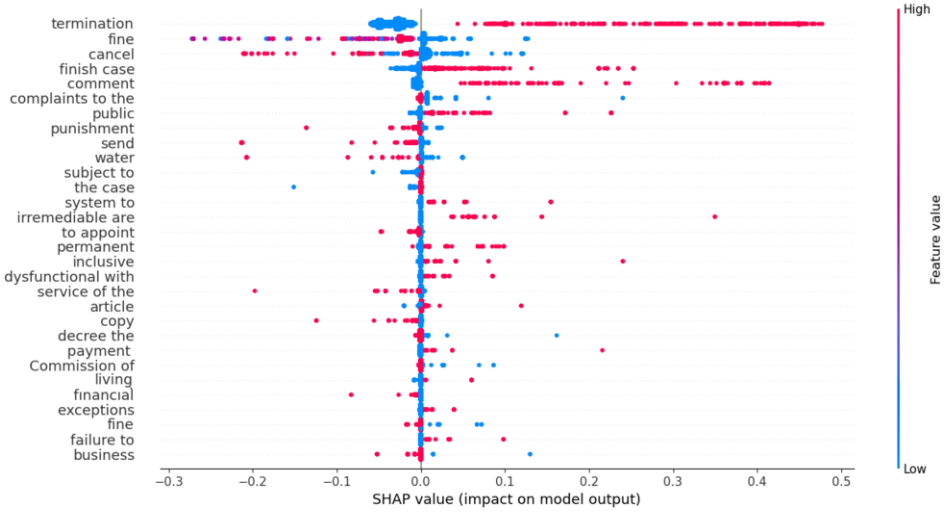


Figure 2. SHAP values for the remission of penalty

1) The feature «Formally» means cancellation of an earlier ruling and termination of proceedings when the violation was formal but not significant (e.g. a citizen moved a little more away from home than allowed).
2) The feature «intractable» supports the conclusion that the courts tended to terminate the cases due to presumption of innocence when all intractable doubts are interpreted in favor of the offender, which is important if the laws and rules are too confusing or unclear.

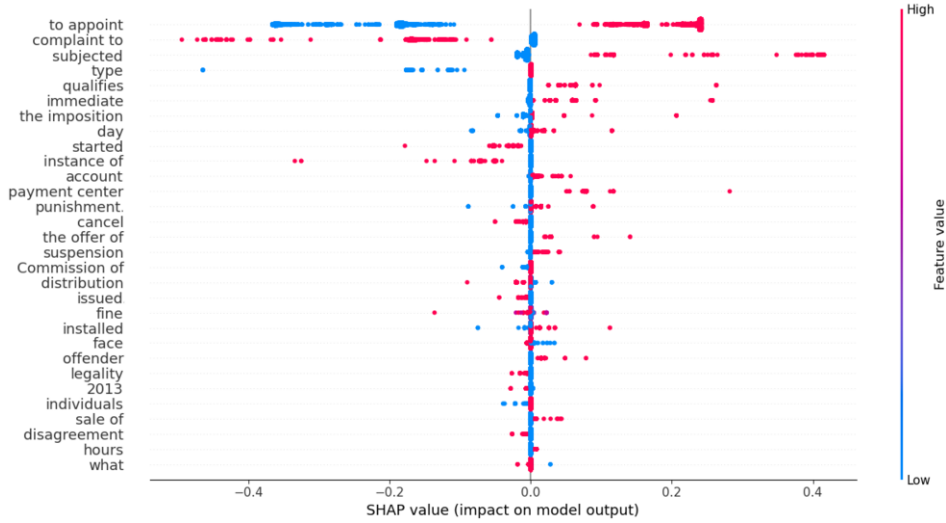


Figure 3. SHAP values for the setting a sentence

Here we can observe the logical result that indicate the correctness of the data extraction. The terms "to appoint", "subjected", "qualify", "impose", "pay", "punish", "suspend", "offender" – are all related to the assigning a punishment, so the red dots (figure 3) are on the right side.

The terms "started", "annul", "disagree" – indicate the termination of the started proceedings and annulment of the earlier decision in the case, so the red dots (figure 3) are on the left side.

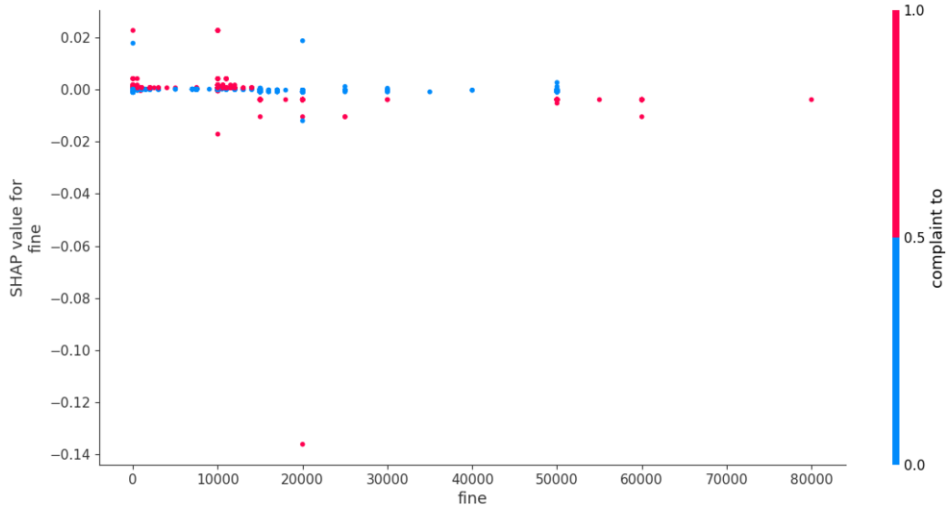


Figure 4. Dependency plot of the amount of the fine and the presence of the complaint

The plot (Figure 4) clearly shows that usually the complaint is associated with the appointment of a fine of fifteen thousand rubles positively on the appointment of punishment while after fifteen thousand rubles it is negative or not related.

4. Discussion

The new legislation came into force only on April 1. Assuming that no more violations were revealed in the first half of the year than a year earlier, we can conclude that for April, May and June (three months) of 2020 about 26685 cases were considered under paragraph 2 and paragraph 3 of Article 6.3 of the Code. The analysis showed that:
1) Almost every sixth case was terminated because there were no signs of an offence or because of the insignificance of the offence;
2) cases against legal entities constitute a negligible proportion (the number of legal entities held liable for violations in the sanitary-epidemiological area has decreased as compared to previous years);
3) Out of 10 thousand individuals who were punished, about 10 percent were individual entrepreneurs and about 5 percent were officials (the rest 85 percent or more - ordinary individuals);
4) warnings and fines were imposed on about 9 thousand individuals;
5) the average fine was about 18 thousand rubles, i.e. for all categories of individuals (individual entrepreneurs, officials, other individuals) the fine is approximately at the minimum defined in the article 6.3;
6) the courts have actively exercised their right to impose a penalty below the minimum statutory penalty, doing so in every tenth case. Although a year earlier, during the same period (quarter), all courts in Russia made such a decision only once. The probability of such condescension in cases on paragraphs 2 and 3 of Article 6.3 of the Code was about 10 percent, while for other violations in the sanitary-epidemiological sphere - about 0.04 percent.

5. Conclusion

Our findings prove that the law norms and explanations are too abstract and uncertain by the number of canceled cases and the amount of the decisions with a minimum fine. The results of the study can be beneficial for the analysis of the legal practice and improving legislation and law enforcement in the situations of pandemic.

6. Acknowledgement

This work financially supported by the government of the Russian Federation through the ITMO fellowship and professorship program. The reported study was funded by RFBR, project number 20-011-00837.

References

[1] Morris DH, Rossine FW, Plotkin JB, Levin SA. Optimal, near-optimal, and robust epidemic control. Communications Physics 2021; 4(1): 1–8. doi:10.1038/s42005-021-00570-y.
[2] Xu C, Luo X, Yu C, Cao S-J. The 2019-nCoV epidemic control strategies and future challenges of building healthy smart cities. Indoor Built Environment 2020; 29(5): 639–644. doi:10.1177/1420326X20910408.

[3] Rosenbaum L. Facing Covid-19 in Italy — Ethics, Logistics, and Therapeutics on the Epidemic's Front Line, N Engl J Med. 2020 May 14; 382(20): 1873-1875. doi: 10.1056/NEJMp2005492. Epub 2020 Mar 18.
[4] Canestrini N. Covid-19 Italian emergency legislation and infection of the rule of law. New Journal of European Criminal Law 2020; 11(2): 116-122. doi:10.1177/2032284420934669.
[5] Pinkas J, Jankowski M, Szumowski L, Lusawa A, et al. Public health interventions to mitigate early spread of SARS-CoV-2 in Poland. Med Sci Monit. 2020 Apr 13;26:e924730. doi:10.12659/MSM.924730.
[6] Gostin LO, Hodge JG. US Emergency Legal Responses to Novel Coronavirus: Balancing Public Health and Civil Liberties. JAMA 2020 Mar 24;323(12):1131-1132. doi: 10.1001/jama.2020.2025.
[7] Pinheiro VM, Ilarraz M, Mestriner MT. The impacts of the COVID-19 crisis on the Brazilian legal system–a report on the functioning of the branches of the government and on the legal scrutiny of their activities. The Theory and Practice of Legislation 2020; 8: 193-212; https://doi.org/10.1080/20508840.2020.1790104
[8] Parmet WE, Sinha MS. Covid-19 — The Law and Limits of Quarantine. N Engl J Med. 2020 Apr 9;382(15):e28. doi: 10.1056/NEJMp2004211. Epub 2020 Mar 18.
[9] Chinazzi M, Davis JT, Ajelli M, Gioannini C, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science. 2020 Apr 24;368(6489):395-400. doi: 10.1126/science.aba9757. Epub 2020 Mar 6.
[10] Stogner J, Miller BL, McLean K. Police Stress, Mental Health, and Resiliency during the COVID-19 Pandemic. Am J Crim Justice. 2020 Jun 26: 1–13. doi: 10.1007/s12103-020-09548-y [Epub ahead of print]
[11] Miller JM, Blumstein A. Crime, Justice & the COVID-19 Pandemic: Toward a National Research Agenda, American Journal of Criminal Justice 2020; 45: 515-524. doi:10.1007/s12103-020-09555-z.
[12] Voigt S. How to Measure the Rule of Law, Kyklos. May 2012; 65(2): 262-284. doi:10.1111/j.1467-6435.2012.00538.x.
[13] Metsker O, Trofimov E, Grechishcheva S. Natural Language Processing of Russian Court Decisions for Digital Indicators Mapping for Oversight Process Control Efficiency: Disobeying a Police Officer Case, in: Communications in Computer and Information Science, 2020; 1135: . doi:10.1007/978-3-030-39296-3_22.
[14] Pedregosa F, Varoquaux G, Gramfort A, Michel V, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011;12(85):2825−2830.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210611

Association Between Physical Activity and Osteoarthritis of Knee with Quality of Life in Community-Dwelling Older Adults

Nachalida YUKALANG[a,b], Niruwan TURNBULL[a,b], Wisit THONGKUM[a,b], Adisorn WONGKONGDECH[a], Kukiat TUDPOR[a,b, 1]
[a]*Faculty of Public Health, Mahasarakham University, Thailand*
[b]*Public Health and Environmental Policy in Southeast Asia Research Unit (PHEP-SEA), Mahasarakham University, Thailand*

Abstract Older adults are relatively physically active compared to other age group. A lack of physical activity (PA) can cause chronic diseases including osteoarthritis of knee (OA knee) and might eventually reduce quality of life (QOL). This present study was aimed to investigate association between levels of PA and OA knee with levels of QOL in community-dwelling older adults. One thousand and sixty-seven community-dwelling older persons were recruited to this descriptive study. PA activity questionnaire was invented. Standardized Oxford knee score and World Health Organization's Quality of Life scale (WHOQOL-BREF) were used to measure OA knee and QOL levels. Results showed that levels of PA in older adults were significantly associated with levels of OKS ($\chi^2 = 78.565$, P-value < .001) and levels of OA knee in older adults were significantly associated with levels of overall QOL ($\chi^2 = 57.738$, P-value < .001). Pearson's correlation also showed interrelation among PA, OA knee, and QOL. In conclusion, PA, OA knee, and QOL are interrelated. Therefore, close monitoring and design of proper PA activity should be implemented in community-dwelling older adults with OA knee.

Keywords. Osteoarthritis of knee, Older adults, Physical activity, Quality of life, Personal health

1. Introduction

Older adults are the least physically active age group with less than 25% performing regular physical activity (PA) [7]. A lack of PA has been considered as an important factor in etiology of diabetes, hypertension, cancer, and osteoarthritis [6]. Osteoarthritis of knee (OA knee) is a leading cause of global disability and is associated with significant economic costs [3]. The incidence of OA knee increases with advancing age, especially in the people of age above 60 years old [8]. OA knee leads to knee joint instability, postural sway, and risk of fall [20]. The Oxford Knee Score (OKS) is well correlated increasing age [4]. Quality of life (QOL) is an individual's perception of position in life in the context of the culture and value systems in relation to one's goals, expectations, standards and concerns [13]. The World Health Organization's Quality of Life scale (WHOQOL-BREF) is a widely used, reliable, valid, and self-report questionnaire which

[1] Corresponding Author: Kukiat Tudpor, Faculty of Public Health, Mahasarakham University, Kham Rieng, Kantarawichai, Maha Sarakham, Thailand 44150; E-mail: kukiat.t@msu.ac.th

assesses 4 domains of QOL in terms of physical health, psychological health, social relationships, and environment [22]. Relationship between PA and QOL or OA knee and QOL have been extensively reported [1; 11; 16; 17; 19; 23]. However, interrelationship among these three domain were still elusive. Therefore, this present study was aimed to investigate association between levels of PA and OA knee with levels of QOL in community-dwelling older adults.

