STUDIES IN HEALTH TECHNOLOGY AND INFORMATICS 289

Informatics and Technology in Clinical Care and Public Health

Editors: John Mantas
Arie Hasman
Mowafa S. Househ
Parisis Gallos
Emmanouil Zoulias
Joseph Liaskos

IOS Press

Data, informatics, and technology are now among the most important aspects inspiring health professionals and informaticians to improve healthcare for the benefit of patients.

This book presents the proceedings of the 19th annual International Conference on Informatics, Management, and Technology in Healthcare (ICIMTH 2021), held as a virtual event due to COVID-19 pandemic restrictions on 16 and 17 October 2021 in Athens, Greece. The ICIMTH conferences are a series of scientific events which bring together scientists working in the field of biomedical and health informatics from around the world. The 2021 conference examined the field of biomedical and health informatics in a very broad framework, presenting the research and application outcomes of informatics from cell to populations, and including a number of technologies such as imaging, sensors and biomedical equipment, as well as management and organizational aspects, including legal and social issues and the setting of research priorities in health informatics. A significant number of the papers included here relate to the COVID-19 pandemic.

Providing an insight into the latest developments in biomedical and health informatics, the book will be of interest to all those working in the field.

ISBN 978-1-64368-250-1 (print)
ISBN 978-1-64368-251-8 (online)
ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

INFORMATICS AND TECHNOLOGY IN CLINICAL CARE AND PUBLIC HEALTH

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 289

Recently published in this series

Vol. 288 B.L. Humphreys, R.A. Logan, R.A. Miller and E.R. Siegel (Eds.), Transforming Biomedical Informatics and Health Information Access – Don Lindberg and the U.S. National Library of Medicine

Vol. 287 J. Delgado, A. Benis, P. de Toledo, P. Gallos, M. Giacomini, A. Martínez-García and D. Salvi (Eds.), Applying the FAIR Principles to Accelerate Health Research in Europe in the Post COVID-19 Era – Proceedings of the 2021 EFMI Special Topic Conference

Vol. 286 R. Marcilly, L. Dusseljee-Peute, C.E. Kuziemsky, X. Zhu, P. Elkin and C. Nøhr (Eds.), Context Sensitive Health Informatics: The Role of Informatics in Global Pandemics

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

Informatics and Technology in Clinical Care and Public Health

Edited by

John Mantas

Health Informatics Laboratory, National and Kapodistrian University of Athens, Greece

Arie Hasman

Department of Medical Informatics, AMC – University of Amsterdam, The Netherlands

Mowafa S. Househ

College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

Parisis Gallos

Health Informatics Laboratory, National and Kapodistrian University of Athens, Greece

Emmanouil Zoulias

Health Informatics Laboratory, National and Kapodistrian University of Athens, Greece

and

Joseph Liaskos

Health Informatics Laboratory, National and Kapodistrian University of Athens, Greece

IOS Press

Amsterdam • Berlin • Washington, DC

ISBN 978-1-64368-250-1 (print)
ISBN 978-1-64368-251-8 (online)
Library of Congress Control Number: 2021951438
doi: 10.3233/SHTI289

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

Preface

This volume contains accepted papers from the ICIMTH (International Conference on Informatics, Management, and Technology in Healthcare), the scientific outcomes of which the Scientific Programme Committee is pleased to present to the academic and professional community of Biomedical and Health Informatics. The conference was held virtually on 16 and 17 October 2021 in Athens, Greece.

The ICIMTH 2021 Conference is the 19th Annual Conference in this series of scientific events, which brings together scientists working in the field of Biomedical and Health Informatics from all continents.

As was also the case last year, this year's conference was held as a virtual event by means of interactive teleconferencing platforms due to the COVID-19 pandemic and the consequent restrictions on gatherings and travel in many parts of the world.

The conference examines the field of Biomedical and Health Informatics in a very broad framework, presenting the research and application outcomes of informatics from cell to populations, and including a number of technologies, such as imaging, sensors and biomedical equipment, as well as management and organisational aspects, including legal and social issues and setting research priorities in health informatics. Essentially, data, informatics and technology inspire health professionals and informaticians to improve healthcare for the benefit of patients. As was expected this year, a significant number of papers relate to the COVID-19 pandemic.

It should be noted that these proceedings are published with open access in the Studies in Health Technology and Informatics (SHTI) series of IOS Press, with e-access for ease of use and browsing without the loss of any of the advantages of indexing and citation in the biggest scientific literature databases, such as Medline and Scopus.

By the deadline for papers, we had more than 170 submissions, of which we have accepted 120 after review as papers to be included in this volume of proceedings. This year, due to the shifting of the conference dates by a few months from the traditional time to later dates due to pandemic issues, the proceedings were not available at the time of the conference, but will be published by the end of the year.

The Editors would like to thank the Members of the Scientific Programme Committee, the Organising Committee, and all those reviewers who performed a very professional, thorough and objective refereeing of the scientific work in order to achieve this high-quality publication for a successful scientific event.

Athens, 02.11.2021
The Editors,
John Mantas, Arie Hasman, Mowafa S. Househ, Parisis Gallos, Emmanouil Zoulias, and Joseph Liaskos

This page intentionally left blank

ICIMTH 2021
Scientific Programme Committee and Reviewers

John Mantas, National and Kapodistrian University of Athens, Greece
Arie Hasman, AMC, University of Amsterdam, The Netherlands
Emmanouil Zoulias, National and Kapodistrian University of Athens, Greece
Andrianna Magdalinou, National and Kapodistrian University of Athens, Greece
Joseph Liaskos, National and Kapodistrian University of Athens, Greece
Mowafa S. Househ, Hamad Bin Khalifa University (HBKU), Doha, Qatar
Parisis Gallos, National and Kapodistrian University of Athens, Greece
Reinhold Haux, University of Braunschweig - Institute of Technology, Germany
Theo Arvanitis, Institute of Digital Healthcare, University of Warwick, UK

Reviewers List

Bakheet Aldosari
Yasser Alsafadi
Elske Ammenwerth
Konstantinos Antypas
Muzawir Arief
Mansoor Baig
Charalampos Balis
Alireza Banaye Yazdipour
Franziska Bathelt
Petr Berka
Antonis Billis
Stefano Bonacina
Taxiarchis Botsis
Zahid Butt
Georgi Chaltikyan
Chi-Chang Chang
Martha Charalampidou
Oana Sorina Chirila
Mihaela Crișan - Vida
George Demiris
George Despotou
Marianna Diomidous
Verena Dornauer
Leila Erfannia
Mina Fallah
Parisis Gallos
Michael Georg Grasser
Vincenzo Guardabasso
Arie Hasman
Mira Hercigonja-Szekeres
Felix Holl
Mowafa Househ
Marianna Isaakidou
Pashalina Lialiou
Joseph Liaskos
Andrianna Magdalinou
John Mantas
Ivana Ognjanovic
Kaija Saranto
Ramo Šendelj
Michael Shifrin
Vasile Stoicu-Tivadar
Lacramioara Stoicu-Tivadar
Orsolya Varga
Stefan Vogel
Patrick Weber
Dimitrios Zikos
Spiros Zogas
Emmanouil Zoulias

This page intentionally left blank

Contents

This page intentionally left blank

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210844

Using Artificial Intelligence to Develop a Lexicon-Based African American Tweet Detection Algorithm to Inform Culturally Sensitive Twitter-Based Social Support Interventions for African American Dementia Caregivers

Peter BROADWELL[a], Nicole DAVIS[b] and Sunmoo YOON[c,1]
[a] *Center for Interdisciplinary Digital Research, Stanford University, USA*
[b] *School of Nursing, Clemson University, USA*
[c] *General Medicine, Department of Medicine, Columbia University, USA*

Abstract. We extracted 3,291,101 Tweets using hashtags associated with African American-related discourse (#BlackTwitter, #BlackLivesMatter, #StayWoke) and 1,382,441 Tweets from a control set (general or no hashtags) from September 1, 2019 to December 31, 2019 using the Twitter API. We also extracted a literary historical corpus of 14,692 poems and prose writings by African American authors and 66,083 items authored by others as a control, including poems, plays, short stories, novels and essays, using a cloud-based machine learning platform (Amazon SageMaker) via ProQuest TDM Studio. Lastly, we combined statistics from log likelihood and Fisher's exact tests as well as feature analysis of a batch-trained Naive Bayes classifier to select lexicons of terms most strongly associated with the target or control texts. The resulting Tweet-derived African American lexicon contains 1,734 unigrams, while the control contains 2,266 unigrams. This initial version of a lexicon-based African American Tweet detection algorithm developed using Tweet texts will be useful to inform culturally sensitive Twitter-based social support interventions for African American dementia caregivers.

Keywords. unigram, lexicon, social media, caregiver, disparity, dementia

1. Introduction

The prevalence of dementia is disproportionately higher among African Americans than Whites in the United States, and family caregivers of persons with dementia suffer significant physical and psychological symptoms due to caregiving burdens and stress. According to a systematic review, aspects of caregiving and as well as the caregiver's quality of life have been found to vary with the race and ethnicity of the family caregivers of persons with dementia; these differences manifest at the intrapersonal, interpersonal and environment levels. In particular, evidence from a large psychosocial intervention trial indicated that African American family caregivers received no measurable benefit from the intervention (null results), while Hispanic/Latino and white caregivers showed

[1] Corresponding Author, Sunmoo Yoon, General Medicine, Department of Medicine, Columbia University Irving Medical Center, 630W 168 Street, PH105, New York, NY, 10032, USA; E-mail: sy2102@cumc.columbia.edu

positive responses to the intervention [1]. Twitter has the potential to be a platform for providing culturally sensitive social support interventions among African American family caregivers of dementia; this was especially true during the Covid-19 pandemic, as the role of online technology and networking became critical to facilitate the exchange of knowledge and to alleviate loneliness among caregivers. Yet, racial and ethnic demographic information about Twitter users is largely unavailable. Although deep learning and other statistical techniques have been widely applied to infer demographic information on Twitter, those demographic detection algorithms tend to be unavailable to open science communities and/or require access to account details that would compromise individuals' privacy [2]. The purpose of this study is to develop a lexicon-based African American Tweet detection algorithm to inform culturally sensitive Twitter-based social support interventions for African American dementia caregivers.

2. Methods

In order to gain a broad perspective on African American-related discourse that would guide the development of a social media search lexicon, we collected texts from two data sources: historical literature authored by African American writers (e.g., poems, plays, novels) provided a source of "old" terms, while recent Tweets captured "new" terms (**Figure 1**). For the literature corpora, an American history librarian at Columbia University first identified databases and collections authored by African Americans; the titles of these collections included African-American Studies, Black Short Fiction and Folklore, Black Women Writers, Black Drama, Black Thought and Culture, The Black Panther Newspaper and Black Music. From those collections, we extracted 14,692 poems and prose writings by African American authors using a cloud-based machine learning platform (Amazon SageMaker) provided via ProQuest Text and Data Mining (TDM) Studio from March 15, 2019 to December 31, 2019 with the help of the project director at TDM Studio. We used this same platform to collect a larger "control" corpus of 66,083 writings by authors from other ethnicities, including poems, plays, short stories, novels and essays. Due to copyright regulations in the United States, most of the materials in both sets were written prior to the mid-20th century.

For the Tweet corpora, we first extracted 3,291,101 Tweets using top hashtags associated with African American-related discourse (#BlackTwitter, #BlackLivesMatter, #StayWoke). We also collected 1,382,441 Tweets as a control set (these were Tweets without the above hashtags; they could contain other hashtags or none at all) from September 1, 2019 to December 31, 2019, using the Twitter application programming interface (API). As a first step of corpus processing, we removed redundant messages (retweets) and applied a bot detection algorithm to exclude irrelevant Tweets which were automatically generated by non-human accounts. A total of 803,495 Tweets (24.41%) associated with African American-related discourse and 369,348 Tweets (26.71%) in the control group were identified as unique and non-bot generated Tweets [3]. Lastly, we combined statistics from log likelihood and Fisher's exact tests as well as feature analysis of a batch-trained Naive Bayes (NB) classifier to select lexicons of terms most strongly associated with the target or control Tweets. We used 10-fold cross validation to evaluate the consistency and accuracy of the NB classifiers. Resources including de-identified data, videos on how to extract data, and the corpus processing and analytical code are available at https://osf.io/qruf3/. The larger study was approved by the Institutional Review Board (IRB).

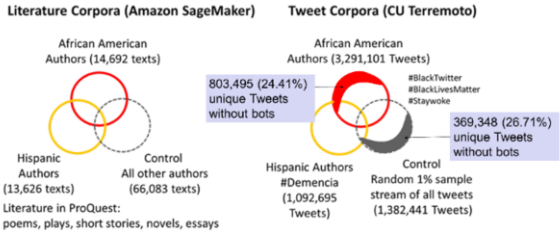


Figure 1. Quantities of texts used to create an African American lexicon and African American discourse-associated Tweet detection algorithm after applying a Twitter-bot detection process

3. Results

The "historical" African American literary-derived lexicon contains 1,643 unigrams (**Table 1**), while the Twitter-derived African American lexicon contains 1,734 unigrams and the control contains 2,266 unigrams. From 10-fold cross validation evaluation, the F-measure of the performance of the literary lexicon-based NB African American detection classifier was noted as 0.91 (area under the receiver operating characteristic curve: 0.97).

Table 1. Statistics of log likelihood, Fisher's exact tests, and feature analysis of a batch-trained Naive Bayes classifier on sample terms in African American lexicon (*p-value =0.00)

term	aa_freq	all_freq	log_ratio	fisher_ratio	bayes_nll	doc_freq	log_inv_freq
black	28873	52819	1.85*	2.89*	-6.22	29880	1.71
white	22301	51368	1.47*	1.84*	-6.5	31968	1.64
old	20267	63743	1.08*	1.12*	-6.69	40211	1.41
woman	18606	40940	1.54*	2*	-6.59	24799	1.9
hair	7693	23581	1.11*	1.16*	-7.37	18011	2.22
hell	3786	10955	1.17*	1.26*	-7.88	8891	2.92
hurt	2756	7413	1.26*	1.42*	-8.11	6175	3.29
Jesus	1687	4298	1.33*	1.55*	-8.45	2973	4.02
Harlem	1586	1778	3.03*	19.77*	-8.33	1116	5
hoped	1140	2658	1.46*	1.8*	-8.8	2517	4.18
cops	796	1840	1.47*	1.82*	-9.04	1284	4.86
afro	656	757	2.94*	15.54*	-9.1	566	5.68

aa: African American; freq: frequency; nll: negative log likelihood; inv_freq: inverse document frequency

4. Discussion and Conclusiond

Although our lexicon-based classifiers for the historical literary corpus and the recent Tweet corpus were quite similar in their effectiveness at classifying held-out samples of pre-1950s literature and Tweets from 2019, respectively, each was less successful when applied to the other corpus, or when the training sources were combined to produce a unified classifier. This argues against the usefulness of historical texts for the classification of modern social media, at least with the methods employed in this study.

We also found that a lexicon composed of unigrams was more accurate in differentiating texts from held-out test samples of the two groups than lexicons composed of n-grams of various lengths. This result contrasts with recent studies reporting superior sentiment classification performance from n-gram based approaches [4]. Our finding is

likely due to the binary nature of the classification task and to the relatively terse nature of Tweets. In addition, although two domain experts in our team reviewed the final lexicon, further cleaning of an n-gram lexicon by multiple experts to retain culturally meaningful phrases (e.g., "family caregiver" as bi-gram) while removing irrelevant n-grams (e.g., "look at"), may improve its performance. Ongoing experiments involving classifiers using more sophisticated language models, specifically Bidirectional Encoder Representations from Transformers (BERT), have a strong likelihood of significantly altering the findings above.

This study developed a lexicon-based classifier to identify Tweets authored by African Americans to inform culturally sensitive Twitter-based intervention design for family caregivers of persons with dementia. As mentioned earlier, African Americans as a group experience disproportionately high rates of dementia. Relevant research literature urges the development of culturally sensitive educational interventions that carefully incorporate cultural traditions into family-oriented care strategies to deal with caregiving issues among African American caregivers, such as navigating not knowing the true wishes of a dementia patient. Furthermore, the Covid-19 pandemic has highlighted the role of social media in the propagation of health-related information. Nevertheless, the most effective content and framing strategies for culturally sensitive interventions using social media remain unclear. Identifying existing African American communities and discourse patterns on social media platforms like Twitter is the first basic step towards understanding a community and culture necessary to develop culturally sensitive interventions (e.g., terms, norms, culturally sensitive expressions, and topics). As a limitation, it is important to consider the ethical issues regarding the use of Twitter data in mental health surveillance [5]. Researchers must be aware of the unresolved distrust towards scientists and health professionals among African Americans in the U.S. due to historical factors. The Tuskegee experiment is perhaps the most infamous when evaluating the benefits and risks inherent to social media-based health research among African Americans. With these concerns in mind, we believe that our first version of a lexicon-based African American Tweet detection algorithm developed using literature and Tweet texts can be used both effectively and ethically to inform culturally sensitive Twitter-based social support interventions for African American dementia caregivers and future studies are needed to refine this algorithm.

Acknowledgments

U.S. federal grant TweetS2 R01AG060929 (P.I.: Yoon).

References

[1] Alzheimer's Association. Alzheimer's disease facts and figures. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2020.
[2] Sloan L, Morgan J, Burnap P, Williams M. Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. PloS one. 2015 Mar 2;10(3):e0115545.
[3] Subrahmanian VS, Azaria A, Durst S, Kagan V, Galstyan A, Lerman K, Zhu L, Ferrara E, Flammini A, Menczer F. The DARPA Twitter bot challenge. Computer. 2016 Jun 13;49(6):38-46.
[4] Dey A, Jenamani M, Thakkar JJ. Senti-N-Gram: An n-gram lexicon for sentiment analysis. Expert Systems with Applications. 2018 Aug 1;103:92-105.
[5] Mikal J, Hurst S, Conway M. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. BMC medical ethics. 2016 Dec;17(1):1-1.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210845

A Comparison of Word Embeddings to Study Complications in Neurosurgery

Gleb DANILOV[a,1], Konstantin KOTIK[a], Michael SHIFRIN[a], Yulia STRUNINA[a], Tatyana PRONKINA[a], Tatyana TSUKANOVA[a], Timur ISHANKULOV[a], Maria SHULTS[a], Elizaveta MAKASHOVA[a], Yaroslav LATYSHEV[a], Rinat SUFIANOV[a], Oleg SHARIPOV[a], Anton NAZARENKO[a], Nikolay KONOVALOV[a], and Alexander POTAPOV[a]
[a] *Laboratory of Biomedical Informatics and Artificial Intelligence, National Medical Research Center for Neurosurgery named after N.N. Burdenko, Moscow, Russian Federation*

Abstract. Our study aimed to compare the capability of different word embeddings to capture the semantic similarity of clinical concepts related to complications in neurosurgery at the level of medical experts. Eighty-four sets of word embeddings (based on Word2vec, GloVe, FastText, PMI, and BERT algorithms) were benchmarked in a clustering task. FastText model showed the best close to the medical expertise capability to group medical terms by their meaning (adjusted Rand index = 0.682). Word embedding models can accurately reflect clinical concepts' semantic and linguistic similarities, promising their robust usage in medical domain-specific NLP tasks.

Keywords. Neurosurgery, complications, NLP, word embeddings, clustering

1. Introduction

Word embeddings enable capturing useful semantic properties and linguistic relationships between words which might be important for information extraction, classification, and more complex natural language processing (NLP) tasks. In our opinion, word embeddings might help study distinct clinical concepts and discover relationships between them as soon as the underlying models accurately reflect clinical semantics. Thus, it seems reasonable to test this capability of word embeddings in modeling the known relationships between clinical concepts well-recognized by medical experts.

The specific domain we focused on was complications in neurosurgery. There are well-known types of complications accepted by many experts. However, the wide spectrum has not been described and agreed upon between neurosurgeons. Furthermore, the formal definitions of complications in neurosurgery are vague. Therefore, we hypothesized that word embeddings learned from narrative clinical notes can contribute to understanding the spectrum of complications in neurosurgery and a more rigorous definition of this concept. Our study aimed to compare the capability of different word

[1] Corresponding author, Gleb Danilov, N.N. Burdenko Neurosurgery Center, 4th Tverskaya-Yamskaya str. 16, Moscow 125047, Russian Federation; E-mail: glebda@yandex.ru.

embeddings to capture the semantic similarity of clinical concepts related to complications in neurosurgery at the level of medical experts.

2. Methods

To accomplish the research task, all unstructured textual data potentially containing the information on complications in neurosurgery were obtained from the electronic health records (EHR) of the National Medical Research Center of Neurosurgery named after academician N.N. Burdenko (Moscow, Russia) for the period between 2000 and 2017. The source documents reflected the initial assessment, past medical history, history of present illness, laboratory tests, physical and neurological examination, studies, medication, operative reports, daily notes, discharge summaries, etc. All the texts were typed in by doctors and other medical personnel on a keyboard. The corpus was preprocessed as follows: all the characters except for Cyrillic symbols and single spaces removed; texts tokenized with a space separator; stop-words, meaningless tokens (single letters, artifacts, etc.) and words occurred less than 6 times in the corpus eliminated; spelling corrected with the method we proposed in our previous work and tokens lemmatized (1). A medical expert then screened the resulted vocabulary of unique lemmas to select maximum terms potentially related to any adverse events (with broad inclusion criteria to capture diseases, symptoms, syndromes, accidents, medical errors, etc.).

All the words in the initial corpus were substituted by their lemmas to train word embeddings with Word2vec, GloVe, FastText, and pointwise mutual information (PMI) algorithms (2–5). When appropriate, we varied model type (CBOW/skip-gram), context window size (5-20), vector size (50-300), and the number of iterations over data across the models. The unprocessed clinical corpus was used to train RoBERTa (Robustly Optimized BERT Pretraining Approach) masked language model (6). It was trained during 5-10 epochs using base architecture. Different aggregated techniques were applied to get word embeddings of the vocabulary: mean average and maximum calculation of representations from the encoder-layers. The intersection of all sets of nouns showing positive cosine similarity with the word "complication" in every vector space obtained was further screened and labeled (when possible) by the type of clinical entity (symptom, syndrome, disease), body system, organ involved and ICD10 code for each term. A fully labeled subset of nouns was grouped by 4 aforementioned features to shape benchmark clusters. A k-means clustering algorithm was then applied to cluster each set of word embeddings with k equal to the number of benchmark clusters. Finally, we judged the clustering quality comparing to benchmark clusters using an adjusted Rand index.

The data were processed, and most word embeddings were learned within the R programming environment (version 4.0.3) in RStudio Server IDE (version 1.3.1093) using *tidyverse, tidytext, dplyr, Matrix, text2vec, word2vec, widyr, irlba, SnowballC, furrr and fossil* packages. FastText and RoBERTa vector representations were obtained with Python programming language (version 3.6.10) in Jupyter Notebook (version 6.1.4) using *fasttext* and HuggingFace *tokenizers* and *transformers* libraries.

3. Results

To create a clinical corpus, 588 text fields from 78 tables of the EHR database were identified. The corpus was compiled of 13 060 326 narrative text records containing data for 90 688 complete cases of neurosurgical treatment. Text preprocessing and tokenization produced 229 019 413 raw word tokens ending up with 40 121 unique lemmas. Of these, the expert selected 5 853 terms, potentially relevant for the concepts of complications/adverse events. A total of 84 vector spaces with different word embedding engines and varying learning parameters were obtained in the study. After finding the intersection of all sets of nouns showing a positive cosine similarity to the word "complication" in every vector space, it was possible to completely label 258 words, which were grouped into 40 benchmark clusters by 4 features (see the "Methods" section). The results of vector clustering with the k-means algorithm in benchmarking are shown selectively for 10 types of word embeddings in Table 1.

Table 1. Benchmarking of word embeddings clustering assessed with ARI. CBOW – continuous bag of words, SG – skip-gram, NI – number of iterations, ARI – adjusted Rand index.

	Model	CBOW/SG	Window size	Vector size	NI	ARI
1	FastText	skipgram	10	100	-	0.682
2	FastText	cbow	5	200	-	0.677
3	RoBERTa	-	-	-	-	0.330
4	GloVe	-	10	100	20	0.157
5	GloVe	-	10	300	20	0.116
6	PMI	-	10	100	-	0.081
7	Word2vec	cbow	10	300	10	0.013
8	Word2vec	skipgram	10	300	10	0.013
9	Word2vec	cbow	10	100	10	0.005
10	Word2vec	skipgram	10	100	10	0.005

Figure 1 shows a word cloud of medical terms semantically similar to the word “complication” in a high dimensional space produced by the best FastText model in our experiment (ARI = 0.682) and projected in 3-dimensional space by the t-SNE algorithm (perplexity = 8, learning rate = 10) with TensorBoard Embedding Projector. All the terms in Russian were automatically translated with the https://translate.yandex.ru/ service for international presentation purposes. Some of the terms containing misspellings were transliterated. The best word embedding approach demonstrates a reasonable spatial distribution of the related concepts that occur in the context of the "complication" term.

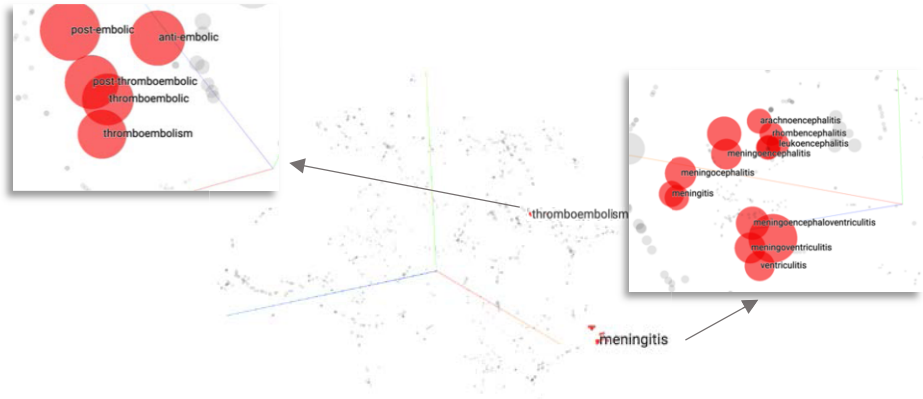


Figure 1. Word clusters of complications in neurosurgery derived from high dimensional FastText word embeddings and represented in 3 dimensions by the t-SNE algorithm with TensorBoard Embedding Projector (http://projector.tensorflow.org/). For example, word clusters for intracranial inflammatory complications are scaled right, and thromboembolism is shown left.

4. Discussion

In our study, models leveraging sub-word information from morphologically rich Russian language performed better compared to those treating words as atomic units. Interestingly, the BERT-based model demonstrated worse results than FastText in our domain-specific task, possibly due to the isolation of words from their contexts. Source words misspellings, expert-dependent benchmark cluster labeling, and a fixed set of models might be the limitations of our study. Generally, our results support those of the authors from other medical domains (7). Y. Wang et al. (2018) showed that word embeddings trained from EHR and medical literature can capture the semantics of medical terms better, and find semantically relevant medical terms closer to human experts' judgments than those trained from general domain data (8). The authors also importantly concluded that no global ranking of word embeddings for all biomedical NLP applications exists (8).

5. Conclusion

Word embedding models can accurately reflect clinical concepts' semantic and linguistic similarities, promising their robust usage in medical domain-specific NLP tasks. *This project was supported by the RFBR grants 18-29-01052 (data preprocessing) and 18-29-22085 (adverse events clustering).*

References

[1] Danilov G, Shifrin M, Strunina U, Pronkina T, Potapov A. An Information Extraction Algorithm for Detecting Adverse Events in Neurosurgery Using Documents Written in a Natural Rich-in-Morphology Language. Stud Health Technol Inform. 2019 Jul;262:194–7.

[2] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings. International Conference on Learning Representations, ICLR; 2013.

[3] Pennington J, Socher R, Manning CD. GloVe: Global vectors for word representation. In: EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference. 2014. p. 1532–43.

[4] Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching Word Vectors with Subword Information. Trans Assoc Comput Linguist. 2017 Jul;5:135–46.

[5] Silge J. Tidy word vectors, take 2! [Internet]. 2017 [cited 2020 Aug 20]. Available from: https://juliasilge.com/blog/word-vectors-take-two/

[6] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. 2019 Jul; Available from: http://arxiv.org/abs/1907.11692

[7] Kalyan KS, Sangeetha S. SECNLP: A survey of embeddings in clinical natural language processing. Vol. 101, Journal of Biomedical Informatics. Academic Press Inc.; 2020. p. 103323.

[8] Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, et al. A comparison of word embeddings for the biomedical natural language processing. J Biomed Inform. 2018 Nov;87:12–20.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210846

Gulf Cooperation Council Clinical Trials in the Pursuit of Medications for COVID-19

Bassam Ali JABER[a], Mowafa HOUSEH[a], Zubair SHAH[a], M.Sohel RAHMAN[c], Mounir HAMDI[a], Kabir H BISWAS[b] and Tanvir ALAM[a,1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*
[b] *College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar*
[c] *Department of Computer Science & Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh*

Abstract. Tremendous changes have been witnessed in the post-COVID-19 world. Global efforts were initiated to reach a successful treatment for this emerging disease. These efforts have focused on developing vaccinations and/or finding therapeutic agents that can be used to combat the virus or reduce its accompanying symptoms. Gulf Cooperation Council (GCC) countries have initiated efforts on many clinical trials to address the efficacy and the safety of several therapeutic agents used for COVID-19 treatment. In this article, we provide an overview of the GCC's clinical trials and associated drugs' discovery process in the pursuit of an effective medication for COVID-19.

Keywords. Coronavirus, COVID-19, Drug, GCC, Qatar.

1. Introduction

The fast and widespread COVID-19 pandemic compelled the medical society to use all resources to combat the virus as quickly as possible. One approach was to use certain medications already related to inhibiting virus replications, human cell invasions, or preventing inflammatory reactions that can lead to further deterioration or failure in numerous body organs [1]. In this article, we investigated the clinical trials that are currently under different phases of completion in the Gulf Cooperation Council (GCC) countries, which include Qatar, Kuwait, Oman, Saudi Arabia, UAE and Bahrain.

2. Methods

Detailed information on the clinical trials aimed towards finding therapeutic solutions for COVID-19 was collected from https://clinicaltrials.gov/ct2/covid_view (download date: May 07, 2021). Each clinical trial was then analyzed to understand the medications under investigation in the GCC countries.

[1] Corresponding Author, Tanvir Alam, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar; E-mail: talam@hbku.edu.qa.

3. Results and Discussions

In total, we found 27 clinical trials currently underway in GCC countries. Among them, 12, 8, and 5 clinical trials are being conducted in Saudi Arabia (KSA), Qatar, and Kuwait, respectively. Only one trial each was enrolled from the UAE and Bahrain. The medications investigated mainly fall into antiviral, anti-inflammatory, and anti-coagulator categories. We also note additional drugs that have been repurposed as they exhibited some anti-viral, immune-modulating, or other effects. We also summarize the drugs in Table 1 and provide a brief of ongoing trials in GCC countries.

Table 1. List of drugs that are under clinical trials in GCC countries

Drug Family	Drug name	Clinical Trials
Anti-viral (Direct)	Favipiravir, Remdesivir, Darunavir/ Lopinavir	NCT04387760, NCT04529499, NCT04464408, NCT04302766, NCT04425382
Anti-viral (Repurposed)	Hydroxychloroquine, Artemisinin/ Artesunate, ACE Inhibitors and Angiotensin II Receptor Blockers	NCT04437693, NCT04349592, NCT04394442, NCT04387240, NCT04357535
Anti-inflammatory	Pioglitazone, Anakinra, Zafirlukast, Tocilizumab/Siltuximab, Omega 3 Oil	NCT04604223, NCT04643678, NCT04486521, NCT04553705, NCT04836052
Anti-coagulant	Bivalirudin	NCT04445935
Miscellaneous	Iloprost, Oestrogen, Alpha One Antitrypsin, Colchicine, Interferon-Beta, Aspirin, Rivaroxaban	NCT04445246, NCT04853069, NCT04385836, NCT04324463, NCT04468139

3.1 Antiviral Drugs

Favipiravir: This antiviral is being used to treat unresponsive or re-emerging viral influenza strains in contrast to conventional anti-viral treatment [2]. Favipiravir (Avigan® as a trade name) is currently the subject of 2 clinical trials being conducted in Kuwait and KSA (NCT04529499 and NCT04464408, respectively). While both clinical trials are evaluating the efficacy and safety of Favipiravir in COVID-19 treatment, they differ in the outcome measurements. The first one explores "Time to resolution of hypoxia" as a primary outcome while the other one is being conducted on "Time from randomization to negativity in the RT-PCR nucleic acid test for COVID-19 within 15 days of randomization".

Remdesivir: Remdesivir is similar to Favipiravir as it also selectively inhibits RNA-dependent RNA polymerase (RdRp) [3]. The Kuwait-based trial (NCT04302766) has not determined the outcome measures of the interventions being applied throughout the clinical trial.

Darunavir/Lopinavir: Both of these antivirals are HIV protease inhibitors employed in COVID-19 treatment due to their ability to prevent further infection of the virus. The clinical trial (NCT04425382) taking place in Qatar aims to compare the efficacy and safety outcomes of Darunavir/Cobicistat versus Lopinavir/Ritonavir in treating pneumonia associated with COVID-19.

3.2 Antiviral Drugs (Repurposed)

Hydroxychloroquine (HCQ): This anti-malarial drug, which is also used in rheumatoid arthritis and other autoimmune diseases, blocks viral entry into the cells through endosomal acidification [4]. In Bahrain, clinical trial NCT04387760 is a pilot study

exploring the ability to create a comparison between HCQ, Favipiravir, and the routine care using the time frame until hospital discharge or death as the measure of the primary outcome. HCQ was also evaluated for its safety, efficacy, and effectiveness in Post Exposure Prophylaxis (PEP) for health care providers in Qatar (NCT04437693). Another clinical trial in Qatar, NCT04349592, addressed the value of adding Azithromycin, an antibiotic in the Macrolide family that exhibits anti-inflammatory effects, to HCQ in COVID-19 treatment.