Table 1 Questionnaires of physical activity over the last one month

Items
1 .You exercise 3-5 times a week.
2 .You exercise at least 30 minutes at a time.
3. You can walk until you feel moderately tired or breathe a little harder than usual, but can still talk to others until the end of the sentence at least 3- 5days a week
4. You can jog until you feel moderately tired or breathe a little harder than usual, but can still talk to others until the end of the sentence at least 3- 5days a week
5 .You warm up 5- 10minutes before exercising or playing sports.
6 .You stretch after exercising or playing sports for 5-10 minutes.
7 .You do housework such as sweeping and mopping the house by yourself.
8 .You walk up and down stairs by yourself.

2. Methods

2.1. Study design and participants

One thousand and sixty-seven community-dwelling older persons from Nonghi district, Roi Et, Thailand, voluntarily participated in this descriptive research. Inclusion criteria was age ≥ 60 years old. All procedures have been voted by 2 reviewers, approved by the Ethics Committee for Research Subjects of Roi Et Provincial Public Health Office, Ministry of Public Health and endorsed by the chairperson Mr.Pitak Payuha (No.COE 063256).

2.2. Questionnaires

The 8-item questionnaire of physical activity (Table 1) was approved by index of item-objective congruence (IOC), a procedure used in test development for evaluating content validity at the item development stage. Only the items with IOC scores ≥ 0.5 were qualified for the questionnaire. The items with Cronbach's α (test score reliability coefficient) ≥ 0.7 were acceptable and used in the questionnaire. Each item was scored 1-4 (never, sometimes, frequently, and always, respectively), given total score of 32. The total score was further ranked as physical activity levels from low, moderate, and high (< 17, 17-24, and ≥ 25, respectively). The Oxford Knee Score (OKS) and WHOQOL-BREF are standardized questionnaires [12; 13]. The WHOQOL-BREF consists of 4 subscales with 26 items that measure different health concepts. Responses are rated from 1 (very poor/very dissatisfied/not at all) to 5 (very good/very satisfied/completely). The total score ranging from 0-100 is calculated by averaging all item responses. Score ≥ 60 and < 60 indicate good and poor QOL, respectively [18].

2.3. Statistical Analyses

The data was employed descriptive statistics, then the association of categorical and continuous data were analyzed by χ^2 and Pearson's correlation tests (a measure of linear correlation between two sets of data). The level of statistical significance was $p<0.05$. All data were analyzed by SPSS Statistics version 18.

Table 2 Personal characteristics of participants

Characteristics	Number	Percentage
Sex		
Male	425	39.8
Female	642	60.2
Age (years old)		
60 - 69	529	54.3
70 - 79	315	32.3
> 80	130	13.3
Educational level		
Primary school	7	0.7
Secondary school	973	92.7
High school	53	5.0
Diploma	5	0.5
Bachelor and higher	12	1.1
Marital status		
Single	43	4.1
Married	609	58.2
Divorced/widowed	364	34.5
Separated	34	3.2
Average income per month		
less than 5,000 baht	964	90.4
5,000 – 10,000 baht	82	7.7
10,001 – 15,000 baht	8	0.8
More than 15,000 baht	12	1.1
Body mass index level		
Underweight	644	60.8
Normal	370	34.9
Overweight	19	1.8
Obese	26	2.5
Living status		
Alone	64	6.0
With spouse	545	51.1
With children/grandchildren	491	46.0
With other relatives	33	3.1

3. Results

3.1. Association between levels of physical activity and levels of OA knee

Personal characteristics of the participants are shown in Table 2. Results showed that the levels of PA of 1,067 older adults were significantly associated with levels of OKS ($\chi 2 = 78.565$, P-value < .001) (Table 3).

Table 3 Association between levels of physical activity and levels of OA knee

Physical activity level	OA knee level					
	Excellent	Good	Moderate	Poor	χ^2	P-value
High	277	102	20	8	78.565	<.001

Moderate	292	114	34	25
Low	83	51	28	33

3.2. Association between Levels of OA Knee and Levels of Quality of Life

Chi-square analysis showed that the levels of OA knee in older adults were significantly associated with levels of overall QOL ($\chi^2 = 57.738$, P-value < .001) (Table 4).

Table 4 Association between levels of OA knee and levels of quality of life

OA knee level	Quality of life level			
	Good	Poor	χ^2	P-value
Excellent	597	55	57.738	<.001
Good	232	35		
Moderate	75	7		
Poor	40	26		

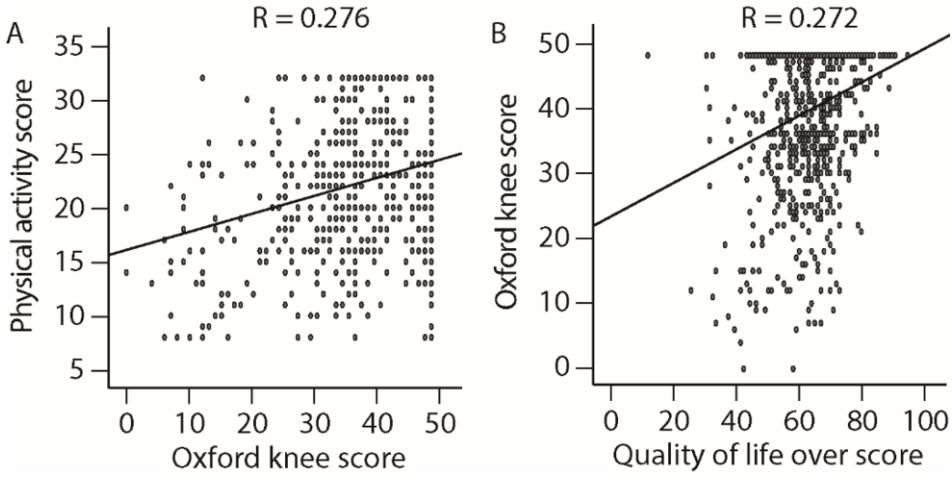


Figure 1. Pearson correlation between levels of physical activity and levels of OA knee (A); and levels of OA knee and levels of quality of life (B)

3.3 Correlation of scores of physical activity, osteoarthritis of knee, and quality of life

Pearson's correlation analysis revealed that PA, OKS, and WHOQOL-bref scores were interrelated (Pearson's correlation = .275 and .273; R = 0.276 and 0.272; and P-value < .001 and .001, in PA vs OKS (Figure 1A) and OKS vs WHOQOL-bref (Figure 1B), respectively).

4. Discussion

4.1. Characteristics of the Older Adults

The characteristics studied were the sex of the participants, age, educational levels, marital status, income, body mass index, and living status. Most of the participants low educational levels and low income (< 5,000 baht). Low educational levels are inversely

associated with higher QOL scores [15]. Similarly, low income is also associated low QOL level [10]. Therefore, low educational levels and low income found in this present study might be also the factors for low QOL.

4.2. Physical Activity on Osteoarthritis of Knee

This study revealed that PA levels were associated with OA knee levels. PA and OA knee levels are interrelated. Previous study showed that pain and disability in OA knee individuals limited an ability to perform exercises [17]. The PA with low impact on knee joints was reportedly beneficial to OA outcomes [19]. For OA knee patients, low-intensity exercise was recommended [16]. Therefore, it can be implied that appropriate level of PA might be able to prevent OA and the low-intensity with low impact on lower extremities should be implemented in older adult population.

4.3. Osteoarthritis of Knee and Quality of Life

Previous study showed that QOL was associated with self-reported disability in patients with OA knee [1; 11; 23]. OA knee also influences mental health and QOL [14]. Both conservative and surgical treatments were found to improve QOL in individuals with OA knee [5; 9; 21]. Interestingly, the older adults with OA knee reported their good QOL when they participated voluntary work and received proper recommendation for physical activity, especially resistance exercises [2]. This finding indicates that better QOL of the persons with OA knee can be obtained through appropriate design of physical activity.

5. Conclusion

In conclusion, physical activity, osteoarthritis of knee, and quality of life are interrelated. Therefore, in order to improve their quality of life, appropriate levels of physical activity should be tailor-made for individual needs of the older adults with osteoarthritis of knee. Further analysis for health service accessibility among osteoarthritis of knee patients as well as follow-up of its impacts on quality of life are of interest in the field of community public health and personalized health.

References

[1] Alkan BM, Fidan F, Tosun A, Ardicoglu O. Quality of life and self-reported disability in patients with knee osteoarthritis, Mod Rheumatol 2014; 24: 166-171.
[2] Alves JC, Bassitt DP. Quality of life and functional capacity of elderly women with knee osteoarthritis, Einstein (Sao Paulo) 2013; 11: 209-215.
[3] Araujo-Castillo RV, Culquichicon C, Solis Condor R. Burden of disease due to hip, knee, and unspecified osteoarthritis in the Peruvian social health insurance system (EsSalud), 2016, F1000Res 2020; 9: 238.
[4] Bremner-Smith AT, Ewings P, Weale AE. Knee scores in a 'normal' elderly population, Knee 2004; 11:, 279-282.
[5] Dervin GF, Stiell IG, Rody K, Grabowski J. Effect of arthroscopic debridement for osteoarthritis of the knee on health-related quality of life, J Bone Joint Surg Am 2003; 85: 10-19.
[6] Diet W. Nutrition and the prevention of chronic diseases. Report of a joint WHO/FAO expert consultation, WHO Technical report series 2003; 916: 34-38.
[7] Fakhouri TH, Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of Obesity Among Older Adults in the United States, 2007-2010, NCHS data brief 2012: 1-8.

[8] Felson DT. Clinical practice. Osteoarthritis of the knee, N Engl J Med 2006; 354: 841-848.
[9] Fransen M, Crosbie J, Edmonds J. Physical therapy is effective for patients with osteoarthritis of the knee: a randomized controlled clinical trial, J Rheumatol 2001; 28: 156-164.
[10] Georgellis Y, Tsitsianis N, Yin YP. Personal values as mitigating factors in the link between income and life satisfaction: Evidence from the European Social Survey, Social Indicators Research 2009; 91: 329-344.
[11] Kawano MM, Araujo IL, Castro MC, Matos MA. Assessment of quality of life in patients with knee osteoarthritis, Acta Ortop Bras 2015; 23: 307-310.
[12] Khatri K, Lakhotia D, Sharma V, Kiran Kumar GN, Sharma G, Farooque K. Functional Evaluation in High Energy (Schatzker Type V and Type VI) Tibial Plateau Fractures Treated by Open Reduction and Internal Fixation, Int Sch Res Notices 2014 (2014), 589538.
[13] World Health Organization. WHOQOL-BREF: introduction, administration, scoring and generic version of the assessment: field trial version, December 1996. In: World Health Organization, 1996.
[14] Park HM, Kim HS, Lee YJ. Knee osteoarthritis and its association with mental health and health-related quality of life: A nationwide cross-sectional study, Geriatr Gerontol Int 2020; 20: 379-383.
[15] Raggi A, Corso B, Minicuci N, Quintas R, Sattin D, et al. Determinants of Quality of Life in Ageing Populations: Results from a Cross-Sectional Study in Finland, Poland and Spain, PLoS One 2016; 11: e0159293.
[16] Regnaux JP, Lefevre-Colau MM, Trinquart L, Nguyen C, Boutron I, et al. High-intensity versus low-intensity physical activity or exercise in people with hip or knee osteoarthritis, Cochrane Database Syst Rev (2015), CD010203.
[17] Rosemann T, Kuehlein T, Laux G, Szecsenyi J. Osteoarthritis of the knee and hip: a comparison of factors associated with physical activity, Clin Rheumatol 2007; 26: 1811-1817.
[18] Silva PAB, Soares SM, Santos JFG, Silva LB. Cut-off point for WHOQOL-bref as a measure of quality of life of older adults, Revista de saude publica 2014; 48: 390-397.
[19] Sutton AJ, Muir KR, Mockett S, Fentem P. A case-control study to investigate the relation between low and moderate levels of physical activity and osteoarthritis of the knee using data collected as part of the Allied Dunbar National Fitness Survey, Ann Rheum Dis 2001; 60: 756-764.
[20] Tudpor K, Kanjanawanishkul K, Kam-Ard S, Intarak T, Traithip W, et al. Star Excursion Balance Test as an Exercise to Improve Static and Dynamic Balance in Community-Dwelling Persons with Unilateral Osteoarthritis of Knee, Editorial Advisory Board 2021; 15: 30.
[21] Usman Z, Maharaj SS, Kaka B. Effects of combination therapy and infrared radiation on pain, physical function, and quality of life in subjects with knee osteoarthritis: A randomized controlled study, Hong Kong Physiother J 2019; 39: 133-142.
[22] Vahedi S. World Health Organization Quality-of-Life Scale (WHOQOL-BREF): Analyses of Their Item Response Theory Properties Based on the Graded Responses Model, Iran J Psychiatry 2010; 5: 140-153.
[23] Xie Y, Yu Y, Wang JX, Yang X, Zhao F, et al. Health-related quality of life and its influencing factors in Chinese with knee osteoarthritis, Qual Life Res 2020; 29: 2395-2402.

Posters

This page intentionally left blank

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210613

Factors Predicting Sexual Risk Behaviors of Adolescents in North-Eastern Thailand

Athiwat BUTDABUT[1] and Pissamai HOMCHAMPA
Health System Science Department,Faculty of Medicine, Mahasarakham University, Maha Sarakham, Thailand

Abstract. The number of young people who have had sex at an early age increases in proportion, it concerns unsafe sexual behaviors, teenage pregnancy, HIV aids and sexually transmitted infections (STIs). This study examines the health behaviors and factors predicting sexual risk behaviors pertaining to teenage pregnancy among adolescents in Thailand. Adolescents consulted the reproductive health center about problems with the same gender. The factors of adolescent reproductive behaviors were significantly associated with age, education level, and the perception of peer norms. Receiving social support from media information also significantly correlated with those behaviors. The results recommend that to prevent premature pregnancy, adolescents should protect themselves. Parents should take the issue of social media use by their teenagers very seriously.

Keywords: Teen pregnancy, teen sexual health, pregnancy prevention, reproductive health

1. Introduction

Adolescence is a unique age of transition from childhood to adulthood in many aspects [7]. The subjects change many varieties including physical, mental, and social shifts. These changes bring both risks and opportunities that influence the development into adulthood. Teenagers today are growing up in societies with developed economies, social systems and technologies, especially communication technology. Advancements such as social media with mobile phones, tablets, and other wireless devices have revolutionized communication styles and the way of accessing knowledge. According to WHO, approximately 12 million girls aged 15–19 years and at least 777,000 girls under 15 years give birth each year in developing regions [1]. At least 10 million unintended pregnancies occur each year among adolescent girls aged 15–19 years in the developing world [5]. Complications during pregnancy and childbirth are the leading cause of death for 15–19-year-old girls globally [6]. Adolescent mothers (ages 10–19 years) face higher risks of eclampsia, puerperal endometritis, and systemic infections, compared to women aged 20 to 24 years, and their babies face higher risks of low birth weight, preterm delivery and severe neonatal conditions [10]. This study investigates the behaviors that may play a role in the reproductive health of teenagers in Health Region7 in Northeast of Thailand, aiming at determining which behaviors enhance reproductive health

[1] Corresponding Author: Athiwat Butdabut, Faculty of Medicine, Mahasarakham University, MahaSarakham, Thailand. Email: athiwatbutdabut@gmail.com

promotion among teens with the goal of further developing, supporting, and encouraging them.

2. Methods

2.1. Study Design and Participants

The sample comprised 15–19 years old secondary school teenagers from 4 provinces under the responsibility of the 7th Regional Health Promotion Centre including Khon Kean, Roi-Et, Mahasarakham, and Kalasin. Selected by a randomization method, we included 389 teenagers in our study.

2.2. The Questionnaire Ethical Concern and Statistical Analysis

The questionnaire consisted of 137 questions on following topics: 1) Socio-economic and characteristic, 2) behaviors related to reproductive health, 3) attitudes towards self-control ability, 4) gender, 5) peer norms perception, 6) social support in media information, confirmed by Index of Item-Objective Congruency (IOC) scores ≥ 0.5 Cronbach's $\alpha \geq 0.7$. The participants were protected, being very aware of the individual rights and taking action to prevent inadvertent negative effects based on ethical research principles, accredited by the Human Research Ethics Committee of the Mahasarakham University (No. 78/2018). Data analysis used descriptive statistics (Pearson Chi-square), then testing correlation of each categorical data by χ^2 test. The level of statistical significance was $p<0.05$. All data were analyzed by SPSS Statistics version 18.

3. Results

Table 1. Correlation between other Factors and Teen Reproductive Health Behavior (n = 389)

Factors	Adolescent Reproductive Behaviors			Chi-Square	*p*-value
	Poor	Medium	Good		
Gender				8.489[a]	0.075
- Male	17(25.6)	28(35.4)	34(43.0)		
- Female	56(19.6)	139(48.8)	90(31.6)		
- LGBT	7(28.0)	14(56.0)	4(16.0)		
Age				27.217[a]	<0.001**
- Less than 16 years	12(13.3)	28(31.1)	50(55.6)		
- Over 16 years	68(22.7)	153(51.2)	78(26.1)		
Education Level				47.366[b]	<0.001**
- High school	22(12.0)	71(38.8)	90(49.2)		
- Voc. Cert.	55(28.1)	103(52.6)	38(19.3)		
- High Voc. Cert.	3(30.0)	7(70.0)	0(0.0)		
GPA				3.544[a]	0.170
- Less than 3.00	12(32.4)	15(40.6)	10(27.0)		
- More than or equal to 3.00	68(19.3)	166(47.2)	118(33.5)		
Monthly Expense				1.737[a]	0.420
- Less than2,000 baht	24(24.7)	45(46.4)	28(28.9)		
- More than 2,000 baht	56(19.2)	136(46.6)	100(34.2)		
Peer Norms Perception				80.233[a]	<0.001**
- Very risky level	69(30.3)	115(50.4)	44(19.3)		
- Moderate risk level	7(16.3)	26(60.5)	10(23.2)		
- Low risk level	4(3.4)	40(33.9)	74(62.7)		
Receiving Social Support Information from Media				11.468[b]	0.013*
- Internet/TV programme	68(18.8)	170(47.1)	123(34.1)		
- Printed media	10(50.0)	8(40.0)	2(10.0)		
- Health personnel	2(25.0)	3(37.5)	3(37.5)		

[a] Pearson Chi-square, [b] Fisher's Exact Test, * Significance level 0.05, ** Significance level *p*-value<0.001

The gender distribution among the adolescent showed females (73.3%), males (20.3%), and LGBT (Lesbian, Gay, Bisexual, and Transgender) (6.4%) with an average age of 16.55± 1.18. Most of the students studied at the first year vocational certificate level (27.0%). The most family received an income less than 15,000 baht per month (36.4%). Most teen caregiver's relationships were parents (76.9). Most people lived in parent's house (88.9%), mostly in adolescent's family with 4-6 people (68.6%). Adolescents and friends consulted on reproductive health problems mostly hold the same gender (88.9). The factors significantly associated with adolescent reproductive behaviors were: 1) age group (p-value <0.001), 2) educational level (p-value <0.001), 3) perception of peer group norms (p-value <0.001), and 4) social support in media information (p-value = 0.013).

4. Discussion

4.1. Characteristics of the Adolescent and their Reproductive Behaviors

The result revealed that age group and educational level were significantly associated with adolescent reproductive behaviors. An earlier study reported that the age group is of relevance. Females who reported indirect aggression toward peers had earlier ages at first sexual intercourse, while females who were more victimized in adolescence experienced later ages at first sexual intercourse [9]. The education also affected several things for females in Pakistan. Literacy, for instance, is lower for women than for men. Only 20% of all females have attended primary school. Although most women know at least 1 contraceptive method, it is the urban educated woman who is twice as likely to know a source of supply and 5 times more likely to be a user [2]. Thai culture values virginity in the female gender, and marriage is the method that makes early pregnancy more acceptable. The research results eventually prompted health officials to advise teenagers on pregnancy prevention. Most of the peer norms perception are at high and medium risk level regarding adolescent reproductive behaviors. Van de Bongardt et al. [8] performed a meta-analysis to investigate the associations between three types of peer norms (descriptive norms (peer sexual behaviors), injunctive norms (peer sexual attitudes)), the peer pressure to have sex, and two adolescent sexual behavior outcomes (sexual activity and sexual risk behavior). Adolescent sexual activity was stronger associated with descriptive norms than with injunctive norms or peer pressure. Compared with the sexual activity outcome, the effect size of descriptive norms (peer sexual risk behavior) for sexual risk behavior was smaller. They also approved that age, gender, peer type, and socio-cultural context significantly moderated these associations. Another study explains that perceived peer norms supporting safer sex were inversely associated with recently having two or more sexual partners after controlling for demographic characteristics. Perceived peer norms around safer sexual behavior contribute to a lower likelihood of engaging in two HIV/STI risk behaviors: inconsistent condom use and multiple partnering [4].

4.2. Social Media Support Information

The adolescent's behaviors are also correlated to social support received by gathering information from any kinds of media. Regarding the use of information especially in the

context of health education such as safe sex, we recommend that adults should select supporting media. A study [3] conducted a systematic review of the literature to examine the effectiveness of social media among young adults aged 15 through 24 years, which indicated that social media and text messaging can increase knowledge regarding the prevention of STDs. These interventions may also affect behavior, such as screening/testing for STDs, sexual risk behaviors, and STD acquisition, but the evidence for effect is weak.

5. Conclusion

In conclusion, age group and educational level were significantly associated with adolescent reproductive behaviors. It also verified that the adolescent is influenced by peer norms perceptions in both the very risk level and medium level of adolescent reproductive behaviors. These findings emerge empirical factors for the risk behavior especially of peer norms and hold important implications for reproductive health of teenagers.

References

[1] Darroch JE, Woog V, Bankole A, Ashford LS. Adding it up: costs and benefits of meeting the contraceptive needs of adolescents. New York: Guttmacher Institute, 2016, https://www.guttmacher.org/report/adding-it-meeting-contraceptive-needs-....
[2] Greenspan A. Effects of education on reproductive behavior: lessons from Pakistan. Asia Pac Pop Policy. 1992 Dec;(23):1-4.
[3] Jones K, Eathington P, Baldwin K, Sipsma H. The impact of health education transmitted via social media or text messaging on adolescent and young adult risky sexual behavior: a systematic review of the literature. Sex Transm Dis. 2014 Jul;41(7):413-9. doi: 10.1097/OLQ.0000000000000146.
[4] Kapadia F, Frye V, Bonner S, Emmanuel PJ, Samples CL, Latka MH. Perceived peer safer sex norms and sexual risk behaviors among substance-using Latino adolescents. AIDS Education and Prevention 01 Feb 2012, 24(1):27-40. DOI: 10.1521/aeap.2012.24.1.27 PMID: 22339143
[5] Kiani MA, Ghazanfarpour M, Saeidi M. Adolescent pregnancy: a health challenge. International Journal of Pediatrics 2019; 7(7): 9749-9752.
[6] Neal S, Matthews Z, Frost M, Fogstad H, Camacho AV, Laski L. Childbearing in adolescents aged 12–15 years in low resource countries: a neglected issue. New estimates from demographic and household surveys in 42 countries. Acta Obstet Gynecol Scand. 2012 Sep;91(9):1114-8. doi: 10.1111/j.1600-0412.2012.01467.x.
[7] Spear LP. Neurobehavioral changes in adolescence. Current Directions in Psychological Science 2000; 9(4): 111-114.
[8] Van de Bongardt D, Reitz E, Sandfort T, Deković M. A meta-analysis of the relations between three types of peer norms and adolescent sexual behavior. Pers Soc Psychol Rev. 2015 Aug;19(3):203-34. doi: 10.1177/1088868314544223.
[9] White DD, Gallup AC, Gallup Jr GG. Indirect peer aggression in adolescence and reproductive behavior. Evolutionary Psychology 2010; 8: 147470491000800106.
[10] WHO U, Mathers C. Global strategy for women's, children's and adolescents' health (2016-2030). Organization 2016; 201: 4-103.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210614

Conceptual Model for Behaviour Change Progress – Instrument in Design Processes for Behaviour Change Systems

Helena LINDGREN[1] and Saskia WECK
Department of Computing Science, Umeå University, Umeå, Sweden

Abstract. The aim of the work presented in this article was to develop a conceptual model for behavior change progress, which could be used for automated assessment of reasons for progress or non-progress. The model was developed based on theories for behavior change, and evaluated by domain experts. The information models of two prototype systems of a digital coach under development for preventing cardio-vascular diseases and stress respectively, were evaluated by comparing the content of the prototypes with concepts in the model. The conceptual model was found useful as instrument to evaluate to what extent the prototypes are based in theories for behavior change, whether some vital information is missing, and to identify mechanisms for short and long time goal setting. Moreover, the connection between the ontology underpinning the prototypes and the conceptual model could be defined. Future work includes the integration of the conceptual model to function as a meta-ontology, which could be used for capturing causal relationships between information collected by the applications at baseline and at runtime.

Keywords. behavior change, ontology, participatory design, knowledge engineering, personalisation

1. Introduction

Theories of behavior change (e.g. [1-5]), and design models for behavior change systems, partly based on these theories [6-7], are increasingly applied in research on persuasive technology for behavior change [8]. When applying a participatory design process to develop such systems, a central assumption is that the participating stakeholders are experts in the domain in which the persuasive technology is aimed to be used. The design process aims to translate this knowledge into digital tools, which an individual can use for health coaching. Moreover, the knowledge needs to be translated into formal, computational models, which the system can use for tailoring support and automatically and intelligently adapt its behavior to the individual (e.g. [9-10]).

The purpose of this study is to develop and evaluate a conceptual model of behavior change progress, which can be used in a participatory design process of behavior change systems, and for automated analysis of progress over time as a complement to everyday assessments at runtime.

[1] Corresponding Author, Helena Lindgren, Department of Computing Science, Umeå University, 901 87 Umeå, Sweden; E-mail: helena.lindgren@umu.se.

2. Methods

Theories on behavior change explain the determinants of human behavior, motivations for behavior, and how to guide behavior change. The conceptual model is based on the Theory of Planned Behavior [1], the Self-Determination Theory [2], Prochaska's Transtheoretical Model of Change [3], the Goal Setting Theory [4], and the Fogg Behavior Model [6]. Concepts from the Health Belief Model [5] were also embedded.

The model was created in the following steps [11]. Relevant concepts were selected from behavior change theories, and defined. Next, based on the theoretical frameworks, the relationships between the concepts were defined and explained. In a third step, the initial model was validated by five experts through interviews and adjusted accordingly. Participating experts represented expertise in epidemiology, nutrition, sociology and nursing. The model was applied and further evaluated as part of two ongoing participatory design processes [10,12]. One aims at developing a digital coach application for preventing cardio-vascular diseases and promoting change of behavior to improve health [13], another to support stress rehabilitation. The preliminary information models of the two prototypes, which had emerged in focus group discussions and workshops, were evaluated by connecting the information to the conceptual model.

3. Results

The experts interviewees suggested changes of the conceptual model based on their knowledge of the concepts, theories and experience from clinical practice, gave elaborations on the concepts, and proposed additional concepts and relationships, partly based on the Health Belief Model (e.g., *Risk Perception, Previous Attempts*) [5,11].