Artemisinin/Artesunate: An antimalarial combination that shows antiviral activity. It also has an anti-inflammatory effect, which is attributed to its ability to reduce the level of inflammatory cytokines. The reduction avoids both the cytokine storm as well as inflammatory organ injuries for high-risk patients [5]. A clinical trial taking place in KSA is evaluating the efficacy of this combination.

Angiotensin-Converting Enzyme Inhibitors (ACE-I) and Angiotensin II Receptor (ARB) Blockers: The role of ACE and ARB in COVID-19 prognosis is still controversial. The hypothesis suggests that the SARS-CoV-2 uses ACE2 protein expressed on alveolar cell surfaces to enter the cell. Accordingly, patients who are hypertensive and receiving ACEs or ARBs are at high risk for potentially developing Acute Respiratory Distress Syndrome (ARDS). Another hypothesis denies these claims as they are unsupported by virus binding. Additionally, internalization may not require the catalytic activity of the protein [6]. These inhibitors may cause alteration in the levels of the protein on the cell membrane, which can affect virus binding. Additionally, a clinical trial is taking place in KSA and targeting only hypertensive patients who are on ACE or ARB to assess the impact of ARB and ACE-I on a COVID-19 prognosis. The primary outcome would be the time frame from the date of enrollment until discharge from hospital or death; whichever comes first.

3.3 Anti-Inflammatory Drugs

Pioglitazone: Researchers have claimed that this insulin sensitizing, anti-diabetic agent reduces chronic inflammation in T2DM patients. Moreover, it enhances the plasma level of anti-inflammatory agents. This feature has been used to improve the clinical outcome (NCT04604223) of the COVID-19 disease in T2DM patients.

Anakinra: An interleukin (IL)-1 receptor antagonist used in many inflammatory diseases such as rheumatoid arthritis, familial Mediterranean fever. In the clinical trial NCT04643678, it is used to avoid Cytokine storm syndrome, thus reducing the time for mechanical ventilation and improving the clinical outcome of patients with moderate to severe COVID-19 symptoms [7].

Zafirlukast: A leukotriene receptor antagonist used for treating chronic asthma. Its efficacy is being investigated in a clinical trial in KSA (NCT04871828). The primary investigation on this trial is time frame of clinical improvement.

Tocilizumab/Siltuximab: Interleukin-6 Antagonists (Anti IL6) have been used as antirheumatic. Tocilizumab and Siltuximab Inhibition of IL-6 of receptors may lead to the reduction of cytokine and acute phase reactant production [8]. The efficacy of such an approach is being investigated in a KSA-based clinical trial, NCT0448652, which explores a primary outcome of ventilator-free days.

Omega 3 Oil: Under a clinical trial (NCT04836052) in Qatar, Omega-3-oil has been shown to have fewer proinflammatory mediators that may have immunomodulating, anti-inflammatory and antiviral effects. These features can be employed to improve the clinical outcomes of patients admitted to ICU due to COVID-19.

3.4 Anticoagulation Drugs

Bivalirudin: In the ARDS associated with COVID-19 infection is contributed to the activation of the coagulation system. Administrating Bivalirudin, an anticoagulant with a direct thrombin inhibition effect, can have positive effects in improving a patient's respiratory status. An ongoing clinical trial (NCT04445935) in Qatar is investigating the effects of Bivalirudin.

3.5 Miscellaneous

Hydrogen Peroxide, Povidone Iodine, Cetylpyridinium Chloride: Sodium Hypochlorite (0.1% solution) is an oral mouthwash used as a local antiseptic in dental procedures to clear saliva's viral load in COVID-19 patients. A clinical trial is being conducted to test the efficacy of this combination in KSA, with the change in viral load as the primary outcome. Additionally, these local antiseptics are being used to prevent the transmission of the virus rather than treatment. Miscellaneous drugs such as Iloprost, Oestrogen, Alpha One Antitrypsin, Colchicine, Interferon-Beta, Aspirin, and Rivaroxaban, etc., as well as plasma therapy (NCT04474340) are also in clinical trials utilized for exploring treatment plans of COVID-19. We considered clinical trials registered till May 2021 and this is a limitation of this study. The full list of the clinical trials can be found at: https://github.com/tanviralambd/GCC_ClinicalTrials.

4. Conclusions

GCC countries are working to curb the spread of COVID-19 and find a therapeutic solution. While the vaccination efforts are continuing in the GCC countries at a reasonable speed, no particular drug exists that may be considered a solution for this disease. A major stream of the drugs in clinical trials are antivirals (known or repurposed) that can combat the virus with acceptable levels of patient safety. Moreover, anti-inflammatory drugs are also under trial to keep patients free from inflammation, leading to ARDS, cytokine storm, and other consequences. Ideally, these clinical trials conducted in the GCC will provide more insight into the efficacy of drugs against COVID-19.

References

[1] Khan JY, et al. Toward Preparing a Knowledge Base to Explore Potential Drugs and Biomedical Entities Related to COVID-19: Automated Computational Approach. JMIR Med Inform. 2020;8:e21648.
[2] Agrawal U, Raju R, Udwadia ZF. Favipiravir: A new and emerging antiviral option in COVID-19. Armed Forces Med J India. 2020;76:370–6.
[3] Rochwerg B, et al. Remdesivir for severe covid-19: a clinical practice guideline. BMJ. 2020;370:m2924.
[4] Schaper C. A Mechanism of Action for Hydroxychloroquine and Azithromycin to Inhibit Coronavirus Disease COVID-19 [Internet]. Available from: http://dx.doi.org/10.26434/chemrxiv.12148914.v1.
[5] Uzun T, Toptas O. Artesunate: could be an alternative drug to chloroquine in COVID-19 treatment? Chin Med. 2020;15:54.
[6] Javanmard SH, Heshmat-Ghahdarijani K, Vaseghi G. Angiotensin-converting-enzyme inhibitors (ACE inhibitors) and angiotensin II receptor blocker (ARB) use in COVID-19 prevention or treatment: A paradox [Internet]. Infection Control & Hospital Epidemiology. 2021. p. 118–9.

[7] Navarro-Millán I, Sattui SE, Lakhanpal A, Zisa D, Siegel CH, Crow MK. Use of Anakinra to Prevent Mechanical Ventilation in Severe COVID-19: A Case Series [Internet]. Arthritis & Rheumatology. 2020. p. 1990–7. Available from: http://dx.doi.org/10.1002/art.41422

[8] Khan FA, Stewart I, Fabbri L, Moss S, Robinson K, Smyth AR, et al. Systematic review and meta-analysis of anakinra, sarilumab, siltuximab and tocilizumab for COVID-19. Thorax. 2021;76:907–19.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210847

Hazards for the Implementation and Use of Artificial Intelligence Enabled Digital Health Interventions, a UK Perspective

Stuart HARRISON[a,1], George DESPOTOU[a] and Theodoros N. ARVANITIS[a]
[a]*Institute of Digital Healthcare, WMG, University of Warwick, UK*

Abstract. Background: Artificial Intelligence (AI) has seen an increased application within digital healthcare interventions (DHIs). DHIs use entails challenges about their safety assurance. Exacerbated by regulatory requirements, in the UK, this places the onus of safety assurance not only on the manufacturer, but also on the operator of a DHI. Clinical Safety claims and evidencing safe implementation and use of AI-based DHIs require expertise, to understand and act to control or mitigate risk. Current health software standards, regulation, and guidance do not provide the insight necessary for safer implementation. Objective: To interpret published guidance and policy related to AI and justify clinical safety assurance of DHIs. Method: Assessment of UK health regulation policy, standards, and AI institution insights, utilizing a published Hazard Assessment framework, to structure safety justifications, and articulate hazards relating to AI-based DHIs. Results: AI enabled DHI hazard identification, relating to implementation and use within healthcare delivery organizations. Conclusion: By application of the method, we postulate that UK research of AI DHIs highlighted issues that may affect safety, in need of consideration to justify safety of a DHI.

Keywords. Digital health, safety, justification, health system, artificial intelligence, hazard analysis

1. Introduction

Digital Health Intervention manufacturers collate evidence and justification about the safety of their products. This has also been presented as a safety case [1], to aid the interpretation, understanding and communication across stakeholders (e.g., regulatory assessment, client engagement). In addition to safety, effectiveness and security are also foundational elements of the lifecycle of all DHIs [2]. In addition to the manufacturer justifying safety, in the UK, the operator of DHI also needs to justify safe implementation and use. AI hazards are unsubstantiated in guidance, standards, policy and overall practice [2]. Although they are often acknowledged, there is currently little published providing a comprehensive insight in the safe implementation and use of AI DHIs [3]. Many implementations of AI-based DHI's rely on expert resources and insights from the manufacturer, to provide safety assurance prior to implementation or use of the DHI [4]. The dynamic nature of the involved algorithms, in AI, poses a challenge to regulation, requiring "real time" post-market product change, implementation, and use of the DHI.

[1] Corresponding Author, S. Harrison, Institute of Digital Healthcare, WMG, University of Warwick, CV4 7AL, UK; E-mail: stuart.harrison@warwick.ac.uk

This will involve feedback and regulatory insight during post-deployment of the intervention, thus, to demonstrate the ongoing applicability of any safety justifications made. We investigate current recommendations in this field, aiming to collate best practice recommendations from sources. We identify a list of high level (generic) AI hazards, through the application of a framework for safety justification of DHIs [5].

2. Method

This study followed a 2-step approach. Firstly, the identification of UK-based standards, regulation, legislation, health policy and guidance sources, and concerns regarding AI were consolidated. The sources selected from UK Government Regulators, policy makers and national institutions provide standards, regulation and insight to AI DHI implementation and use. We document issues from each source, while sorting and grouping to align to common themes, technical and operational areas. Secondly, we identify high-level hazards relating to AI, along with their contributing factors. This was achieved through a hazard identification and safety justification process as part of the method applied [5], where issues related to hazards (e.g., directly affecting the patient care), or as classes of failures that constitute contributing factors to hazards were retrieved. Analyzing issues, presented in literature, enables the postulation of hazard, hazardous situation, harm, and effect. Assumptions, about likelihood of hazard, are out of scope. Medical device risk management standards, such as ISO 14971 [6], provide suitable defined terms.

3. Results

Table 1 sees the safety related issues, summarized by the application of the hazard assessment framework, with a rationale from sources. Extrapolation and application of this framework to express issues in the form of hazard table is presented in Table 2.

Table 1. Safety related recommendations or issues

Recommendation	Summary	
Access to sensitive data	Provide trust and confidence in data sharing. Reduction in transaction costs of accessing data [7–12].	Data
Improve the availability of data. Data Quality and Maturity.	Access to AI training data, in compliance with regulation [2,7–14].	
Bias and Discrimination. Minimizing Bias.	Training data and algorithms need to be verified for objectivity and inclusivity, extrapolated to the real world [2,12–14].	
Representativeness	Misrepresentation of groups within data samples [12,14].	
Fit-for-purpose and sufficiency. Self-fulfilling prediction.	Justification of data quality and quantity for intended purpose. Reinforcement learning bias [7,9,10,12,14,15].	
Source integrity and Measurement Accuracy	Ensuring data sources have reliable and impartial methods of collection [11,12,14].	
Timeliness and recency.	Accuracy and currency of datasets [11,12,14,15].	
Relevance, Appropriateness and Domain Knowledge	Utilization of domain experts [10–12,14].	
Outcome fairness	Methods to ensure unbiased deployment of AI based systems [14].	Operational use
Decision-Automation Bias/The Technological Halo Effect. Automation complacency.	Over reliance on the system and inability to respond to failures [14,15].	
Automation-Distrust Bias	Reluctance to trust AI based system decisions [14].	
Stakeholder Engagement. Trustworthiness. Accountability.	Aimed to build trust and confidence. Improved regulation and health care monitoring of AI systems [2,7,9,11–14].	

Human centered implementation processes	AI development specific Human factors considerations [10–12,14].	
Accuracy and performance metrics. Unsafe Failure Modes.	Error rates for AI generated outputs. Data to predict an output and performance (e.g., accuracy) [14,15].	Performance
Reliability. Negative side effects. Reward hacking.	Consistency of behavior. Supervision for longer-term operational reliability goals and control [11,12,14,15].	
Robustness	A measure of a systems integrity [11,14].	
End-to-end AI Safety. Unscalable oversight.	Regular verification and validation of AI throughout its lifecycle [2,10–12,14].	
Concept Drift or Distributional Shift	Training data mismatch over time [10–12,14,15].	
Brittleness	Undetectable changes in input data leading to failures [14].	
Model hardening	Securing the AI system to combat adversarial attack [14].	
Misdirected Reinforcement Learning Behavior. Insensitivity to impact. Unsafe exploration.	Insufficient controls placed upon trial-and-error processing methods. Inefficient supervision or monitoring impacts outcome and efficiency aims [14,15].	
Transparency and explain-ability.	Explanation of processes, services and decision making by AI [7,9–15].	

Table 2 summarizes associated failures as a hazard table. Hazardous situation is the circumstance exposing patients/users hazard(s) aligned to medical device standards [6].

Table 2. Hazard Analysis & Results

	Hazard	Hazardous Situation	Harm	Effects
1	Incorrect clinical decision result or diagnosis.	Incorrect treatment plan or decision selected.	Incorrect diagnosis	Availability and quality of data; Insufficient data sample. Training error (e.g., bias). Incorrect usage of intervention; Inconsistent intervention performance; Recall, precision accuracy training for users.
2	Failure to operate as intended (annunciated).	Software failure of data error.	Delay in diagnosis and treatment.	Corrupt files and data, hardware failure, algorithmic (expected) errors.
3	Incorrect use of intervention as intended.	Effectiveness and suitability of intervention.	Delay or ineffective impact of patients.	Communication and validation of intervention. Insufficient documentation of use within the pathway. Poor adherence to guidelines. Inadequate supervision. Lack of algorithmic explain-ability.
4	Ineffective use of digital intervention.	Ineffective provision of healthcare services.	Delay or ineffective provision of service.	Unsuitable planning of use. Intervention suitability justification. Lack of buy-in from users. Wrong timeframe of evidence basis. Lack of co-production of intervention.

4. Discussion and Conclusion

The results, Table 2, highlight the increased clinical safety risk AI presents to Healthcare Delivery Organizations (HDO). The timeliness and accuracy of decision-making are two specific hazards highlighted by this analysis. By presenting the potential causes of such hazards, we can direct effort applied in risk mitigation to relevant sources with greater likelihood of success. Inadequate utilization of data, social acceptance or trust of AI technology influences clinical effectiveness and operational benefit to larger patient cohorts. Issues with algorithmic functions and data specific facets of bias, validation and revalidation have the potential to cause direct harm to a patient. There is risk in assuming increased accuracy/efficiency from AI decision-making and limited human involvement. A commissioning HDO has a direct relationship with effectiveness, safety, and clinical outcomes. By application of the method, we postulate that UK sources have highlighted

issues that may affect justification of safety of a DHI, Table 2. The complexity and dynamic nature of risk, together with increased HDO control of potential hazards, enhances the need for more effective methods for communicating safety justification. Addressing the causes of these issues, a body of evidence (sources) will support the clinical safety of the AI-based DHI. The correct use of DHIs adds greater burden on organizations, as unintended operation / use exposes risk beyond the immediate decision support function, and into future decision-making and algorithmic learning of the DHI. Operating AI-based DHIs as intended, correct implementation within the clinical/patient pathway and effective periodic review of clinical outcomes would enhance safety claims. Manufacturers must consider collaborative engagement with HDOs to establish proven in use safety, reliability, and efficacy claims. Simplifying the explanation of decision-making, diagnosis and foundation of operation may enable intelligent supervision to mitigate the clinical risk. The results align with current EU policy and industry body recommendations for the use of AI DHI, including medical devices [16].

References

[1] Royal Academy of Engineering. Establishing high-level evidence for the safety and efficacy of medical devices and systems. 2013;(January 2013):1-40.
[2] Rowley A. The emergence of artificial intelligence and machine learning algorithms in healthcare: Recommendations to support governance and regulation. BSI Gr. 2019;1-18.
[3] Glauner P. An Assessment of the AI Regulation Proposed by the European Commission. 2021;(May 2021).
[4] Shaw J. Artificial Intelligence and the Implementation Challenge. J Med Internet Res. 2019;21(7).
[5] Despotou G, Ryan M, Arvanitis TN, Rae AJ, White S, Kelly T, et al. A framework for synthesis of safety justification for digitally enabled healthcare services. Digit Heal. 2017;3(April):205520761770427.
[6] ISO. ISO - ISO 14971:2019 - Medical devices - Application of risk management to medical devices.
[7] UK Gov. Accelerating Artificial Intelligence in health and care: results from a state of the nation survey. 2018;1–64.
[8] Hall W, Pesenti J. Growing the Artificial Intelligence Industry in the UK. 2017;
[9] Thorpe GJ. The U.K.: ready, willing, and able. J Palliat Care. 1988;4(1–2):26–8.
[10] UK Gov. A guide to good practice for digital and data-driven health technologies. 2021.
[11] UK NHSx. A Buyer's Guide to AI in Health and Care.
[12] NHSX. Artificial Intelligence: How to get it right. 2019;(October):1–55.
[13] Harwich EE. Thinking on its own: AI in the NHS. Reform. 2018; Jan:1–60.
[14] Leslie D. Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector. SSRN Electron J. 2020;
[15] Challen R. Artificial intelligence, bias and clinical safety. BMJ. 2019.
[16] Gudeppu M. Medical device regulations. In: Trends in Development of Medical Devices. Elsevier Inc.; 2020. p. 135–52.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210848

An Evaluation of Pretrained BERT Models for Comparing Semantic Similarity Across Unstructured Clinical Trial Texts

Jessica PATRICOSKI[a], Kory KREIMEYER[b], Archana BALAN[b], Kent HARDART[b], Jessica TAO[b], the Johns Hopkins Molecular Tumor Board Investigators[b], Valsamo ANAGNOSTOU[b] and Taxiarchis BOTSIS[a,b,1]
[a] *Biomedical Informatics and Data Science Section, Johns Hopkins University School of Medicine, Baltimore, MD*
[b] *Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD*

Abstract. Processing unstructured clinical texts is often necessary to support certain tasks in biomedicine, such as matching patients to clinical trials. Among other methods, domain-specific language models have been built to utilize free-text information. This study evaluated the performance of Bidirectional Encoder Representations from Transformers (BERT) models in assessing the similarity between clinical trial texts. We compared an unstructured aggregated summary of clinical trials reviewed at the Johns Hopkins Molecular Tumor Board with the ClinicalTrials.gov records, focusing on the titles and eligibility criteria. Seven pre-trained BERT-Based models were used in our analysis. Of the six biomedical-domain-specific models, only SciBERT outperformed the original BERT model by accurately assigning higher similarity scores to matched than mismatched trials. This finding is promising and shows that BERT and, likely, other language models may support patient-trial matching.

Keywords. Clinical trial, word embeddings, bidirectional coder representations

1. Introduction

Clinical texts often contain unstructured information that requires applying advanced text processing methods to support specific tasks like patient-trial matching. Various language models have the potential to process biomedical and clinical texts to aid in these challenges. The Bidirectional Encoder Representations from Transformers (BERT) family has shown promise in solving multiple problems, including semantic similarity. BERT models have been trained on multiple corpora, including PubMed abstracts, PMC full-text articles, clinical notes, and synthetic vocabularies [1-6].

In this study, we report the performance of pre-trained BERT-based language models by assessing the level of similarity between clinical trial raw text (official titles and eligibility criteria) drawn from two sources: an Institutional unstructured aggregated summary of clinical trials and ClinicalTrials.gov.

[1] Taxiarchis Botsis, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, 1550 Orleans St, CRB2-Rm 153, Baltimore, MD 21287, USA; E-mail: tbotsis1@jhmi.edu.

2. Methods

We created an unstructured aggregated summary of clinical trials reviewed at the Johns Hopkins (JH) Molecular Tumor Board at the Sidney Kimmel Cancer Center. Our corpus included all clinical trials registered as of June 1st, 2021. We chose to focus on trials with no provided National Clinical Trial identifier (NCT ID), incorrectly formatted identifiers, or unrecognized identifiers, in order to capture the challenges of missing data. Figure 1 outlines a breakdown of the dataset and the final subset used to create a data frame of the official study titles and eligibility criteria (EC) for the 689 trials with non-null EC fields.

To create pairs of clinical trial data between the clinical trials in our registry and those registered with ClinicalTrials.gov, we downloaded a corpus of official study titles (and associated NCT IDs) for all clinical trials registered with ClinicalTrials.gov from the Clinical Trials Transformation Initiative's Aggregate Analysis database (AACT) [7]. After basic preprocessing, embeddings (i.e., real-valued vector representations for each word based on context) for each querying title (from our clinical trials) and the 387,486 corpus titles were created using Sentence-Transformers [6, 8].

For each querying title, the corpus title with the highest cosine similarity score was assigned as a potential match. After manually reviewing the potential matches, each pair was labeled as either belonging to the same trial (hereafter, a "match") or different trials (hereafter, a "mismatch"). Of the pairs created, 603 pairs were matches and 86 were mismatches. Finally, the eligibility criteria for each paired ClinicalTrails.gov trial were retrieved through ClinicalTrials.gov's Application Programming Interface URLs using the linked NCT IDs from the AACT database.

The first pre-trained model used to calculate the similarity between the created pairs was an uncased version of BERT-Base model [1]. We also used six more models from the BERT family with the same architecture as the Base model but domain-specific pre-training and fine-tuning using different corpora. Each of the BioBERT [2], BlueBERT [3], Clinical BioBERT [4], SciBERT [4], PubMedBERT [5], and CODER [6] models created embeddings for and checked the cosine similarity between pair titles and between pair EC. The difference in mean similarity scores for matched and mismatched pairs was compared across all models to assess performance. It is important to note that because treatment information in our aggregated summary was not separated from the EC, the EC field contained noise that might have affected similarity calculations.

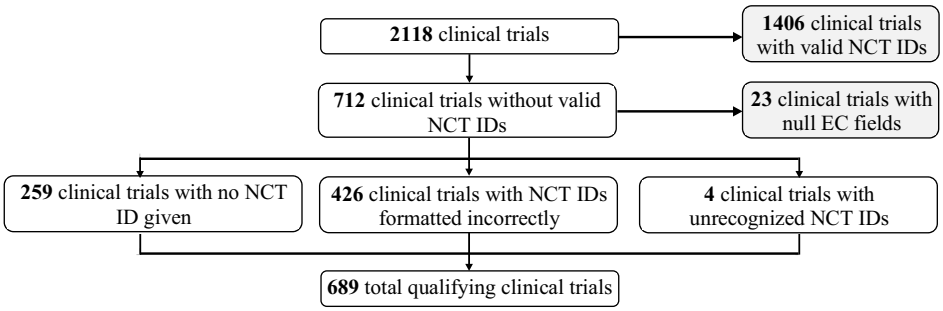


Figure 1. A visual breakdown of the clinical trial summaries. Grey-shaded boxes indicate excluded trials.

3. Results

Visual representations of the similarity scores calculated by the BERT-Base and other pre-trained models for titles and eligibility criteria are shown in Figure 2. SciBERT, while ranking lowest in mean similarity of matching pairs, demonstrated the largest overall difference in mean similarity between matched and mismatched pairs (0.141 for titles and 0.064 for EC) and assigned lower values for mismatched pairs than all other models. Thus, we consider that SciBERT had the highest performance for the given task. Outside of SciBERT, no other model outperformed BERT, which had an overall difference in mean similarity between matched and mismatched pairs of 0.102 for titles and 0.054 for EC. Table 1 includes two examples of matched and mismatched trial titles and the similarity scores for the seven BERT models. As expected, the mean similarity scores for trial EC were generally lower than those for title scores, likely due to the noise present in the EC field of our aggregated clinical trial summaries.

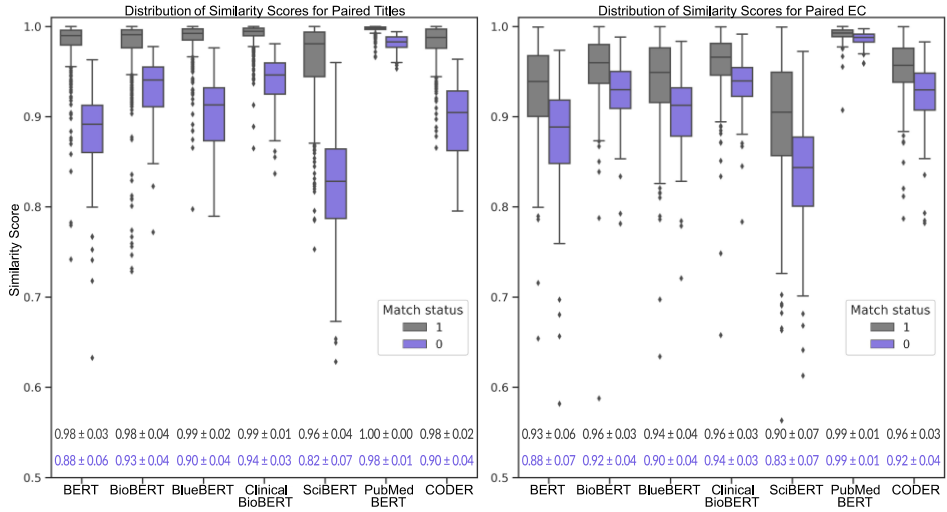


Figure 2. Boxplots showing the spread of similarity scores for paired clinical trial titles (left) and eligibility criteria (right), grouped by match status. A match status of 1 indicates the content assessed belonged to the same clinical trial, while a match status of 0 indicates the content belonged to separate trials. The mean similarity scores and standard deviations, also grouped by match status, are listed along the x-axis.

Table 1. Examples of compared trial titles and their similarity scores for matched and mismatched pairs.

Example Match (Title Similarity)		
"A Ph. II Study of the Efficacy and Safety of SU011248 in Patients with Metastatic Breast Cancer"	BERT: 0.986 BioBERT: 0.973 BlueBERT: 0.973 Clin. BioBERT: 0.974	SciBERT: 0.977 PubMedBERT: 0.997 CODER: 0.984
"A Phase 2 Study Of The Efficacy And Safety of SU011248 In Patients With Metastatic Breast Cancer"		
Example Mismatch (Title Similarity)		
"Donor Lymphocyte Infusions (DLI) plus Rapamycin to Decrease Toxicity Associated with DLI"	BERT: 0.858 BioBERT: 0.901 BlueBERT: 0.875 Clin. BioBERT: 0.916	SciBERT: 0.649 PubMedBERT: 0.957 CODER: 0.862
"Rapamycin in Relapsed Acute Lymphoblastic Leukemia"		

4. Discussion

Overall, SciBERT performed best in terms of distinguishing between sets of trial texts belonging to the same clinical trial and sets of trial text belonging to different clinical trials. The largest difference between SciBERT and most of the other models is its use of SciVocab, which overlaps with BaseVocab used in the BERT-Base by only 42% [9]. SciBERT was trained from scratch and did not use BERT's weights as initialization [4, 9]. After SciBERT, BERT had the best performance. Although trained with biomedical texts, the remaining models did not show promise in efficiently supporting our matching task.

A significant limitation of this study was that due to the inconsistent format of our corpus, treatment information could not be separated from the EC, resulting in an intangible level of interference in similarity assessment. However, the handling of noisy information is a standard challenge in free-text processing and comparison that was successfully handled by some of the selected models. The second major limitation is that the BERT architecture has a maximum token length of 512, and as a result, most EC texts were not compared in full. We acknowledge that this limitation may have introduced a bias to the similarity calculations but did apply to all models that used the same token length.

Future research is needed to address the above limitations and investigate other biomedical language model architectures, the impact of pre-training on clinical trials data, and the feasibility of integrating semantic similarity techniques for comparing clinical trial identities and best utilizing them into the patient matching process.

References

[1] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Burstein J, Doran C, Solorio T, editors. Proceedings of NAACL-HLT; 2019 Jun 2-7; Minneapolis, MN. Stroudsburg (PA): ACL; c2019;4171-4186.

[2] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 2020 Feb 15;36(4):1234-40.

[3] Peng Y, Yan S, Lu Z. Transfer learning in biomedical natural language processing: An evaluation of BERT and ELMo on ten benchmarking datasets. In: Demner-Fushman D, Cohen KB, Ananiadou S, Tsujii J, editors. Proceedings of the BioNLP workshop; 2019 Aug 1; Florence, IT. Stroudsburg (PA): ACL; c2019. p. 58-65.

[4] Khattak FK, Jeblee S, Pou-Prom C, Abdalla M, Meaney C, Rudzicz F. A survey of word embeddings for clinical text. Journal of Biomedical Informatics. 2019;100:100057.

[5] Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, Naumann T, Gao J, Poon H. Domain-specific language model pretraining for biomedical natural language processing. arXiv: 2007.15779v5 [Preprint]. 2020 [cited 2021 Aug 20]: [24 p.]. Available from: https://arxiv.org/abs/2007.15779.

[6] Yuan Z, Zhao Z, Sun H, Li J, Wang F, Yu S. CODER: Knowledge infused cross-lingual medical term embedding for term normalization. arXiv: 2011.02947 [Preprint]. 2017 [cited 2021 Aug 20]: [11 p.]. Available from: https://arxiv.org/pdf/2011.02947.pdf.

[7] Tasneem A, Aberle L, Ananth H, Chakraborty S, Chiswell K, McCourt BJ, Pietrobon R. The database for aggregate analysis of ClinicalTrials.gov (AACT) and subsequent regrouping by clinical specialty. PLoS One. 2012;7(3):e33677.

[8] Reimers J, Gurevych I. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In: Inui K, Jiang J, Ng V, Wan X, editors. Proceedings of the Conference on Empirical Methods in Natural Language Processing; 2019 Nov 3-7; Hong Kong, CN. Stroudsburg (PA): ACL; c2019;3892-3992.

[9] Beltagy I, Lo K, Cohan A. SciBERT: A pretrained language model for scientific text. In: Inui K, Jiang J, Ng V, Wan X, editors. Proceedings of the Conference on Empirical Methods in Natural Language Processing; 2019 Nov 3-7; Hong Kong, CN. Stroudsburg (PA): ACL; c2019;3615-33620.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210849

AI Enhanced Person-Centred Care Services for Monitoring Stroke Outpatient Rehabilitation

Ramo ŠENDELJ[a,1], Ivana OGNJANOVIĆ[a], Milovan ROGANOVIĆ[b] and Dalia KRIKŠČIŪNIENĖ[c]
[a] *University of Donja Gorica, Montenegro*
[b] *Department of Neurology, Clinical Centre of Montenegro*
[c] *Vilnius University, Lithuania*

Abstract. Development of person-centred care (PCC) services require adjustment to specific domain of application and integration with existing processes implemented in healthcare institution. This poster present PCC services for monitoring stroke outpatient rehabilitation, enhanced by modern ICT technologies (thus enabling adjustments to different kind of patients, which is especially relevant due to potential consequences of the stroke and caused degree of disability).

Keywords. person-centred care services, tool support, stroke rehabilitation

1. Introduction

The PCC as innovative approach in health care has attracted attention as a proven concept for improvement of health outcomes by advancing cooperation and shared decision making between doctors and patients, with simultaneous increase of patient satisfaction [4][1]. Development of PCC services for stroke outpatient rehabilitation follows the following specific characteristics [3]: it is a chronic condition and thus the whole cycle of recovery is needed to be covered, after hospital treatment special focus shall be put on rehabilitation and further prevention of recurrent stroke through increased quality of life, self-efficacy, etc. The developed services are elaborated over evidence from neurology department of Clinical Centre in Montenegro, within scientific research project which is funded by the Ministry of Science of Montenegro.

2. Methods

Implemented methodology includes step-wised approach, consisting of several major steps: (i) analysis of best practices related to development of PCC services, (ii) analysis of the context for applications at Neurology department of the Clinical Centre of Montenegro [1], (iii) creation of the framework covering business processes and PCC concept, built upon well known concepts of model-driven engineering and software

[1] Corresponding author, Ramo Šendelj, University of Donja Gorica, Oktoih 1, 81000 Podgorica, Montenegro; E-mail: ramo.sendelj@{gmail.com, udg.edu.me}.

product families [2], and (iv) prototyping and testing of specific IT services (in the form of mobile applications for patients and doctors). This approach ensures replicability to other domains of applications, thus being fully in line with strategical orientation for development of the Centre for PCC at Clinical Centre of Montenegro.

3. Results

Developed PCC services support identified features in outpatient rehabilitation business process model developed in [2]. It is implemented in the form of mobile application for patients (connected with external devices, like smart phones, etc.) and web/mobile application for doctors supporting detailed view on rehabilitation process of patients, definition of corrective measures (highly relevant for exercises for motoric rehabilitation, pain reduction medications, etc.) (currently available on Montenegrin language) (Fig.1., Fig.2.). Specific module for prediction of the rehabilitation process, as well as for prediction of recurrent strokes for each specific patient is developed. The predictive system is developed by using existing databases and stroke patient registries [1][2], while it is expected to increase the accuracy by exploring data collected by active system usage.