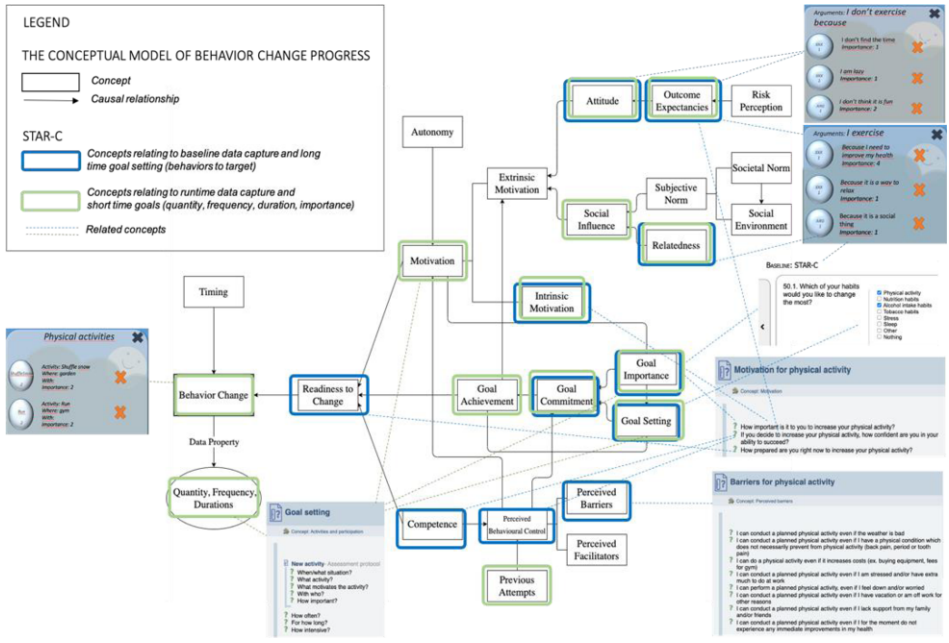


Figure 1. STAR-C information mapped to the conceptual model of behavior change progress.

The baseline questions for the cardio-vascular intervention cover most of the factors relating to a person's own attitude and intention to change, while less information is collected on how the social environment may influence the factors for behavior change. The runtime version contains mechanisms that may capture also conflicting motives for activities and more on perceived barriers, which are expected to illuminate reasons for non-adherence to goals specified, and be used for explaining and motivating changes in behaviors.

When mapping the contents of the baseline and ecological momentary assessment (EMA) questions for the coaching system for managing stress (Jonglera) to the conceptual model, it is clear that stress management requires a different approach to design than the previous example, in particular, since mobilizing the physical and mental effort required to changing behavior is part of the behavior change progress, and targeted as part of the digital intervention (Figure 2). In this case, the EMA questions are posed to assess daily engagement and experiences, forming a basis for the system to translate this into motivating trajectories of positive trends, which can be communicated to the person as an extrinsic motivator. As such, it will target also the factors in the behavior change progress model such as *outcome expectancies*, *attitudes*, *barriers* and *facilitators*.

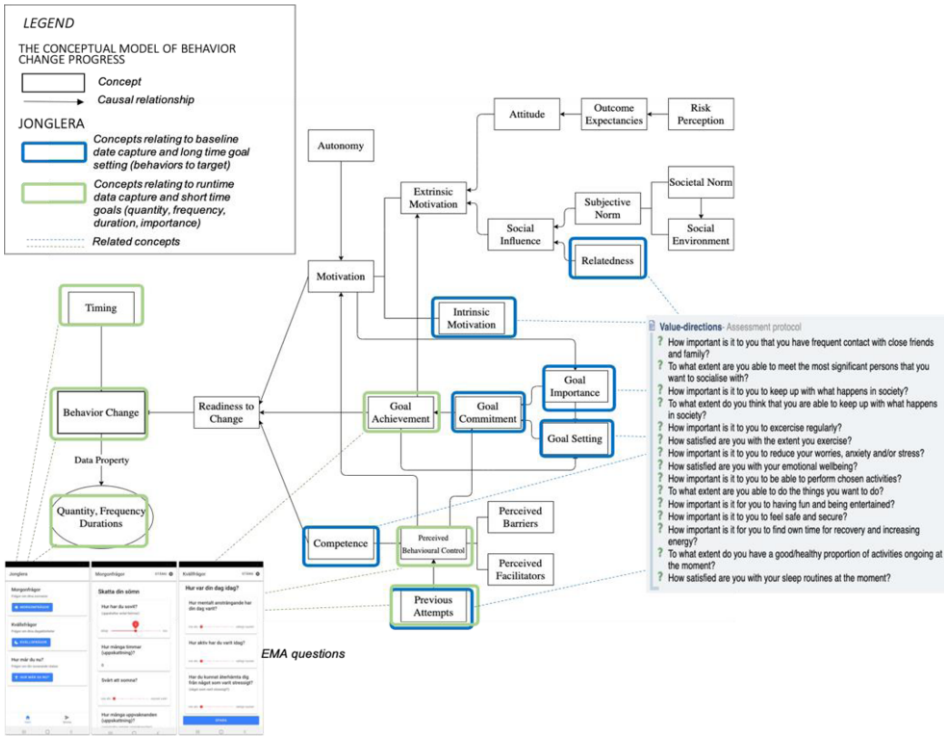


Figure 2. Jonglera information mapped to the conceptual model of behavior change progress.

To summarize, the conceptual model illuminates the complexity of behavior change for improving emotional wellbeing relating to stress and exhaustion syndrome, and the importance to design mechanisms, which can translate small efforts into positive outcome expectancies and attitudes, which can in turn, reinforce sense of competence and self-efficacy.

4. Discussion and Conclusions

A conceptual model of behavior change progress was developed based on theories of behavior change, evaluated by experts on behavior change and applied to evaluate the information models of two prototype systems for supporting behavior change. It was concluded that the model is useful for assessing to what extent an information model is capturing vital factors for behavior change, and how personalized intervention needs to be designed in order to promote behavior change.

Ontologies have been developed for different purposes, e.g., behavior change interventions, or identifying barriers for change [14-15]. The proposed conceptual model for behavior change progress aims to capture the vital concepts of the relevant theories, and their relationships in the development of persuasive systems. As such, the model complements existing models.

Further evaluation studies need to be done, in particular for exploring how the model can be used as instrument by the participating domain experts and stakeholders in a design process. Future work includes also the integration of the conceptual model to function as a meta-ontology in the system architectures of the two prototype systems, which could be used for capturing causal relationships between information collected by an application at baseline and at runtime.

References

[1] Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991;50:179-211.

[2] Ryan RM, Deci LM. Self-Determination Theory and the facilitation of intrinsic motivation, social development and well-being. American Psychologist 2000;55:68-78.

[3] Prochaska JO, Redding CA, Evers KE, The transtheoretical model and stages of change. In: Glanz A, Rimer BK, Viswanath K (Edrs.) Health behavior and health education: Theory, research, and practice. pp. 97–121. Jossey-Bass; 2008.

[4] Locke EA, G.P. Latham GP. Goal setting theory. In: Motivation: Theory and Research 2012:23-40.

[5] Janz NK, Becker MH. The Health Belief Model: A Decade Later. Health Education Quarterly 1984;11,1: 1-47.

[6] Fogg BJ. A behaviour model of persuasive design, In Proc. of the 4th International Conference on Persuasive Technology (Persuasive'09). Association for Computing Machinery, New York, NY, USA, 2009:1-7.

[7] Oinas-Kukkonen H, Harjumaa M. Persuasive Systems Design: Key Issues, Process Model, and System Features. Communnications for the Association for Information Systems 2009;24,28: 485-500.

[8] Orji R, Moffatt K. Persuasive technology for health and wellness: State-of-the-art and emerging trends. Health Informatics Journal March 2018:66-91.

[9] Paramythis A, Weibelzahl S, Masthoff J. Layered evaluation of interactive adaptive systems: framework and formative methods. User Modeling and User-Adapted Interaction 2010;20:383-453.

[10] Lindgren H, Kampik T, Guerrero E, Blusi M, Nieves JC. Argumentation-based Health Information Systems : A Design Methodology. IEEE Intelligent Systems 2021;36,2:72–80.

[11] Weck S. A Conceptual Model of Behavior Change Progress for the Application within Coaching Systems to Support Sustainable Lifestyle Changes. [Dissertation]. Umeå University; 2021.

[12] Janols R, Lindgren H. A method for co-designing theory-based behaviour change systems for health promotion. Studies in Health Technology and Informatics 2017;235:368–72.

[13] Lindgren H, Guerrero E, Jingar M, Lindvall K, Ng N, Richter Sundberg L, et al. The STAR-C Intelligent Coach : a Cross- Disciplinary Design Process of a Behaviour Change Intervention in Primary Care. Studies in Health Technology and Informatics 2020;273:203-208.

[14] Wright AJ, Norris E, Finnerty AN et al. Ontologies relevant to behaviour change interventions: a method for their development. Wellcome Open Res 2020;5:126

[15] Alfaifi Y. An Ontology to Support Behaviour Change Interventions: Barriers to Activity. Doctoral dissertation, University of Liverpool; 2020.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210615

Automatic Extraction and Decryption of Abbreviations from Domain-Specific Texts

Michil EGOROV [a,1] and Anastasia FUNKNER [a,2]
[a] *ITMO University, Saint Petersburg, Russia*

Abstract. This paper explores the problems of extraction and decryption of abbreviations from domain-specific texts in Russian. The main focus are unstructured electronic medical records which pose specific preprocessing problems. The major challenge is that there is no uniform way to write medical histories. The aim of the paper is to generalize the way of decrypting abbreviations from any variant of text. A dataset of nearly three million medical records was collected. A classifier model was trained in order to extract and decrypt abbreviations. After testing the proposed method with 224,307 records, the model showed an F1 score of 93.7% on a valid dataset.

Keywords. Clinical text, medical records, natural language processing, abbreviations

1. Introduction

Electronic health records (EHR) are widely used to build models for predicting the process of healthcare provision [1]. Such texts contain terms, specific abbreviations, and acronyms, whose decryption depends on the field of usage, particular medical institution, or even particular specialist [2]. These factors make research of the task of extraction complicated. This paper presents a method of automatic detection and decryption of abbreviations.

Recognition of well-established abbreviations and acronyms is usually carried out with the help of dictionaries [3] and marked data [4], and mainly addresses data in the English language. MeDAL [3] contains medical texts with abbreviations and their possible decryptions. Models pre-trained on this dataset improve their metrics by 0.2-2%.

For the Russian language, as one of the low-resource languages, this area is not well researched. The author of [5] considers the problem of extraction and decryption of abbreviations from the Corpus of Legislative Acts of the Russian Federation. The paper considers approaches to topic modeling of texts to identify words that are similar in terms of use and contexts. As the author highlighted in their work, automatic decryption of abbreviations using such approaches is not accurate enough.

[1] Corresponding author. e-mail: egorovmichil9@gmail.com.
[2] Corresponding author. e-mail: funkner.anastasia@gmail.com.

2. Methods

This section sequentially describes the steps taken to extract and decrypt abbreviations.

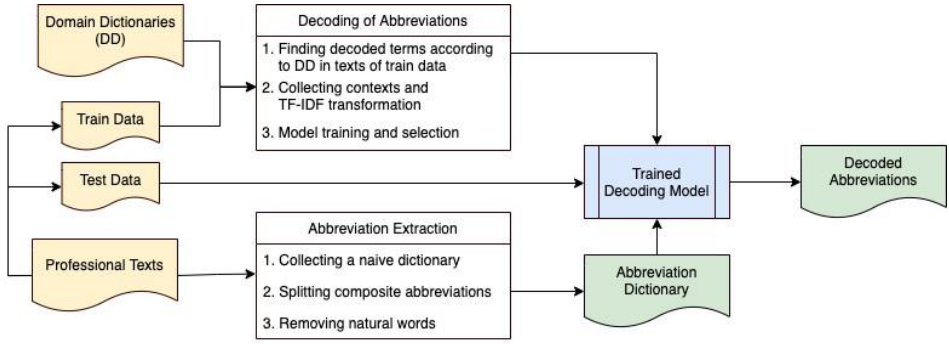


Figure 1. Diagram of methods for extraction and decryption of abbreviations.

Firstly, a dictionary of abbreviations needs to be collected (Fig. 1). Keep in mind the assumption that a word is an abbreviation, if at least one half of its letters are upper-case. This allows to compile a naive abbreviation dictionary. However, this dictionary can include composite abbreviations, for example "BP.HR". Such abbreviations may be distinguished by punctuation. Moreover, the naive dictionary also contains natural words that were written in upper case. As an abbreviation often contains unnatural character pairs, the average entropy of character pairs in a word allows to remove natural words from the dictionary using the Shannon entropy formula. This means the more entropy, the more natural the word is. It remains to choose the threshold value of entropy.

The algorithm has two stages: training and applying. For the training stage, one should make a corpus with encoded abbreviations. Words in the corpus need to be normalized and filtered using a list of stop words. In the stage of applying, the trained model is used with an abbreviation and its context as input and a decoded term as output.

In order to train the decryption model, the input text is divided into three parts: context before and after an abbreviation, and its characters. Contexts are collected with a window and transformed into two Term Frequency-Inverse Document Frequency (TF-IDF) vectors. The stack of bag-of-characters vectors and two transformed TF-IDF contexts become features. Decryption models can only be trained on two TF-IDF vectorizer, since using bag-of-characters vectors often leads to model over-fitting. Any multiclassifier model can be used as a predicting model.

3. Result

For this research, we used medical recommendations from EHRs provided by the Almazov National Medical Research Centre, Saint Petersburg, Russia. The total volume of the dataset is approximately 3 million unstructured records with 440,000 unique words.

After the naive extraction, 36,856 unique abbreviated words were obtained and reduced to 26,989 after splitting by punctuation marks. To collect the statistics for entropy, the Leo Tolstoy novel "War and Peace" with 54,294 unique words was

processed. A dataset of 48,530 natural words and 23,346 abbreviations was collected to choose the threshold. The natural word classifier accuracy was 89% with 0.2 threshold.

The resulting dictionary contains 4,658 unique abbreviations after discarding words whose entropy was above 0.2. Unfortunately, some abbreviations like "АД" (short for "blood pressure", but the word "ад" is translated as "hell") have very large entropy and cannot be divided from natural words, because it is already a separate word.

The recommendation records were normalized using pymorphy2 [6] and stop words were filtered using the NLTK Python module. Unencrypted medical terms were identified in the domain dictionary in the recommendation corpus. The result was 224,307 texts with 19,980 unique terms. The validation set contained 30% of data. The selected domain dictionary was the "Encyclopedic dictionary of medical terms". TF-IDF context vectors (context_window=5, max_features=200) were counted. The projections of the context feature vectors with t-distributed Stochastic Neighbor Embedding (t-SNE) [7] are shown in Fig. 2. As can be seen, TF-IDF vectorizers grouped by context together even if the abbreviation columns in the features were skipped.

After applying difference models to the validation dataset, the Support Vector Classifier emerges as the best model. Its accuracy is 94.5%, ROC-AUC score is 99.7%, and F1 score is 93.7% (Tab. 1).

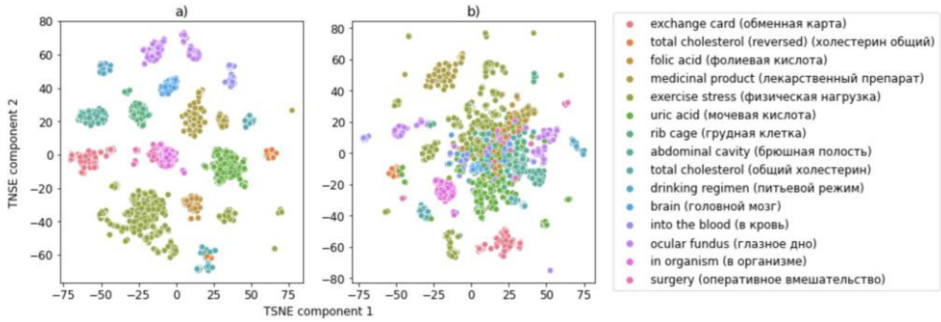


Figure 2. Context feature maps for top 15 terms a) with abbreviation columns, b) without abbreviation columns.

Table 1. Results of abbreviations decryption.

Model	Accuracy	ROC-AUC score	F1 score
RandomForestClassifier	0.938	0.968	0.933
LogisticRegression	0.939	0.985	0.930
XGBClassifier	0.939	0.973	0.935
SGDClassifier	0.940	0.983	0.930
CatBoostClassifier	0.940	0.986	0.932
SVC	0.945	0.997	0.937

Discussion & Conclusion

This work presents an algorithm for automatic extraction and decryption of abbreviations from medical records. To achieve this, we applied a naive hypothesis and then improved the algorithm via entropy of words. The abbreviation filtering algorithm is implemented. After testing the decryption method with randomly chosen records, the classifier shows high accuracy (94.5%).

However, this work does not consider the abbreviations that are written in lower case, for example "mmHg" (millimeter of mercury) and frequent abbreviations (they are virtually always written in their short form, so we were unable to find their context using a domain dictionary). We intend to develop more abbreviation rules and expand the naive dictionary.

In our future research, we intend to apply the algorithm to different corpora, especially where abbreviations are ambiguous. This module is aimed to help data scientists improve their models that use free-form records for predicting processes associated with healthcare.

Acknowledgments

This research is financially supported by The Russian Science Foundation, Agreement #19-11-00326

References

[1] Gamal A, Barakat S, Rezk A. Standardized electronic health record data modeling and persistence: A comparative review. Journal of Biomedical Informatics 2021; 114.

[2] Funkner AA, Egorov MP, Fokin SA et al. Citywide quality of health information system through text mining of electronic health records. Applied Network Science 2021; 6.

[3] Wen Z, Lu XH, Reddy S. MeDAL: Medical Abbreviation Disambiguation Dataset for Natural Language Understanding Pretraining. In Proceedings of the 3rd Clinical Natural Language Processing Workshop. Association for Computational Linguistics; 2020.

[4] Zweigenbaum P, Deleger L, Lavergne T, Neveol A, Bodnari A. A Supervised abbreviation resolution system for medical text. CEUR Workshop Proceedings 2013; 1179.

[5] Shilov IM. Automatic identification and decryption of abbreviations in the text [dissertation on the Internet]. Saint-Petersburg: Saint-Petersburg State University; 2016. Available from: https://nauchkor.ru/pubs/avtomaticheskoe-vyyavlenie-i-rasshifrovkaabbreviatur-i-sokrascheniy-v-tekste-587d36365f1be77c40d58984

[6] Korobov M. Morphological analyzer and generator for Russian and Ukrainian languages. In: Khachay MY, Konstantinova N, Panchenko A, Ignatov DI, Labunets VG (Edrs.) Analysis of images, social networks and texts. Berlin: Springer; 2015.

[7] Laurens M. Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine Learning Research 2008; 15:3221-3245.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210616

Mapping of OpenEHR Archetypes to FHIR Resources in Use Case Oncology

Abdul Mateen RAJPUT [a,1] and Ina BRAKOLLARI [a]
[a] *Universitätsklinikum Köln, Kerpener Str. 62, 50937 Köln, Germany*

Abstract. Unambiguous data exchange among healthcare systems is essential for error-free reporting and improved patient care. Mapping of different standards plays a crucial role in making different systems communicate with each other and have an efficient healthcare systems. This work focuses on exploring the possibilities of semantic interoperability between two widely used clinical modelling standards, OpenEHR and FHIR (Fast Healthcare Interoperability Resources). A manually curated map is being developed where the same semantically meaning OpenEHR Archetypes are mapped to the relevant FHIR Resources.

Keywords. OpenEHR, FHIR, OpenEHR Templates, Semantic Interoperability, HiGHmed

1. Introduction

OpenEHR is a technology for e-health, consisting of open specifications, clinical models, and software that can be used to create standards, and build information and interoperability solutions for healthcare [1]. It is an open-source healthcare information modeling standard that enables modelling of interoperable Electronic Health Records (EHRs). It started in response to the absence of an open-source platform for the exchange of clinical data. The standard is maintained by a community of healthcare professionals and software developers. Since its inception, it has been used widely across the globe [2–4].

On the other hand, FHIR (Fast Healthcare Interoperability Resources) is also an emerging open standard in healthcare. It is considered as "HTML" of healthcare [5], it facilitates extensive data modeling and data exchange, irrespective of the necessity of a common Electronic Health Record (EHR) system. Transitioning the patient data to FHIR would make the data able to connect various applications and increases interoperability significantly.

HiGHmed Consortium [6] uses OpenEHR as a standard for clinical data exchange. However, there are some use cases where usage of FHIR is foreseen and required. For that purpose mapping between both those emerging standards was needed.

[1] Abdul-Mateen RAJPUT, Universitätsklinikum Köln, Kerpener Str. 62, 50937 Köln, Germany ; E-mail: Abdul.mateen@uni-koeln.de.

2. Methods

The archetypes associated with the use case Oncology in OpenEHR Clinical Knowledge Manager (CKM) [7] have been identified and downloaded. The components of the archetypes were analyzed and extracted into an excel file. Each data element of OpenEHR archetypes was searched on the HL7 FHIR website [8] manually for the corresponding resource term. The FHIR resource elements, which had the same semantic meaning as OpenEHR archetypes elements, were mapped into the excel file accordingly.

3. Results

A mapping table was created that contains the archetype elements mapped to resource elements with the web address but also the archetypes elements that can not be mapped when the subsequent resource elements were not found. Figure 1 shows the table of mapped terms.

Archetype	Archetype Element	FHIR Resource Element	Webpage
openEHR-EHR-EVALUATION.problem_diagnosis.v0	Date/time of onset	Condition.recordedDate	https://www.hl7.org/fhir...Condition.code
openEHR-EHR-EVALUATION.problem_diagnosis.v0	Severity	Condition.severity	https://www.hl7.org/fhir...Condition.severity
openEHR-EHR-EVALUATION.problem_diagnosis.v0	Status	Condition.clinicalStatus	https://www.hl7.org/fhir...Condition.clinicalStatus
openEHR-EHR-EVALUATION.problem_diagnosis.v0	Comment	Condition.note	https://www.hl7.org/fhir...Condition.note

Figure 1. The table shows the OpenEHR archetype in first column, its captured elements in second column, corresponding FHIR resource elements and web address in third and fourth columns respectively.

Our findings show that archetypes were more specific as they were designed by clinical domain experts and they are consist of a formal model. Their elements could only be mapped to only one FHIR resource element. On the contrary, FHIR resource elements could be mapped to more than one archetype element.

In addition, different elements of a single archetype can be mapped to multiple resources because a single FHIR resource does not cover all the information that a single archetype contains. It was also noticed that there are FHIR resources that overlap with each other and the boundaries are not strict in FHIR.

4. Discussion and Conclusion

A mapping between different standards is crucial for the efficient data exchange among different standards. It is a complex and time-consuming process. The emerging new health data standards, FHIR and OpenEHR, require to communicate as each of them has limitations. The manual mapping between both standards not only make it easier to understand the data representation in a different context but also streamline the data exchange. The mutual understanding of the limitations of each standard and the development of tools designed to facilitate communication among different standards is highly desirable.

Acknowledgment

The project is funded by the German Federal Ministry of Education and Research (BMBF, grant id: 01ZZ1802U).

References

[1] https://www.openehr.org/about/what_is_openehr (Accessed 15 January 2021).
[2] Maia TA, De Muylder CF, Reis ZSN. Archetype Development Process: A Case Study of Support Interoperability among Electronic Health Record in the State of Minas Gerais, Brazil. J Med Syst 2019; 43: 57
[3] Min L, Tian Q, Lu X, et al. Modeling EHR with the openEHR approach: an exploratory study in China. BMC Medical Informatics and Decision Making 2018; 18: 75.
[4] Mascia C, Frexia F, Uva P, et al. The openEHR Genomics Project. Stud Health Technol Inform 2020; 270: 443–447.
[5] Munro D. Setting Healthcare Interop On Fire. Forbes, https://www.forbes.com/sites/danmunro/2014/03/30/setting-healthcare-interop-on-fire/ (accessed 4 March 2020).
[6] Home | HiGHmed, https://www.highmed.org/ (accessed 14 January 2021).
[7] Clinical Knowledge Manager, https://ckm.openehr.org/ckm/ (accessed 29 July 2020).
[8] Index - FHIR v4.0.1, http://hl7.org/fhir/ (accessed 28 January 2021).

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210617

Detection of COVID-19 from Chest CT Images Using CNN with MLP Hybrid Model

Sakthi Jaya Sundar RAJASEKAR[a], Vasumathi NARAYANAN[b], Varalakshmi PERUMAL[b,1]
[a] *Melmaruvathur Adhiparasakthi Institute of Medical Sciences and Research, Melmaruvathur, Tamil Nadu, India*
[b] *Department of Computer Technology, MIT Campus, Anna University, Chennai, India*

Abstract. COVID-19 when left undetected can lead to a hazardous infection spread, leading to an unfortunate loss of life. It's of utmost importance to diagnose COVID-19 in Infected patients at the earliest, to avoid further complications. RT-PCR, the gold standard method is routinely used for the diagnosis of COVID-19 infection. Yet, this method comes along with few limitations such as its time-consuming nature, a scarcity of trained manpower, sophisticated laboratory equipment and the possibility of false positive and negative results. Physicians and global health care centers use CT scan as an alternate for the diagnosis of COVID-19. But this process of detection too, might demand more manual work, effort and time. Thus, automating the detection of COVID-19 using an intelligent system has been a recent research topic, in the view of pandemic. This will also help in saving the physician's time for carrying out further treatment. In this paper, a hybrid learning model has been proposed to identify the COVID-19 infection using CT scan images. The Convolutional Neural Network (CNN) was used for feature extraction and Multilayer Perceptron was used for classification. This hybrid learning model's results were also compared with traditional CNN and MLP models in terms of Accuracy, F1-Score, Precision and Recall. This Hybrid CNN-MLP model showed an Accuracy of 94.89% when compared with CNN and MLP giving 86.95% and 80.77% respectively.

Keywords. COVID-19, CNN, Classification, Deep Learning, Multilayer Perceptron

1. Introduction

Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV-2) which causes the Corona Virus Disease 2019 (COVID-19) was identified in China in late 2019. In March 2020, the World Health Organization (WHO) labelled the disease a pandemic. As of 22nd July 2021, around 191 million COVID-19 cases and 4 million deaths due to COVID-19 have been reported [1]. The most predominant manifestation of COVID-19 is pneumonia, which affects the lungs, and presents with symptoms like high temperature, cough and dyspnea [2]. CT scan examination for abnormalities is by far the most rapid way to evaluate the patients.

[1] Corresponding Author: Varalakshmi Perumal, PhD, Associate Professor; Dept of Computer Technology, Anna University - MIT Campus; Email: varanip@gmail.com

Over the last decade, deep learning and Artificial Neural Networks (ANNs) have had an accelerating study focus. On a number of important benchmarks, deep ANNs have surpassed other conventional models. As a result, ANNs have shown to be the state-of-the-art technology in a variety of application domains, including speech recognition, image processing, biological sciences, and Natural Language Processing (NLP), as well as academic fields. The development of ANNs for smart healthcare advancement, particularly in medical predictive analytics, diagnosis through medical image processing and analysis, has enormous promise [3,4,5]. As has been witnessed recently, several sections of the world are experiencing a healthcare shortage, both in relation to the required number of healthcare personnel and electronic components. In this study, CNN was used to create an automatic diagnosis method that uses CT scan image results to determine if a person has acquired the COVID-19 infection or not. This study's initial findings in terms of reliability and other quality metrics to evaluate the condition in a cost-effective and time-efficient manner have showed encouraging outcomes. The accuracy of COVID-19 CT scan image categorization was improved in this work by combining CNN and MLP. The CNN architecture in neural networks is specifically designed to deliver two-dimensional image operations.