Figure 1. PCC service for stroke outpatient rehabilitation: Module for doctors

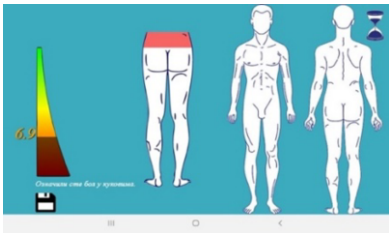


Figure 2. PCC service for stroke outpatient rehabilitation: Module for patients

4. Discussion

The developed services present initial step in creation of PCC healthcare services in Clinical Centre of Montenegro, which implementation has already shown several issues and challenges: necessity for (i) raising awareness among patients and development of partnership with patients, (ii) significant increase of IT health literacy, as well as (iii) holistic support during implementation process, (iv) development of tailored evaluation framework, etc. However, the established close cooperation between patients and doctors during the stroke rehabilitation process ensures greater confidence and a sense of security which directly increase the quality of life and healthcare prevention. Future work will be focused on making comprehensive evaluation and further tool support development.

References

[1] Ognjanović I, Šendelj R, Mantas J, Roganović M. Development of ICT Enhanced Person-Centred Care Services for Stroke Outpatient Rehabilitation. In: Proceedings of the 2021 10th Mediterranean Conference on Embedded Computing (MECO), 2021, p. 1-4.

[2] Ognjanović I, Lewandowski R, Šendelj R, Krikščiunien D, Eraković J. Model Driven Approach for Development of Person-Centred Care in Stroke Rehabilitation. In: J. Mantas et al., editors. The Importance of Health Informatics in Public Health during a Pandemic 272: IOSPress; 2020.
[3] Hubbard IJ, et al. Adherence to Clinical Guidelines Improves Patient Outcomes in Australian Audit of Stroke Rehabilitation Practice. Archives of Physical Medicine and Rehabilitation. 2012; 93: 965-71.
[4] Ekman N, et al. A state-of-the-art review of direct observation tools for assessing competency in person-centred care. International Journal of Nursing Studies. 2020; 109: 103634.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210850

FAIR and Quality Assured Data -The Use Case of Trueness

Jürgen STAUSBERG[a,1], Sonja HARKENER[a], Ekkehart JENETZKY[b,c], Patrick JERSCH[d], David MARTIN[b,e], Rüdiger RUPP[d] and Martin SCHÖNTHALER[f]
[a] *University Duisburg-Essen, Faculty of Medicine, Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), Essen, Germany*
[b] *Faculty of Health/School of Medicine, Witten/Herdecke University, Witten, Germany*
[c] *Department of Child and Adolescent Psychiatry and Psychotherapy, University Medical Center of the Johannes-Gutenberg-University, Mainz, Germany*
[d] *Heidelberg University Hospital - Spinal Cord Injury Center, Heidelberg, Germany*
[e] *Department of Pediatrics, Eberhard-Karls University Tübingen, Tübingen, Germany*
[f] *Department of Urology, Medical Centre - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany*

Abstract. The FAIR Guiding Principles do not address the quality of data and metadata. Therefore, data collections could be FAIR but useless. In a funding initiative of registries for health services research, trueness of data received special attention. Completeness in the definition of recall was selected to represent this dimension in a cross-registry benchmarking. The first analyses of completeness revealed a diversity of its implementation. No registry was able to present results exactly as requested in a guideline on data quality. Two registries switched to a source data verification as alternative, the three others downsized to the dimension integrity. The experiences underline that the achievement of appropriate data quality is a matter of costs and resources, whereas the current Guiding Principles quote for a transparent culture regarding data and metadata. We propose the extension to FAIR-Q, data collections should not only be findable, accessible, interoperable, and reusable, but also quality assured.

Keywords. Completeness, data quality, health services research, registries, validity.

1. Introduction

The FAIR Guiding Principles for scientific data management and stewardship [1] had been published in view of the scientific use of already recorded data. Data collections should be findable, accessible, interoperable, and reusable. Beside legal constraints, achieving the FAIR Guiding Principles is mainly a matter of provision and culture. The principles especially demand the willingness to share detailed information about a data collection with the public, a demand that might jeopardize economic and scientific gains of the primary data holder. However, even if a data collection fulfills the FAIR Guiding Principles, the data collection could be useless, because it does not offer a quality of data

[1] Corresponding Author, Institute for Medical Informatics, Biometry and Epidemiology, Faculty of Medicine, University Duisburg-Essen, Hufelandstrasse 55, 45122 Essen, Germany; E-mail: stausberg@ekmed.de.

needed to support the use case at hand. Furthermore, the re-use of data with inappropriate data quality can establish a hazard. In electronic health records, a wrong lab value such as an erroneously low serum potassium concentration - caused by a mix-up of patient identities - can lead to harm via inappropriate therapy. A lack in recording of adverse events in clinical trials could mask the risk of a drug therapy that is then misleadingly recommended in a clinical guideline. In the following, we will report experiences regarding the measurement of trueness from a funding initiative of registries for health services research.

2. Methods

2.1. Quality Indicators Related to Trueness

The guideline on data quality in medical research [2] organized indicators within three dimensions derived from Donabedian's approach: integrity, organization, and trueness. High quality in the first two dimensions is a necessary, but not a sufficient precondition to achieve a high quality in the dimension of trueness. The guideline defined six indicators for trueness: agreement with source data referring to data elements (TMF-1044), agreement with source data referring to observational units (TMF-1045), compliance with operating procedures (TMF-1047), representativeness (TMF-1048), accuracy (TMF-1043), and completeness (TMF-1046). Two indicators (1044, 1045) correspond with a source data verification well established in clinical trials. The assessment of the compliance with operating procedures requires the existence of such standards. Representativeness checks the agreement of frequencies and distributions from data with expectations, e.g. published in the literature. Accuracy and completeness capture precision and recall: did an event recorded in the data occur in reality (precision), was a real event recorded in the data (completeness)? Many proposals were made to operationalize those aspects also denoted as validity [3].

2.2. Sample

In a funding initiative, five indicators from the guideline [4] were selected for a benchmarking of registries. Trueness was represented by indicator completeness (TMF-1046). According to Arts et al. [5], completeness was defined as "the extent to which all necessary data that could have been registered have actually been registered". The calculation method of the indicator was adapted to each registry. This adaptation became part of the standard operating procedure (SOP) for the benchmarking of the registries. For example, the SOP listed the particular data elements considered for the measurement of completeness as well as the individual approach used to identify the clinical events that "could have been registered" in order to calculate the individual denominator. The quality target was adapted from the guideline as a completeness of 95% or higher.

The benchmarking of trueness for the year 2020 comprised five registries dealing with 1) fever in childhood, 2) lifelong monitoring of patients with spinal cord injury, 3) treatment exit options for uveitis, 4) hereditary breast and ovarian cancer, and 5) safety of living kidney donors. The indicator completeness was calculated by each registry itself accordingly to the SOP. An accompanying project was responsible for the receipt of the results and the preparation of the benchmarking report. The responsibility of the accompanying project did not include any monitoring or data management.

3. Results

The five registries delivered their results for the indicator completeness until February 2021. The determination of the rate's numerator and denominator was up to the registries. In comparison with the guideline's definition, the deliverables were quite diverse.

- Registry A reported a completeness of 112% for pre-existing-illnesses consulting medical reports.
- Registry B compared the relatively frequencies of three subtypes from the qualifying disease with expectations from two observational studies. With one exception, the 95% confidence limits included the literature results.
- Registry C counted adverse events that lead to a re-admission. The result was compared with secondary data. The calculated standardized ratio was 3.5.
- Related to the observational unit, but not controlled for the single event, registry D reported a completeness of 53% (contacts), 62% (drugs) and 87% (symptom).
- Registry E compared the recorded calendar date of an event with the calendar date of the event available in the patient chart. It found no discrepancies.

Based on these deliverables, the accompanying project did not prepare a benchmarking report for this indicator as specified in the SOP. Instead, the results were described and commented in the report without a comparison between the registries.

4. Discussion

Two out of the five registries achieved the intended data quality with a completeness of 95% or higher. However, none of the registries was able to calculate the indicator as specified in the guideline. The measurement of trueness seemed to be a big challenge in the monitoring of registries' data quality! Taking into account the whole set of indicators defined in the guideline [6], the registries made use of some other measures.

- Registry A performed a source data verification as specified in TMF-1044.
- Registry B assessed the evidence of known correlations (TMF-1027).
- Registry C also assessed the evidence of known correlations (TMF-1027).
- Registry D analyzed the concordance between two different data collections covering the same observational units (indicator TMF-1002).
- Registry E performed a source data verification as specified in TMF-1044.

Only two registries switched to another indicator from the dimension trueness, three fell back to the dimension integrity. Indicators from that dimension could be calculated without considering the origin and the context of data thus making data management substantially easier and cheaper. Considering external comparators is at least time- and labor-consuming. Furthermore, the establishment of a realistic gold standard could be expensive by creating organizational workflows in health care, solely related to this task.

Registry-based research is confronted with serious challenges. A main concern is related to a not-verified validity of important clinical events. The judgement of clinical outcomes is up to directly involved physicians [7], patients [8] or carers. In clinical trials, regulatory bodies recommend the establishment of a data monitoring committee [9], among others responsible for the verification of serious adverse events or outcomes. The recording of clinical events in daily practice is biased by the workload as well as the motivation of the responsible staff.

Precisely because the standards of clinical trials are out of the scope of observational research, data management of registries must have a clear understanding of alternative options to assure trueness. Medicine will not benefit from data collections fulfilling the FAIR Guiding Principles on the one hand but delivering an unsubstantiated data quality on the other hand. Therefore, we call for an extension of FAIR to FAIR-Q. In contrary to the already included requirements, to provide **Q**uality assured data is a matter of resources, efforts and procedures. There must be resources for efforts implementing procedures of a quality oriented data management. The presented use case of trueness demonstrated that there is still a long way to achieve this goal. In analogy to the original proposal [1], "**Q**" can be characterized by requirements for data collections:

Q1. metadata are annotated regarding quality management procedures
Q2. indicators are available about the quality of (meta)data
Q2.1 a statement on the trueness of data is present
Q3. (meta)data are released with evidence about their potential impact

In our initiative, the registries could start with a weak-point analysis and subsequently with a redesign of their processes in case of insufficient results. The accompanying project will readjust the representation of the trueness to improve the cross-registry benchmarking. Two preferred options exist, to replace the indicator completeness with another indicator of that dimension or to rethink its specification for each registry. The calculation of the indicator by the registries themselves is a limitation of the benchmarking. So far, reliability and validity of the results remain open.

Acknowledgements

The work was funded by the German Federal Ministry of Education and Research within the "Model Registers Funding Scheme" (FeverApp 01GY1905, ParaReg 01GY1904, RECUR 01GY1902, REGISVF-APB 01GY1917B).

References

[1] Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018.
[2] Nonnemacher M, Weiland D, Neuhäuser M, Stausberg J. Adaptive management of data quality in cohort studies and registers: proposal for a guideline. Acta Informatica Medica. 2007;15:225-30.
[3] Stausberg J, Kuklik N, Jöckel KH. Opportunities and Pitfalls in the Definition of Data Validity. Stud Health Technol Inform. 2018;247:566-70.
[4] Stausberg J, Bauer U, Nasseh D, et al. Indicators of data quality: review and requirements from the perspective of networked medical research. GMS Med Inform Biom Epidemiol. 2019;15:Doc05.
[5] Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc. 2002;9:600-11.
[6] Harkener S, Stausberg J, Hagel C, Siddiqui R. Towards a core set of indicators for data quality of registries. Stud Health Technol Inform. 2019;267:39-45.
[7] Roovers J-PWR. Registries: what level of evidence do they provide? Int Urogynecol J. 2007;18:1119–20.
[8] Martin D, Wachtmeister J, Ludwigs K, Jenetzky E. The FeverApp registry - ecological momentary assessment (EMA) of fever management in families regarding conformity to up-to-date recommendations. BMC Med Inform Decis Mak. 2020;20:249.
[9] U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Clinical Trial Sponsors. Establishment and Operation of Clinical Trial Data Monitoring Committees. March 2006. https://www.fda.gov/media/75398/download (access 2021-06-21).

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210851

Using Machine Learning to Improve Personalised Prediction: A Data-Driven Approach to Segment and Stratify Populations for Healthcare

Will YUILL[a,b,1] and Holger KUNZ[a,2]
[a] *Institute of Health Informatics, University College London, UK*
[b] *Hertfordshire County Council, UK*

Abstract. Population Health Management typically relies on subjective decisions to segment and stratify populations. This study combines unsupervised clustering for segmentation and supervised classification, personalised to clusters, for stratification. An increase in cluster homogeneity, sensitivity and positive predictive value was observed compared to an unlinked approach. This analysis demonstrates the potential for a cluster-then-predict methodology to improve and personalise decisions in healthcare systems.

Keywords. Population Health Management, segmentation, stratification, clustering, machine learning

1. Introduction

Population Health Management (PHM) is increasingly being adopted in England to improve outcomes for individuals by personalising services to address their health and care needs in a way that recognises that health is determined more by socio-economic factors than by healthcare provision itself [1]. However, personalisation of interventions on an individual level is not feasible because of resource constraints [2]. Instead, PHM proposes designing systems around defined segments of the population, further targeting individuals through stratification by risk of an adverse event, such as hospital readmission or onset of disease. Segmentation and stratification can improve patient care and management and inform the design of care systems [3,4]. The foundations of segmentation and stratification in healthcare have already been defined by Garfield [5], but the increasing availability of data provides opportunity for new methods to be used.

There are two broad approaches to segmentation [6]. The traditional approach is to use *a priori* groups based on expert knowledge, for instance segmenting patients by morbidity, age, or disease. This approach can have limitations:

- The number of features by which to segment must remain small because the number of groups can expand rapidly, even exponentially in some cases.

[1] Corresponding Author, Will Yuill, Institute of Health Informatics, University College London, London, UK, & Hertfordshire County Council, UK, E-mail: will.yuill@hertfordshire.gov.uk
[2] Corresponding Author, Holger Kunz, Institute of Health Informatics, University College London, London, UK, E-mail: h.kunz@ucl.ac.uk

- With few features, important differentiators between people are missed.
- Variables such as age require arbitrary breakpoints to be defined.
- With few segments and features, entropy is likely to be high.

The second, more data-driven approach has been adopted recently and addresses many of the limitations of *a priori* methods, allowing more variables to be used, and for feature breakpoints to be derived naturally [6,7]. However, this approach requires a large amount of data and may result in less readily comprehensible clusters. For example, while *a priori* segmentation allows precise classification of segments [2], data-driven methods can result in more nuanced groupings, requiring interpretation [6]. Therefore, the purpose of the segmentation is likely to influence the method chosen.

Segmentation can often group patients of similar types but differing magnitudes of need. This is where stratification can assist. Typically, stratification is performed with ordinary regression models (developed on a wholly different population) and applied equally to every segment. However, by using segmentation, we postulate there are meaningful differences between groups and therefore predictors of risk may differ. Not accounting for this could reduce the performance of stratification models [8,9].

This study therefore explored the potential for data-driven segmentation, coupled with stratification personalised to each segment, to achieve better performance than existing models when predicting the risk of emergency readmission within 30 days of discharge. This cluster-then-predict method has been effective in other settings [10].

2. Methodology

The study population comprised 78,786 admission episodes in the NHS Secondary Uses Service dataset for patients registered with a single Clinical Commissioning Group in England in the fiscal year 2020–2021. For each admission the following data was extracted with exclusions applied in accordance with Billings et al. [4]:

- Person: age, deprivation assessed via Index of Multiple Deprivation 2019 [11].
- Health: morbidities in inpatient and outpatient records in the three years prior to admission classified in the Charlson Comorbidity Index [12].
- Care: NHS Provider Trust, count of emergency admissions in the year prior, emergency admission in the past 30 days, whether the current admission was an emergency, emergency readmission within 30 days (target variable).

Three methods were implemented using the methodology summarised in Figure 1.

2.1. Model 1: a priori Segmentation and Traditional Stratification

Morbidities were assigned a score [12] and these were summed to split episodes into groups ‘0’, ‘1-2’, ‘3-4’, ‘5+’. The PARR-30 algorithm [4] was applied to calculate readmission risk. The cost of a false negative was set at three times the cost of a false positive and used to select a threshold to predict readmission [13]. Homogeneity of clusters was assessed through Silhouette scores performed on a 30,000-record subset.

Figure 1. Flowchart of methods used to cluster-then-predict using both supervised and unsupervised learning.

2.2. Model 2: Traditional Segmentation and Personalised Stratification

Patients were segmented as in Model 1, but a generalised logistic mixed model (GLMM) defined in Eq. (1) was used to calculate readmission risk with the threshold optimised as in Model 1. Rather than fitting regressions to each segment, GLMMs allow for pooling between segments, mitigating small groups, and allowing greater scaling. Intercept and slope were allowed to vary by segment. Only feature used in Model 1 were included to facilitate comparison. Providers with fewer than 1,000 instances were recoded as 'other'.

$$Readmission = f(Age, CurrentEmergAdmi, EmergAdmiLast30d, EmergAdmiLast1yr, CHF, PVD, CPD, ChronicDM, Renal, SolidCancer, OtherCancer, MildLiver, Dementia, ModOrSevereLiver, HemiParaPlegia, ProviderTrust, IndexMultipleDeprivationScore) \quad (1)$$

2.3. Model 3: Data-Driven Segmentation and Personalised Stratification

To investigate whether discovering natural clusters in the data would result in improved predictions, unsupervised clustering was undertaken using k-prototypes in order for both binary and continuous variables to be used [14]. Features were limited to those in PARR-30, excluding NHS Provider Trust because of its high cardinality. Continuous features were scaled and centered. The number of clusters was set to maximise Silhouette score and a GLMM as defined in Model 2 was fitted to the result.

3. Results

Table 1 summarises the performance for each model. Performance was similar across accuracy, area under the curve (AUC) and specificity. Small but significant improvements were seen in positive predictive value (PPV) when comparing Models 1 and 2, with a further significant improvement in sensitivity when comparing Models 1 and 3. Performance between segments varied, for instance PPV in Model 3 ranged from 0.19–0.44. The clustering of Models 1 and 2 was poor with a mean Silhouette score of -0.12 compared to 0.18 in Model 3, which found five clusters to be optimum.

Table 1. The performance of each model with 95% confidence intervals in brackets.

Model	Accuracy	AUC	Sensitivity	Specificity	PPV
1	0.88 (0.88-0.88)	0.73 (0.72-0.74)	0.20 (0.19-0.21)	0.95 (0.95-0.95)	0.30 (0.29-0.32)
2	0.89 (0.88-0.89)	0.75 (0.74-0.76)	0.22 (0.21-0.23)	0.96 (0.95-0.96)	0.34 (0.33-0.35)
3	0.88 (0.88-0.89)	0.75 (0.75-0.76)	0.25 (0.24-0.26)	0.95 (0.95-0.95)	0.34 (0.34-0.35)

4. Discussion

This study implemented three models to improve PHM segmentation and stratification through a cluster-then-predict methodology. Personalising risk stratification as in Model 2 resulted in small but significant improvements to predictive performance over Model 1, suggesting that integrating segmentation and stratification approaches can improve understanding of patient risk by personalising prediction to segments. The variance between the performance of different segments remained, suggesting that the data used

may not provide sufficient information to accurately predict risk or that segments were not sufficiently homogenous, as indicated by Silhouette score. The creation of more homogenous clusters in Model 3 provided further small but significant improvements to stratification, suggesting that increases in homogeneity result in better predictions, even if these increases are small and homogeneity remains poor.

These results suggest that, with further optimisation, both more homogenous clustering methods and personalisation of stratification can provide more accurate risk prediction than traditional techniques. Given the similarity in performance between Models 2 and 3, a sufficiently homogenous *a priori* segmentation could provide both well-defined segments and improved risk prediction and, where the understandability of segments is important, this may be preferred. There is clear evidence that the addition of primary care and other datasets can result in both more homogenous segments (be this *a priori* or data-driven) and accurate classification [3,7], and future work could include this data where feasible. Further improvements may also be possible with methods that bind segmentation and stratification more closely together, such as ToPs/R [15]. The development of these methods offers to inform the development care that better meet the diverse needs of the population. Personalised medicine could be supported with the identification of appropriate clusters and to address the specific needs of each group.

References

[1] Hood CM, Gennuso KP, Swain GR, Catlin BB. County health rankings: Relationships between determinant factors and health outcomes. Am J Prev Med. 2016 Feb 1;50(2):129–35.
[2] Lynn J, Straube BM, Bell KM, Jencks SF, Kambic RT. Using population segmentation to provide better health care for all: The "Bridges to Health" model. Milbank Q. 2007 Jun;85(2):185–208.
[3] Vuik SI, Mayer E, Darzi A. Enhancing risk stratification for use in integrated care: a cluster analysis of high-risk patients in a retrospective cohort study. BMJ Open. 2016 Dec 1;6(12):e012903.
[4] Billings J, Blunt I, Steventon A, Georghiou T, Lewis G, Bardsley M. Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). BMJ Open. 2012 Aug 10;2(4):e001667.
[5] Garfield SR. The delivery of medical care. Sci Am. 1970 Apr;222(4):15–23.
[6] Nnoaham KE, Cann KF. Can cluster analyses of linked healthcare data identify unique population segments in a general practice-registered population? BMC Public Health. 2020 May 27;20(798):1–10.
[7] Vuik S. On the application of data-driven population segmentation to design patient-centred integrated care. Imperial College London; 2017. 189 p.
[8] Li Y, Sperrin M, Belmonte M, Pate A, Ashcroft DM, van Staa TP. Do population-level risk prediction models that use routinely collected health data reliably predict individual risks? Sci Rep. 2019 Aug 2;9(1):11222.
[9] Xia E, Du X, Mei J, Sun W, Tong S, Kang Z et al. Outcome-driven clustering of acute coronary syndrome patients using multi-task neural network with attention. In: Studies in Health Technology and Informatics. Amsterdam: IOS Press; 2019. p. 457–61.
[10] Soni R, Mathai KJ. Improved Twitter Sentiment Prediction through Cluster-then-Predict Model. Int J Comput Sci Netw. 2015 Aug;4(4):559–63.
[11] Ministry of Housing, Communities & Local Government. English indices of deprivation [Internet]. 2019. Available from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019 [Accessed 01 Sep 2021].
[12] Gasparini A. comorbidity: An R package for computing comorbidity scores. J Open Source Software. 2018;3(23):648.
[13] Billings J, Dixon J, Mijanovich T, Wennberg D. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. BMJ. 2006 Aug 12;333(7563):327–30.
[14] Szepannek G. clustMixType: User-friendly clustering of mixed-type data in R. R J. 2019;10(2):200–8.
[15] Yoon J, Zame WR, Banerjee A, Cadeiras M, Alaa AM, van der Schaar M. Personalized survival predictions via Trees of Predictors: An application to cardiac transplantation. PLoS One. 2018 Mar 28;13(3):e0194985.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210852

Prediction of Postoperative Speech Dysfunction Based on Cortico-Cortical Evoked Potentials and Machine Learning

Timur ISHANKULOV[a,1], Gleb DANILOV[a], David PITSKHELAURI[a], Oleg TITOV[a], Anna OGURTSOVA[a], Svetlana BUKLINA[a], Evgeniy GULAEV[b], Tatyana KONAKOVA[a] and Andrey BYKANOV[a]
[a] *Laboratory of Biomedical Informatics and Artificial Intelligence, National Medical Research Center for Neurosurgery named after N.N. Burdenko, Moscow, Russian Federation*
[b] *National Medical Research Center of Traumatology and Orthopedics named after N.N. Priorov, Moscow, Russian Federation*

Abstract. The possibility of postoperative speech dysfunction prediction in neurosurgery based on intraoperative cortico-cortical evoked potentials (CCEP) might provide a new basis to refine the criteria for the extent of intracerebral tumor resection and preserve patients' quality of life. In this study, we aimed to test the quality of predicting postoperative speech dysfunction with machine learning based on the initial intraoperative CCEP before tumor removal. CCEP data were reported for 26 patients. We used several machine learning models to predict speech deterioration following neurosurgery: a random forest of decision trees, logistic regression, support vector machine with different types of the kernel (linear, radial, and polynomial). The best result with F1-score = 0.638 was obtained by a support vector machine with a polynomial kernel. Most models showed low specificity and high sensitivity (reached 0.993 for the best model). Our pilot study demonstrated the insufficient quality of speech dysfunction prediction by solely intraoperative CCEP recorded before glial tumor resection, grounding our further research of CCEP postresectional dynamics.

Keywords. CCEP, cortico-cortical evoked potentials, machine learning, artificial intelligence, neuro-oncology, glial tumors, speech function

1. Introduction

Structural and functional neural networks underlying such human brain functions as speech are a permanent research subject for modern brain connectomics [1]. Intraoperative preservation of speech function is one of the most important goals in neurosurgery of intracerebral tumors located near eloquent areas [2]. The monitoring of the effective connections through language pathways during brain tumor surgery can be achieved by recording cortico-cortical evoked potentials (CCEPs) [3–5]. Nowadays the number of such studies is very limited.

[1] Corresponding Author, Timur Ishankulov, N.N. Burdenko Neurosurgery Center, 4th Tverskaya-Yamskaya str. 16, Moscow 125047, Russian Federation; E-mail: tishankulov@nsi.ru.

This pilot study aimed to predict speech function deterioration in the early postoperative period based on intraoperative cortico-cortical evoked potentials [6] recorded before glial tumor removal. The hypothesis we tested was that the baseline CCEPs might contain predictors of postoperative speech disorders. The exploration of CCEPs patterns might contribute to refine the criteria for the extent of intracerebral tumor resection and preserve patients' quality of life.

2. Methods

Intraoperative registration of CCEPs [7] was performed using a 32-channel intraoperative monitoring system "Neuro-IOM" (Neurosoft LLC, Russia) and a pair of subdural electrode strips. One electrode was placed in the frontal speech region (Broca's area); the second electrode was located on the surface of the upper temporal gyrus in its posterior parts and the supramarginal gyrus. The CCEPs were registered before and after resection of the tumor.

The dataset obtained contained recordings of 26 patients with brain gliomas in eloquent areas. The number of CCEP recordings for each patient was not set in this pilot study (varied from 1 to 8). The dataset included a set of files (n = 105, 1 file for each recording) with intraoperative CCEPs records before tumor removal. Each record contained 8 or 16 signal channels with high correlation between them.

The duration of the signal recording after stimulation was 300 ms. Each signal record included 7,500 discrete values in 8 or 16 channels. A vector of 7500 values was averaged and smoothed by the moving average method, transforming into a new vector consisting of 300 average values. Stimulation artifacts were automatically removed by comparing with the amplitude of the remaining signal part multiplied by 1.25. If this value was exceeded, then the starting index was shifted up to 10 ms to the right. In addition, the starting index of the signal was always shifted by 1 ms, even if the artifact was not observed, in order to exclude the influence of the first millisecond of the signal.

The basic set of signal features included signal amplitude, wave type, latency up to a peak (positive or negative) value [7]. Neurophysiologists typically used them to describe the CCEPs records. A medical expert indicated the characteristics of speech dysfunctions before and after surgery for each patient.

The average value across the entire signal was calculated and used as an additional feature. The peak values (local extremums) were calculated with a minimum distance between the peaks equal to 20 ms and a minimum peak height of 5 μV.

We formed the target variable based on the changes in the cumulative assessment of the patient's speech dysfunctions after surgery ranging from 0 to 45 (0 is the norm). The binary target variable took a value of 1 if the speech dysfunctions estimate increased after surgery (speech worsened) and a value of 0 otherwise (speech preserved).

Several machine learning models were used to predict the deterioration of speech functions in the postoperative period: a random forest of decision trees (RF), logistic regression (LR), support vector machine (SVM) with different types of kernel – linear (Lin), radial basis function (RBF) and polynomial (Poly). Each test was performed after the data were randomly sampled into training (80%) and testing (20%) subsets with stratification. The model was trained on a training subset; 5-fold cross-validation (CV) was applied to evaluate the model's quality before the final testing. Each machine learning model was tested 300 times with stratified resampling (1500 tests in total). This approach allowed to do the calculations with low margin of error (<0.005).

We used standard metrics to evaluate the test results: accuracy on validation samples within the cross-validation (CV), specificity (Spec), sensitivity (Sens), the proportion of correct classifier responses (Acc), precision (Prec), recall (Rec), F1-score (F1) and the area under the receiver operating characteristic curve (AUC). The results for particular machine learning model were averaged across all metrics to exclude the influence of the "by chance" data split and to reduce the margin of error. We separated data by patients — all the recordings of each patient were included into the only one subset: train or test.

3. Results

The results of our classification experiments are presented in Table 1.

Table 1. Classification of CCEPs data by binary outcome with 5 machine learning models.

Model	CV	Spec	Sens	Prec	Rec	Acc	F1	AUC
RF	0.680	0.319	0.809	0.569	0.564	0.606	0.530	0.564
LR	0.687	0.168	0.965	0.555	0.566	0.649	0.500	0.566
SVM (Lin)	0.674	0.098	0.944	0.411	0.521	0.612	0.432	0.521
SVM (RBF)	0.730	0.324	0.973	0.649	0.649	0.716	0.604	0.649
SVM (Poly)	**0.747**	**0.370**	**0.993**	**0.683**	**0.681**	**0.747**	**0.638**	**0.681**

The results of cross-validation were expectedly higher (equal in case of SVM (Poly) model) compared to the accuracy of test results. The difference between mentioned metrics varied from 0 to 0.074 (up to 12% decrease).

The best result for the F1-score metric was 0.638 using the SVM (Poly) model. A high sensitivity index was observed in most tests, reaching 0.993 in the best model. The specificity of the best solution was 0.370 — the model correctly identified only 37% of patients with improved/preserved speech functions.

4. Discussion

Our pilot research considered methods for predicting the deterioration or improvement/preservation of speech functions in the postoperative period using machine learning algorithms. This is the pioneering study to apply machine learning for predicting speech dysfunctions based on CCEP data to the best of our literature knowledge.

Researchers rely on such parameters as the amplitude of value fluctuation and the latency to the signal peak in the analysis of CCEP data (7–11). We utilized the average for all signal values, the latency to the signals' peak states (local extremums), and their values in μV in addition to common parameters.

The obvious limitations of this study are a relatively small number of patients (n = 26), several recordings per patient in one sample and the insufficient number of CCEP recordings after tumor removal. Increasing the amount of data may lead to a higher classification quality.

In our classification approach, we used a binary target variable. Thus, the dataset was split with a smaller possible imbalance compared to using the target variable broke down into several categories according to speech disorders degree (in the latter case, there was a significant imbalance between classes). It will be possible to test CCEPs classification by speech dysfunctions severity with the increased number of patients.

This pilot study demonstrated the insufficient quality of speech dysfunction prediction by solely intraoperative CCEP recorded before glial tumor resection. Our future work will be related to testing new methods for predicting speech disorders, focusing on postresectional CCEP dynamics, adding new features to existing models, and developing new machine learning models, including ensembles.

5. Conclusion

In this pilot study, the quality of speech dysfunction prediction after the neurosurgical interventions in the eloquent area was demonstrated using traditional machine learning methods based on the CCEP data registered before the main stage of surgery. Early detection of the speech dysfunction precursors, according to the CCEP data, can significantly affect the results of such neurosurgery. Thus, it is necessary to continue the research of CCEPs that contributes to a better understanding of speech dysfunctions resulting from surgical interventions and greater surgery safety.

The research was supported by the Russian Foundation for Basic Research grant 19-29-01231.

References

[1] Sporns O. The human connectome: A complex network. Annals of the New York Academy of Sciences. 2011:109–25.
[2] Yamao Y, Matsumoto R, Kikuchi T, Yoshida K, Kunieda T, Miyamoto S. Intraoperative Brain Mapping by Cortico-Cortical Evoked Potential. Vol. 15, Frontiers in Human Neuroscience. 2021.
[3] Leisman G, Moustafa A, Shafir T. Thinking, Walking, Talking: Integratory Motor and Cognitive Brain Function. Front Public Heal. 2016;4.
[4] Kunieda T, Yamao Y, Kikuchi T, Matsumoto R. New approach for exploring cerebral functional connectivity: Review of cortico-cortical evoked potential. Neurologia Medico-Chirurgica. 2015;55:374–82.
[5] Huang C-C, Chang C-S, Hsin Y-L. Time–frequency spectral analysis of cortico-cortical evoked potentials by means of Hilbert–Huang transform. Brain Stimul. 2015;8(2):388.
[6] Matsumoto R, Nair DR, LaPresto E, Bingaman W, Shibasaki H, Lüders HO. Functional connectivity in human cortical motor system: A cortico-cortical evoked potential study. Brain. 2007;130(1):181–97.
[7] Bykanov AE, Pitskhelauri DI, Titov OY, Lin MC, Gulaev E V., Ogurtsova AA, et al. Broca's area intraoperative mapping with cortico-cortical evoked potentials. Zhurnal Vopr Nejrokhirurgii Im NN Burdenko. 2020;84(6):49–56.
[8] Conner CR, Ellmore TM, DiSano MA, Pieters TA, Potter AW, Tandon N. Anatomic and electro-physiologic connectivity of the language system: A combined DTI-CCEP study. Comput Biol Med. 2011;41(12):1100–9.
[9] Silverstein BH, Asano E, Sugiura A, Sonoda M, Lee MH, Jeong JW. Dynamic tractography: Integrating cortico-cortical evoked potentials and diffusion imaging. Neuroimage. 2020;215.
[10] Saito T, Tamura M, Muragaki Y, Maruyama T, Kubota Y, Fukuchi S, et al. Intraoperative cortico-cortical evoked potentials for the evaluation of language function during brain tumor resection: Initial experience with 13 cases. J Neurosurg. 2014;121(4):827–38.
[11] Kubota Y, Enatsu R, Gonzalez-Martinez J, Bulacio J, Mosher J, Burgess RC, et al. In vivo human hippocampal cingulate connectivity: A corticocortical evoked potentials (CCEPs) study. Clin Neurophysiol. 2013;124(8):1547–56.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210853

Early Prediction of Neoplasms Using Machine Learning: A Study of Electronic Health Records from the Ministry of National Guard Health Affairs in Saudi Arabia

Asma ABDULLAH ALFAYEZ [a,b,c], Alvina GRACE LAI [a] and Holger KUNZ [a,1]
[a] *Institute of Health Informatics, University College London, London, United Kingdom*
[b] *King Abdullah International Medical Research Center, Riyadh, Saudi Arabia*
[c] *King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia*

Abstract. The early detection and treatment of neoplasms, and in particular the malignant, can save lives. However, identifying those most at risk of developing neoplasms remains challenging. Electronic Health Records (EHR) provide a rich source of "big" data on large numbers of patients. We hypothesised that in the period preceding a definitive diagnosis, there exists a series of ordered healthcare events captured within EHR data that characterise the onset and progression of neoplasms that can be exploited to predict future neoplasms occurrence. Using data from the EHR of the Ministry of National Guard Health Affairs (MNG-HA), a large healthcare provider in Saudi Arabia, we aimed to discover health event patterns present in EHR data that predict the development of neoplasms in the year prior to diagnosis. After data cleaning, pre-processing, and applying the inclusion and exclusion criteria, 5,466 patients were available for model construction: 1,715 cases and 3,751 controls. Two predictive models were developed (using Decision tree (DT), and Random Forests (RF)). Age, gender, ethnicity, and ICD-10-chapter (broad disease classification) codes as predictor variables and the presence or absence of neoplasms as the output variable. The common factors associated with a diagnosis of neoplasms within one or more years after their occurrence across all the models were: (1) age at neoplasms/event diagnosis; (2) gender; and patient medical history of (3) diseases of the blood and blood-forming organs and certain disorders involving immune mechanisms, and (4) diseases of the genitourinary system. Model performance assessment showed that RF has higher Area Under the Curve (AUC)=0.76 whereas the DT was less complex. This study is a demonstration that EHR data can be used to predict future neoplasm occurrence.