2. Proposed Work

In this section, the proposed model has been discussed elaborately. The architecture of the proposed hybrid model is shown in Figure. 1. Here, CNN is used for feature extraction by the removal of the fully connected layers from the CNN model and MLP is used for classification.

2.1. Convolutional Neural Network (CNN)

Convolutional Neural Network is comprised of an input layer, hidden layers, and an output layer. Input CT scan images are fed as input to the CNN model through the input layer. The hidden layers of a CNN include convolutional layers and pooling layers. The ReLU activation function is used in this model, and it is typically the Frobenius inner product. The convolution operation is performed by sliding the convolutional kernel over the input matrix of the CT scan images, thereby the feature map is generated, which then imparts to the input of the succeeding layers. Other subsequent layers such as pooling layers performs the dimensionality reduction, fully connected layers perform the classification task, and the normalization components are also stacked with this model. The convolutional filter of size 3*3*4 is being used for first convolutional layer followed by convolutional operation with size 2*2*8. The max pooling layer of size 2*2 is being used to reduce the dimensions. 8 neurons are being used as dense layer followed by dropout of 0.25 is being constructed. ReLu is used as activation function for the convolutional and max pooling layers. SoftMax is used as an activation function for dense layers.

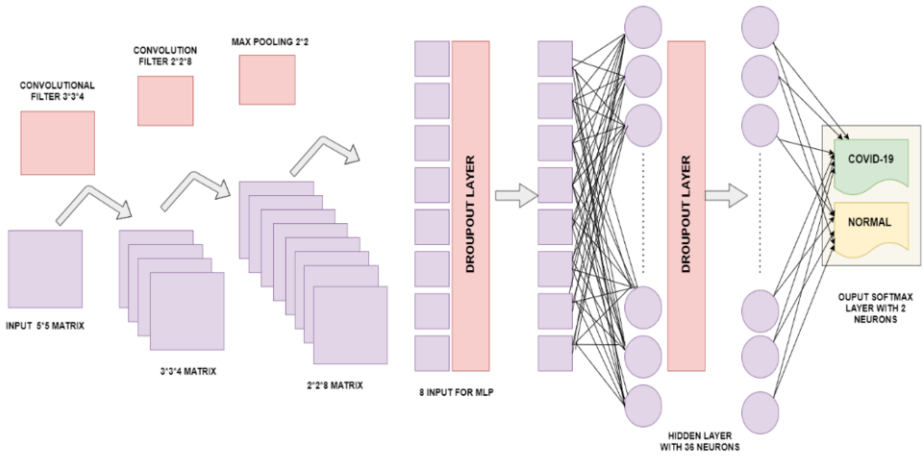


Figure 1. Architecture of the proposed CNN-MLP model

2.2. Multilayer Perceptron (MLP)

Multilayer Perceptron model is one of the widely used supervised learning model in real time applications. It is comprised of input layer, followed by multiple hidden layers with an output layer. It has two phases; one is forward phase and other is the backward phase. Forward phase is used for classification and the backward phase is used for error propagation in the backward direction for the learning process. Any nonlinear activation function will be used in the hidden and the output layers to solve any real time nonlinear problems.

3. Experimental Results

The dataset required for carrying out the experiment has been gathered from GitHub [6] and Kaggle's TPU machine was utilized to run the models. The models were constructed using TensorFlow and Keras provided by Python programming language. The training images and testing images for both classes (COVID-19 and normal CT scan images) were separated in 80:20 ratio with number of images in each class being 1200 (960 for training and 240 for testing). The metrics used for comparing the performance includes accuracy, precision, recall, F1-Score and specificity. The confusion matrix and the performance measures of the CNN, MLP and the proposed Hybrid Model is given in Table 1. From the results, the proposed Hybrid model showcases a better performance compared to the CNN and the MLP models.

Table 1. Confusion Matrix and Classification Report for the models

Models	Category	COVID-19	Normal	Total	Category	Precision	Recall	F1 Score	Specificity
CNN-MLP	COVID-19	114	6	120	COVID-19	0.95	0.93	0.94	0.95
	Normal	9	111	120	Normal	0.92	0.95	0.94	0.95
	Total	123	117	240	Average	0.94	0.94	0.94	0.95
CNN	COVID-19	102	18	120	COVID-19	0.85	0.86	0.86	0.86
	Normal	16	104	120	Normal	0.86	0.85	0.86	0.86
	Total	118	122	240	Average	0.86	0.86	0.86	0.86
MLP	COVID-19	96	24	120	COVID-19	0.80	0.82	0.81	0.80
	Normal	21	99	120	Normal	0.80	0.80	0.80	0.80
	Total	117	123	240	Average	0.80	0.81	0.81	0.80

Conclusion

In conclusion, our research revealed the viability of using trained hybrid model to help clinicians determine whether or not a patient is infected with COVID-19. The proposed hybrid model could enable rapid identification. With the deployment of the proposed model in patient care, clinicians could make more trustworthy decisions by attaining good performance on diagnoses.

Acknowledgement

The authors would like to acknowledge Tamil Nadu State Council for Science and Technology for funding this work under the project TNSCST/STP/Covid-19/2020-2021/3696.

References

[1] WHO, Coronavirus Disease 2019 (COVID-19) https://covid19.who.int/ , World Health Organization, Accessed on 22nd July 2021.

[2] Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, Fan Y, Zheng C. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. The Lancet infectious diseases. 2020 Apr 1;20(4):425-34.

[3] Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J. Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: relationship to negative RT-PCR testing. Radiology. 2020 Aug;296(2):E41-5.

[4] Zhu Z, Xingming Z, Tao G, Dan T, Li J, Chen X, Li Y, Zhou Z, Zhang X, Zhou J, Chen D. Classification of COVID-19 by compressed chest CT image through deep learning on a large patients cohort. Interdisciplinary Sciences: Computational Life Sciences. 2021 Mar;13(1):73-82.

[5] Shah V, Keniya R, Shridharani A, Punjabi M, Shah J, Mehendale N. Diagnosis of COVID-19 using CT scan images and deep learning techniques. Emergency radiology. 2021 Jun;28(3):497-505.

[6] Mohammad Rahimzadeh, Abolfazl Attar, Seyed Mohammad Sakhaei, A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset, Biomedical Signal Processing and Control, Volume 68, 2021, 102588, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2021.102588 .

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210618

An openEHR Virtual Patient Template for Pancreatic Cancer

Samer ALKARKOUKLY[a, 1] and Abdul-Mateen RAJPUT[a]
[a] *CECAD, Faculty of Medicine and University Hospital Cologne, University of Cologne, Joseph-Stelzmann-Strasse 26, 50931 Cologne, Germany*

Abstract. openEHR is an open-source technology for e-health, aims to build data models for interoperable Electronic Health Records (EHRs) and to enhance semantic interoperability. openEHR architecture consists of different building blocks, among them is the "template" which consists of different archetypes and aims to collect the data for a specific use-case. In this paper, we created a generic data model for a virtual pancreatic cancer patient, using the openEHR approach and tools, to be used for testing and virtual environments. The data elements for this template were derived from the "Oncology minimal data set" of HiGHmed project. In addition, we generated virtual data profiles for 10 patients using the template. The objective of this exercise is to provide a data model and virtual data profiles for testing and experimenting scenarios within the openEHR environment. Both of the template and the 10 virtual patient profiles are available publicly.

Keywords. openEHR, data model, data architecture, openEHR Templates, semantic interoperability, HiGHmed, virtual patient, test data, oncology.

1. Introduction

openEHR is a technology for e-health, consisting of open specifications, clinical models, and software that can be used to create standards, and build information and interoperability solutions for healthcare (1) . It provides an open-source healthcare information modelling environment that enables modelling of interoperable Electronic Health Records (EHRs). The standard is maintained by a community of healthcare professionals and software developers. Since its inception, it has been used widely across the globe (2–4).

openEHR maintains a vendor-independent standard based on two-level modelling: Reference Model and Archetype Model. In addition, Terminology binding is considered as a third major model of the openEHR (3). Archetypes are combined and nested in templates to represent specific use cases. Typically, templates express entire clinical documents containing different information modelled as several archetypes such as discharge letters, result reports, or medical history (5).

HiGHmed (6) is one of the medical informatics projects funded by the German Federal Ministry of Education and Research (BMBF) and it perceives semantic interoperability as a prerequisite for enabling a meaningful exchange of the data in

[1] Samer Alkarkoukly, CECAD, Faculty of Medicine and University Hospital Cologne, University of Cologne, Joseph-Stelzmann-Strasse 26, 50931 Cologne, Germany; E-mail: mohammad.abboud-alkarkoukly@uk-koeln.de.

federated cross-institutional settings by adopting the openEHR approach for semantic modelling (7).

To test the capabilities of an openEHR system, a complete profile of an individual patient was required. The available templates on Clinical Knowledge Manager (CKM) (8) can be used to generate partial clinical data and it has been shown that large datasets can be generated by following the semi-automated approach (9) but a complete representation of a virtual patient in a particular disease is still missing. The complete profile would provide the insights of a clinical course of the disease and all the available data elements associated with a particular disease in an openEHR format. The modelling of the virtual representation in biomedical sciences is not new. There have been many initiatives e.g., Virtual Brain (10), Virtual Cell (11), etc. to either simulate the patient phenotypes or generate data for further processing.

The objective of our current work in this paper is to build a virtual patient data model in openEHR format, that could perform as a generic data architecture for testing purposes and also for virtual environments of electronic medical records. To the author's knowledge, the approach described here is new to this field.

2. Methods

An openEHR template for a virtual patient of pancreatic cancer is developed by following the guidelines and using the available archetypes associated with use-case oncology in HiGHmed CKM (8).

The template was created to represent the minimal dataset of oncology use-case within HiGHmed (8).

The online platform "openEHR Archetype Designer" (12) was used to design the template. All the used archetypes, clusters, sections are available publicly and mentioned later in this paper. Figure 1 illustrates the architecture of the template in Archetype Designer.

The template was used as a base to generate a full set of test data for 10 virtual patients by automating the REST API calls using locally installed Better platform Think!EHR (13) endpoint and following the same methodology as described in our previous paper (9). The patient's data were generated automatically and were not exported from any medical information system.

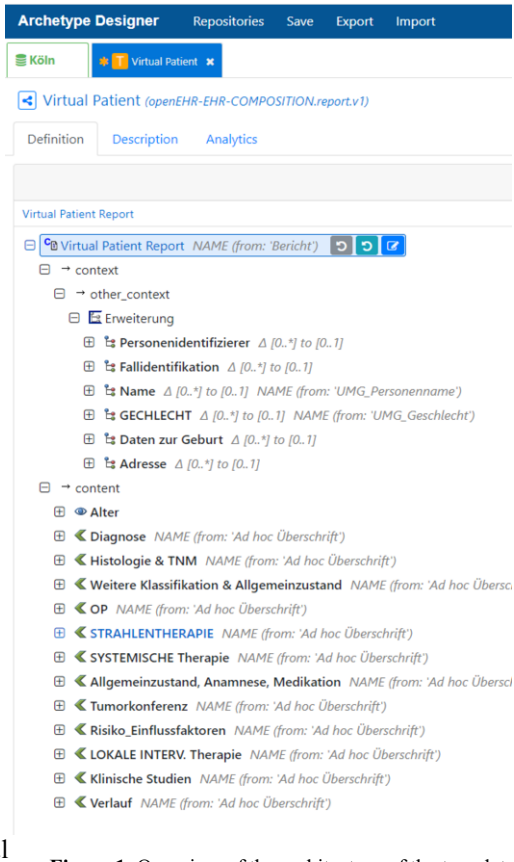


Figure 1. Overview of the architecture of the template

3. Results

The template was created using the German language, basically to follow the language of the source dataset. It covers the following data modules: patient personal information, diagnosis (ICD, ICD-O, TNM), histology grading and analysis, clinical history, surgical operations, radiotherapy, systemic therapy, local interventional therapy, medications, risk factors, disease progress, tumour board conferences, and clinical studies.

The virtual patient template, and also the used archetypes, are now available publicly on GitHub: "https://github.com/abdulmateenraj/ VirtualPatient_openEHR" and could be used as a testing structure in any openEHR system.

A test data set for 10 virtual pancreatic cancer patients was created based on the template, and also available on the same URL mentioned above in .json format. Those files could be used also for testing purposes not only in openEHR systems but also in any other electronic health record environment.


```
{
  "ctx/language" : "de",
  "ctx/territory" : "US",
  "ctx/composer_name" : "Silvia Blake",
  "ctx/id_namespace" : "HOSPITAL-NS",
  "ctx/id_scheme" : "HOSPITAL-NS",
  "ctx/participation_name" : "Dr. Marcus Johnson",
  "ctx/participation_function" : "requester",
  "ctx/participation_mode" : "face-to-face communication",
  "ctx/participation_id" : "199",
  "ctx/participation_name:1" : "Lara Markham",
  "ctx/participation_function:1" : "performer",
  "ctx/participation_id:1" : "198",
  "ctx/health_care_facility|name" : "Hospital",
  "ctx/health_care_facility|id" : "9091",
  "csv_export_test/context/personenidentifizierer/pid:0" : "bd3b63fa-d598-4c41-a2b7-a0f0e6082544",
  "csv_export_test/context/personenidentifizierer/pid:0|issuer" : "Issuer",
  "csv_export_test/context/personenidentifizierer/pid:0|assigner" : "Assigner",
  "csv_export_test/context/personenidentifizierer/pid:0|type" : "Prescription",
  "csv_export_test/context/personenidentifizierer/geografischer_geltungsbereich:0|code" : "at0011",
  "csv_export_test/context/fallidentifikation/fallnr" : "FALLNR 31",
  "csv_export_test/context/name/name_strukturiert/patienten_vornamen" : "Patienten_Vornamen 7",
  "csv_export_test/context/name/name_strukturiert/patienten_nachname" : "Patienten_Nachname 92",
  "csv_export_test/context/gechlecht/patienten_geschlecht" : "Sonstiges/Intersexuell",
  "csv_export_test/context/daten_zur_geburt/patienten_geburtsdatum" : "2020-07-16",
  "csv_export_test/context/adresse/adresse:0/art|code" : "at0011",
  "csv_export_test/context/adresse/adresse:0/strukturierte_adresse/ort:0" : "ORT 7",
  "csv_export_test/context/adresse/adresse:0/strukturierte_adresse/land:0" : "Land 96",
  "csv_export_test/context/adresse/adresse:0/patienten_plz" : "Patienten_PLZ 53",
  "csv_export_test/alter/patientenalter" : "P2DT12H43M",
  "csv_export_test/diagnose/diagnose/primaertumor_icd_code" : "Primaertumor_ICD_Code 97",
  "csv_export_test/diagnose/diagnose/primaertumor_diagnosetext" : "Primaertumor_Diagnosetext 60",
  "csv_export_test/diagnose/diagnose/diagnosedatum" : "2020-07-16T12:52:17.869827+02:00",
  "csv_export_test/diagnose/diagnose/codesystem_icd/primaertumor_icd_version" : "Primaertumor_ICD_Version 75",
  "csv_export_test/diagnose/diagnose/tumorklassifikation_icd-o/primaertumor_topographie_icd_o" : "Primaertumor_To
  "csv_export_test/diagnose/diagnose/tumorklassifikation_icd-o/primaertumor_topographie_icd_o_freitext" :
  "Primaertumor_Topographie_ICD_O_Freitext 1",
```


Figure 2. Example of a json representation of one test patient

4. Discussion and Conclusion

In this paper, we address a challenge that EHR systems face, especially during the establishment phase, which is the need to test the infrastructure with the patient data. The dataset of 10 virtual patients covers the various aspects of the healthcare system (starting from patient demographics and clinical history, through diagnostics and diagnosis, ending in treatment and clinical studies). It includes data coming from different clinical systems, and on the other hand covers the different data types (Boolean, integer, string, date, datetime) for a pancreatic cancer use-case.

Additionally, the virtual patient template represents a full set of data parameters and can perform as a generic data architecture for testing and virtual environments.

Through our work, we made sure to maintain a high level of interoperability with the other EMR systems at both semantic (by providing a nationally used dataset for oncology) and syntactic levels (by following the international architecture and rules of openEHR).

The limitation of this virtual patient dataset is that it is not a real clinical data thus it is not suitable for some clinically oriented scenarios such as Natural Language Processing, however, it has the similar characteristics of real patient data and can be used in the other scenarios. Although the template is based on a pancreatic cancer profile, it could be used for similar cancers. Additionally, it represents the first step for future work on patient profiles from the other diseases by following the same methodology and implementing modifications to the dataset elements.

Acknowledgement

This work was partially supported by the HiGHmed Consortium funded by the German Federal Ministry of Education and Research (BMBF, funding code [01ZZ1802U]).

References

[1] openEHR International. What is openEHR?; 2021 [cited 2021 Jan 29]. Available from: URL: https://www.openehr.org/about/what_is_openehr.

[2] Maia TA, Muylder CF de, Reis ZSN. Archetype Development Process: A Case Study of Support Interoperability among Electronic Health Record in the State of Minas Gerais, Brazil. J Med Syst 2019; 43(3):57.

[3] Min L, Tian Q, Lu X, Duan H. Modeling EHR with the openEHR approach: an exploratory study in China. BMC Med Inform Decis Mak 2018; 18(1):75.

[4] Mascia C, Frexia F, Uva P, Zanetti G, Pireddu L, Giacomelli G et al. The openEHR Genomics Project. Stud Health Technol Inform 2020; 270:443–7.

[5] Wulff A, Mast M, Hassler M, Montag S, Marschollek M, Jack T. Designing an openEHR-Based Pipeline for Extracting and Standardizing Unstructured Clinical Data Using Natural Language Processing. Methods Inf Med 2020; 59(S 02):e64-e78.

[6] Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O et al. HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. Methods Inf Med 2018; 57(S 01):e66-e81.

[7] Wulff A, Haarbrandt B, Marschollek M. Clinical Knowledge Governance Framework for Nationwide Data Infrastructure Projects. Stud Health Technol Inform 2018; 248:196–203.

[8] HiGHmed Modelling Group. Project: Use Case Onkologie: HiGHmed; 2020 [cited 2021 Jan 29]. Available from: URL: https://ckm.highmed.org/ckm/projects/867.116.5.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210619

Psychological Stress Is a Risk Factor for Type 2 Diabetes Mellitus in College Students

Kukiat TUDPOR[1], Chatcharida SAISAWAI, Sasina KAEWPHO and Amaraporn JITSEEDA
Faculty of Public Health, Mahasarakham University, Maha Sarakham, Thailand

Abstract Type 2 diabetes mellitus (T2DM) is multifactorial disease. This cross-sectional study was aimed to investigate relationship between stress and risk for T2DM in college students. Seven-hundred participants (350 T2DM risk and 350 non-T2DM risk groups). Stress index levels and heart rate variability (HRV) were respectively measured as primary and secondary outcomes. Results showed that both T2DM-risk and non-T2DM-risk groups had temporary stress, but the T2DM-risk group had significantly higher level of psychological stress ($P < .001$). For the HRV, the T2DM-risk group had significantly lower levels of parasympathetic proxies (lnHF, SDNN, and RMSSD) ($P < .001$). Chi-square (χ^2) test showed significant correlation of the stressful state with T2DM risk ($\chi^2 = 159.372$, $P < .001$, odds ratio (OR) = 9.326). In conclusion, psychological stress is a risk factor for T2DM in college students. Early detection, monitoring, and treatments of psychological stress should be implemented in this group of population.

Keywords Psychological stress, Type 2 diabetes mellitus, Young adulthood

1. Introduction

Type 2 diabetes mellitus (T2DM) is major non-communicable disease worldwide [2]. According to Clinical Practice Guideline (CPG) for Diabetes 2017 of the Diabetes Association of Thailand, risk score checklist for T2DM in Thai people has been developed based on advancing age, masculinity, family history, body mass index (BMI), central obesity, and hypertension [1]. Young adults (18-26 years old) are in a critical period in life as they have to think about their future career paths and economic security [12]. Hence, they are prone to depression, and psychological stress [8]. However, their needs during this transitional period were not concerned. Psychological stress might cause T2DM via stress hormone-induced insulin insensitivity [7]. This study investigated relationship between psychological stress and T2DM risk in young adult individuals by using a novel non-invasive technology device for stress index measurement on a fingertip.

[1]Corresponding Author: Kukiat Tudpor, Faculty of Public Health, Mahasarakham University, Kham Rieng, Kantarawichai, Maha Sarakham, Thailand 44150; E-mail: kukiat.t@msu.ac.th

2. Methods

2.1. Study Design and Participants

Seven hundred college students from Mahasarakham University were recruited into this cross-sectional study. Inclusion criteria were being bachelor students and signed consent form. All procedures were approved by 2 reviewers of the Ethical Review Committee for Human Research, Mahasarakham University, and endorsed by the chairperson Dr.Ratree Sawangjit (No.276/2563). Exclusion criteria were psychiatric disease, metabolic disease, and recent major trauma. T2DM risk score was obtained from a checklist from the CPG [4]. The risk score was ranged from 0 - 17. The participants with score < 6 and ≥ 6 were divided into non-T2DM-risk and T2DM-risk groups (Table 1).

Table 1 Risk factors for T2DM and risk score

Risk factors for T2DM		Risk score
Age (years old)	34 – 39	0
	40 – 44	0
	45 – 49	1
	≥ 50	2
Sex	Female	0
	Male	2
Body mass index	< 23 kg/m^2	0
	23-27.4 kg/m^2	3
	≥ 27.5 kg/m^2	5
Waist circumference	Male < 90 cm, Female < 80 cm	0
	Male ≥ 90 cm, Female ≥ 80 cm	2
Hypertension	No	0
	Yes	2
Type 2 diabetes mellitus in first-degree relative	No	0
	Yes	4

2.2. Outcome Measurements

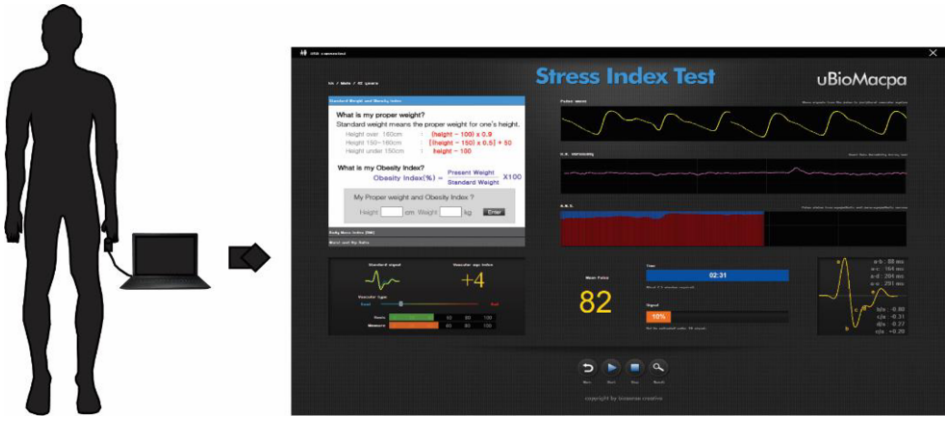


Figure 1 Non-invasive technology device for stress index measurement on a fingertip

Primary outcome (stress index) and secondary outcomes (HRV) were measured by photo-plethysmography (PPG) on the participant's left index finger tips connected to the uBioMacpa software as previously detailed [11] (Figure 1). The stress index scores were classified as 5 groups: no stress (< 25), temporary stress (25-34), primary stress (35-44),

accumulative stress (45-59), and chronic stress (≥ 60). The HRV parameters include sympathetic proxies [low frequency (lnLF) and low/high frequency ratio (LF/HF ratio)] and parasympathetic proxies [high frequency (lnHF), standard deviation of all normal R-R intervals (SDNN), and square root of the mean of the squared successive differences in R-R intervals (RMSSD)]. Statistical analyses were independent t-test and χ^2 test.

3. Results

3.1. T2DM-risk group had higher level of psychological stress

Both T2DM-risk and non-T2DM-risk groups had temporary stress, but T2DM-risk group had significantly higher level of psychological stress (P < .001). When further subdivided into the stressful subgroups (stress index ≥ 25), the T2DM-risk group had primary stress whilst the non-T2DM-risk group had temporary stress. Moreover, the T2DM-risk group also had statistically faster pulse rate. For the HRV, the T2DM-risk group had significantly lower lnHF, SDNN, and RMSSD (P < .001) (Table 2).

Table 2 Stress index and heart rate variability parameters in non-risk-T2DM and risk-T2DM

Parameter	Non-T2DM risk (n = 350)	T2DM risk (n = 350)	P-value
Stress index	25.20 ± 5.06	33.96 ± 8.06	<.001
Non-stressful	21.93 ± 2.37	22.37 ± 2.41	.25
Stressful	30.00 ± 4.03	35.79 ± 7.03	<.001
Pulse rate (beat/min)	76.59 ± 7.47	80.43 ± 10.99	<.001
lnLF (ms^2)	7.97 ± 0.61	8.57 ± 7.88	.16
lnHF (ms^2)	7.20 ± 0.48	6.93 ± 0.64	<.001
LF/HF ratio	1.11 ± 0.08	1.12 ± 0.09	.17
SDNN (ms)	65.74 ± 17.15	54.60 ± 23.34	<.001
RMSSD (ms)	51.25 ± 16.42	42.52 ± 23.09	<.001

3.2. Stressful State was Correlated T2DM Risk

Further analysis revealed that χ^2 test revealed that a stressful state was significantly correlated T2DM risk (χ^2 = 159.37, P < .001, OR = 9.32) (Table 3).

Table 3 Association between stress index and risk for T2DM

Stress index	T2DM risk (%)	Non-T2DM risk (%)	Total	χ^2	P-value	OR (95% CI)
Non-stressful	141 (40.29)	302 (86.29)	443	159.37	<0.001	9.32 (6.42-13.52)
Stressful	209 (59.71)	48 (13.71)	257			
Total	350	350	700			

4. Discussion and Conclusion

Risk for Type 2 Diabetes Mellitus in Transitional Age Youth (TAY): T2DM is caused by both genetic and non-genetic factors. The non-genetic factors consist of modifiable (physical activity, nutrition, smoking, alcohol drinking etc.) and non-modifiable factors (race, age, and sex etc.) [3]. Psychological stress is a modifiable factor causing T2DM in adults and elderly persons [5]. This present study was first to show a high psychological stress level in college students with high-risk score of T2DM. Age of the participants in this study (19-22 years old) is considered as TAY. An MRI study shows that prefrontal cortex (part of the brain for decision making) in TAY is still developing, implying stressful challenges of making decision in college students [6].

Possible Mechanisms of Psychological Stress-Induced Insulin Resistance in TAY: Psychological stress can be measured by psychological marker (stress questionnaire), biochemical marker (plasma cortisol), or physiological marker (HRV) [9]. Here we found that parasympathetic proxies of the HRV in the T2DM-risk group were lower than the non-T2DM-risk group. An increase in stress index in the T2DM-risk group might be due to reduced parasympathetic tone and increased sympathetic tone as shown by rapid pulse rate. Lastly, the participants with high stress index had 9 times greater likelihood of T2DM risk. Mechanisms of T2DM risk might be explained by psychological stress-induced insulin resistance via phosphorylated protein kinase B downregulation [10].

Conclusion: To conclude, psychological stress is a risk factor for T2DM in college students. Therefore, early detection, monitoring, and treatments of psychological stress should be implemented in this group of population.