Keywords. Machine learning, decision tree, random forest, neoplasms, cancer, artificial intelligence, big data.

[1] Corresponding author, Holger Kunz, Institute of Health Informatics, University College London, 222 Euston Road NW12DA, London, United Kingdom; E-mail: h.kunz@ucl.ac.uk

1. Introduction

The objective of this study was to discover health event patterns present in Electronic Health Record (EHR) data that predict the development and occurrence of neoplasms (malignant and benign) in one year prior to diagnosis based on the patient's medical history by developing predictive models. This will improve neoplasms prediction to support the early diagnosis and as a warning system. After developing machine learning models, we compared models' performances and complexity to discuss their likely applicability in practice.

2. Methodology

This study used data from the Ministry of National Guard Health Affairs (MNG-HA) in Saudi Arabia. It is one of the largest healthcare organisations and has pioneered healthcare advances in Saudi Arabia [1, 2]. The data covered all patient encounters: in-patient, out-patient, and emergency department (ED) visits in the primary and secondary care settings between 2015 to the beginning of 2019 were included. Patient data were extracted from the EHR and included: patients' demographics and diseases history. We selected patients according to the inclusion and exclusion criteria to obtain the case group (patients with neoplasms) and control group (patients without neoplasms). We chose the first neoplasm incident/diagnosis as the index event before which the medical history was retrieved for the case group. For the control group, the last disease diagnosis was used as the index since there was no mutual diagnosis/event for the entire control group that could be used as the index event. For modelling the impact of prevalent comorbidities on early neoplasms prediction, we modelled the medical history using the International Classification of Diseases, Tenth Revision (ICD-10) codes in the EHR. We then implemented two machine learning predictive algorithms to build the models then compared their performances and complexities. The algorithms were: Decision Tree (DT) and Random Forests (RF). We chose these models as they are used to predict binary outcomes based on a set of observed characteristics and explore the effect of these characteristics on the outcome. The outcome was the presence or absence of neoplasm whereas the characteristics were the medical history beside the social demographic variables: age, gender, and ethnicity. We used the following performances metrics: Area Under the Curve (AUC); balanced accuracy; sensitivity; specificity; positive predictive values; negative predictive values; and F-score for accuracy.

3. Results

The total population was 17,631 patients including both: the case and control groups. The case group has 11,204 patients and the control group has 6,427 patients. After data cleaning, pre-processing, and applying the inclusion and exclusion criteria, 5,466 patients remained: 1,715 cases and 3,751 controls. Table 1 shows the top ten neoplasms represented in the study sample. It shows the ICD10 codes, the corresponding neoplasms types for each ICD10 code according to the World Health Organisation (WHO) [3], and the number of cases for each.

Table 1. Top ten neoplasm types represented in our study sample

ICD10 code	Type	N(%)
C50	malignant neoplasms of breast	153(9%)
C73	malignant neoplasms of the thyroid gland	147(8.6%)
C22	malignant neoplasms of the liver and intrahepatic bile ducts	140(8.2%)
C18	malignant neoplasms of the colon	99(5.8%)
C64	malignant neoplasms of the kidney, except renal pelvis	87(5%)
D43	neoplasm of uncertain or unknown behaviour of brain and central nervous system	85(5%)
C61	malignant neoplasms of the prostate	61(3.6%)
C67	malignant neoplasms of the bladder	58(3.4%)
D24	benign neoplasms of the breast	55(3.2%)
C25	malignant neoplasms of the pancreas	46(2.7%)

The mean and standard deviation of the follow-up time for the case and control groups one year before the event were 624 days (SD: 188) and 731 days (SD: 231) respectively. Table 2 shows the 10-fold cross-validation prediction performance assessment results for the models. RF had a slightly higher predictive performance with identical performance in terms of NPV (0.90) with the DT.

Table 2. 10-fold cross-validation model performance assessment results. AUC: area under the curve; BA: balanced accuracy; SE: sensitivity; SP: specificity; PPV: positive predictive values; NPV: negative predictive values; and F1-score

Model	AUC	BA	SE	SP	PPV	NPV	F1-score
DT	0.74	0.77	0.73	0.82	0.58	0.90	0.64
RF	**0.76**	0.79	0.74	0.84	0.61	0.90	0.67

4. Discussion

The aim of this study was to develop models that predict the occurrence of a neoplasm within one year using medical history data obtained from the patients' EHRs. Two models: DT and RF for neoplasm prediction were used. The common factors associated with a diagnosis of neoplasms within one or more years after their occurrence across all the models were: (1) age at neoplasms/event diagnosis; (2) gender; and patient medical history of (3) diseases of the blood and blood-forming organs and certain disorders involving immune mechanisms; and (4) diseases of the genitourinary system. These factors were associated with subsequent neoplasm development but predicted differently: age and male gender increased the likelihood of developing neoplasms, which are known risk factors as older people and males are more likely to develop neoplasms compared to younger and female patients [4-6].

This study used a small subset of the data available in the EHR. Therefore, this analysis is unlikely to have captured all of the variables that predict neoplasm occurrence. For example, while smoking and alcohol status were available and are known risk factors

for malignant neoplasms [7], we did not use them as most of their values in the EHR were either not recorded after being collected from the patient, were missing, or were not collected from the patients at all.

RF showed a better performance in terms of AUC. RF model, by definition, used all the provided predictors, twenty-seven in our case, making it more complex in this study. In contrast, the DT model used only four of the input predictors, making it less complex model. Therefore, while RF showed better performance it was also more complex, whereas DT showed lower performance but was less complex model.

5. Conclusion

In this study, two ML models have been used to predict neoplasms (malignant and benign) one year prior to diagnosis. The two ML algorithms used to build the models were: (1) DT and (2) RF. The medical history was obtained from the EHR to construct a feature space for training and testing the two models. The study showed that ML algorithms applied to data from EHR's could improve and support the early diagnosis of neoplasms.

References

[1] King Abdulaziz Medical City in Riyadh [Internet]. Kingdom of Saudi Arabia Ministry of National Guard Health Affairs. 2020 [cited 2021Sep17]. Available from: https://www.ngha.med.sa/English/MedicalCities/AlRiyadh/Pages/default.aspx

[2] Kingdom of Saudi Arabia Ministry of National Guard Health Affairs. 2018 [cited 2021Sep15]. Available from: https://www.ngha.med.sa/English/AboutNGHA/Pages/profile.aspx

[3] International Statistical Classification of Diseases and Related Health Problems 10th Revision [Internet]. ICD-10 Version:2010. World Health Organisation (WHO); [cited 2021Jul14]. Available from: https://icd.who.int/browse10/2010/en

[4] White MC, Holman DM, Boehm JE, Peipins LA, Grossman M, Henley SJ. Age and cancer risk: a potentially modifiable relationship. American journal of preventive medicine. 2014 Mar;46(3):S7-15.

[5] Laconi E, Marongiu F, DeGregori J. Cancer as a disease of old age: changing mutational and microenvironmental landscapes. British journal of cancer. 2020 Mar;122(7):943-52.

[6] García-Martín JM, Varela-Centelles P, Gonzále M, Seoane-Romero JM, Seoane J, García-Pola MJ. Epidemiology of Oral Cancer. In: Oral Cancer Detection. Cham: Springer International Publishing; 2019. p. 81–93.

[7] Hydes TJ, Burton R, Inskip H, Bellis MA, Sheron N. A comparison of gender-linked population cancer risks between alcohol and tobacco: how many cigarettes are there in a bottle of wine?. BMC public health. 2019 Dec;19(1):1-8.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210854

Medical Informatics in a Tension Between Black-Box AI and Trust

Murat SARIYAR[a,1] and Jürgen HOLM[a]
[a] *Bern University of Appl. Sciences, Department of Medical Informatics, Switzerland*

Abstract. For medical informaticians, it became more and more crucial to assess the benefits and disadvantages of AI-based solutions as promising alternatives for many traditional tools. Besides quantitative criteria such as accuracy and processing time, healthcare providers are often interested in qualitative explanations of the solutions. Explainable AI provides methods and tools, which are interpretable enough that it affords different stakeholders a qualitative understanding of its solutions. Its main purpose is to provide insights into the black-box mechanism of machine learning programs. Our goal here is to advance the problem of qualitatively assessing AI from the perspective of medical informaticians by providing insights into the central notions, namely: explainability, interpretability, understanding, trust, and confidence.

Keywords. Artificial intelligence, Explainable AI, Luhmann, Confidence, Trust

1. Introduction

Artificial intelligence (AI) is now essential part of many activities in the field of medical informatics, not only in research but also in the healthcare setting [1, 2]. The first FDA-approved medical device that relies on AI was the BodyGuardian ® Remote Monitoring System from Preventice Solutions in 2012, which detects cardiac rhythm abnormalities, using small wearable monitors paired with a dedicated smart-phone [3]. Main reason for the success story of AI is the boost in prediction accuracy due to advances in digital documentation, computing power, (deep learning) algorithms, and wearable medical devices. Relevant application areas are disease diagnosis, image classification, natural language processing of electronic health records, biomarker discovery, and drug development.

Even for medical informaticians not developing AI-based decision support systems, it became crucial to assess the benefits and disadvantages of AI-based solutions as promising alternatives for many traditional tools. Besides quantitative criteria such as accuracy and processing time, healthcare providers are often interested in qualitative explanations of the solutions. For the former issue, medical informaticians could rely more and more on the results of explainable AI (XAI) approaches [4]. However, there are still many trust problems. As all three terms “explainability”, “responsibility”, and “trust” are often not clarified, qualitative assessments of AI frequently are frequently not satisfying, independently of relying on XAI or not.

[1] Corresponding Author, Murat Sariyar, Bern University of Applied Sciences, Quellgasse 21, CH2502 Biel/Bienne, Switzerland; E-mail: murat.sariyar@bfh.ch.

Our goal here is to advance the problem of qualitatively assessing AI from the perspective of medical informaticians by providing insights into the central notions and their relations. As a use case, we will refer to biomarker discovery for pharmacogenes as a step towards a comprehensive decision support systems concerning drug therapies. Pharmacogenetics investigates the association between genetic variations and the drug response to help tailoring pharmacotherapy to the individual patients' characteristics, thereby avoiding unnecessary adverse drug events (ADE) and increasing therapeutic efficacy [5]. Especially stratification according to the origin of individuals is an urgent issue, which requires augmenting the current biomarkers with the help of AI.

2. Methods

A first step for the qualitative assessment of AI is the definition of AI as well as XAI and the listing of approaches for XAI relevant for making the distinction between explanation and interpretation. After that, the term explanation is defined with reference to scientific theory. Understanding, trust, and confidence are defined through reference to system theory of Luhmann, as it provides a holistic foundation of these notions, which allows to detect discrepancies in their empirical use [6,7]. The impacts of these definitions are discussed with respect to our use case.

3. Results

AI is the "theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages" [8]. Usually weak AI is assumed: computer systems are made to act as if they were intelligent, not to be intelligent. Two high-level approaches for achieving such intelligence are rule-based and non-rule-based. The latter is associated with machine learning, i.e., a computer program learns it tasks from available data and improves its performance with further data from the same context. AI methods are used as a basis for many clinical expert systems, which suggest solutions to problems that were previously solved by human experts alone. Such a genealogy and the relevance of medical decisions raise the demand for interpretations of AI solutions.

XAI provides methods and tools, which are interpretable enough that it affords different stakeholders a qualitative understanding of its solutions [9]. Its main purpose is to provide insights into the black-box mechanism of machine learning programs. The term "black-box" refers to the opaqueness related to the mechanisms responsible for producing solutions. Approaches for opening the black-box can be differentiated into model-agnostic and model-specific. Examples for the former ones are sensitivity analysis and local interpretable models (LIME). Both are not looking under hood, but tweak the input and observe the resulting effects, thereby gaining insights into the relevance of features. LIME is often used in practice, as it can model interactions by fitting surrogate linear models to the results of multifeatured perturbations around certain predictions (hence local). Model-specific XAI methods rely either on the mechanism of the algorithm itself – e.g., decision trees allow to trace the decision paths and to extract the key determinants – or provide means for looking under the hood in a certain class of methods, e.g., relevance propagation in the case of deep neural networks.

What should be achieved by XAI? Explainability as the name indicates, interpretability as the definition above suggests or understandability? Is the final goal trust or confidence? We locate many acceptance problems of AI in the lack of clarity and relevance assessment with respect to these notions. Explainability refers to the possibility of providing reasons (explanans) for the outcome (explanandum) of a system. There are many different forms for explanation: the deductive-nomological, inductive-statistical, causal mechanical, etc. [10]. In all cases, it is central to provide justification (reasons) that are comprehensible and transparent (details) at the appropriate level for the addressed audience. For example, in terms of causal mechanical explanation, an explanandum X explains an explanans C, if (i) X increases the probability of C, given the other explanatory factors F (statistical relevance $P(C \mid F \,\&\, X) > P(C \mid F)$), and (ii) X fits into the causal nexus of the explanans C. An explanation for "Why does this *deep neural net* provide the *best solutions*?" would generally refer to components of the deep neural net at different levels that increases the probability of making the right predictions. The problem is that they are multiple explanations, and most of them won't be satisfying for a non-expert. Hence, pure explainability is not sufficient for the goals of XAI.

Interpretability is a property of an explanation that describes the extent to which the cognitive capacities of the addressed audience is taken into consideration. Hence, this notion highlights, that explanation is a social process for which understanding of the addressed audience should come into play. According to Luhmann, understanding is the result of constructing a distinction between the information provided and its form (as a text or verbally), which is only retrospectively perceivable in following communication events [11]. In other words, understanding is only measurable through ensuing activities and utterances of the addressed audience. The advantage of this definition is, that it can ignore the unsolvable problem of how to avoid the case of where someone can explain something without having understood it. If the following communication is compatible with the goal of XAI, we are fine. This does not provide a solution to the appropriate level of explanations; it just indicates the necessity for an – often iterative – social process for achieving satisfying interpretability.

The most important result of understanding in the context of XAI should be either trust or confidence. Following Luhmann again, trust is a decision to rely on one's own expectations with respect to certain mechanisms in view of the involved risks and alternatives. Confidence on the other hand is a reliance on one's own expectations concerning mechanisms without taking alternatives or the risks into consideration. In both cases, complexity is reduced by taken something for granted. Within the context of AI, confidence is sufficient, if the decision-making system is related to non-sensitive data, for example in the context of search engines or recommender systems. Especially, for clinical AI application, it can be important to arrive at trust. Health care providers know at least one alternative to an AI system: the human decision making. One should distinguish two levels of trust: layman trust, for which high-level explanations are sufficient and expert trust, which require many details. Table1 summarizes the insights concerning XAI in terms of central properties.

Table 1. Properties of different AI goals with one example for each of the goals.

Goals of XAI	Black box	Alternatives	Examples
Confidence	yes	Not Considered	"The system is already in use in hospital X"
Layman trust	Opened slightly	Considered	List of features that have significant impacts (LIME)
Expert trust	Opened fully	Considered fully	Results of relevance propagation for deep nets

For our use case, confidence means that a new pharmacogenetic biomarker signature for a certain subgroup produced by an AI algorithm will be accepted by physicians not familiar with biomarker discovery if, for example, working groups such as CPIC (Clinical Pharmacogenetics Implementation Consortium) will validate it. Layman trust is necessary, if other biomarkers for the subgroup are known to be inferred through classical statistical methods, in which case we provide high-level insights for physicians through LIME. Expert trust can be required for the CPIC validation, which must ensure that there are no biases in using certain AI implementations.

4. Discussion

An implication of our results is the requirement for more stakeholder involvement, especially in the case of translational research. There is no one-fits-all solution of XAI. It needs to be adapted to the context and via a (n iterative) social process, which means to augment the available methods by diversified qualitative explanations. We are aware of the fact, that scientists are not eager to invest time in such processes, but instead of regulating it or creating working groups for discussing how to foster such a culture, there should be an intrinsic motivation for appropriate explanations. Physicians can provide much better feedback if they understand mechanism behind solutions.

Further research in qualitative XAI should focus on informing quantitative XAI methods and vice versa. A comprehensive categorization of available XAI methods in terms of their usefulness for qualitative explanations would also be useful for these methods themselves, as this could foster an understanding for which audiences these methods are developed and how they should be improved. In addition to that, system theory provides a rich arsenal of explanations regarding social settings of trust and understanding, which should be referred to more often. We are confident that AI will gain more trust through adapted social practices, not by ex-cathedra statements.

References

[1] Emanuel EJ, Wachter RM. Artificial Intelligence in Health Care: Will the Value Match the Hype? JAMA 321 (2019), 2281–2282.
[2] Choudhury A, Asan O. Role of Artificial Intelligence in Patient Safety Outcomes: Systematic Literature Review. JMIR Med Inform 8 (2020), e18599.
[3] Teplitzky BA, McRoberts M, Ghanbari H. Deep learning for comprehensive ECG annotation. Heart Rhythm 17 (2020), 881–888.
[4] Holzinger A, Langs G, Denk H, et al. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 9 (2019), e1312.
[5] Wake DT, Ilbawi N, Dunnenberger HM, et al. Pharmacogenomics: Prescribing Precisely. Med Clin North Am 103 (2019), 977–990.
[6] Luhmann N. Trust and Power. Chichester: Wiley, 2017.
[7] Pieters W. Explanation and trust: what to tell the user in security and AI? Ethics Inf Technol 13 (2011), 53–64.
[8] Walsh K. Artificial intelligence and healthcare professional education: superhuman resources for health? Postgrad Med J 96 (2020), 121–122.
[9] Duell J, Fan X, Burnett B, et al. A Comparison of Explanations Given by Explainable Artificial Intelligence Methods on Analysing Electronic Health Records. In: 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI). 2021, pp. 1–4.
[10] Dammann O. Evidence Mapping to Justify Health Interventions. Perspect Biol Med 64 (2021), 155–172.
[11] Luhmann N. Social Systems. Stanford: Stanford University Press, 1996.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210855

Development of an Expert Knowledge-Based Genomic Variant Prioritisation Platform

Aurora SUCRE[a,b], Gregory MACLAIR[a], Iride MARTINEZ[c], Concha VIDALES[c], Gorka EPELDE[a,b] and Alba GARIN-MUGA [a,b]
[a] *Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), Mikeletegi 57, Donostia-San Sebastián, Spain*
[b] *Biodonostia Health Research Institute, eHealth Group, Donostia-San Sebastián, Spain*
[c] *I+D+I department, DNAData, Donostia-San Sebastián, Spain*

Abstract. Considering the growing interest towards next generation sequencing (NGS) and data analysis, and the substantial challenges associated to fully exploiting these technologies and data without the proper experience, an expert knowledge-based user-friendly analytical tool was developed to allow non-bioinformatics experts to process NGS genomic data, automatically prioritise genomic variants and make their own annotations. This tool was developed using a user-centred methodology, where an iterative process was followed until a useful product was developed. This tool allows the users to set-up the pre-processing pipeline, filter the obtained data, annotate it using external and local databases (DBs) and help on deciding which variants are more relevant for each study, taking advantage of its customised expert-based scoring system. The end users involved in the project concluded that CRIBOMICS was easy to learn, use and interact with, reducing the analysis time and possible errors of variant prioritisation for genetic diagnosis.

Keywords. Genomics, genetic variants prioritisation, user-centred design, standalone tool.

1. Introduction

Genetic variants refer to the variations found on a DNA sequence that may alter the individual's anatomy, physiology, psychology, disease predisposition and drugs susceptibility. The detection and screening of these variants is gaining a good reputation, as it is becoming a very powerful tool in personalised medicine, specifically for diagnosis. Different tools that help researchers doing semiautomatic variant prioritisation exist in the market, but most of them simply do not fulfil all the users' requirements that were detected during this project. Therefore, many researchers and biologists are relying on time-consuming manual mechanisms for variant prioritisation, so their approaches are far from being optimised, having a huge impact on the geneticists' productivity.

In this paper, we present CRIBOMICS, a novel variant prioritisation tool based on expert knowledge. This tool was designed following a user-centred-design (UCD) methodology. UCD methodologies have been applied extensively to develop clinical applications, and have been found to be successful in this field (1). CRIBOMICS allows to process raw NGS data, call genomic variants (SNPs and CNVs), filter them by various

features, annotate them using well-known DBs and automatically prioritise the kept variants according to their clinical relevance. Additionally, a customizable DB is embedded in the tool to allow experts to record their own findings in CRIBOMICS and have them available in future analysis. The automatic prioritisation process is based on the formalisation of experts' knowledge into rules. Expert end-users confirmed that the tool facilitates variant prioritisation for genetic disease diagnosis.

2. Methods

The main objective of CRIBOMICS was to create an automatic tool for variant prioritisation that could facilitate the genetic expert workflow. To do so, an iterative UCD methodology was followed, where a multidisciplinary expert committee was involved in the entire development process, in order to conduct an evaluation after every iteration, to provide feedback regarding the tool usability and effectiveness in real conditions. The panel was established at the beginning of the project following Nielsen recommendations (2).

During the initial stages, the main requirements of the end-users were collected and contrasted with the variant analysis guidelines defined by the ACMG (3), to make sure the tool would be compliant with existing standards. Afterwards, the engineering experts were able to define the user task scenarios, the different needed functionalities, and the overall system architecture. Regarding the algorithm for automatic prioritisation, its logic was determined by having the geneticists explain their decision-making process and the developers formalise it into rules that can be interpreted by the software. All the relevant parameters, thresholds, decision-making conditions were collected and translated to be embedded in the knowledge core of the system.

On every development iteration, the experts analysed the novel features and the modifications in the GUI, and provide their feedback following a preestablished protocol, so developers can exploit the data and implement a list of enhancements on every iteration, until the tool fulfilled all the user requirements. The initial versions of the software focused on the secondary analysis, which comprises the filtering, annotation, and prioritization of SNPs. Then, after several iterations the committee detected the need to include new modules for primary analysis (alignment, quality control, variant calling), which was achieved by implementing the GATK pipelines for data pre-processing and germline variant calling (4). Finally, the same happened for CNV detection and analysis, and the necessary modules were developed following GATK best practices.

3. Results

The UCD methodology allowed us to tailor the solution to the experts' needs and enabled the automatization of the process to ease their workflow by formalising their reasoning into a rule-based system. The iterative methodology encouraged having constant interactions with end-users and facilitated the development of a functional user-friendly standalone application that automates the prioritisation of genomic variants and can be exploited by specialists with no bioinformatics knowledge.

The obtained solution combines a GUI developed using QT and Python, different analytical modules based on Python and other third-party tools for genome data processing and prioritisation (i.e. GATK, Samtools, Picard and BWA).

The tool allows users to analyse the data associated to a single patient's genetic sample and also includes modules for duo (5) and trio analyses. In single analyses, it is possible to analyse both SNPs and CNVs and the process varies based on the maturity of the input data and given file format (BCL, FQ, BAM, VCF). CRIBOMICS includes pre-processing modules to perform all the steps from base calls deconvolution to variant calling, in order to obtain the patients' SNPs and CNVs in VCF format. This pipeline is based on the GATK best practices for germline variant discovery (4). Users can tune the analytical pipeline by modifying several parameters; however, default parameters are set to ease the analysis.

Once the variants are stored in the VCF format, the tool extracts all the relevant data regarding the detected variants and performs an initial annotation step to obtain basic information that may be missing (RS IDs, genes, genomic location and the effect they may have in the proteins). Regarding CNV analysis, also the genes that were completely or partially affected by each variant are collected.

Then, the data is shown in summarised tables, and users are able to filter and annotate the detected variants according to various categories.

The variants can be annotated from various sources:

- NCBI: Clinical significance according to ClinVar, diseases associated with the variant or with the gene and publications found in PubMed.
- Ensembl: SIFT prediction, Polyphen prediction and HGVSc IDs.
- dbNSFP: MT2 prediction.
- Internal DB: Users can store their findings in a customisable DB and their annotations will be available for future analyses.

Initially, the SNPs can be filtered according to their 1) genotype, 2) chromosome, 3) affected genes, 4) different quality parameters, 5) genomic location and 6) protein consequence. Once they are annotated, they can also be filtered by their 7) clinical significance and 8) predicted pathogenicity.

CRIBOMICS also includes a proprietary scoring system that helps prioritize the variants during each analysis. This system is based on the experts' decision-making process, that was successfully formalised into logical rules that can be interpreted by the software.

The scoring system set of rules considers 1) the diseases of interest, 2) the clinical significance of each variant, 3) the pathogenicity predictions, and 4) the diseases that have been found associated to the variants or the genes, to objectively classify all detected variants and give them a high, medium, or low relevance tag. Depending on the available information for each analysis, the relevance tags are assigned following different rules. For instance, if no diseases of interest are defined during the analysis, the algorithm only focuses on the clinical significance and prediction results, but if the user provides any disease of interest, the related diseases feature gains relevance and constrains the output of the algorithm.

CRIBOMICS final output includes all the initial relevant data, the annotated information and the relevance score assigned to all the kept variants. This can be stored as an MS Excel file or in a CRIBOMICS-specific format, to be further analysed in the future.

The current analytical workflow was designed for the existing standalone application (Figure 1), but as it was conceived as a modular, customizable solution, it is currently being migrated to a cloud environment without requiring extensive resources.

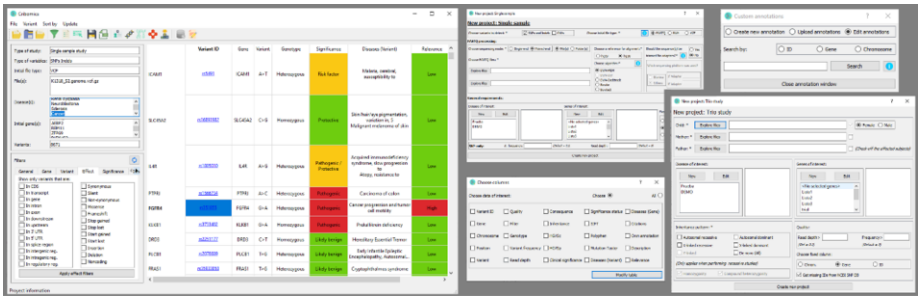


Figure 1. CRIBOMICS graphical user interface

4. Discussion

The UCD approach that was followed proved to be effective when developing this kind of solution, where the final-user contributions are so relevant in order to meet their needs.

During development it was mandatory to exploit standard or at least the most commonly-used tools in the field, therefore the GATK best practices (4) were followed during primary analysis, and the NCBI and Ensembl DBs were used for data annotation.

According to the end-user's evaluation, CRIBOMICS is indeed an intuitive tool that ease the SNP and CNV detection and prioritisation processes, and the proprietary scoring system seems to be the most useful feature, that simplifies their interpretation of the data and allow them to reach conclusions faster, even without any bioinformatic skill. Also the usability of the DB was emphasised, as it helps on keeping track of new findings.

Regarding future work, an authoring tool is currently being developed to allow experts to easily formalise their own prioritisation rules, and make the system customisable.

It is important to highlight that this tool does not make decisions regarding diagnosis but provides experts with recommendations based on the available information and needs.

Even if the tool has been tested by the expert committee and they have provided promising feedback, the tool it is still not publicly available, as it first need to overcome validation. This process will be performed by a larger group of experts that will evaluate: 1) software usability, 2) results accuracy and 3) software efficiency.

References

1. Luna D, Quispe M, Gonzalez Z, Alemrares A, Risk M, Otero C, et al. User-centered design to develop clinical applications. Literature review. Stud Health Technol Inform. 2015;216:967–967.
2. Nielsen J, Landauer TK. A Mathematical Model of the Finding of Usability Problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 1993. p. 206–13. (CHI '93). Available from: https://doi.org/10.1145/169059.169166
3. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405–23.
4. Van der Auwera GA, O'Connor BD. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media; 2020.
5. Sucre A, Garcia-Longarte S, Garin-Muga A. DONOR-FINDER: A Web Tool for the Automatic Comparison of Genomic Profiles in ART. Stud Health Technol Inform. 2020 Jun 26;272:139–42.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210856

Pro-Active Detection of Potentially Wrong Diagnoses Due to Substantial Changes of Laboratory Measurements

Motaas ABU BAKER [a], Fabienne CALDERARA [a,] Jürgen HOLM [a] and Murat SARIYAR [a,1]
[a] *Bern University of Appl. Sciences, Dept. Medical Informatics, Switzerland*
[b] *Center for laboratory medicine (ZLM) St Gallen, Switzerland*

Abstract. For guiding decisions on medical diagnoses and diagnoses, it is crucial to receive valid laboratory test results. However, such results can be implausible for the physician, even if the measurements are within the range of known reference values. There are technical sources of implausible results that are related to the laboratory environment, which are frequently not detected through usual measures for ensuring technical validity. Here, we describe the development of a quality assurance tool that tackles this problem and replaces the current manual statistical analyses at the Center for Laboratory Medicine in St Gallen (ZLM). Further analysis of the factors responsible for shifts in laboratory test results requires to collect and analyze data related to reagents as well as calibration or reference probes. Due to a lack of standard operating procedures in many laboratories with respect to these processes, this remains one of the big challenges.

Keywords. Clinical validation, laboratory tests, medical diagnosis, data quality

1. Introduction

Physicians rely on valid results of laboratory tests, when making decisions on diagnoses and therapies, especially in times of COVID testing. However, results can be implausible for the physician, even if the measurements are within the range of known reference values. In a previous work [1], we have dealt with the analysis of factors for erroneous laboratory test results outside the laboratory context. This kind of analysis is referred to as *medical validation* and deals with non-technical sources for implausible values, e.g., a value of 6.0 pmol/L for free triiodothyronine (fT3) resulting from a contaminated blood test of a patient without hyperthyroidism. In addition to that, there are also technical sources of implausible results that are related to the laboratory environment, which are frequently not detected through usual measures for ensuring technical validity [2,3]. For example, reagent lot changes (e.g., of antibodies) can be accompanied by a change of the calibration material that together led to a calibration curve within the same ranges as before, even though the related results for the patients will be shifted. In other words, even if measured values in a technical quality control procedure are within the range of

[1] Corresponding Author, Murat Sariyar, Bern University of Applied Sciences, Quellgasse 21, CH2502 Biel/Bienne, Switzerland; E-mail: murat.sariyar@bfh.ch.

reference values, implausible test results can occur in practice. Physicians should not be confronted with such test results.

Here, we describe the development of a quality assurance tool with R Shiny that replaces the current manual statistical analyses at the Center for Laboratory Medicine in St Gallen (ZLM). To avoid complaints of physicians and reduce the number of suboptimal decisions due to biased laboratory test results, the ZLM wanted to establish a new statistical quality assurance tool to be used in the laboratory only. Two key requirements were, (i) that a user with limited IT knowhow should be able to use the tool, and (ii) that automated statistical evaluations can be triggered. In the end, the current manual statistical analyses are replaced using a standard frontend to perform the necessary validation work. Further, interactive, and annotated figures, which can be produced on the fly, facilitate verification of the computed statistical values.

2. Methods

First, we derived user requirements of the ZLM through user stories. A user story describes the features of a software system from the perspective of the user [4], for example: User KJ (who) wants a moving average graph for all fT3 test results from 2019-01-01 to 2019-12-31 (what) in order detect significant shifts in the measurements (goal). After that, we developed a mock-up to ensure to have captured all requirements. Finally, we developed a web-based prototype based on the R shiny framework (see [5]) and anonymized data related to laboratory devices stemming from laboratory information system of the ZLM. The decision for R as programming environment was motivated by the large number of packages for statistical analysis and the sophisticated visualization capabilities.

In addition to that, we conducted a literature search using the PubMed database. We were only interested in studies that aimed at technical validation and the calibration process by means of statistical quality assurance. Following combination of keywords were used:

- "technical validation" AND "calibration" AND "quality assurance"
- "technical validation" AND "medical validation"
- "technical validation" AND "clinical validation"
- "technical validation" AND "clinical validation" AND "statistics"
- "technical validation" AND "statistical quality control"
- "technical validation" AND "calibration process"
- "technical validation" AND "statistical"

3. Results

In the literature research, we did not find any study that addresses with the issue of proactively detecting substantial changes of laboratory test results caused by technical processes that are not detected by common technical validation procedures. The focus is either on technical validation or on medical validation, neglecting the impacts of laboratory device changes on laboratory parameter measurements not detected in standard validation procedures are not tackled. We assume two central reasons for this unexpected result: (i) reliance on the technical validation, once it is established and (ii)

a high number of unreported cases in practice, especially when the shifts in the test results are rather minor (unknown unknowns cannot be investigated). With respect to the development of our tool, the outcome of the literature research meant, that we had no further inputs other than the requirements of our stakeholder. To facilitate generalizability of our prototype, we concentrated on statistical analysis that are easy to interpret and allowed many options for going into details, e.g., by selecting date ranges.

Our prototype has three main tabs: monitoring (automatic evaluation), manual evaluation and settings. Per default, the monitoring tab is selected when accessing the web application. For all laboratory parameters, daily, weekly, and monthly changes of the median measurement values are computed and visualized with a density plot (see Figure 1). The median value changes are computed as (representing relative changes)

$$\frac{median_t - median_{t-1}}{median_{t-1}},$$

with t ranging over days, weeks, or months. An automated generated email is sent out by the prototype, when the absolute values exceed a certain threshold. Furthermore, reactive moving median curves are generated in a subtab of the monitoring tab, allowing to inspect the development of measurement values. Within the manual evaluation tab, a broader range of statistical analyses becomes available, e.g., hierarchical clustering and statistical tests, allowing the user to investigate possible reasons for certain changes.

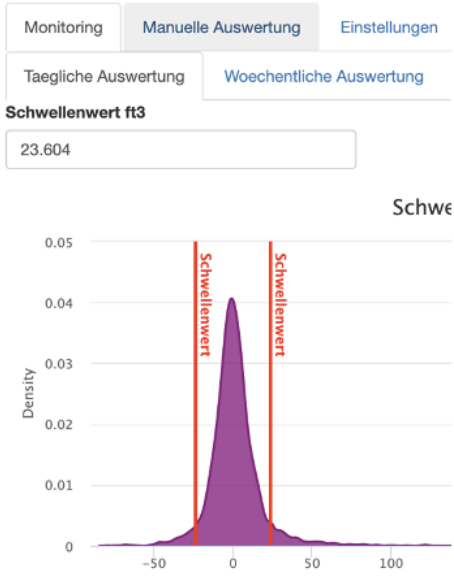


Figure 1. Density plot of relative changes of the median (y) for fT3 values with corresponding thresholds (Schwellenwert), for which only the right one (here the 90% percentile) is relevant in our datasets.

We discovered on our data, that the most important insights were given by the moving median curves of the measurement values summarized per week (daily data were too granular and monthly data too coarse). The user of our prototype is enabled to change the smoothing parameter to determine whether to focus on short- or long-time effects (trends). In Figure 1, the relevant shift of the fT3 measurements in December 2015 is

clearly discernible. Due to the interactivity of the graphs, by using the R-package plotly, the user gets annotation of data points and can zoom in and out. The reason for the peak in January 2015 is not known.

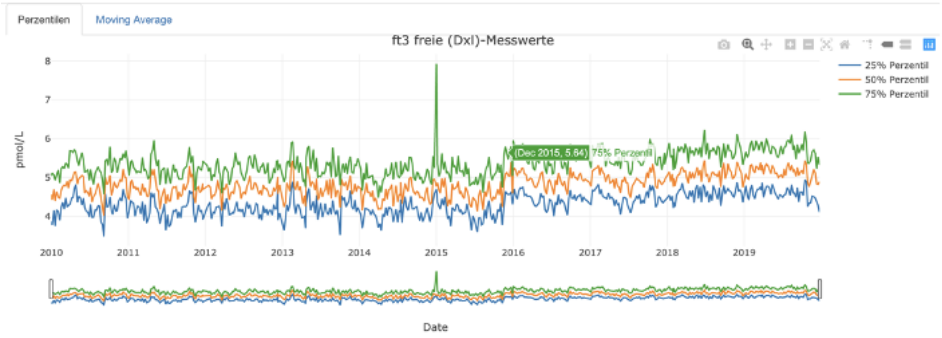


Figure 2. Moving median curve for the fT3 values in a reactive plotly graph.

4. Discussion

Proactively detecting substantial changes of laboratory test results caused by technical processes that are not detected by common technical validation procedures is still uncommon, as there is often reliance on the technical validation. For the analysis of factors outside standards technical validation responsible for shifts in laboratory test results, it is crucial to include valid data related to reagents, calibration, or reference probes in addition to time series of test result values. Due to a lack of standard operating procedures in laboratories, this remains one of the big challenges.

An implication of substantial changes of laboratory test results is the adjustment of the corresponding reference intervals. At the ZLM, there is a standardized procedure for setting a reference interval. The prototype will increase its usefulness, when it integrates and enriches this procedure, allowing to address all issues around the shifts of measurement values. Following insight should be considered in this context [6]: "Analytical factors and pre-analytical factors need to be considered, along with partitioning on the basis of sex or age, either as part of a reference interval study or when interpreting other studies."

References

[1] Degen GA, Günther V, Holm J, Bürkle T, Sariyar M. Using Business Intelligence Tools to Support Medical Validation of Laboratory Tests. *Stud Health Technol Inform*. 2020; **270**:494–498.

[2] Jennings L, Van Deerlin VM, Gulley ML, College of American Pathologists Molecular Pathology Resource Committee. Recommended principles and practices for validating clinical molecular pathology tests. *Arch Pathol Lab Med*. 2009; **133**:743–755.

[3] Kay JF. Combining ISO/IEC 17025:2005 and European Commission Decision 2002/657 audit requirements: a practical way forward. *Drug Test Anal 4 Suppl*. 2012; **1**:25–27.

[4] Kannan V, Basit MA, Bajaj P, et al. User stories as lightweight requirements for agile clinical decision support development. *J Am Med Inform Assoc*. 2019; **26**:1344–1354.

[5] Kraus M, Mathew Stephen M, Schapranow MP. Eatomics: Shiny Exploration of Quantitative Proteomics *Data. J Proteome Res*. 2019; **20**:1070–1078.

[6] Jones G, Barker A. Reference Intervals. *Clin Biochem Rev.* 2008; **29**: S93–S97.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210857

The Clinical and Biochemical Parameter Data: Effect of High Blood Glucose on Diabetic Nephropathy Among Non-Insulin-Dependent Diabetes Mellitus (NIDDM), Thailand

Chitkamon SRICHOMPOO[a], Warantorn POTARIN[b], Supattra NOO-In[c], Patcharin PHUWILERT[c] and Santisith KHIEWKHERN[c,1]
[a] *Mahachanachai Hospital, Yasothorn, Thailand*
[b] *Srimahasarakham Nursing collage, Faculty of Nursing, Praboromarajchanok, Thailand*
[c] *Faculty of public health, Mahasarakham University,Thailand*

Abstract. Background: Diabetic nephropathy (DN) is one of the long-term complications of patients with type 2 diabetes. The leading causes of DN are high blood glucose and hyper systolic blood pressure. Methods: A retrospective cohort study was performed to explore the effects of high blood sugar and hyper systolic blood pressure on DN among 660 non-insulin-dependent diabetes mellitus. Data was collected from the HosXP program and medical records from 2016 to 2020. The Forest plot was used to examine the effect of hypertriglyceridemia and hyper systolic blood pressure with DN. Results: The results confirmed that the factors associated with DN were male, age ≥ 60years, diabetic duration ≥ 10 years, systolic BP ≥ 130 mmHg, and HbA1c ≥ 6%. Conclusion: The health promotion program should be comprised of the control of blood glucose and systolic blood pressure procedure especially male patients with age ≥ 60 years and diabetic duration ≥ 10 years.

Keywords. Chronic kidney disease, albuminuria, glomerular filtration rate, diabetic duration

1. Introduction

Diabetic nephropathy (DN) is one of the long-term complications of patients with both type 1 and type 2 diabetes mellitus. Diabetic nephropathy affects the function of kidney failure to remove waste products and superfluous fluid from the body. The leading causes of DN are hypertriglyceridemia and hyper systolic blood pressure [1,3]. The best way to prevent or delay diabetic nephropathy is by maintaining a healthy lifestyle and controlling blood glucose and blood pressure. In this study, the effects of hypertriglyceridemia and hyper systolic blood pressure on diabetic nephropathy among

[1] Corresponding Author, Santisith Khiewkhern, Faculty of Public Health, Mahasarakham University, Thailand; 44150. E-mail: santisith.k@msu.ac.th.

Thai patients with non-insulin-dependent (NIDDM) are explored for the improved management of this disease.

2. Methods

Study design: A retrospective cohort study was used to examine the effects of hypertriglyceridemia and hyper systolic blood pressure to diabetic nephropathy among Thai patients with non-insulin-dependent.

Population: The population included 2,303 type non-insulin dependent diabetes mellitus who attended the diabetes care clinic of Mahachanachai Hospital, Yasothorn Province in the North-East region of Thailand. The characteristics, clinical and biochemical data of the patients were recorded during the period from October 1st, 2016 to September 30th, 2020.

Sample: 660 eligible cases of type non-insulin dependent diabetes mellitus attending the diabetes care clinic of Mahachanachai Hospital, Thailand.

Inclusion criteria: Inclusion criteria included patients with type 2 diabetes over the age of 18 years, free of DN and free from diabetic complications like retinopathy, neuropathy, peripheral vascular disease, or coronary artery disease at baseline.

Exclusion criteria: Exclusion criteria were type 2 diabetes patients who had additional complications approximating retinopathy, neuropathy, peripheral vascular disease, or cardiovascular disease during the follow up period and who did not have complete data for kidney diseases assessment and patients who did not have complete data regarding the biochemical parameter such as, triglyceride and systolic blood pressure.

Outcome measures: The identification of DN in type 2 diabetes patients was based on the Thai National Kidney Foundation 2002 criteria for diagnosis, which used an eGFR of less than 60 ml/min/1.73 m2, a urinary albumin/creatinine ratio (UACR) of30-300 mg/g, or positive Microalbuminuria dipstick test ≥ 1+ the same criteria as followed in the other studies [2,3].

Data collection: This study used secondary data which collected from the HosXP program and medical records from 2016 to 2020 at Mahachanachai Hospital, Thailand.

Statistical analysis: Descriptive statistics were used to define the characteristics of the subjects. The 95% CI of adjusted OR from multiple logistic regressions were used to examine the effects of hypertriglyceridemia and hyper systolic blood pressure with DN. Backsword method was used to examine risk factors of DN. These effects were presented by the forest plot.

Ethic Declarations: Approval from the Ethics Committee for Research Involving Human Subjects of Mahasarakham University, Thailand was attained before the study was carried out (PH057/2561). The information of respondents that was obtained through medical records and medical electronic data were kept confidential and anonymous.

3. Results

The results from the multiple logistic regression with a backward method confirmed that the factors associated with diabetic nephropathy were male, age ≥ 60years, diabetic duration ≥ 10 years, systolic BP ≥ 130 mmHg, and high blood sugar (HbA1c ≥ 6%).

However, smoking and alcohol consumption were not statistically significantly associated with DN (Table 1).

Table 1 Factors associated with DN.

Factors	adjusted OR	P-value	95%CI of OR	
			Lower	Upper
male	1.69	.003	1.28	2.14
age ≥ 60years	1.58	.005	1.18	2.00
diabetic duration≥ 10 years	1.35	.008	1.04	1.92
BP ≥130 mmHg	1.41	.025	1.06	1.80
HbA1c ≥ 6%	1.94	.020	1.44	2.11

The radar plot was created by using the adjusted OR from multiple logistic regression with a backward method to examine risk factors of DN. The results showed that high blood sugar (HbA1c ≥ 6%) is a higher effect on DN, the second is male and follow by age ≥ 60years, BP ≥130 mmHg, and diabetic duration≥ 10 years. Figure 1.

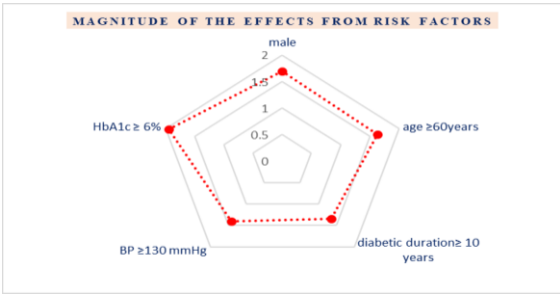


Figure 1 Radar plot present the magnitude of the effect from risk factors to DN.

The adjusted OR and 95%CI of OR were used to create the Forest plot. A Forest plot was used to explain approximately the accuracy of the effect of risk factors on DN. The results showed that the effect of high blood sugar (HbA1c ≥ 6%) is the most far from 1 and followed by male (HbA1c ≥ 6%) Figure 2.

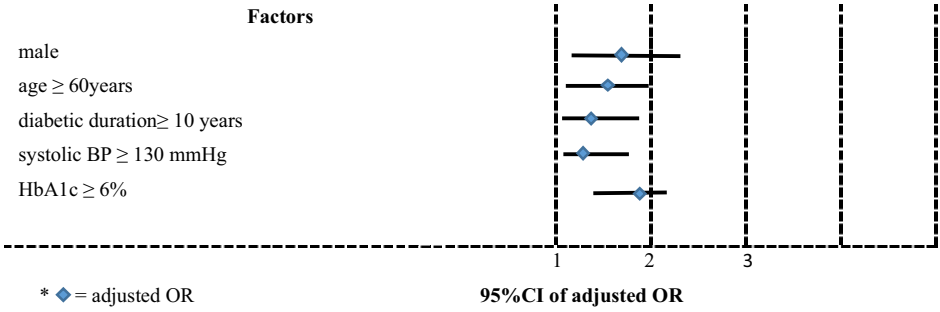


Figure 2. Forest Plot of 95%CI of adjusted OR.

4. Discussion

The results confirmed that the significant factors associated with diabetic nephropathy were male, age ≥ 60years, diabetic duration ≥ 10 years, systolic BP ≥ 130 mmHg, and high blood glucose level (HbA1c ≥ 6%) are the high magnitude and strongest

associated with DN in non-insulin-dependent diabetes mellitus patients. These results are similar to previous studies of Elmarakby et al. [1] and Alrawahi et al. [4]. High blood sugar levels can damage the small blood vessels of the nephron especially those who cannot control blood sugar for a long time [1,5]. In addition, the blood vessels of the elderly patients must be weekend by age [5] including quite a lot of patients with diabetes who also have high blood pressure, which can cause damage to blood vessels of the nephron [4]. All so, blood pressure is the strength of blood pushing against blood vessels, including in the kidneys. When the blood vessels in the kidneys are damaged, this stretching weakens blood vessels in the nephron, they cannot function to remove wastes and extra fluid from the body. However, smoking and alcohol consumption were not statistically significantly associated with DN could be due to inequalities in several factors including study design, source of population, definition, and suitable information.

5. Conclusions

The screening approaches should be emphasized especially in male patients and those with age more than 60 years including duration of diabetes ≥ 10 years. Control of blood sugar and systolic blood pressure should be included in health promotion programs for the patients to result in delaying nephropathy in non-insulin-dependent diabetes mellitus patients.

Acknowledgments

We would like to thank you the nurses and doctors at the Diabetic Care Centre in Mahachanachai Hospital, Thailand. This study was supported by the Faculty of Public Health, Mahasarakham University, and Mahachai Hospital, Thailand.

References

[1] Elmarakby AA, Sullivan JC. Relationship between oxidative stress and inflammatory cytokines in diabetic nephropathy. Cardiovasc Ther. 2012 Feb;30(1): 49-59.
[2] National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation,classification, and stratification. Am J Kidney Dis Off J Natl Kidney Found. 2002;39(2 Suppl 1):S1-266.
[3] Khiewkhern S, Yoosook W, Srichomphoo C, Thongkum W, Thitisutthi S. Risk Factors of Diabetic Nephropathy Development in Type 2 Diabetic Patients: A Cross-sectional Retrospective Study. Journal of Clinical and Diagnostic Research. 2021;15(2):7-12.
[4] Alrawahi AH, Rizvi SG, Al-Riyami D, Al-Anqoodi Z. Prevalence and risk factors of diabetic nephropathy in omani type 2 diabetics in Al-dakhiliyah region. Oman medical journal. 2012 May;27(3):212.
[5] Juraschek SP, Chang AR, Appel LJ, Anderson CAM, Crews DC, Thomas L, et al. Effect of glycemic index and carbohydrate intake on kidney function in healthy adults. BMC Nephrol. 2016 08;17(1):70.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210858

Public Sentiment Towards Vaccination After COVID-19 Outbreak in the Arab World

Rafiul BISWAS[a], Tanvir ALAM[a], Mowafa HOUSEH[a] and Zubair SHAH[a,1]
[a]College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

Abstract. Public perception about vaccines is imperative for successful vaccination programs. This study aims to measure the shift of sentiment towards vaccines after the COVID-19 outbreak in the Arab-speaking population. The study used vaccine-related Arabic Tweets and analyzed the sentiment of users in two different time frames, before 2020 (T1) and after 2020 (T2). The analysis showed that in T1, 48.05% of tweets were positive, and 16.47% of tweets were negative. In T2, 43.03% of tweets were positive, and 20.56% of tweets were negative. Among the Twitter users, the sentiment of 15.92% users shifted towards positive, and the sentiment of 17.90% users shifted towards negative. Public sentiment that have shifted towards positive may be due to the hope of vaccine efficacy, whereas public sentiment that have shifted towards negative may be due to the concerns related to vaccine side effects and misinformation. This study can support policymakers in the Arab world to combat the COVID-19 pandemic by utilizing tools to understand public opinion and sentiment.

Keywords. Sentiment analysis, Twitter, Arab, Arabic tweet, COVID-19

1. Introduction

Vaccinating people is thought to be one of the crucial factors that can mitigate the spread of the COVID-19 pandemic. With a large number of people being vaccinated, people can return to a more normal and likely cautious lifestyle as restrictions are eased [1]. Researchers and healthcare professionals worldwide are working to develop an effective vaccine to prevent novel coronavirus strains. However, the success of vaccination programs relies upon their public acceptance. Vaccination intentions are influenced by vaccine efficacy, side effects, mistrust in health care, political concern, health inequality in an ethnic minority and rapid development of vaccines [2]. Exposure to misinformation about the safety and effectiveness of vaccines through online social media often leads people to feel pessimistic and hesitant [3].

Twitter is one of the most popular microblogging social media platforms in the world [4]. Sentiment analysis based on tweets can measure public opinions, attitudes and emotions toward an entity [5]. However, it often becomes difficult to find out the exact meaning of the text when facing short and ambiguous Twitter messages. In this study, sentiment analysis was performed on vaccine-related Arabic tweets to observe public attitudes towards vaccines after the outbreak of COVID-19. It is hoped that this study

[1] Corresponding Author. Zubair Shah, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar. Email: zshah@hbku.edu.qa.

will support understanding the public sentiment about COVID-19 vaccines in the Arab world.

2. Methods

2.1 Data collection

Twitter Academic Research API [6] was used to access vaccines-related public tweets. Tweets were searched using key Arabic words. The search terms included "الطفولة" OR "لقاحات" OR "تطعيم" OR "تطعيمات" OR "لقاح". Python Tweepy library was used to download the data into the PostgreSQL database. The downloaded data was processed to extract tweets and their metadata such as ID, timestamp, text, author ID, number of likes, retweets and replies, and author profile information, including followers. The IDs of tweets were used as primary key of the table in the database to avoid duplicate entries of tweets.

2.2 Data Preprocessing

The downloaded data was preprocessed before the analysis. First, retweets were removed using string "RT @|RT@" that exists in tweet. Next, tweets of verified users were removed as these users are often government entities and celebrities which may not represent public opinion. After that, punctuation, links, mentions, hashtags, and stopwords were removed. Further, text was processed to remove emojies and special characters. Next, the data was separated into T1 (01 August 2009 to 31 December 2019) and T2 (01 January 2020 to 15 February 2021). Finally, users who posted at least one tweet in both periods were included in the study data. Figure 1 demonstrates the steps included in data preprocessing.

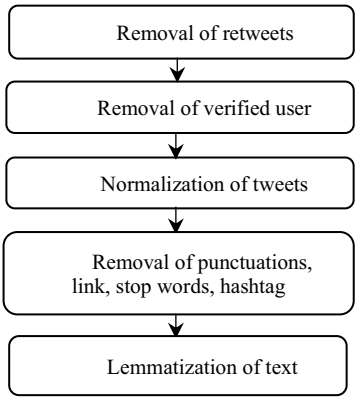


Figure 1. Data preprocessing

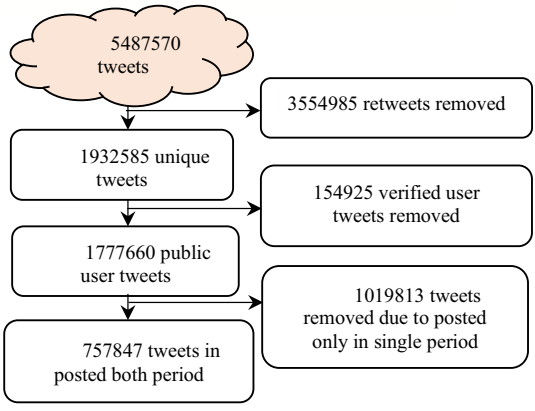


Figure 2. Flowchart of selection of tweets

2.3 Data Analysis

Python Textblob library [7] was used to find sentiment of each word within a range from -1 to +1. Tweets were labelled as positive (>0 to +1), negative (<0 to -1), and neutral

(exact 0). If all the words in a sentence are positive, the polarity is high. Similarly, if all the words in a sentence are negative, the polarity is low. If any sentence has both positive and negative words, polarity is calculated by the summing positive and negative scores. The sentiment shift of users is measured from T1 towards T2. The shift in sentiment is defined as Negative to Positive (N2P), Neutral to Positive (U2P), Negative to Neutral (N2U), Positive to Negative (P2N), Neutral to Negative (U2N) and Positive to Neutral (P2U). The positive shift in sentiment is measured by (N2P, U2P, N2U), and the negative shift in sentiment is measured by (P2U, P2N, U2N). No change in sentiment is measured by (U2U, P2P, N2N).

3. Results

3.1 Collection of Tweets after filtering steps

The search returned nearly 5.5M (5,487,570) tweets. It was found that most of the tweets were from the Kingdom of Saudi Arabia (KSA). People from Egypt, United Arab Emirates, Jordan, and Qatar also shared vaccine information through tweets. About 3.5M (3,554,985, 64.78%) retweets were removed, which left about 2M (1,932,585, 35.72%) unique tweets. A total of 154K (154,925, 8.01%) tweets were removed as they belonged to verified users. After these exclusions, about 1.78M (1,777,660) public tweets were left. After that, tweets from 74,240 users who posted at least one tweet in both time periods were selected, which were 757,847 tweets. Finally, sentiment analysis was performed on these selected tweets. Figure 2 illustrates the overall filtering steps.

3.2 Shift of Sentiment in Arabic Tweets

Table 1. Summary of Sentiment analysis results on Arabic tweets

Period	Positive	Negative	Neutral
T1	112933 (48.05%)	70897 (16.47%)	101359 (35.59%)
T2	203204 (43.03%)	97159 (20.56%)	172063 (36.41%)

Table 1 shows the result of the overall sentiment analysis in both periods. In T1, 112,933 (48.05%) tweets were positive, 70897 (16.44%) were negative and 101,359 (35.59%) were neutral. In T2, 203204 (43.03%) tweets were positive, 97,159 (20.56%) were negative and 172,063 (36.41%) were neutral. Among the 74,240 Twitter users, the sentiment of 15.92% (n=11819) users shifted towards positive from T1 to T2, with 5645 users shifted as U2P, 1,928 users shifted as N2P and 4,246 users shifted as N2U. Whereas the sentiment of 17.90% (n =13,294) users shifted towards negative from T1 to T2, with 3142 users shifted as P2N, 6,567 users shifted as P2U and 3,585 users shifted as U2N. The sentiment of 66.17% (n =49,127) users did not change (P2P, U2U, N2N) during these two periods.

4. Discussion

This study found that almost half of the tweets regarding vaccines were positive before COVID pandemic. However, a decrease in positive sentiment and an increase in negative

sentiment were observed after the COVID-19 outbreak. Users with positive sentiment mainly posted tweets about vaccine effectiveness, and they were optimistic about the role of government and healthcare authorities. Some of the users posted negative tweets regarding the new vaccines for COVID-19, and lack of informations about the side effects of the newly developed vaccines. Negative sentiment users also claimed that the overall vaccine effectiveness has not been studied and expressed concerns to be vaccinated. Some users were neutral in expressing their thought in both periods, and continuously posted neutral tweets. These neutral tweets mainly included vaccine development and vaccine distribution information. In general, users discussed vaccine efficacy and side effects in their tweets.

This study has a few limitations. The method calculated polarity based on linguistic analyses, which may not always be reliable. The tweets were collected in February 2021. As public sentiment may change over time, data collection, analysis and continuous monitoring are necessary to keep track on the latest sentiment.

Twitter dataset used in this study and other relavant files are available at online appendix https://github.com/rafiulbiswas/PSVA.

5. Conclusions

From a public health perspective, vaccination is the only way to fight against the COVID-19 pandemic. This study described how vaccination intention varied within two-time frames among the Arab-speaking people based on their native language tweets. The reasons for shifting sentiment downwards need to be analyzed broadly in future studies. We believe this study would support policymakers in taking the necessary steps to be more informed about developing a successful vaccination program in the Arab world.

References

[1] Williams L, Gallant AJ, Rasmussen S, Brown Nicholls LA, Cogan N, Deakin K, Young D, Flowers P. Towards intervention development to increase the uptake of COVID‐19 vaccination among those at high risk: Outlining evidence‐based and theoretically informed future intervention content. British Journal of Health Psychology. 2020 Nov;25(4):1039-54.

[2] Melton CA, Olusanya OA, Ammar N, Shaban-Nejad A. Public Sentiment and Topic Modeling Regarding COVID-19 Vaccines on Reddit Social Media Platform: A Call to Action For Strengthening Vaccine Confidence. Journal of Infection and Public Health.

[3] Sallam M, Dababseh D, Eid H, Al-Mahzoum K, Al-Haidar A, Taim D, Yaseen A, Ababneh NA, Bakri FG, Mahafzah A. High rates of COVID-19 vaccine hesitancy and its association with conspiracy beliefs: A study in Jordan and Kuwait among other Arab countries. Vaccines. 2021 Jan;9(1):42.

[4] Hawkins CM, Duszak R, Rawson JV. Social media in radiology: early trends in Twitter microblogging at radiology's largest international meeting. Journal of the American College of Radiology. 2014 Apr 1;11(4):387-90.

[5] Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal. 2014 Dec 1;5(4):1093-113.

[6] Twitter. Twitter API for academic research | twitter developer platform," https://developer.twitter.com/en/products/twitter-api/academic-research.(accessed on 29 August 2021).

[7] Adhaamehab. (n.d.). Adhaamehab/Textblob-Ar: Arabic support for textblob. GitHub. [Online]. Available from: https://github.com/adhaamehab/textblob-ar (accessed on 29August 2021).

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210859

ABiMed: Towards an Innovative Clinical Decision Support System for Medication Reviews and Polypharmacy Management

Abdelmalek MOUAZER[a], Romain LEGUILLON[a], Bertille LEROY[d], Karima SEDKI[a], Christian SIMON[b], Hector FALCOFF[c], Matthieu SCHUERS[d,a], Sophie DUBOIS[e], Cédric BOUSQUET[a], Lina SOUALMIA[f,a], Catherine LETORD[a], Stéfan DARMONI[a,g], Jacques BOUAUD[a], Brigitte SÉROUSSI[a], Rosy TSOPRA[h,i] and Jean-Baptiste LAMY[a,1]

[a] *Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé – LIMICS, Université Sorbonne Paris Nord, Sorbonne Université, INSERM UMR 1142, F-93000, Bobigny, France*
[b] *EIG Santé, 49070 Beaucouzé, France*
[c] *Société de Formation Thérapeutique du Généraliste (SFTG), Paris, France*
[d] *Département de Médecine Générale, Université de Rouen, France*
[e] *CPTS Paris 13, 75013 Paris, France*
[f] *Normandie Univ, UNIROUEN, LITIS EA 4108, F-76000 Rouen, France*
[g] *Department of Biomedical Informatics, Rouen University Hospital, France*
[h] *INSERM, Université de Paris, Sorbonne Université, Centre de Recherche des Cordeliers, Information Sciences to support Personalized Medicine, F-75006 Paris, France*
[i] *Department of Medical Informatics, Hôpital Européen Georges-Pompidou, AP-HP, Paris, France*

Abstract. Polypharmacy in elderly is a public health problem with both clinical (increase of adverse drug events) and economic issues. One solution is medication review, a structured assessment of patients' drug orders by the pharmacist for optimizing the therapy. However, this task is tedious, cognitively complex and error-prone, and only a few clinical decision support systems have been proposed for supporting it. Existing systems are either rule-based systems implementing guidelines, or documentary systems presenting drug knowledge. In this paper, we present the ABiMed research project, and, through literature reviews and brainstorming, we identified five candidate innovations for a decision support system for medication review: patient data transfer from GP to pharmacists, use of semantic technologies, association of rule-based and documentary approaches, use of machine learning, and a two-way discussion between pharmacist and GP after the medication review.

Keywords. Polypharmacy, Medication review, Clinical decision support system

[1] Corresponding Author, Jean-Baptiste LAMY; E-mail: jean-baptiste.lamy@univ-paris13.fr.

1. Introduction

Many elderly patients suffer from multiple chronic disorders and receive polypharmacy (*i.e.* 5 drugs or more). It is both a public health and an economic issue. Each new drug administered in polypharmacy increases the risk of adverse events by 12-18% [1], *e.g.* recently, many drugs (including anti-inflammatory) increased the symptoms of Covid-19. Both the drugs and the hospitalization caused by the adverse events have a high cost for health insurances. Finally, reducing polypharmacy is also beneficial for achieving the UNO goal of a more sustainable consumption.

One solution is medication reviews (MR) [2], a structured assessment of patients' drug orders for optimizing the drug use and to improve health outcomes. In MR, the pharmacist assesses the patient observance and tolerance, and wrote a synthesis with recommendations for the GP. MR can reduce polypharmacy and save money, without lowering the quality of care [3]. However, few pharmacists are engaged in MR, because it is a tedious and highly cognitive task that requires specific knowledge in geriatrics. It requires to collect the patient data (often available in the GP's patient records, but not in the pharmacist's ones), to assess the interactions and adverse effects of 5-20 drugs, to identify inappropriate or missing drugs (*e.g.* by manually applying the numerous rules in the existing recommendations [4]) and to write the synthesis.

Few clinical decision support systems (CDSSs) were proposed for MR. Rule-based systems aim at detecting inadequate prescriptions, either by implementing guidelines [5] or by letting users build their rule base [6]. Documentary systems provide information on demand. They typically present the drugs the patient takes on a table, and displays their interactions and potential adverse effects, *e.g.* [7]. Semantic web technologies are known to improve CDSSs [8], but have not yet been applied to MR.

ABiMed (*Aide au Bilan de Médication*, medication review support in English) is a research project funded by the French research agency. The objectives of ABiMed are to design and evaluate a clinical decision support system for facilitating medication reviews and reducing polypharmacy, which combines rule-based and documentary approaches. The system is destined to pharmacists (for MR) and GPs (for preventing polypharmacy when prescribing). In this paper, we present the project and the five candidate innovations we identified for decision support for medication reviews.

2. Methods

Candidate innovations for a CDSS devoted to MR and polypharmacy were identified through literature reviews and multidisciplinary collective brainstorming sessions. Two systematic literature reviews were performed, the first targeting guidelines and recommendations for MR and the second targeting CDSSs for MR. These reviews are still under progress before publication.

Brainstorming sessions involved all partners of the project: an academic research lab in medical informatics (LIMICS), an enterprise developing medical software (EIG Santé), and two clinical partners: an association performing continuing medical education (SFTG) and a territorial professional health community (CPTS 13). Participant backgrounds include pharmacy, primary care, hospital medicine, and computer science; with several persons having a double competence. Due to Covid-19, meetings were performed online.

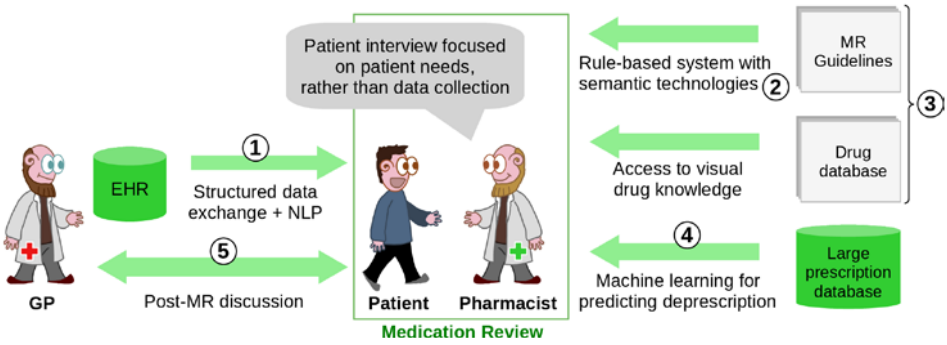


Figure 1. General overview of ABiMed (numbers match those in the text).

3. Results

We identified five candidate innovations that we will include in our CDSS (Figure 1):

(1) We will automatically extract patient data from GP electronic health records (EHR) and transfer it to the pharmacist. Current CDSSs are not linked to EHR when they target pharmacists, due to the wide number of EHR vendors. We will rely on the French shared health records (DMP) for accessing data, and on natural language processing (NLP) tools [9] for extracting concepts from plain text medical reports.