References

[1] Aekplakorn W, Bunnag P, Woodward M, Sritara P, Cheepudomwit S, etal. A risk score for predicting incident diabetes in the Thai population. Diabetes Care 2006; 29: 1872-1877.
[2] Apidechkul T. Prevalence and factors associated with type 2 diabetes mellitus and hypertension among the hill tribe elderly populations in northern Thailand. BMC Public Health 2018; 18: 694.
[3] Deshpande AD, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related complications. Phys Ther 2008; 88: 1254-1264.
[4] Diabetes Association of Thailand. Clinical Practice Guidelines for Diabetes Diabetes Association of Thailand, 2017.
[5] Diez Roux AV, Jacobs DR, Kiefe CI, Coronary S. Neighborhood characteristics and components of the insulin resistance syndrome in young adults - The coronary artery risk development in young adults (CARDIA) study. Diabetes Care 2002; 25: 1976-1982.
[6] Giedd JN. The teen brain: insights from neuroimaging. J Adolesc Health 2008; 42 (4): 335-343.
[7] Hackett RA, Steptoe A. Type 2 diabetes mellitus and psychological stress - a modifiable risk factor. Nat Rev Endocrinol 2017; 13(9): 547-560.
[8] Hafner H, Maurer K, Loffler W, Riecher-Rossler A. The influence of age and sex on the onset and early course of schizophrenia. Br J Psychiatry 1993; 162: 80-86.
[9] Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. Psychiatry Investig. 2018; 15(3): 235-245.
[10] Li L, Li X, Zhou W, Messina JL. Acute psychological stress results in the rapid development of insulin resistance. J Endocrinol. 2013; 217(2): 175-184.
[11] Shin Y, Ham J, Cho H. Experimental Study of Thermal Comfort Based on Driver Physiological Signals in Cooling Mode under Summer Conditions. Applied Sciences 2021; 11(2): 845.
[12] Wilens TE, Rosenbaum JF. Transitional aged youth: a new frontier in child and adolescent psychiatry. J Am Acad Child Adolesc Psychiatry 2013; 52(9), 887-890.

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210620

Treatment Trajectories Graph Compression Algorithm Based on Cliques

Svetozar MILYKH[1] and Sergey KOVALCHUK
ITMO University, Saint Petersburg, Russia

Abstract. Learning treatment methods and disease progression is significant part of medicine. Graph representation of data provides wide area for visualization and optimization of structure. Present work is dedicated to suggest method of data processing for increasing information interpretability. Graph compression algorithm based on maximum clique search is applied to data set with acute coronary syndrome treatment trajectories. Results of compression are studied using graph entropy measures.

Keywords. Graph compression, treatment trajectory, graph entropy

1. Introduction

Medicine is one of the most significant areas of technologies development nowadays. Several areas such as disease outbreaks modeling, pathogens evolutionary and exploring, illness spreading prediction often use computational methods to find better solution.

Moreover, assistance systems are used to speed up and ease analysis of disease course, make information more complete according to already accumulated data. For example, in work [1] machine learning techniques are applied for type of hepatitis determination.

During the analysis of treatment trajectories we deal with heterogeneous objects. Graph structure is one of possible variants to work with such data set [2]. Objects can be distanced from each other with sequence of edges and similar values will be grouped together. The more data is considered the larger structure becomes. Consequently large structure turns more difficult for visualization and future analysis. Present work is intended to propose method of optimization graph structure in term of interpretability using graph compression technique.

2. Methods

2.1. Initial Graph Structure

The graph used in this paper can be classified as *similarity graph*. In order to construct such object *metric function* should be defined. Further, graph can be constructed with help of *threshold value* - value that is used to determine if edge between value is exists.

[1] Corresponding Author: Svetozar Milykh; ITMO University, St Petersburg, Russia; Email: svt2404@gmail.com

Combination of metric function, threshold value and graph structure will produce object where existing edge means that connected elements is similar enough to be considered.

This type of structure is often applied for NNS (nearest neighbour search) problem and used in areas such as template classification, morphology and computer vision, web graphs compression.

2.2. Compression Algorithm

The idea of suggested algorithm is based on finding cliques in graph and replacing them with representative vertex. Common schema of algorithm in shown in Figure 1.

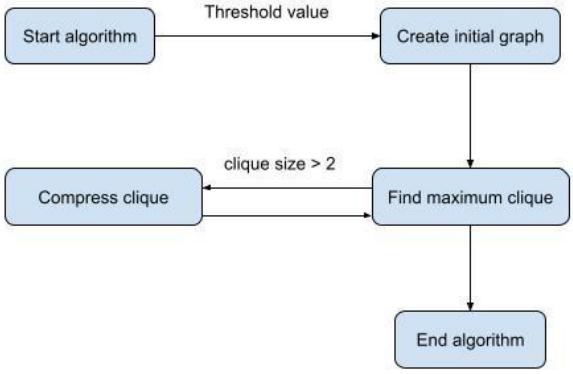


Figure 1. Common algorithm schema

Initial set of edges is usually filterer with threshold value. Then process of finding and compressing cliques is starting. Maximum clique search problem can be solved with any existing method. In current work evolutionary approach was applied. Algorithm finishes when no cliques with appropriate size is found in structure.

3. Results

3.1. Experiment Data

Given data contains information about treatment trajectory of acute coronary syndrome. Each episode is described with sequence of letter each of that is corresponding to certain department. Departments that are used in initial data is following:

1. A: Admission department 1
2. D: Admission department 2
3. E: Cardiology department
4. I: Surgery
5. F: Reanimation
6. N: Surgery (coronary angiography)

For example, we can have elements like this: AFIFD, AEFID, AFED.

All 3500 items in data set was initially divided into 7 clusters according to similarity of pathways.

Another important thing that need to be defined is similarity metric. According to initial data format, normalized Levenshtein distance was chosen as metric between vertices.

With all defined values we can construct *graph of episodes*, where every vertex is treatment trajectory and edges between elements contain information about their similarity.

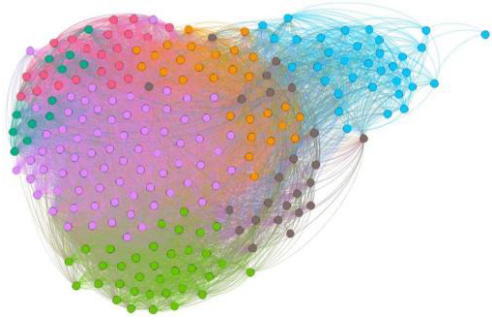


Figure 2. Example of graph of episodes with threshold value 0.5

3.2. Compression of Structure

During the compression operation several important details should be taken into account:

1. Mono and multi clustering compression
2. Threshold value effect

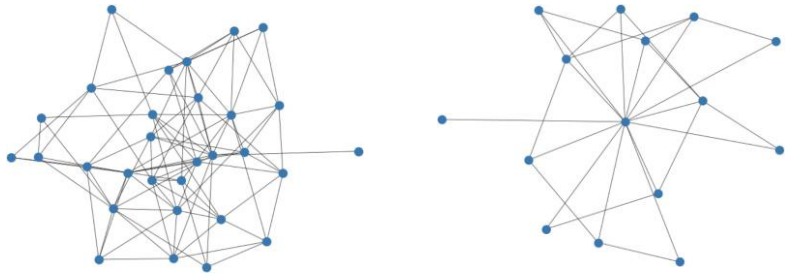
First experiment with initial data was performed in order to determine number of cliques that can be found with or without restriction on embedding elements from different clusters. Results are shown in Table 1.

Table 1. Number of found cliques depending on threshold value

Threshold value	Mono cluster	Multi cluster
0.5	0	17
0.55	0	26
0.6	2	27
0.65	3	35
0.7	7	37
0.75	10	42
0.8	11	33
0.85	6	10
0.9	0	4

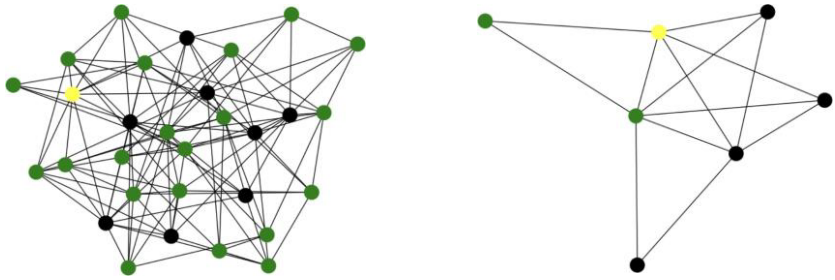
As expected, quantity of cliques with elements from same cluster is significantly lower that multi cluster variant. The main information that can be found from calculated values is following: varying threshold value can improve or worsen results of compression, because this has direct influence on number of found cliques.

Example of graph compression was illustrated using graph with 30 vertices and 75% of edges. Results are shown on Figure 3 and Figure 4. Figure 3(a) and 4(a) shows initial graph structure where color of vertex means cluster property, figure 3(b) and 4(b) shows compressed graph results.



(a) (b)

Figure 3. Graph with single cluster compression example



(a) (b)

Figure 4. Graph with 4 clusters compression example

The difference between single and multi cluster compression is based on clique structure. Embedding elements with different properties should be carefully handled or just prohibited as it was done in out work.

Another important moment is that during compression we do not break initial graph connectivity because of edge keeping strategy. Example of clique compression is shown on Figure 5. Group of vertices 3, 4, 5 was replaced with representative vertex 7 with all 3 edges kept.

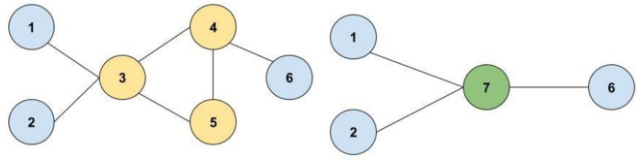


Figure 5. General clique compression example

Overall compression ratio of algorithm also was measured for different threshold values. Results of this experiment is present in Table 2.

Table 2. Compression ratio depending on

Threshold value	Number of compression	Compression ratio
0.5	20	85
0.6	25	76
0.7	11	33
0.8	3	8
0.9	0	0

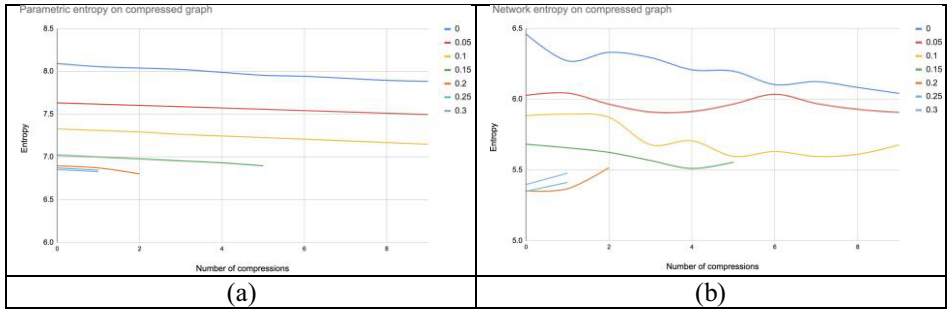
This table acknowledges that threshold value for initial graph construction can vary compression process.

4. Analysis

Interpretation of results is important part of work with graph. Often this structure is used for visualization of initial data and simplification the receipt of new conclusions during analysis. In paper [3] authors describe block structure simplification and use tests for people who is not involved in graphs during everyday work. This experiment showed that for 59% of people simplified structure is preferable to initial one.

In our work we suggest numerical method of interpretability check. Graph *entropy measure* shows the level of uncertainty in structure. It means decreasing entropy value during compression algorithm work will show growing level of certainty in structure. This parameter can be measured differently [4], but is this paper we will use two variant: parametric entropy (PE) [5] and network entropy (NE) [6].

Experiments was performed on initial data with different threshold value and calculated results are shown in Figure 6.



(a) (b)

Figure 6. Entropy dynamics during graph compression

According to the pictures we can say that parametric entropy has more consistent decreasing tendency, but overall values are getting smaller and certainty of structure is growing.

5. Conclusion and Future Works

Convolution of treatment trajectories is worse learning because analysis of collected data about treatment methods and possible disease progression can simplify and make therapy more predictable and accurate.

Our work propose structure and compression algorithm description for graph based structure that can be optimized in term of interpretability. Shown results can be applied not only for medical data, but for every data set, where similarity or distance metric between elements can be calculated.

Question of interpretability also can be considered from different angles because automatic tests have both advantages and disadvantages comparing to human tests. Nevertheless, purpose of decreasing information under consideration is reached and can be used in other related optimization works.

Acknowledgement

This research is financially supported by The Russian Science Foundation, Agreement #19-11-00326.

References

[1] Kaya Y, Uyar M. A hybrid decision support system based on rough set and extreme learning ma-chine for diagnosis of hepatitis disease. Applied Soft Computing Journal, Aug. 2013; 13(8): 3429–3438, doi: 10.1016/j.asoc.2013.03.008.

[2] Metsker O, et al. Modelling and Analysis of Complex Patient-Treatment Process Using Graph-Miner Toolbox. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS Jun. 2019; 11540: 674–680, doi: 10.1007/978-3-030-22750-0 65.

[3] Dwyer T, Riche NH, Marriott K, Mears C. Edge Compression Techniques for Visualization of Dense Directed Graphs. IEEE Transactions on Visualiz. 2013; 19(12): 2596– 2605.

[4] Dehmer M, Mowshowitz A. A history of graph entropy measures. Information Sciences Jan. 2011; 181(1): 57–78, doi: 10.1016/j.ins.2010.08.041.

[5] Dehmer M. Information processing in complex networks: Graph entropy and information functionals. Applied Mathematics and Computation 15 July 2008; 201(1–2): 82-94. doi: 10.1016/j.amc.2007.12.010.

[6] v Solé R, Valverde S. Information Theory of Complex Networks: on evolution and architectural constraints. Lecture Nots in Physics. LNP 2004; 650:189-207. Available: http://citeseer.nj.nec.com/450707.html.

This page intentionally left blank

pHealth 2021
B. Blobel and M. Giacomini (Eds.)

Subject Index

This page intentionally left blank

B. Blobel and M. Giacomini (Eds.)

Author Index