(2) We will use semantic technologies, such as formal ontologies and the SPARQL query language. They facilitate the management of several granularity levels, *e.g.* specific disorders in patient data *vs* more general disorder categories in rules from guidelines.

(3) We will combine a rule-based system implementing the recommendations from guidelines for MR with a documentary system presenting drug knowledge. It will rely on the visual analytics we developed previously for clinical trial data [10]. Preliminary results on the visualization of drug-drug interactions with graphs are promising [11].

(4) We will also rely on machine learning for learning deprescribing behaviors from large prescription databases such as the reimbursement data of the French health insurance. The objective is to learn real-life deprescription rules, possibly different from those in the guidelines. For instance, guidelines recommend stopping benzodiazepines, but GPs said that patients often ask for them at the next consultation. Thus, we may learn that benzodiazepines are not the best candidates for deprescription.

(5) We will experiment a two-way discussion between the pharmacist and the GP after the MR, instead of the traditional one-way communication mode, through the textual synthesis sent by the pharmacist to the GP.

4. Discussion and conclusion

Through literature reviews and brainstorming, we identified five innovations that have the potential of improving decision support for MR and polypharmacy management. There is a mix of technical innovation (*e.g.* 4), combination of existing technics (3) or application of known technics to MR (1, 2), and organizational innovation (5). These innovations still have to be properly implemented and evaluated in clinical settings. The

ABiMed project includes both *in vitro* and *in vivo* evaluation, aimed at showing a positive impact of the proposed CDSS on medical prescriptions.

In the literature, EHR data extraction was proposed for CDSSs for physicians, but rarely for pharmacists, because the data is not at the pharmacist office. Semantic technologies are commonly used for CDSSs, but they have rarely been implemented for MR. A few systems tried to associate rule-based and documentary approaches, such as PRIMA-eDS [12]. Finally, no machine learning approaches have been proposed for MR and very few for drug therapy in general, despite a recent study showed that machine learning was better than traditional methods (*i.e.* if-then rules implemented from guidelines) for the automatic detection of at-risk prescriptions at hospital [13].

Acknowledgments

This work was funded by the French Research Agency (ANR) through the ABiMed project [grant number ANR-20-CE19-0017-02].

References

[1] Calderón-Larrañaga A, Poblador-Plou B, González-Rubio F, et al. Multimorbidity, polypharmacy, referrals, and adverse drug events: are we doing things well? The British journal of general practice: the journal of the Royal College of General Practitioners. 2012;62(605):e821-6.
[2] Silva RDOS, Macêdo LA, Santos GAD, Aguiar PM, de Lyra DP. Pharmacist-participated medication review in different practice settings: Service or intervention? An overview of systematic reviews. PloS one. 2019;14(1):e0210312.
[3] Zermansky A G, Petty D R, Raynor D K, Freemantle N, Vail A, Lowe C J. Randomised controlled trial of clinical medication review by a pharmacist of elderly patients receiving repeat prescriptions in general practice. BMJ (Clinical research ed). 2001;323(7325):1340-3.
[4] Roux B, Berthou-Contreras J, Beuscart JB, Charenton-Blavignac M, Doucet J, Fournier JP, et al. REview of potentially inappropriate MEDIcation pr[e]scribing in Seniors (REMEDI[e]S): French implicit and explicit criteria. Eur J Clin Pharmacol. 2021.
[5] Meulendijk MC, Spruit MR, Drenth-van Maanen AC, Numans ME, Brinkkemper S, Jansen PAF, et al. Computerized Decision Support Improves Medication Review Effectiveness: An Experiment Evaluating the STRIP Assistant's Usability. Drugs & aging. 2015;32(6):495-503.
[6] Bindoff I, Stafford A, Peterson G, Kang B H, Tenni P. The potential for intelligent decision support systems to improve the quality and consistency of medication reviews. J Clin Pharm Ther. 2012;37(4):452-8.
[7] Koskela T, Sandström S, Mäkinen J, Liira H. User perspectives on an electronic decision-support tool performing comprehensive medication reviews - a focus group study with physicians and nurses. BMC medical informatics and decision making. 2016;16:6.
[8] Dissanayake PI, Colicchio TK, Cimino JJ. Using clinical reasoning ontologies to make smarter clinical decision support systems: a systematic review and data synthesis. J Am Med Inform Assoc. 2019.
[9] Siefridt C, Grosjean J, Lefebvre T, Rollin L, Darmoni S, Schuers M. Evaluation of automatic annotation by a multi-terminological concepts extractor within a corpus of data from family medicine consultations. Int J Med Inf. 2020;133:104009.
[10] Lamy JB. A data science approach to drug safety: Semantic and visual mining of adverse drug events from clinical trials of pain treatments. Artif Intell Med. 2021;115:102074.
[11] Mouazer A, Sedki K, Tsopra R, Lamy JB. Visualization of drug interactions for supporting medication review. In: Studies in health technology and informatics (ICIMTH). vol. 272; 2020. p. 107-10.
[12] Rieckert A, Sommerauer C, Krumeich A, Sönnichsen A. Reduction of inappropriate medication in older populations by electronic decision support (the PRIMA-eDS study): a qualitative study of practical implementation in primary care. BMC family practice. 2018;19(1):110.
[13] Corny J, Rajkumar A, Martin O, Dode X, Lajonchère JP, Billuart O, et al. A machine learning-based clinical decision support system to identify prescriptions with a high risk of medication error. J Am Med Inform Assoc. 2020;27(11):1688-94.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210860

Machine Learning Approaches for Early Prostate Cancer Prediction Based on Healthcare Utilization Patterns

Joseph FINKELSTEIN [1], Wanting CUI, Tiphaine C. MARTIN and Ramon PARSONS
Icahn School of Medicine at Mount Sinai, New York, NY, USA

Abstract. The goal of this study was to build a machine learning model for early prostate cancer prediction based on healthcare utilization patterns. We examined the frequency and pattern changes of healthcare utilization in 2916 prostate cancer patients 3 years prior to their prostate cancer diagnoses and explored several supervised machine learning techniques to predict possible prostate cancer diagnosis. Analysis of patients' medical activities between 1 year and 2 years prior to their prostate cancer diagnoses using XGBoost model provided the best prediction accuracy with high F1 score (0.9) and AUC score (0.73). These pilot results indicated that application of machine learning to healthcare utilization patterns may result in early identification of prostate cancer diagnosis.

Keywords. Prostate Cancer, Big Data Analytics, Machine Learning

1. Introduction

Early discovery of cancer has crucial ramifications on the disease prognosis. In the recent meta-analysis of seven major cancers, even a 4-week delay in cancer treatment was associated with increased mortality across systemic treatment, surgical and radiotherapy indications [1]. A number of observational studies reported differing patterns in healthcare utilization before and after cancer diagnosis [2-3]. Significant differences were found in healthcare utilization patterns preceding cancer diagnosis as compared to matched non-cancer patients [4] including patients with prostate cancer [5]. Previous studies demonstrated utility of claims-based approaches for building predictive models in prostate cancer [6]. Based on these resent reports, we hypothesized that healthcare consumption preceding a diagnosis of prostate cancer may exhibit specific patterns which can be used for early cancer prediction. As machine learning approaches have been shown to be particularly instrumental in identifying characteristic data patterns in building predictive models, supervised machine learning techniques have been employed in this project. Thus, the goal of this pilot study was to analyze the frequency and pattern changes of patients' medical activities 3 years prior to their prostate cancer diagnoses, and to build machine learning models based on their medical activities to predict possible prostate cancer diagnosis within the near future.

[1] Joseph Finkelstein, Icahn School of Medicine at Mount Sinai, 1425 Madison Ave, New York, NY, USA, 10029; E-mail: Joseph.Finkelstein@mssm.edu.

2. Method

A de-identified analytical dataset has been constructed from electronic health record at Mount Sinai Health System in New York City comprising patients who were diagnosed with prostate cancer between 01/2009 and 12/2019. Since we aimed to monitor patients' medical activity frequency and pattern prior to their prostate cancer diagnoses, we extracted these patients' medical activities such as medical procedures, lab tests and radiology, 3 years prior to their cancer diagnoses. We only included patients who have at least 3 procedures or tests.

In predictive modeling, we used 84 predictive variables, which spanned through 3 groups: lab test, radiology and procedures representing the entire spectrum of care utilization. All predictive variables were continuous, and the value indicated the number of times a patient had done a test. XGBoost model was used to select most informative features out of the initial 187 types of lab tests. We included all features where the feature importance was over 0.05. In the end, 67 variables were selected. We also summarized lab tests into 4 parent groups using standard Logical Observation Identifier Names and Codes (LOINC) codes. The 4 parents groups are: Chem_drug_tox_chal_sero_allergy, urine, MassMolConc and CellDiffCoun. There were 5 variables related to radiology: X-Ray, MRI, CT, PET/CT and ultrasound. In addition, we also categorized all procedures into 8 groups using Current Procedural Terminology (CPT) codes. The procedure groups we included were: chemistry procedures, hematology procedures, organ disease panel, urinalysis procedures, immunology procedures, microbiology procedures, therapeutic drug assays and cardiovascular procedures.

We constructed 4 datasets with different time cut off points for target variables. The target variable was binary with indication of cancer. In the first dataset, we defined lab tests or procedures performed greater than 2 years and less than 3 years prior to diagnoses, as the timeframe for no cancer. We defined tests or procedures performed within 1 year of cancer diagnoses as time period with cancer. In the second dataset, we defined lab tests or procedures performed greater than 1 year and less than 3 years prior to diagnoses, as the timeframe for no cancer; and tests or procedures performed within 1 year of cancer diagnoses as time period with cancer. In the third dataset, we defined lab tests or procedures performed greater than 2 year and less than 3 years prior to diagnoses, as the timeframe for no cancer; and tests or procedures performed within 2 years of cancer diagnoses as time period with cancer. In the last dataset, we defined lab tests or procedures performed greater than 1 year and less than 2 years prior to diagnoses, as the timeframe for no cancer; and tests or procedures performed within 1 year of cancer diagnoses as time period with cancer.

In model training, each dataset was randomly divided into 80% training and 20% testing. We compared 3 machine learning models: Support Vector Machine (SVM), Random Forest (RF) and XGBoost (XGB). We tuned hyper parameters and performed 3-fold cross validation to choose the best hyper parameters using the training set. In the end, we applied the best-tuned model for each algorithm to the testing set and calculated accuracy, precision, recall, F1 score and area under the curve (AUC) accordingly.

3. Results

There are 2916 number of records in the first dataset; 724 patients had records of lab tests and procedures that are greater than 2 years and less than 3 years prior to the

diagnoses; and 2174 patients had records of tests and procedures within 1 year of prostate cancer diagnoses. XGBoost model performed the best (Table 1), since it has the highest F1 score (0.9) and AUC score (0.73).

In the second dataset, 1108 patients had medical activities greater than 1 years and less than 3 years prior to the prostate cancer diagnoses. In contrast 2159 patients had medical activities within 1 year of cancer diagnoses. Both XGBoost model and random forest model performed well in this subset (Table 1), as they both have high AUC score (0.69) and F1 score (0.82).

In the third dataset, 742 patients had medical activities greater than 2 years and less than 3 years prior to their cancer diagnoses, and 2196 patients had medical activities within 2 years of their diagnoses. According to Table 1, although SVM model produced the highest AUC score (0.74), the F1 score (0.83) and recall (0.77) were both low, compared to the other 2 models (F1 score = 0.88, recall 0.92).

In the fourth dataset, 918 patients had medical activities greater than 1 year and less than 2 years prior to their diagnoses, and 2174 patients had medical activities within 1 year of their cancer diagnoses. All 3 models produced the AUC score (0.68), with random forest generated the highest F1 score (0.86).

Table 1. Results of predictive models.

	Accuracy	Precision	Recall	F1	AUC
Dataset 1					
XGB	0.84	0.87	0.93	0.90	0.73
SVM	0.72	0.89	0.73	0.80	0.72
RF	0.83	0.85	0.95	0.89	0.70
Dataset 2					
XGB	0.75	0.80	0.84	0.82	0.69
SVM	0.67	0.83	0.65	0.73	0.68
RF	0.75	0.80	0.84	0.82	0.68
Dataset 3					
XGB	0.8	0.84	0.92	0.88	0.69
SVM	0.76	0.89	0.77	0.83	0.74
RF	0.8	0.84	0.92	0.88	0.69
Dataset 4					
XGB	0.77	0.80	0.90	0.85	0.68
SVM	0.69	0.83	0.70	0.76	0.68
RF	0.79	0.80	0.94	0.86	0.68

4. Discussion

Overall, models in the first dataset has the highest AUC scores, and models in the fourth dataset had the lowest AUC scores. The time cut off point for the first dataset was within 1 year for cancer and greater than 2 years and less than 3 years prior to the diagnosis for no cancer in this time period. In contrast the time cut off point for the last dataset was within 1 year for cancer and greater than 1 year and less than 2 years prior to the diagnosis for no cancer. Thus, patterns of patients' medical activities were more likely to change

between 1 year and 2 years prior to their prostate cancer diagnoses. In addition, the frequency of patients using medical services increased significantly when closer to the diagnoses.

XGBoost model performed well for all four datasets. Although the SVM model from the third dataset had the highest AUC score, this model produced relatively low F1 score and recall. Since we aimed to find the best overall model, we selected the XGBoost model from the first dataset. It has the highest F1 score (0.9) and second highest AUC score (0.73) among all models. By examining the feature importance of this model, we found that PSA (Prostate-Specific Antigen) test, microbiology procedures, chemistry procedures, organ disease panel and aPTT (activated partial thromboplastin time) blood test were the top 5 factors, which indicates that a change of these tests and procedures' frequency and patterns was highly associated with possible prostate cancer diagnoses within 1 or 2 years.

In future studies, we plan to optimize our predictive features. We will explore various methodologies to summarize LOINC codes and CPT codes. And we will also examine medication intake prior to cancer diagnoses. In addition, we will expand our studies to 5 years prior to patients' prostate cancer diagnoses and explore various time cut off points in relationship to the pattern of patients' medical activities.

5. Conclusion

In this study, we examined the frequency and pattern changes of patients' medical activities 3 years prior to their prostate cancer diagnoses and built machine learning models based on their medical activities to predict possible prostate cancer diagnosis within the near future. XGBoost model from the first dataset performed the best, with high F1 score (0.9) and AUC score (0.73). Frequency and patterns of patients' medical activities would change between 1 year and 2 years prior to their prostate cancer diagnoses. The results indicated that further exploration of this approach is warranted.

Reference

[1] Hanna TP, King WD, Thibodeau S, Jalink M, Paulin GA, Harvey-Jones E, O'Sullivan DE, Booth CM, Sullivan R, Aggarwal A. Mortality due to cancer treatment delay: systematic review and meta-analysis. BMJ. 2020;371:m4087.

[2] Shen C, Dasari A, Xu Y, Zhou S, Gu D, Chu Y, Halperin DM, Shih YT, Yao JC. Pre-existing Symptoms and Healthcare Utilization Prior to Diagnosis of Neuroendocrine Tumors: A SEER-Medicare Database Study. Sci Rep. 2018;8(1):16863.

[3] Jones LE, Doebbeling CC. Primary care utilization patterns before and after lung cancer diagnosis. Eur J of Cancer Care 2009;18:165-173.

[4] Park J, Look KA. Health Care Expenditure Burden of Cancer Care in the United States. Inquiry. 2019;56:46958019880696.

[5] Sun M, Marchese M, Friedlander DF, et al. Health care spending in prostate cancer: An assessment of characteristics and health care utilization of high resource-patients. Urol Oncol. 2021 Feb;39(2):130.e17-130.e24.

[6] Riviere P, Tokeshi C, Hou J, Nalawade V, Sarkar R, Paravati AJ, Schiaffino M, Rose B, Xu R, Murphy JD. Claims-Based Approach to Predict Cause-Specific Survival in Men with Prostate Cancer. JCO Clin Cancer Inform. 2019;3:1-7.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210861

Semiautomatic Identification of Pulmonary Embolism in Electronic Health Records Through Sentence Labeling

Gleb DANILOV[a,1], Timur ISHANKULOV[a], Alexandra KOSYRKOVA[a], Maria SHULTS[a], Semen MELCHENKO[a], Tatyana TSUKANOVA[a], Michael SHIFRIN[a] and Alexander POTAPOV[a]
[a] *Laboratory of Biomedical Informatics and Artificial Intelligence, National Medical Research Center for Neurosurgery named after N.N. Burdenko, Moscow, Russian Federation*

Abstract. In this study, we tested the quality of the information extraction algorithm proposed by our group to detect pulmonary embolism (PE) in medical cases through sentence labeling. Having shown a comparable result (F1 = 0.921) to the best machine learning method (random forest, F1 = 0.937), our approach proved not to miss the information of interest. Scoping the number of texts under review down to distinct sentences and introducing labeling rules contributes to the efficiency and quality of information extraction by experts and makes the challenging tasks of labeling large textual datasets solvable.

Keywords. Natural Language Processing, Pulmonary Embolism, Neurosurgery, Machine Learning

1. Introduction

The reliability of classification or regression by machine learning based on natural language processing (NLP) largely depends on the primary textual data quality and the target variable's validity. Data labeling in medicine is usually performed by medical experts. The primary extraction of information from a body of unstructured narrative texts is a non-trivial task, especially for non-professionals in NLP [1]. Extracting information from tens of thousands of case records, even by a group of experts, can take an unreasonably enormous amount of time, and the reliability of the final result is still questionable. Our previous study demonstrated a moderate level of agreement between experts assessing the same medical records [2]. Therefore, the parallel extraction of information from one massive textual source by different healthcare professionals necessitates resolving the disagreements, which is also laborious.

Our group has previously proposed an algorithm for semi-automatic information extraction (IEA, *information extraction algorithm*) from unstructured texts written in natural language (patent RU2751993C1) [3]. The essence of the algorithm is in the decoding of distinct phrases - microcontexts (for example, n-grams) containing words

[1] Corresponding Author, Gleb Danilov, N.N. Burdenko Neurosurgery Center, 4th Tverskaya-Yamskaya str. 16, Moscow 125047, Russian Federation; E-mail: glebda@yandex.ru.

from the lexicon, used to describe the sought phenomenon. When retrieving information about relatively rare facts, the size of the micro contexts can be expanded to complete sentences. Thus, significantly narrowing the scope of the texts to analyze down to certain sentences dramatically saves time and other costs for specific information extraction. We assume it is possible to extract the information on various phenomena related to a medical case summarizing the set of distinct sentences labeled with our algorithm. In this study, we assessed the quality of information extraction about pulmonary embolism (PE) following neurosurgical interventions using the proposed method of information extraction based on sentences labeling.

2. Methods

The data for the study were obtained from the Electronic Health Records (EHR) of the National Medical Research Center of Neurosurgery, named after academician N.N. Burdenko (Moscow, Russia) for the period between 2000 and 2017 (90 688 cases). We queried the database with keywords related to the PE lexicon in Russian. Thus, the clinical narrative texts in which PE was mentioned or suspected were retrieved. The clinical corpus was constructed with postoperative daily notes, examinations, postmortem findings, and other relevant documents typed on a keyboard with free text. All health records were screened by three similarly skilled neurosurgeons independently. The experts with an overall working experience of 7-11 years had been previously trained in neurosurgical residency at N.N. Burdenko Neurosurgery Center. These coders had small comparable experience in labeling data from approximately 2000 health records. However, no expert was specially trained for PE detection task. The texts were presented to the experts in a special software designed to focus on medical texts only. Each expert was asked to label the cases with either “PE detected” or “No PE detected” or “the fact of PE could not be well-verified.” The disagreements between the experts were resolved with the involvement of the fourth expert.

All the records were also independently screened by the first author of this article, with the methodology described in our previous publications [2–5]. The text was tokenized into sentences and then into lemmatized words. The unique lemmas were screened to select those likely related to or certainly used in the PE description. All the sentences containing the initial words matching all selected lemmas were then reviewed to label with “PE detected” or “No PE detected” or “the fact of PE could not be well-verified” categories and score by 1, 0 and 0.5, respectively. To apply our information extraction algorithm (IEA) as described previously, we labeled the whole medical case with “PE detected” if at least one sentence in a case was assigned with the “PE detected” label. We marked the case with the “the fact of PE could not be well-verified” tag if the same sentence label occurred and no "PE detection" labels were met. We rejected PE in a case if no labeled sentences were present or “No PE detected” was the only sentence label found.

We also decided to test whether it is possible to improve decision-making on PE detection in the entire medical case with labeled sentences using machine learning. To accomplish this, the scores of labeled sentences were further summarized for each clinical case into a set of 33 parameters (maximum scores for each type of clinical document, maximum score for all documents, the simultaneous 0.5 and 1 maximums in different documents of the same case). The resulting matrix of 621 rows (clinical cases) and 34 columns (aggregated numeric features and the target variable - the

presence/absence/uncertainty of PE, agreed between experts) was then used to train and test machine learning algorithms to classify cases by PE labels.

A total of 6 machine learning models were tested: random forest (RF), logistic regression (LR), support vector machine (SVM) with different kernel types (linear (lin), radial (rad), and polynomial (poly)), and K-nearest neighbors (KNN). Each model was trained in 300 experiments with resampling. We kept the training/testing sampling ratio as 80% / 20% and applied stratification by the target variable. The average sensitivity (SENS, also referred to as recall), specificity (SPEC), accuracy (ACC), positive predictive value (PPV, also referred to as precision), and F-score (F1) were computed for each model across all the experiments. To fairly compare the machine learning results with our rule-based PE detection (IEA), we calculated these indicators for the entire set of cases and reproduced the average metrics for our algorithm performance on 300 testing samples simulated as 20% of the initial sample randomly with stratification.

3. Results

A set of documents for 621 medical cases was obtained from EHR for the study and reviewed by three neurosurgeons. A total of 3161 sentences containing preselected PE lexicon was labeled by the fourth expert independently. The results of the IEA application for PE detection compared to machine learning performance are demonstrated in Table 1.

Table 1. The quality metrics of the rule-based algorithm (IEA) proposed to detect PE in medical records by sentence labeling compared to machine learning over labeled sentences.

Model	SENS	SPEC	PPV	ACC	F1
IEA (full set)	0.966	0.974	0.891	0.936	0.921
IEA (resampling)	0.967	0.975	0.893	0.937	0.921
RF	0.959	0.976	0.920	0.950	0.937
SVM (lin)	0.960	0.974	0.920	0.949	0.937
SVM (poly)	0.917	0.959	0.909	0.934	0.911
SVM (rbf)	0.739	0.903	0.859	0.873	0.759
LR	0.720	0.897	0.806	0.864	0.729
KNN	0.602	0.838	0.726	0.797	0.624

A confusion matrix for PE detection with the proposed IEA in the entire set of cases is shown in Figure 1.

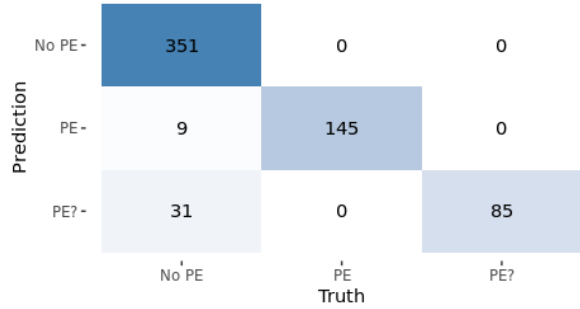


Figure 1. A confusion matrix to assess PE detection quality with the proposed IEA. “PE” = “PE detected”; “No PE” = “No PE detected”; “PE?” = “the fact of PE could not be well-verified”.

4. Discussion

We found the PE lexicon in only 621 cases out of 90688, proving PE rarity in neurosurgery (<1%) [6]. RF and SVM (lin) methods slightly outperformed our algorithm in metrics (F1 = 0.937 vs. 0.921). However, as one can notice from Fig. 1, no false rejections on PE were met, which means the algorithm is highly specific not to miss any case of interest. This fundamental property becomes especially important when searching for rare events through tens and hundreds of thousands of cases, the overwhelming majority of which does not contain the required information a priori. Moreover, that minor part of cases falling under the "suspicion" of the algorithm can be verified by a team of experts in a reasonable time. Our method demonstrates high quality considering the results of other authors [7,8]. This approach enables reliable information extraction from thousands of case histories and shaping trustworthy target variables, significantly reducing time costs. For example, assuming that a set of 621 medical cases we studied contain most PEs detected in 2000 - 2017 at N.N. Burdenko Neurosurgery Center, applying IEA to the entire database, enables labeling all the 90 688 cases at a little added time cost, which is a significant achievement.

5. Conclusions

Scoping the number of texts under review down to distinct sentences and introducing labeling rules contributes to the efficiency and quality of information extraction by experts and makes the challenging tasks of labeling large textual datasets solvable. *This project was supported by the RFBR grant 18-29-22085.*

References

[1] Spasic I, Nenadic G. Clinical Text Data in Machine Learning: Systematic Review. JMIR Med. Informatics 2020; 8:. doi:10.2196/17984.

[2] Danilov G, Kosyrkova A, Shults M, Melchenko S, Tsukanova T, Shifrin M, Potapov A. Inter-Rater Reliability of Unstructured Text Labeling: Artificially vs. Naturally Intelligent Approaches. Stud. Health Technol. Inform. 2021;281:118–122. doi:10.3233/SHTI210132.

[3] Danilov G, Shifrin M, Strunina U, Pronkina T, Potapov A. An Information Extraction Algorithm for Detecting Adverse Events in Neurosurgery Using Documents Written in a Natural Rich-in-Morphology Language. Stud. Health Technol. Inform. 2019; 262: 194–197. doi:10.3233/SHTI190051.

[4] Danilov G, Shifrin M, Strunina Y, Kotik K, Tsukanova T, Pronkina T, Ishankulov T, Makashova E, Kosyrkova A, Melchenko S, Zagidullin T, Potapov A. Semiautomated approach for muscle weakness detection in clinical texts, in: Stud. Health Technol. Inform., IOS Press, 2020;55–58. doi:10.3233/SHTI200492.

[5] Danilov G, Shifrin M, Strunina Y, Kotik K, Tsukanova T, Pronkina T, Ishankulov T, Makashova E, Kosyrkova A, Potapov A. Detection of muscle weakness in medical texts using natural language processing, in: Stud. Health Technol. Inform., IOS Press, 2020;163–167. doi:10.3233/SHTI200143.

[6] Khan NR, Patel PG, Sharpe JP, Lee SL, Sorenson J. Chemical venous thromboembolism prophylaxis in neurosurgical patients: An updated systematic review and meta-analysis. J. Neurosurg. 2018;129:906–915. doi:10.3171/2017.2.JNS162040.

[7] Tvardik N, Kergourlay I, Bittar A, Segond F, Darmoni S, Metzger M-H. Accuracy of using natural language processing methods for identifying healthcare-associated infections. Int. J. Med. Inform. 2018;117:96–102. doi:10.1016/j.ijmedinf.2018.06.002.

[8] Shen F, Larson DW, Naessens JM, Habermann EB, Liu H, Sohn S. Detection of Surgical Site Infection Utilizing Automated Feature Generation in Clinical Notes. J. Healthc. Informatics Res. 2019;3:267–282. doi:10.1007/s41666-018-0042-9.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210862

Utilizing Shared Big Data to Identify Liver Cancer Dedifferentiation Markers

Kirill BORZIAK[a,1] and Joseph FINKELSTEIN[a]
[a] *Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, New York 10029 USA*

Abstract. Big data reanalysis has the potential to generate novel comparative analyses which aim to generate novel hypotheses and knowledge. However, this approach is underutilized in the realm of cancer research, particularly for cancer stem cells (CSCs). CSCs are a rare subset of tumor cells, which dedifferentiate from healthy adult cells, and have the potential for self-renewal and treatment resistance. This analysis utilizes two publically available single-cell RNA-seq datasets of liver cancer and adult liver cell types to demonstrate how reanalysis of big data can lead to valuable new discoveries. We identify 519 differentially expressed genes between liver CSCs and healthy liver cell types. Here we report the potential novel liver CSC dedifferentiation factor, Msh Homeobox 2, which was significantly upregulated in liver CSCs by 1.36 fold (p-value < 1E-10). These findings have the potential to further advance our knowledge of genes governing the formation of CSCs.

Keywords. Liver, cancer stem cells, single cell transcriptomics, dedifferentiation

1. Introduction

The improvements in high-throughput genomics through the funding of National Institutes of Health and other agencies are driving ever increasing volumes of digital data to be available in the public domain [1]. As the number of peer-reviewed publications that utilize previously deposited data continues to grow, the importance of shared big data in catalyzing knowledge discovery and predictive analytics continues to be further illustrated. Specifically, cancer stem cells (CSC) are a promising target for research which utilizes publically available single cell sequencing archives, due to the needs of isolating CSCs from the surroundings [2]. As one of the leading frontiers of oncology, the study of CSCs provides an interesting insight into the causes of therapeutic resistance, metastasis, and tumor recurrence [3]. Although much work has been done on identifying how the similarities between CSCs and stem cells promote their abilities to self-renew, less is known about how CSCs arise from terminally differentiated cell types and about the key triggers that initiate their stemness [3].

Although a lot research has been dedicate to the origins of CSCs, it is still not entirely clear how adult terminally differentiated cell types are able to transform into cancer cells [3]. In particular, it has been previously shown that adult terminally-differentiated hepatocytes undergo dedifferentiation which then leads to hepatocellular carcinoma [4]. The exact process of dedifferentiation which leads to the establishment of a CSC

[1] Corresponding Author, Kirill Borziak, Icahn School of Medicine, One Gustave L. Levy Place, New York, New York 10029 USA; E-mail: Kirill.Borziak@mountsinai.org.

population, however, remains to be elucidated. To help answer this question, we undertook a reanalysis of publically available single-cell RNA-seq datasets from primary liver cancer samples and healthy adult samples [5, 6]. Due to the deadliness of liver cancer and the difficulty of detecting and treating primary liver cancers, it is important to understand how treatment of liver CSCs can be utilized to promote efficacy of treatment [7]. Using the cost-effective approach of reanalyzing publically available datasets, we examined the differences in expression between liver CSCs and healthy adult liver cell types, with a focus on understanding the dedifferentiation process.

2. Methods

Expression data was obtained for the liver cancer [5] and fetal and adult healthy liver [6] studies. For the purposes of our reanalysis study, gene expression data count matrices were utilized as the starting point. This consideration was done due to the unavailability of the raw data from the liver cancer study, as it is currently still under embargo. As the two studies utilized similar approaches to sequencing, assembly, and gene calling, we do not expect any systematic issues in the gene expression profiles, which cannot be controlled for using the stringent normalization we employed.

The following studies were utilized for liver cancer cell profiles (GSE125449) [5] and healthy liver cell profiles (GSE130473) [6]. Pre-processing normalization steps included filtering out low coverage samples, low coverage genes, and non-protein coding genes. Samples with fewer than 1000 total reads were excluded. In addition to excluding genes with 0 reads in all remaining samples, we excluded all non-protein coding genes.

The normalization and differential expression analysis was performed using the edgeR [8] R package, using the gold standard methodology. To further account for possible systematic differences in mRNA detection between the two datasets, we utilized a batch effect correction in the analysis. Study type was also included in the design matrix as an additional variable. The Bonferroni multiple testing correction was used to control the false discovery rate.

Gene Ontology (GO) analysis was done using the DAVID 6.8 [9] Functional Annotation Tool, using Benjamini multiple testing correction.

3. Results

Using previously published single-cell RNA-seq data for liver cancer [5] and fetal and adult healthy liver [6], we reanalyzed 2434 single-cell samples across 18,263 protein-coding genes. To examine the differences between adult liver cells and liver CSCs, we preformed differential expression analysis between two types of adult liver cells (CD235a-/EpCAM+/ASGPR1+ and CD235a-/EpCAM+) and the liver CSCs. The adult liver cells totaled 444 samples, and the liver CSCs totaled 1990 samples. Of the 18,263 protein coding genes included in the differential expression analysis, we identified 519 genes that were differentially expressed between liver CSCs and adult liver cell types. Of these, 134 were significantly higher expressed in the liver CSCs, while 385 protein coding genes were significantly higher expressed in the adult liver cell types.

Based on Gene Ontology analysis, genes with significantly higher expression in liver CSCs were significantly enriched in several GO terms that are hallmarks of higher rates of cell division seen in cancer cells. This includes structural constituent of ribosome (p-

value = 8.9E-27), translation initiation (p-value = 9.8E-24), and rRNA processing (p-value = 1.1E-20). Additionally, liver CSCs had significant enrichments of mitochondrial related GO terms. These include mitochondrial respiratory chain complex I (p-value = 3.1E-5), NADH dehydrogenase (ubiquinone) activity (p-value = 3.2E-5), and ATP biosynthetic process (p-value = 2.9E-3). CSCs have been previously shown to be more reliant on ATP oxidative metabolism relative to other cancer cell types [10]. Potentially of most interest is the observed enrichment of over-expressed genes that are involved in the extracellular vesicle (p-value = 6.2E-12). It has recently been shown that extracellular vesicles are important factors in driving cell dedifferentiation [11, 12]. Similarly, we also see significant enrichment of ncRNA processing (p-value = 3.2E-14), supporting recent evidence that ncRNAs are able to drive dedifferentiation phenotypes [13].

On the other hand, liver CSCs showed significant decreases in proteins involved in normal liver function. This includes organic acid metabolic process (p-value = 8.1E-18), carboxylic acid metabolic process (p-value = 1.2E-17), lipid metabolic process (p-value = 4.3E-7), and drug metabolic process (p-value = 5.4E-5).

Next we examined if previously identified dedifferentiation protein were more abundant in the liver CSCs, and known differentiation factors more abundant in normal liver cell types. As expected, Hepatocyte Nuclear Factor 4 Alpha (HNF4A), which acts as the primary differentiation factor of liver cell types [14], was significantly higher expressed in healthy liver cell types (2.96X, p-value = 4.43E-5). We found Transforming Growth Factor β1 (TGFB1) to be significantly upregulated in liver CSCs relative to adult differentiated liver cell types (4.74X, p-value = 1.46E-104). This result further confirms the strong implication of TFGB1 in driving the mesenchymal/stemness phenotype observed in hepatocellular carcinomas [15]. Additionally, Msh Homeobox 2 (MSX-2) was significantly upregulated in liver CSCs (1.36X, p-value = 1.99E-17). Although MSX-2 has been previously implicated in dedifferentiation of myotubules [16], MSX-2 has not previously been reported as a potential dedifferentiation factor of liver CSCs.

4. Discussion

With the diversity of single-cell next generation sequencing available from cancer studies, we can begin asking novel questions beyond the scope of the original researchers. Cell gene expression profiles provide us with an important insight into transition to cancerous cells, particularly how dedifferentiation plays a role in generating CSCs. The use of public big data is critical to this aim. Specifically, we aimed to understand how the expression profiles of CSCs compare to healthy adult liver cell types in order to better understand the dedifferentiation capabilities of CSCs, using two publically available single-cell RNA-seq datasets.

We examined the differences in expression profiles of adult liver cells and the liver CSCs, revealing 519 differentially expressed genes. Among these, we see significant upregulation of genes involved in translation, extracellular vesicle proteins, and ncRNA processing in liver CSCs. This is mirrored by significant downregulation of normal liver metabolic proteins. Additionally, we see downregulation of a key liver differentiation factor, HNF4A, and upregulation of dedifferentiation factors, TGFB1 and MSX-2. These results provide further indication of the importance of ncRNAs and TGFB1 in promoting dedifferentiation in CSCs [13, 15]. Further, we report the first evidence of the importance of MSX-2 in the dedifferentiation of liver CSCs. Our work concurs with previous results of liver CSC studies on the importance of dedifferentiation factors [17].

These results provide an insight into cancer biology made possible by utilizing publically available big data. Our results provide a unique insight into the process of dedifferentiation in forming liver CSCs. In particular, revealing novel potential dedifferentiation factors, such as MSX-2.

5. Conclusions

Our analysis presents the power and utility of reanalyzing publically available single-cell RNA-seq datasets to ask novel biomedical questions. Focusing on identifying potential dedifferentiation factors that promote the generation of liver CSCs from adult liver cell types, we have identified increased expression of ncRNA processing and a potentially novel dedifferentiation factor, MSX-2, acting in liver CSCs. Our work demonstrates the value of shared big data to catalyze knowledge discovery and predictive analytics.

References

[1] Toga AW, Foster I, Kesselman C, Madduri R, Chard K, Deutsch EW, Price ND, Glusman G, Heavner BD, Dinov ID, Ames J. Big biomedical data as the key resource for discovery science. Journal of the American Medical Informatics Association. 2015 Nov 1;22(6):1126-31.
[2] Navin NE. The first five years of single-cell cancer genomics and beyond. Genome Res. 2015;25(10):1499-507.
[3] Batlle E, Clevers H. Cancer stem cells revisited. Nat Med. 2017;23(10):1124-34.
[4] Mu X, Español-Suñer R, Mederacke I, Affò S, Manco R, Sempoux C, et al. Hepatocellular carcinoma originates from hepatocytes and not from the progenitor/biliary compartment. J Clin Invest. 2015;125(10):3891-903.
[5] Ma L, Hernandez MO, Zhao Y, Mehta M, Tran B, Kelly M, et al. Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. Cancer Cell. 2019;36(4):418-30.e6.
[6] Segal JM, Kent D, Wesche DJ, Ng SS, Serra M, Oulès B, et al. Single cell analysis of human foetal liver captures the transcriptional profile of hepatobiliary hybrid progenitors. Nat Commun. 2019;10(1):3350.
[7] Liu CY, Chen KF, Chen PJ. Treatment of Liver Cancer. Cold Spring Harb Perspect Med. 2015;5(9):a021535.
[8] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.
[9] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57.
[10] Zhu X, Chen HH, Gao CY, Zhang XX, Jiang JX, Zhang Y, et al. Energy metabolism in cancer stem cells. World J Stem Cells. 2020;12(6):448-61.
[11] Sandiford OA, Donnelly RJ, El-Far MH, Burgmeyer LM, Sinha G, Pamarthi SH, et al. Mesenchymal Stem Cell-Secreted Extracellular Vesicles Instruct Stepwise Dedifferentiation of Breast Cancer Cells into Dormancy at the Bone Marrow Perivascular Region. Cancer Res. 2021;81(6):1567-82.
[12] Sun Z, Wang L, Dong L, Wang X. Emerging role of exosome signalling in maintaining cancer stem cell dynamic equilibrium. J Cell Mol Med. 2018;22(8):3719-28.
[13] Huang T, Alvarez A, Hu B, Cheng SY. Noncoding RNAs in cancer and cancer stem cells. Chin J Cancer. 2013;32(11):582-93.
[14] Martínez-Jiménez CP, Gómez-Lechón MJ, Castell JV, Jover R. Underexpressed coactivators PGC1alpha and SRC1 impair hepatocyte nuclear factor 4 alpha function and promote dedifferentiation in human hepatoma cells. J Biol Chem. 2006;281(40):29840-9.
[15] Fabregat I, Caballero-Díaz D. Transforming Growth Factor-β-Induced Cell Plasticity in Liver Fibrosis and Hepatocarcinogenesis. Front Oncol. 2018;8:357.
[16] Yilmaz A, Engeler R, Constantinescu S, Kokkaliaris KD, Dimitrakopoulos C, Schroeder T, et al. Ectopic expression of Msx2 in mammalian myotubes recapitulates aspects of amphibian muscle dedifferentiation. Stem Cell Res. 2015;15(3):542-53.
[17] Borziak K, Finkelstein J. Identification of Liver Cancer Stem Cell Stemness Markers Using a Comparative Analysis of Public Data Sets. Stem Cells Cloning. 2021;14:9-17.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210863

ALLD: Acute Lymphoblastic Leukemia Detector

Saleh MUSLEH[a], Mohammad Tariqul ISLAM[b], Mohammad Towfik ALAM[c], Mowafa HOUSEH[a], Zubair SHAH[a] and Tanvir ALAM[a,1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*
[b] *Computer Science Department, Southern Connecticut State University, New Haven, CT 06515, USA*
[c] *Department of Vascular Biology and Molecular Pathology, Faculty of Dental Medicine and Graduate School of Dental Medicine, Hokkaido University, Sapporo 060-8586, Japan*

Abstract. Acute Lymphoblastic Leukemia (ALL) is a life-threatening type of cancer wherein mortality rate is unquestionably high. Early detection of ALL can reduce both the rate of fatality as well as improve the diagnosis plan for patients. In this study, we developed the ALL Detector (ALLD), which is a deep learning-based network to distinguish ALL patients from healthy individuals based on blast cell microscopic images. We evaluated multiple DL-based models and the ResNet-based model performed the best with 98% accuracy in the classification task. We also compared the performance of ALLD against state-of-the-art tools utilized for the same purpose, and ALLD outperformed them all. We believe that ALLD will support pathologists to explicitly diagnose ALL in the early stages and reduce the burden on clinical practice overall.

Keywords. Leukemia, Acute lymphoblastic leukemia, Deep learning, Computer aided diagnosis (CAD).

1. Introduction

Leukemia is a malignant, progressive deadly health condition in which abnormal, distorted proliferation of leukocytes and its precursor occur due to error in blood-forming organs. Subsequently, normal blood cell production will cease. The principal type of cell involved in this malignancy is commonly classified as lymphoid or myeloid, and chronic or acute, according to percentage or degree of leukemic cell differentiation in the bone marrow. The environmental risk factors have been well studied; however, the exact cause of leukemia is still unknown [1]. There are four main types of Leukemia classified based on severity level and affected cells type: chronic lymphocytic leukemia, chronic myeloid leukemia, acute myeloid leukemia, and acute lymphoblastic leukemia (ALL). ALL is one of the two classes of acute leukemia that develops from immature forms of lymphocytes cells in the bone marrow. ALL is recorded as the most common type of leukemia among children but the least common type in adults [2]. However, the fatality rate for ALL in adults is relatively higher than other types of acute leukemia. In most

[1] Corresponding Author, Tanvir Alam, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar; E-mail: talam@hbku.edu.qa.

leukemia-affected white blood cells, the average cell often changes to about two times the size of the red blood cells surrounding them. In most ALL affected cells, the nucleus area occupies almost 80-90% of the whole cell and leaves about 20-30% of cell area for the cytoplasm [2]. One of the diagnostic approaches for ALL is the microscopic inspection of blood cells. Both the malformed white blood cells as well outnumbered lymphoblasts may hint at the early onset of ALL and requires the assessment by an 0experienced pathologist [3]. Detecting ALL based on pathological images is tedious work and subjective by nature as it depends upon the evaluation by the pathologist. In some cases, this kind of individual and subjective interpretation may lead to a delay in the proper diagnosis plan for ALL patients.

To overcome these challenges and to support the pathologists, Computer Aided Diagnosis (CAD) systems may play a crucial role in detecting ALL in a precise manner. Many of the included studies focused on the classification of the images of lymphoblast cells from healthy individuals and leukemia patients. Sahlol et al. proposed an automatic ALL classification model using a social spider optimization algorithm. The proposed kNN-based model achieved 95.67% classification accuracy [3]. While Neoh et al. proposed an intelligent decision support system for leukemia diagnosis using microscopic blood cell images. Their Support Vector Machine (SVM)-based model systems attained the highest 96.67% accuracy [4]. Singhal and Singh have used the local binary pattern-based texture features to detect ALL, and the proposed SVM classifier was 93.84% reliable [5]. Recently, deep learning (DL)-based models have gained huge attention for the same task. In 2021, Genovese et al. used the transfer learning-based ResNet18 model to achieve 97.92% accuracy [6]. The same group has used adaptive unsharpening image processing on the images dataset to develop a VGG16-based model achieving 96.84% accuracy [7]. Table 1 summarizes the list of recent publications based on ML models for ALL detection.

Table 1 Existing Tools for the classification of classification of ALL-IDB2 dataset

Reference	Model	Accuracy	Remarks
[5] (2016)	SVM	93.84	Local binary pattern-based texturing
[3] (2020)	kNN	95.67	Social spider optimization
[4] (2015)	SVM	96.67	Texture, color, shape of nucleus and cytoplasm
[7] (2021)	VGG	96.84	Unsharpening of image
[6] (2021)	**ResNet**	**97.92**	Transfer learning based on histopathology database

In this study, we introduce the ALL Detector (ALLD), a deep learning-based model to distinguish ALL patients from healthy persons based on the image of peripheral blood samples.

2. Materials and Methods

2.1. Dataset Collection

We collected microscopic blood cell images from the ALL-IDB2 database [2]. ALL-IDB2 dataset contains 260 images, each having 257*257 resolution. Out of these, half (n=130) of the images were from healthy individuals and the remaining images (n=130) were from leukemia patients. The image of lymphoblast cells was captured using a Canon PowerShot G5 camera. The images were shared in *.tif* format in the publicly available database. Figure 1 shows examples of images from healthy individuals (A-C) and leukemia patients (D-F).

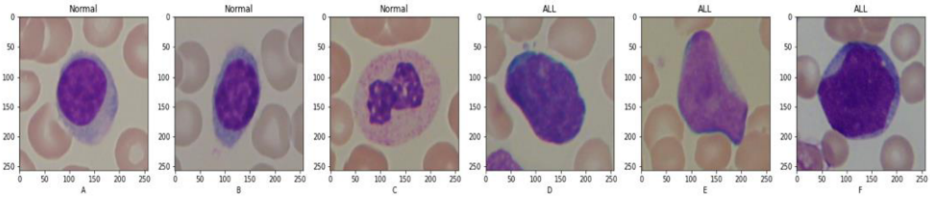


Figure 1 Blood microscopic images. A-C: normal individuals and D-F: leukemia patients.

2.2. Image Preprocessing

To increase the robustness of our experiment, we preprocessed the input images using a variety of image augmentation techniques: cropping, rotation, horizontal and vertical flipping, shearing, brightness and contrast perturbations, etc. Using image augmentation allows the network to become less sensitive to small changes in the inputs. This is possible because in each epoch, the input images undergo a subset of these operations parameterized with a random set of parameters. This forces the network to learn a general representation of the distribution rather than dataset-specific aspects, which reduces overfitting.

2.3. Development of Deep Learning Model ALLD

For the development of a deep learning-based model, we used fast.ai version 2.4.1. We split the data set in an 80-20 fashion. Twenty percent of the images were used for testing. The remaining 80% of images were used for training and validating the trained model. We considered ALL patients as the positive class and the healthy individuals as the negative class for the model. We used VGG16, VGG19, ResNet34, and ResNet50 for distinguishing ALL patients from healthy individuals. The Adam algorithm was used for the optimization of parameters. For momentum parameters (β1 and β2), we used default values of 0.9 and 0.99, respectively. We used a batch size of 64 in our experiments. Additionally, we used the one cycle policy for scheduling the learning rate with a maximum learning rate of 10^{-33} over a total of 20 epochs. The first 2 epochs updated the parameters of only the classification layer, whereas the last 18 updated all layers with differential learning rates. The experiments were carried out on a computer with a Ryzen 5800x CPU, 64 GB of DDR4 memory, and an Nvidia RTX 3090 GPU. For performance evaluation, we considered multiple standard performance evaluation metrics such as (i) Accuracy, (ii) Sensitivity, and (iii) Specificity to measure the effectiveness of the proposed ML models.

3. Results and Discussions

The performance of different DL-based models is highlighted in Table 2. Based on the performance results, we observed that ResNet-based models demonstrated the best performance in classifying the ALL group from the healthy group with the highest accuracy. As the ResNet-based model performed the best, we propose the ResNet-based model as the final ALLD model for ALL detection. ALLD outperformed all the existing machine learning models for the same purpose (Table 1 and Table 2).

Table 2 Machine Learning models performance for developing ALLD

Model	Accuracy	Sensitivity	Specificity
VGG16	94.230	95.652	93.103
VGG19	98.076	95.652	100.00
ResNet34	98.076	100.00	96.551
ResNet50	98.076	100.00	96.551

This study has some limitations worth mentioning. We worked on a relatively smaller-sized dataset of 260 images. In the future, we plan to evaluate the proposed model based on a larger dataset. Moreover, the participants' lifestyle, socio-demographic information, and genetic information were not available in the ALL-IDB2 database. Integrating such information, based on its availability, would provide more insight into ALL diagnoses.

4. Conclusions

Early detection, classification, and identification of ALL are very important for effective treatment to be provided in a timely manner, and, most importantly, on a person's survival. Automated detection and classification of acute leukemia is one of the emerging technology directions previously used. ALLD outperformed multiple existing tools utilized for the same purpose, and we believe it will also support pathologists in making a conclusive diagnosis plan in a treatment setting. We believe our tool ALLD will support both the early detection of ALL as well as the proper diagnosis plan for the patients.

References

[1] Belson M, Kingsley B, Holmes A. Risk Factors for Acute Leukemia in Children: A Review. Environmental Health Perspectives 2007;115:138–45. https://doi.org/10.1289/ehp.9023.

[2] Labati RD, Piuri V, Scotti F. All-IDB: The acute lymphoblastic leukemia image database for image processing. 2011 18th IEEE International Conference on Image Processing 2011. https://doi.org/10.1109/icip.2011.6115881.

[3] Sahlol AT, Kollmannsberger P, Ewees AA. Efficient Classification of White Blood Cell Leukemia with Improved Swarm Optimization of Deep Features. Sci Rep 2020;10:2536.

[4] Chin Neoh S, Srisukkham W, Zhang L, Todryk S, Greystoke B, Peng Lim C, et al. An Intelligent Decision Support System for Leukaemia Diagnosis using Microscopic Blood Images. Sci Rep 2015;5:14938.

[5] Singhal V, Singh P. Texture Features for the Detection of Acute Lymphoblastic Leukemia. Advances in Intelligent Systems and Computing 2016:535–43. https://doi.org/10.1007/978-981-10-0135-2_52.

[6] Genovese A, Hosseini MS, Piuri V, Plataniotis KN, Scotti F. Histopathological Transfer Learning for Acute Lymphoblastic Leukemia Detection. 2021 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA) 2021. https://doi.org/10.1109/civemsa52099.2021.9493677.

[7] Genovese A, Hosseini MS, Piuri V, Plataniotis KN, Scotti F. Acute Lymphoblastic Leukemia Detection Based on Adaptive Unsharpening and Deep Learning. ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2021. https://doi.org/10.1109/icassp39728.2021.9414362.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210864

Application of Natural Language Processing to Learn Insights on the Clinician's Lived Experience of Electronic Health Records

Yalini SENATHIRAJAH[a], Hwayoung CHO[b], Jaime FAWCETT[a], Karla M. MONDEJAR[c], Kenrick CATO[d], Peter BROADWELL[e] and Sunmoo YOON[f,1]
[a] *Biomedical Informatics, School of Medicine, University of Pittsburgh, USA*
[b] *College of Nursing, University of Florida, USA*
[c] *Department of Infectious Disease, Mount Sinai Health System, USA*
[d] *School of Nursing, Columbia University, USA*
[e] *Center for Interdisciplinary Digital Research, Stanford University, USA*
[f] *Department of Medicine, Columbia University Irving Medical Center, USA*

Abstract. We interviewed six clinicians to learn about their lived experience using electronic health records (EHR, Allscripts users) using a semi-structured interview guide in an academic medical center in New York City from October to November 2016. Each participant interview lasted approximately one to two hours. We applied a clustering algorithm to the interview transcript to detect topics, applying natural language processing (NLP). We visualized eight themes using network diagrams (Louvain modularity 0.70). Novel findings include the need for a concise and organized display and data entry page, the user controlling functions for orders, medications, radiology reports, and missing signals of indentation or filtering functions in the order page and lab results. Application of topic modeling to qualitative interview data provides far-reaching research insights into the clinicians' lived experience of EHR and future optimal EHR design to address human-computer interaction issues in an acute care setting.

Keywords. electronic health records, usability, natural language processing

1. Introduction

Poor EHR design causes unintended consequences and harms patients in turn [1]. Clinicians suffer from unnecessarily lengthy documentation and alert fatigue. Clinicians have often been unheard of and ignored from the beginning of the design of EHR in acute care settings in the United States [1]. Further, this is of great importance in the design of the EHR system to ensure all EHR users who need access to patient information can view patient data when and where they need it. While 1st generation EHR designs have been widely used for the past 15 years, more generation X and millennial clinicians have

[1] Corresponding Author, Sunmoo Yoon, General Medicine, Department of Medicine, Columbia University Irving Medical Center, 630W 168 Street, PH105, New York, NY, 10032, USA; E-mail: sy2102@cumc.columbia.edu.

entered the workforce. In machine learning and NLP, topic modeling applies statistical analyses to counts of words or word groups that co-occur within text documents to reveal latent word groupings with potential semantic significance across a collection of documents [2]. Topic modeling is typically applied in health science to facilitate meta-analyses of large collections of research publications or generate high-level discussions among groups of interest. The purpose of the study was to apply natural language processing and topic modeling techniques to qualitative interview data to deeply understand usability and design issues of EHRs and as the foundation to inform future designs of user-centered EHR interfaces for clinicians in acute care settings.

2. Methods

We applied natural language processing (NLP) and topic modeling to an interview discussion on a lived experience of EHR among clinicians in an acute care setting (neurology nurses). We used Python 3.7 and Organization Risk Analyzer (ORA) to apply Latent Dirichlet allocation (LDA) and visualize the clusters. Clinicians including doctors, nurses, and physician assistants with EHR experience in an acute care setting were eligible for this study. We interviewed six clinicians for approximately 60 minutes in a quiet room in a neurology department inpatient unit at an academic medical center from October to November 2016. The sessions combined a standard think-aloud usability protocol. They used their usual EHR to review real patient cases, performing six common review tasks required by the researcher. They were also given a semi-structured interview with broad latitude to talk about their experience with the EHR (Allscripts). First, the audiotapes were transcribed automatically by speech recognition software, and the data collector reviewed the transcript's validity. Second, we preprocessed the text file applying NLP. Third, we applied a clustering LDA algorithm to detect the topics in the interview corpora. Coherences scores and LDA clustering graphs were compared with different numbers of topics to identify the optimal number. Next, we visualized eight topics using a network diagram (Louvain modularity 0.70). Lastly, experts in clinical nursing and EHR usability performed thematic analysis on the detected topics. The Institutional Review Board (IRB) of Columbia University approved the study.

3. Results

A total of eight topics were identified as the optimal number of topics during LDA modeling (8 topics, Perplexity: -7.137, Coherence Score: 0.649). The themes are summarized in **Table 1** and **Figure 1**. Eight topics include: 1) missing user-friendly signals of indentation or filtering functions in the doctor's lengthy order page; 2) admission notes are not linked to the other parts of EHR; 3) automatic note functions to carry in the users pocket are missing, and users need to take manual notes during busy time; 4) missing user control functions in the laboratory results section. As a result, the results are disorganized and lengthy, and a more concise display is needed; 5) user's controlling function to organize orders, medications, and radiology report pages; 6) time consuming and ineffective input functions for pain assessment tools; 7) patient status panel should consider clinician's busy workflow; 8) overall design of the EHR system looks messy and causing user's hate emotion.

Table 1. Topics detected by LDA and representative n-grams with probabilities

Themes	Exemplary quotes, representative terms, and probability (%)
1. Missing signals of indentation or filtering functions in long order	[red 1.2, safe 1.2, look 0.9, shift 0.8, different 0.8, machine 0.7, unsafe 0.7, filtering 0.7, restrain 0.7, indent 0.7]
2.Issues with admission documentation	[understand 3, admission 2.6, click 1.6, cause 1.4, change 1.3, start 1, documentation 0.7, problems 0.5, remember 0.5, gauge 0.5]
3. Room to improve to complement paper note taking behavior during busy time	[system 4.8, move 4.3, improvement 4.1, fix 2.1, hand 2.1, paper 1.2, busy 1.1, doctors 1, note 0.8, takes 0.8]
4. Missing user control functions in lab results section and need for succinct writing.	[control 6.3, important 4, clinical 3.9, results 3.7, function 2.2, show 0.9, labs 0.9, two 0.9, recounts 0.6, succinct 0.4]
5. Needs for user to have flexible organizing on display for orders, medications, radiology reports	[want 5.6, summary 4, go 3.9, point 3, medication 1.9, terms 1.3, orders 1.2, display 1, need 0.6, radiology 0.6]
6. Time consuming and ineffective input functions for pain assessment	[assessment 8.7, pain 5.7, combine 2.7, easy 2, time 2, clicking 1, order 0.6, business 0.6, difficult 0.6, shame 0.6]
7. Patient status panel should consider clinician's busy workflow	[hour 3.4, panel 3.4, diagnosis 3.4, status 3.4, paths 3.4, nutrition 3.4, documents 2.2, else 1.6, harmonic 0.6, workflow 0.6]
8. Design of the system looking messy and causing hate	[design 7.3, replete 3.5, used 2, unions 0.8, mess 0.6, never 0.6, serene 0.6, knot 0.6, hated 0.6, reflection 0.6]

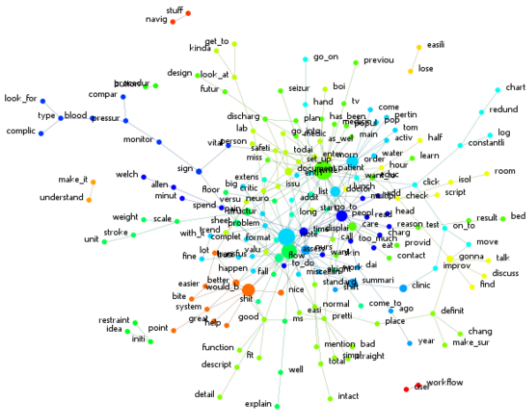


Figure 1. Semantic visualization of topic modeling on in-depth interview data of clinicians' lived experience of EHR

4. Discussion and Conclusion

One of the main findings from our study was a usability deficit due to a lack of necessary signals and functions within the system. The authors recommend using carefully

designed and user-designed alerts/reminders within the EHR system as a solution. Recently scientists and clinicians emphasized urgent needs for developing EHR use measures that are clinically meaningful, trustworthy, scientifically sound, transparent, and feasible for implementation [3]. Our findings contribute to the body of knowledge on priorities and concepts of implementing EHR use measures. The three most cumbersome areas with major usability issues within the system include pain assessment documentation, doctor's order page, and admission documentation [4]. Pain assessment in Allscripts and Epic EHRs currently requires nurses to redundantly enter data in multiple different places (i.e., medication sheet comment, flow sheet, pain management note, care plan note, nursing summary note) for one episode of pain management, which usually requires 18 to 28 or more episodes per shift in an acute care setting. Paper-based documentation only requires one or two places to document (nursing note, pain management note) for several episodes of pain management per person per shift. By contrast, EHR systems including Allscripts and Epic require a substantially large amount of documentation in a fragmented way. Similarly, information in nursing admission notes is not automatically reflected in the system (e.g., weight, height, chief complaints do not trigger suggested care plans) [4]. Auto trigger functions to support clinicians' decisions in care planning using machine learning and thoughtful and user-centered design, which links and updates simultaneously in different systems are suggested for future work. To understand the usability and design issues of EHRs in a real-world setting, we conducted an in-depth interview and exercise in which users carried out major functions in their usual EHR in a think-aloud protocol and interview and applied topic modeling techniques to the qualitative data. The authors recommend comparing the results from manual and topic modeling methods in the future. Despite the rapid proliferation of EHR systems in healthcare, little research evaluates users' (clinicians) systems engagement in everyday clinical settings. The usability of the EHR system is strongly associated not only with system adoption but also with its subsequent use. Usability testing in a real-world setting [5] can help ensure that the EHR system works properly and support clinicians to achieve their goals in real-world clinical settings. Our work enabled us to measure users' actual experiences when interacting with the EHR system, an important strength of this study. In summary, novel NLP methods and topic modeling with semantic network visualization complemented the usual qualitative analysis and contributed to overcoming the subjective bias of qualitative analysis methods. Our findings add the body of knowledge on developing and implementing measurement science for EHR use.

Acknowledgments: US federal grant AHRQ R01HS023708 (PI: Senathirajah)

References

[1] Harrison MI, Koppel R, Bar-Lev S. Unintended consequences of information technologies in health care, an interactive sociotechnical analysis. JAMIA. 2007 Sep 1; 14(5):542-9

[2] Blei DM. Technical perspective: Expressive probabilistic models and scalable method of moments, Communications of the ACM. 2018;61:84.

[3] Melnick ER, Sinsky CA, Krumholz HM. Implementing Measurement Science for Electronic Health Record Use. JAMA. 2021 Apr 5.

[4] Lavin MA, Harper E, Barr N. Health information technology, patient safety, and professional nursing care documentation in acute care settings. Online J Issues Nurs. 2015 May 1;20(6).

[5] Cho H, Yen PY, Dowding D, Merrill JA, Schnall R. A multi-level usability evaluation of mobile health applications: A case study. Journal of biomedical informatics. 2018 Oct 1;86:79-89.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210865

Healthcare Professionals' Perceptions and Opinions on "Do not Attempt Resuscitation" (DNAR) Order and Documentation

Hanna KUUSISTO[a,b,c,1], Tapani KERÄNEN[b] and Kaija SARANTO[b]
[a] *Tampere University Hospital, Neurology, Tampere, Finland*
[b] *University of Eastern Finland, Kuopio, Finland*
[c] *Kanta-Häme Central Hospital, Hämeenlinna, Finland*

Abstract. Insights on end-of-life care decisions, such as do not attempt resuscitation (DNAR), vary between institutions and individual health care professionals. At the era of electronic patient records (EPR), the information of DNAR order may still be recorded in multiple locations making it difficult to find and interpret. A link to a structured web-based questionnaire was sent to all physicians and nurses working in Tampere University Hospital special responsibility area covering a catchment area of 900 000 Finns. Perceptions on DNAR order and documentation was surveyed. In total 934 subjects responded, of which 727 (77%) were nurses and 219 (23%) physicians covering all specialties. We found substantial variation in DNAR order interpretation and documentation among all health care professionals possibly causing information breakdown and compromised end-of-life care.

Keywords. End-of-life decisions, DNAR documentation, electronic patient records

1. Introduction

A "do not attempt resuscitation" (DNAR) order means a process for deciding to withhold cardiopulmonary resuscitation (CPR). The main grounds for the decision are refusal of CPR by the patient with capacity, or known advance decision to refuse treatment, or when the burdens of CPR attempt are thought to outweigh the benefits [1]. Health-care professionals perceive challenges in DNAR order making and understanding DNAR policies [2]. Both nurses and physicians may interpret a DNAR order to mean that in addition to CPR, also other care, such as antibiotics or iv. fluids should be withheld [3-5]. Furthermore, physicians consider that they have lack of knowledge when to issue DNAR order and how patients and their relatives should be involved in the decision process [6]. Disagreement on codification and registration has also been reported [7,8]. Electronic patient records (EPR) have improved the quality of documentation and information management in general [9]. However, it still is possible to document orders, such as DNAR in various locations, increasing the risk of information break down. Here we present the data on how healthcare professionals document, find and interpret the

[1] Corresponding Author, Hanna Kuusisto, Tampere University Hospital po Box 2000. 33520 Tampere, Finland; E-mail: hanna-maija.kuusisto@pshp.fi

DNAR order and what types of information management difficulties there may be in the DNAR process.

2. Methods

A link to a structured web-based questionnaire (webropol) was sent to all physicians and nurses working in Tampere University Hospital special responsibility area, which includes Pirkanmaa, Southern Ostbothnia and Kanta-Häme hospital districts covering a catchment area of ca. 900 000 inhabitants, which represents 16% of Finland's population. The survey included five background questions, 11 multiple-choice questions on DNAR order process and a possibility to give free comments. The results of the multiple choice questions are presented as percentages. Deductive content analysis was used to classify the comments on documentation of a DNAR order.

3. Results

A total of 952 health care professionals covering all medical specialties participated, of which 727 (77%) were nurses and 219 (23%) physicians. Sixty-seven percent were over 45 years of age and majority (87%) worked in shifts or on call. A marked proportion of the respondents considered that documentation (59%) and interpretation (57%) of DNAR order is problematic. Furthermore, 65% of the respondents felt that DNAR order making is far too often dismissed during office hours and left to the doctor on call. The results to the questions: Who can give the DNAR order, where DNAR order should be documented and what does DNAR exclude are presented in tables 1,2, and 3.

Table 1. who can give the DNAR order?

	N	%
Specialist alone without including patient or family in the decision making	188	19,81%
Specialist together with the patient and family	596	62,8%
Doctor in training without consulting a specialist	88	9,27%
Patient alone without consulting a doctor	11	1,16%
Cannot say/ don`t know	66	6,96%

Table 2. Where should DNAR order be documented

	N	%
In "risk information"	846	89,43%
Page of specialty where the patient is being treated	44	4,65%
Anesthesiology page	2	0,21%
In "orders"	19	2,01%
Somewhere else	35	3,7%

Table 3. What does DNAR order exclude?

	N	%
All resuscitation procedures	693	73,18%
Cardio-pulmonary resuscitation (CPR) only	316	33,37%
Treatment and follow up in intensive care unit (ICU)	433	45,72%
Treatment and follow up in STROKE-unit	97	10,24%
Treatment and follow up in cardiac care unit (CCU)	97	10,24%
Administration of intravenous (iv) fluids	14	1,48%
Administration of antibiotic treatments	14	1,48%
All active treatment	72	7,6%
All diagnostic electronic procedures	45	4,75%
Cannot say/ don`t know	26	2,75%

Health care professionals' feedback (n=465) focused on the difficulties in finding the DNAR order (65%), content variation of the DNAR order and lack of responsibility in documenting the order (28%). Respondents (5%) were also concerned about the information given to patients and next of kin**.**

4. Discussion

The European guidelines for resuscitation state that great differences between European countries exist regarding the practice and attitudes towards CPR [10]. According to our study there seems to be major variation also among Finnish health care professionals in understanding the content of DNAR order as well as who can make the order as well as should the patient and the family be included in the process. In Finland, there is no legislation concerning DNAR order, but according to law, all care related decisions such as DNAR order should be discussed with the patient or when the patient is incompetent, with the relatives. The national agency responsible for the supervision of the social and health care (Valvira), states that the DNAR order is a medical decision made by the physician responsible for the care of the patient. Furthermore, similarly as in Sweden, a patient cannot demand CPR if the physician considers it to be against patients benefit [6]. Despite this guidance, in our survey up to 63% of the health care professionals thought that the decision process of the DNAR order should involve the patient and/or relatives and only 20% that the decision should be made by a senior physician only. This was supported by the comments given by the respondents. In a Swedish study almost half of the nurses and physicians considered that patients would not be involved in the DNAR order decision [11].

According to Finnish National institute for Health, a DNAR order should be recorded in the "risk information" part of the EPR by a doctor. Majority (89.4%) of the respondents However, ten percent felt that the order should be recorded somewhere else, such as in "orders" or in, knew this multiple places. According to the comments given, the most frequent information breakdown problem was a missing DNAR order in "risk information".

Only one third of the respondents in our study correctly recognized that the DNAR order excludes CPR only. According to almost half of the respondents, care in the ICU

would be excluded as well. Similar confusion on the meaning of the DNAR order has been previously reported [3-5].

Despite current European and national guidelines, health care professionals consider making and documenting the DNAR order as problematic. Major opinion differences still include variation in the interpretation of the scope of the order and who should be included in the decision-making. Current EPRs allow variation in DNAR documentation process which may lead to uncertainty at the time when this critical information is needed. More precise guidelines as well as systematic education of DNAR order making and documentation already at the basic training level of health care professionals are sorely needed. Furthermore, the information technology experts and clinicians should collaborate in aiming to develop EPRs in a way that recording of critical information such as the DNAR order could be made only in a specific place and the information would be clearly visible.

References

[1] Perkins GD, Griffiths F, Slowther AM, George R et al. Do-not-attempt-cardiopulmonary- resuscitation decisions: an evidence synthesis. NIHR Journals Library; 2016, Health Services and Delivery Research. DOI: 10.3310/hsdr04110

[2] Robinson F, Cupples M, Corrigan M. Implementing a resuscitation policy for patients at the end of life in an acute hospital setting: qualitative study. Palliat Med 2007;2:305-12

[3] Fritz Z, Fuld J, Haydock S, Palmer C. Interpretation and intent: a study of the (mis)understanding of DNAR orders in a teaching hospital. Resuscitation 2010;81(9):1138-41. doi: 10.1016/j.resuscitation.2010.05.014.

[4] Butler MW, Saaidin N, Sheikh AA, Fennell JS. Dissatisfaction with Do Not Attempt Resuscitation Orders: A nationwide study of Irish consultant physician practices. Ir Med J 2006;99(7):208-10.

[5] O'Brien H, Scarlett S, Brady A, Harkin K, Kenny RA, Moriarty J. Do-not-attempt-resuscitation (DNAR) orders: understanding and interpretation of their use in the hospitalised patient in Ireland. A brief report. J Med Ethics 2018;44:201 – 03.

[6] Fallahi M, Mahdavikian S, Abdi A, Borhani F, Taghizadeh P, Hematpoor B. Nurses and physicians' viewpoints about decision making of do not attempt resuscitation (DNAR). Multidiscip Respir Med 2018;13:20. doi: 10.1186/s40248-018-0133-8

[7] Butler JV, Pooviah, PK, Cunningham D, Hasan M. Improving decision-making and documentation relating to do not attempt resuscitation orders. Resuscitation 2003;57:139-44

[8] Bertilsson E, Semak B, Schildmeijer K, Bremer A, Carlsson J Palliat Care Med 2016;6:1 – 8

[9] Priestman W, Sridharan S, Vigne H, Collins R, Seamer L, Sebire NJ. What to expect from electronic patient record system implementation: lessons learned from published evidence. J Innov Health Inform. 2018 Jun 15;25(2):92-104. https://doi.org/10.14236/jhi.v25i2.1007

[10] Bossart et al ERC guidelines for resuscitation 2015. Section 11: the ethics of resuscitation and end-of-life decisions. Resuscitation 2015;95:302-3011.

[11] Pettersson M, Höglund AT, Hedström M. Perspectives on the DNR decision process: A survey of nurses and physicians in hematology and oncology. PLoS One. 2018;13(11):e0206550. doi: 10.1371/journal.pone.0206550.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210866

A Framework for User-Configurable Data Quality Assurance of Electronic Patient Records

Marina ROMANCHIKOVA[a,1] and Jean-Laurent HIPPOLYTE[a]
[a] *National Physical Laboratory, UK*

Abstract. Electronic Patient Records (EPRs) are valuable data resources for clinical and operational research. The heterogeneity of medical software coupled with the changing data formats and long lifespan of the patient datasets stored in EPRs results in data inconsistencies that hinder operational activities and increase personnel efforts for data lookup and cleaning. This study presents an approach for automated data quality reporting that was developed and tested within a real-world hospital setting at Royal Surrey County Hospital NHS Foundation Trust in 2020. 81 data quality tests configurable via spreadsheets were defined and executed to yield standardised human-readable reports in comma-separated value format. The data evaluation and reporting routines provided manyfold improvement over existing data quality reporting mechanisms.

Keywords. Data quality, electronic patient records, automation

1. Introduction

Positioned at the heart of clinical information exchange, Electronic Patient Records (EPRs) are a valuable resource of data for clinical and operational research. However, heterogeneous nature of clinical tools that feed data to EPR leads to variable data quality, making data reuse or exchange a challenging exercise. The demand for efficient data quality assurance mechanisms has been exacerbated by the additional pressure on EPR data management posed by the onset of Covid-19 pandemic that required rapid access and exchange of patient records. This study presents an approach for automated data quality evaluation that was developed and tested within a real-world hospital setting at Royal Surrey County Hospital (RSCH) NHS Foundation Trust in 2020 to aid the migration of patient data into a new EPR system.

Migration-ready EPRs must fulfil the data quality criteria defined by the hospital Data Quality department and the target EPR provider. All patient records that do not meet these criteria need to be corrected pre-migration. Numerous data quality checks (DQCs) must be run to detect the non-compliances. These checks must be executed at least once per day to capture the changes resulting from normal hospital activity and from the ongoing EPR correction efforts. The existing DQC process demanded involvement of several NHS Trusts departments and needed to be adjusted to handle the high number of EPR database reports. This work aimed to create a user-friendly low-

[1] Corresponding Author, Marina Romanchikova, National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, UK; E-mail: marina.romanchikova@npl.co.uk

technology solution for evaluation of EPR compliance with respect to each of the defined DQCs. Techniques, tools, and data flows used to create the solution are described in the following sections.

2. Materials and Methods

2.1. User requirements

The Data Quality team users needed to be able to run individual DQCs (i.e., find all patient records with missing home address) or groups of DQCs (i.e., locate all patient records that do not comply with any of the "red" category DQCs). The application output had to produce human-readable listings of patient records in need of correction, as well as snapshot statistics of compliant and non-compliant patient records.

2.2. Data Quality Checks

A total of 148 DQCs were defined by the RSCH Trust's Data Quality department and refined within this collaboration. The DQCs specified individual EPR data items (e.g., NHS numbers) or sets of items (e.g., whether any data elements in the patient record contain certain characters) to be tested against a set of rules (i.e., an NHS number should be 10 digits long). Every DQC was assigned a Red/Amber/Green risk category described in Table 1.

Table 1. Types of data quality checks and their implications for data readiness for migration to new EPR.

DQC category	# DQC items	Example	Impact of non-compliance
Red	9	Hospital number is used in more than one patient record	Cannot be migrated as is or will be result in incorrect data.
Amber	126	Missing current home address	It is unclear whether target EPR can accept the record.
Green	13	Missing gender specification	Potentially can be migrated as is, but is likely to cause issues in target EPR.

2.3. Software framework

The application used the data quality rules definitions from the Data Quality Vocabulary [1] that describes concepts relevant to data quality such as data, reference data, quality rules and quality metrics. The implemented quality rules included legal/illegal/unique value, property completeness, functional dependency reference and custom rules.
The application presented in Figure 1 was implemented in Python 3.7 using *pypika*[2] library for SQL query generation and *pyodbc*[3] library to connect to the EPR Microsoft SQL Server database.

[2] https://github.com/kayak/pypika
[3] https://github.com/mkleehammer/pyodbc

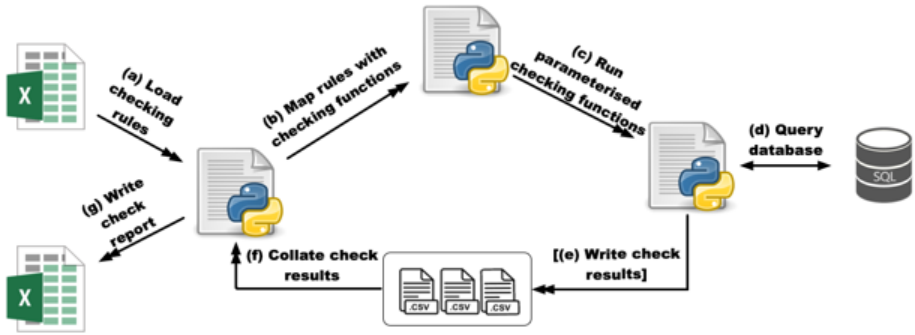


Figure 1. Data quality checking application. (a) The DQCs to be executed are loaded from the spreadsheet provided by the user. (b) Data quality rules associated with each DQC are combined in checking functions. (c) Parametrised SQL queries are generated according to rule types and database mappings. (d) The generated queries are run on the EPR database. (e, f) The query results are collated into human-readable reports and returned to the user.

From the total of 148 DQCs, 81 were mapped to the EPR (i.e., it was known which tables and columns contain the data) and implemented in the software. The input spreadsheet structure for several data quality rule types is illustrated in the Table 2.

Table 2. Configurable spreadsheet with data quality rules definitions. Table names and DQC IDs have been altered. Queries are generated by the software based on the table & column names, rule type and rule constraint.

DQC ID	Table.Column	Rule Type	Rule Constraint	Generated SQL Query
1	Patient.NHS_ID	Valid Value	<pattern>	SELECT Hospital_ID FROM Patient WHERE NHS_ID NOT LIKE <pattern>
2	Patient.DateOfDeath Patient.Status	Conditional Property Completeness	<condition>	SELECT Hospital_ID from Patient WHERE <condition>
3	Patient.Gender	Missing Value	IS NULL	SELECT Hospital_ID from Patient WHERE Gender IS NULL

2.4. Application outputs

The software produced 81 individual DQC reports listing patient identifiers and the elements of the EPR that failed the DQC. The reports enabled location and subsequent correction of the identified records in the EPR. Additionally, the software generated a summary report with overview of failed record statistics against each completed DQC (Table 3).

Table 3. Fragment of the overview report produced by the software. The "#failed" column shows the total number of non-compliant patient records for a given DQC.

DQC ID	RAG	DB tables	Description	#failed
1	Red	Patient.NHS_ID	Invalid NHS number	270
2	Amber	Patient.DateOfDeath; Patient.Status	Missing date of death in deceased patients	334
3	Green	Patient.Gender	Missing gender value	6173

2.5. Performance

Development and tests were performed on a remotely accessed Windows 7 office workstation with 16 GB RAM and Intel Core i7 CPU. From the 81 implemented DQCs, 79 DQCs were completed in less than 7 minutes. The runtime for two less frequently used DQCs that scanned for invalid characters in all columns of all database tables took approximately 17 minutes. The runtime for 81 implemented DQCs was 24 minutes.

3. Discussion

The presented framework was developed to provide the hospital Data Quality personnel with a configurable tool to obtain up-to-date information on completeness and correctness of EPRs. The framework made use of a) user-configurable spreadsheets to instruct the software which reports are required; b) user- and machine-readable data quality checking rules and c) auto-generation of SQL queries from data quality rules and EPR database mappings. While the solution proved useful and fast to implement for simple DQCs such as verifying that a data attribute is present or follows a pattern, the method showed limitations for a small number of complex queries that involved multi-stage evaluation of several tables, for example, counting of NHS numbers by trace status[4]. Such queries were pre-defined in the input spreadsheet and would need additional development time to be fitted into the rule-based query generation mechanism.

Acknowledgements

This study was funded by the department of Business, Energy, and Industrial Strategy through the Covid-19 response programme[5]. The authors would like to thank the Royal Surrey NHS Hospital Foundation Trust's Data Quality and Data Warehousing teams for setting the clinical objectives and for providing the access to EPR database.

References

[1] Fürber C, Hepp M. Towards a vocabulary for data quality management in semantic web architectures. InProceedings of the 1st International Workshop on Linked Web Data Management 2011 Mar 25 (pp. 1-8).

[4] https://www.datadictionary.nhs.uk/attributes/nhs_number_status_indicator_code.html
[5] https://www.npl.co.uk/covid-response

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210867

Predicting the Risk Factors of Second Primary Cancer in Patients with Hepatocellular Carcinoma

Hsien-Hua LIAO[a,b], Chi-Chang CHANG[c,d,1], Yu-Xiang WANG[c] and Chalong CHEEWAKRIANGKRAI[e]
[a]*Department of Surgery, Chung Shan Medical University Hospital, Taichung, Taiwan*
[b]*School of Medicine, Chung Shan Medical University, Taichung, Taiwan*
[c]*School of Medical Informatics, Chung Shan Medical University & IT office, Chung Shan Medical University Hospital, Taichung 40201, Taiwan*
[d]*Department of Information Management, Ming Chuan University, Taoyuan, Taiwan*
[e]*Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand*

Abstract. Screening for cancer and improved treatments have not only improved treatment outcomes and patient survival but have also led to an increase in the number of second primary cancers (SPCs). Hepatocellular carcinoma has been a common occurrence in Taiwan over the past decade. The mortality rate is second only to malignant tumors of lung cancer, and it also represents the fourth highest cancer medical expenditure. This study aimed to use machine learning to identify the risk factors for Hepatocellular carcinoma survivors. Of 378,445 datasets, including 15,251 from patients with SPCs, were collected; 18 predictive variables were considered risk factors for SPCs based on the physician panel discussion. The machine learning techniques employed included support vector machine, C5 decision tree, and random forest. SMOTE (Synthetic Minority Oversampling Technique) sampling method was used to resolve the imbalance problem. The results showed that the top 5 risk factors for SPCs were tumor size, clinical stage, surgery, total bilirubin, and BCLC Stage. The support vector machine method had the highest predicted accuracy (0.7673). The risk factors extracted from the classification models and association rules will be used to provide valuable information for HCC therapy.

Keywords. second primary cancer, hepatocellular carcinoma, machine learning techniques

1. Introduction

According to the latest data from the International Agency for Research on Cancer, hepatocellular carcinoma is the 7th most common cancer and the 3rd most common cancer. [1] Hepatocellular carcinoma (HCC) has been a common occurrence in Taiwan over the past decade. The mortality rate is second only to malignant tumors of lung cancer, and it also represents the fourth highest cancer medical expenditure, as shown in

[1] Corresponding Author: Chi-Chang Chang, School of Medical Informatics, Chung Shan Medical University & IT Office, Chung Shan Medical University Hospital, 110, Sec. 1, Chien-Kuo N. Rd., Taichung, Taiwan; E-Mail: changintw@gmail.com.

Table 1. [2] There are many treatment options for patients with hepatocellular carcinoma. Often, depending on the tumor, liver function, and physical situation. Screening for cancer and improved treatments have not only improved treatment outcomes and patient survival but have also led to an increase in the number of second primary cancers (SPCs). According to the Bureau of Health Promotion, Ministry of Health and Welfare, hepatocellular carcinoma was ranked second with regards to mortality rates of top 10 cancers and was only ranked below lung cancer, and ranked fourth with regards to medical expenditure, as shown in Table 2.

Table 1. 5-Year Survival Rates for Top 10 Cancers in Taiwan

Cancers / Year	1	2	3	4	5	6	7	8	9	10	All cancers
2014	96.9	82.3	61.1	92.8	59.0	80.4	93.1	97.7	92.2	60.1	77.0
2015	93.7	72.4	45.8	85.5	47.0	67.4	88.2	96.6	85.6	45.6	67.2
2016	90.5	65.5	37.3	79.1	39.6	61.3	84.6	95.7	80.0	38.8	61.3
2017	87.5	60.5	32.0	73.6	33.9	57.1	82.3	94.8	75.1	34.7	57.0
2018	85.2	56.9	28.6	68.5	29.9	53.4	80.3	94.1	71.3	31.6	53.9

Source: The Health Promotion Administration (2018).
Note: 1: breast cancer; 2: colorectal cancer; 3: lung cancer; 4: prostate cancer; 5: liver cancer; 6: oral cancer; 7: uterine cancer; 8: thyroid cancer; 9: skin cancer; 10: stomach cancer.

Table 2. Expenditure for top 10 cancers in Taiwan (unit: NTD/100,000)

Cancers / Year	1	2	3	4	5	6	7	8	9	10
2014	10,311	10,808	10,987	8,471	6,631	4,082	3,162	3,923	2,385	2,538
2015	10,923	11,326	11,138	8,811	6,737	4,298	3,518	4,309	2,446	2,576
2016	11,521	11,745	11,323	8,791	5,382	4,569	3,643	2,696	2,506	2,570
2017	13,217	12,829	12,468	9,666	6,000	4,907	4,158	3,015	2,767	2,755
2018	14,355	15,020	13,845	10,952	6,652	5,433	5,397	5,359	3,122	3,066

Source: The Health Promotion Administration (2018).
Note: 1: breast cancer; 2: lung cancer; 3: colorectal cancer; 4: liver cancer; 5: oral cancer; 6: leukemia; 7: prostate cancer; 8: non-Hodgkin's lymphoma; 9: Esophageal cancer; 10: stomach cancer.

Due to continuous improvements in screening, diagnosis and treatment, the survival rate of newly diagnosed cancer patients is increasing. [3] Clinical information regarding the SPCs patients with HCC is important, because it could explain the cause and may verify the necessity of related secondary cancer during patient follow-up. [4,5,6]

2. Methods

Figure 1 depicts the study framework. Herein, an in-depth analysis was performed based on the tumor size and clinical stage. The HCC dataset containing 378,445 valid records, was used in this study. There were 19 variables in the cancer registry and 18 predictive variables were selected by clinical experts and literature review (Table 3) C5 is a

modified iterative dichotomizer 3 (ID3) algorithm. Based on information theory and probability statistics, the greater the information gain and the higher the information entropy. RF (Random Forest) is based on the statistical learning theory and combines several individual classification trees. RF is a supervised machine learning algorithm that considers the unweighted classified votes. SVM (Support Vector Machine) is a machine learning algorithm based on the principle of structural risk minimization and is used to estimate function by minimizing the upper limit of generalization error.

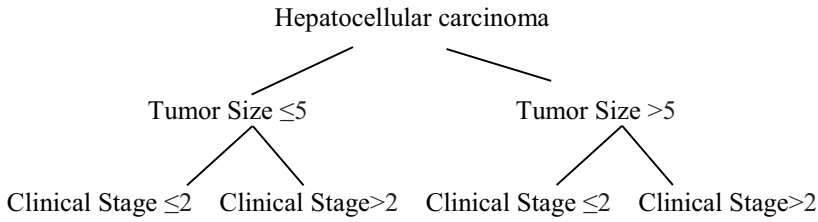


Figure 1. Study framework

Table 3. Important Variables and Coding in this study

Variable	Name	Definition of test data
X1	Age	≤65/>65
X2	Sex	Male/Female
X3	Grade	Well/Moderaterly/Poorly
X4	Tumor number	Single tumor/Multiple tumors
X5	Tumor Size	≤5/>5
X6	Clinical Stage	Stage I、II/Stage III
X7	Pathologic Stage	Stage I、II/Stage III
X8	BCLC Stage	Stage 0、A/Stage B/ Stage C、D
X9	Operation	YES/NO
X10	Surgical Margins	YES/NO
X11	BMI	<18.5/18.5-24/ >=24
X12	Alpha-Fetoprotein	≦400 ng/ml/ >400 ng/ml
X13	Liver Fibrosis	YES/NO
X14	eGFR	≦2 mg/dl/ >2 mg/dl
X15	Total bilirubin	≦2 mg/dl/ >2 mg/dl
X16	INR	<1.5/≧1.5
X17	HBV	YES/NO
X18	HCV	YES/NO
Y	SPCs	YES/NO

3. Results

In Figure 2, the analysis results after stratification of tumor size and clinical stage showed that in the tumor size (>5cm): the accuracy of <= IIb staging was the best with SVM (97.48%), the accuracy for > IIb staging was the best with SVM (90.03%). In the tumor size (<=5cm): the accuracy of <= IIb staging was the best with SVM (81.80%), and the accuracy of > IIb staging was the highest with SVM (83.75%).

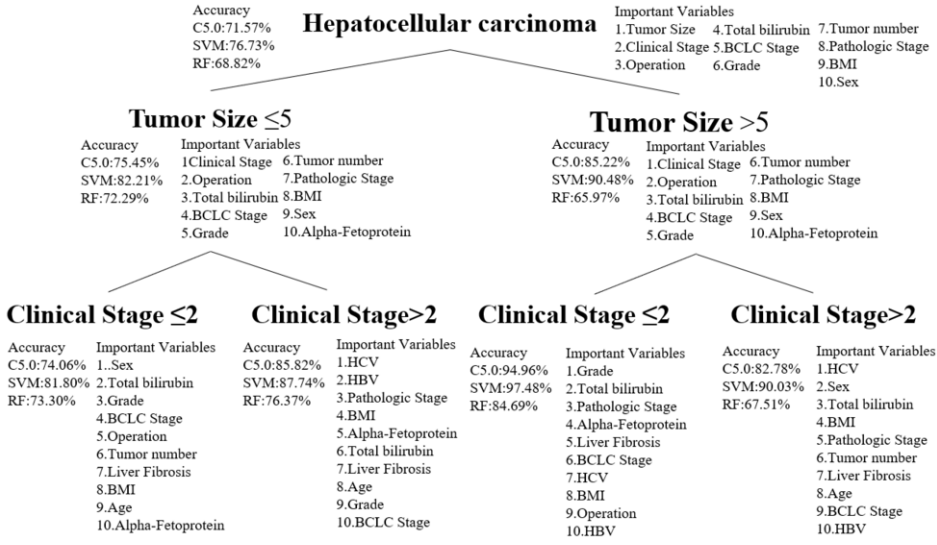


Figure 2. Analysis results after stratification of tumor size and clinical stage

4. Discussion

This study aimed to predict the risk factors of second primary Hepatocellular Carcinoma in survivors of Hepatocellular Carcinoma. The risk factors extracted from the classification models and association rules can be used to provide valuable information for HCC therapy. The models can be used to find the risk factors from database to provide valuable information for improving HCC patients.

Acknowledgements

This work was funded by grant no. CSH-2018-A-015 from the Chung-Shan Medical University Hospital.

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021, 71(3):209-249.

[2] The Taiwan Cancer Registry, The age-adjusted incidence rates: 2018. Availiable on https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=269&pid=13498 [Access date: 2021.08.10]

[3] Sun CC, Chang CC, Multiple primary malignant neoplasms: Results from a 5-Year Retrospective Analysis in a Metropolitan Hospital. Formosan Journal of Surgery. 2017, 50: 209-214.

[4] Chin C, Ting WC, Chang CC, Zhang YX, Prediction of risk factors for Synchronous Colorectal Cancer in Patients with Colorectal Cancer, Journal of Quality. 2020, 27: 23-37.

[5] Ting WC, Chang HR, Chang CC, Lu CJ, Developing a Novel Machine Learning-Based Classification Scheme for Predicting SPCs in Colorectal Cancer Survivors, Applied Sciences. 2020, 10: 1355.

[6] Chang CC, Chen SH, Developing a Novel Machine Learning-based Classification Scheme for Predicting SPCs in Women with Breast Cancer, Frontiers in Genetics. 2019, 10: 848.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210868

A Deep Learning Program to Predict Acute Kidney Injury

Xiaoqiang Li[a,1]
[a]*Heidelberg University, Germany*

Abstract. Acute kidney injury is a dangerous and sometime fatal clinical situation, which can cause irreversible damage. If we can predict it earlier and make appropriate prevention before its outbreak, kidney injury could be avoided. One challenge of early recognition of AKI is that the most e-alerts have focused on creatinine-based algorithms, but the elevation of serum creatinine lags behind renal injury. We use recurrent neural network (RNN) to make data mining on laboratory results of MIMIC-III Database. At first, we transfer the case data into Pandas DataFrame of series framed for supervised learning. Then we can use RNN predicts the next serum creatinine values (SCr) based on the last laboratory test results after emergency admissions. We train the RNN on whole dataset (i.e. multi-cases prediction) with LSTM. As the result shown, this prototype can predict criteria (SCr) of AKI with a RMSE (Root Mean Square Error) of 0.017mg/dL.

Keywords. acute kidney injury, prediction, artificial intelligence, neural network

1. Introduction

AKI (Acute kidney injury) is one of a number of conditions that affect kidney structure and function. AKI is defined as any of the following (Not Graded) [1]:

1. Increase in SCr (Serum creatinine) by >=0.3 mg/dl (>=26.5 lmol/l) within 48 hours; or
2. Increase in SCr to >=1.5 times baseline, which is known or presumed to have occurred within the prior 7 days; or
3. Urine volume <0.5 ml/kg/h for 6 hours.

A recent clinical practice assessment concluded there was an unacceptable delay in recognizing AKI in 43% of those that developed the condition after admission [4]. One challenge of early recognition of AKI is that the most e-alerts have focused on creatinine-based algorithms, but the current rule-based alerts cannot improve survival, because the elevation of serum creatinine lags behind renal injury [7].

[1] Corresponding Author, Xiaoqiang Li, MSc. Medicine Informatics, Heidelberg University, Germany. LIMS Software Engineer, Carl-Sonnenschein-Str 66 65936 Frankfurt Germany; E-mail: livvddcc@gmail.com.

2. Method

2.1. Mathematical methodology of this study

This study focuses on creating a program prototype to predict acute kidney injury by using the deep learning algorithm. Deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks (RNN) and convolutional neural networks have been applied to many fields, where they have produced results comparable to and in some cases surpassing human expert performance [2].

Step 1: We transform Bayes' theorem to a format, which looks like Sigmoid Function:

$$\begin{aligned} P(C_1|x) &= \frac{P(C_1,x)}{P(x)} \\ &= \frac{P(x|C_1)P(C_1)}{p(x|C_1)P(C_1)+P(x|C_2)P(C_2)} \\ &= \frac{1}{1+\frac{P(x|C_2)P(C_2)}{p(x|C_1)P(C_1)}} \\ &= \frac{1}{1+e^{-a}} \qquad [sigmoid] \end{aligned}$$

Where: P(C|x) is conditional probability, “a” represents function “a(x)”:

$$a(x)=ln\frac{P(x|C_1)P(C_1)}{P(x|C_2)P(C_2)}$$

Step 2: We use Gaussian distribution to calculate P(x|C):

$$P(x|C_1)\sim N(x|\mu_1,\Sigma)=\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}exp\{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\}$$

$$P(x|C_2)\sim N(x|\mu_2,\Sigma)=\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}exp\{-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\}$$

$$lnP(x|C_1)=-\frac{D}{2}ln(2\pi)-\frac{1}{2}ln|\Sigma|-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)$$

$$lnP(x|C_2)=-\frac{D}{2}ln(2\pi)-\frac{1}{2}ln|\Sigma|-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)$$

So we can transform the function “a” to a format like (w*x + b), take “a” back to step 1, we get the whole Sigmoid Function now:

$$\begin{aligned} a(x) &= lnP(x|C_1)-lnP(x|C_2)+ln\frac{P(C_1)}{P(C_2)} \\ &= (\mu_1-\mu_2)^T\Sigma^{-1}x-\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1+\frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2+ln\frac{P(C_1)}{P(C_2)} \\ &= w^Tx+w_0 \end{aligned}$$

Where:

$$w = \Sigma^{-1}(\mu_1 - \mu_2)$$

$$w_0 = \frac{1}{2}\mu_2^T \Sigma^{-1}\mu_2 - \frac{1}{2}\mu_1^T \Sigma^{-1}\mu_1 + ln\frac{P(C_1)}{P(C_2)}$$

Step 3: if we only have C1 and C2, then P(C2|x):

$$P(C_2|x) = 1 - P(C_1|x)$$

If we have C1, C2, … Cn, then P(Ck|x):

$$P(C_k|x) = \frac{P(x|C_k)P(C_k)}{\sum_j P(x|C_j)P(C_j)} = \frac{exp(a_k)}{\sum_j exp(a_j)}$$

Where k∈n.

After step 1 to 3, we get the Softmax function based on Bayes' theorem. At last, we build our recurrent neural network for prediction. On each ICU labor test (time series), the input is fed forward and a learning rule is applied. The back-connections save a copy of the previous values of the hidden units in the context units. Thus, it can learn the relation of time series data and make a prediction based on the history. That is why our RNN can predict the serum creatinine elevation, which normally lags behind the renal injury.

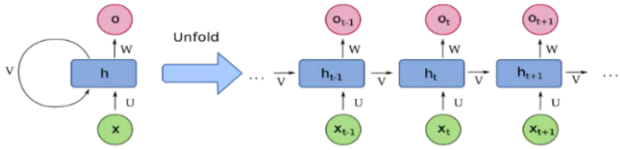


Figure 1. Structure of the recurrent neural network

Step 4: We use our neural network, which is based on Bayes' theorem, to predict real number (e.g., serum creatinine, range 0.00 – 2.00), but Bayesian network uses the probability value (0.00 – 1.00) as input and output. Therefore, we rescale the serum creatinine value to fit the input of Bayesian network and rescale the result (0.00 – 1.00) back to serum creatinine range (0.00 – 2.00) for prediction.

2.2. Data-Set

MIMIC-III (Medical Information Mart for Intensive Care III) is a database comprising de-identified health-related data associated with over forty thousand patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012 [3]. The patient data used in our study contains approximately 60000 admissions of patients including information such as patient demographics, vital signs, laboratory test results. In this prototype, we consider laboratory test results e.g., SCr (Serum creatinine), bicarbonate, blood urea nitrogen (BUN), chloride, international normalized ratio (INR), white blood count (WBC) as features, because these indicators are related to AKI

according to the past studies [4]. In this study the patient data do not have to be labeled manually.

2.3. Programming

Mining all the information from raw Data and dimensionality reduction are two main targets of data processing for neural network. Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning [5]. It cannot only process single data points, but also entire sequences of data. LSTM networks are well suited to making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. We use the ICU case data as multi-Case multivariate time series. At first, we transfer the case data into Pandas DataFrame of series framed for supervised learning. As example of this prototype, this program predicts the next serum creatinine values (SCr), which is the most important indicator of AKI.We use 3 time steps(n=3) for 1 prediction, i.e., we use the laboratory results of 3 time steps of one case to predict the next SCr value of next time step. For multi-cases predictions, the data-structure is as follow:

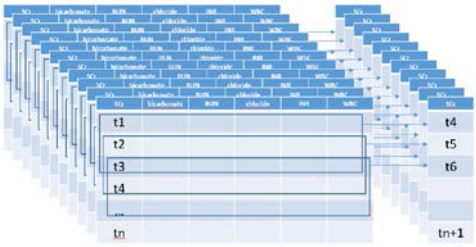


Figure 2. SCr. Prediction of multi-cases

2.4. Training of neural network

Keras API is used to train the dataset. Keras is a deep learning API written in Python, running on top of the machine learning platform TensorFlow [6], which is installed on CUDA GPU. GRU is also used and compared with LSTM to get better results and performance. 60% dataset is used for training and 40% for evaluation of training.

3. Result

Serum creatinine is reported as milligrams of creatinine to a deciliter of blood (mg/dL), n: timesteps for 1 prediction, inv_yhat: forecast, inv_y: actual, RMSE: Root Mean Square Error. We got following results:

n=1: inv_yhat: [1.33949034 1.44861898 1.37046677], inv_y: [0.8 0.8 1.3], Test RMSE: 0.489

n=2: inv_yhat: [1.32214486 1.41522023], inv_y: [0.8 1.3], Test RMSE: 0.378

n=3: inv_yhat: [1.28334119], inv_y: [1.3], Test RMSE: 0.017

If set n=4, we got “out of range” error, so we use n=3 as our result. The reason of “out of range error” is that some of our patients have only 4 labor results during their ICU stays. If n = 4, no prediction can be made. We should not delete such patients’ data

because they can give contribution for prediction if we set n<4. We train the RNN on whole dataset (i.e., multi-cases prediction). As the result shown, this prototype can predict criteria (SCr) of AKI with a RMSE (Root Mean Square Error) of 0.017mg/dL.

4. Discussion

The prediction of AKI is very challenging; therefore, this study is only trying to make a prototype of a prediction approach, following points can be considered for further improvement of the prediction's precision:

- Prediction should be grouped by demographics to avoid Simpson's paradox. E.g., using data of same age/sex group for prediction is better than mixed data.
- Influence of medications and comorbidities should be considered. Vital signs measured at the bedside such as ECG, SpO2, and respiration rate could also be helpful for AKI development prediction. However, because our activation function of neural network is based on Bayesian theorem, the conditional independence of features should be further analyzed and managed before we can use them to minimize overfitting.

5. Conclusion

Because of the complexity of the clinical data, it is difficultly to precisely analyze the clinical features manually by using statistical methods such as linear regression. Our purpose of feature engineering is keeping the original information of data and dimension reduction at same time, our algorithm can achieve these goals automatically, it can predict some important laboratory test results of acute kidney injury, such as serum creatinine, with a RMSE (Root Mean Square Error) of 0.017mg/dL. Therefore, our program can make early alert before AKI occurs.

References

[1] KDIGO Clinical Practice Guideline for Acute Kidney Injury. Kidney International Supplements (2012) 2, 2; doi:10.1038/kisup.2012.2 https://kdigo.org/

[2] Andrew Ng, Jiquan Ngiam, Chuan Yu Foo, Yifan Mai, Caroline Suen, Adam Coates, Andrew Maas http://ufldl.stanford.edu/tutorial/supervised/MultiLayerNeuralNetworks/

[3] MIMIC-III, Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). DOI: 10.1038/sdata.2016.35. Available at: http://www.nature.com/articles/sdata201635

[4] Predictive Modeling of the Risk of Acute Kidney Injury in Critical Care: A Systematic Investigation of The Class Imbalance Problem. Zhenxing Xu, Yujuan Feng, Yun Li, Anand Srivastava, Prakash Adekkanattu, Jessica S. Ancker, Guoqian Jiang, Richard C. Kiefer, Kathleen Lee, Jennifer A. Pacheco

[5] Sepp Hochreiter; Jürgen Schmidhuber (1997). "Long short-term memory". Neural Computation. 9 (8): 1735–1780. doi:10.1162/neco.1997.9.8.1735. PMID 9377276. S2CID 1915014.

[6] Tomašev, N., Glorot, X., Rae, J.W. et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 572, 116–119 (2019). https://doi.org/10.1038/s41586-019-1390-1

[7] Lachance P, Villeneuve PM, Rewa OG, Wilson FP, Selby NM, Featherstone RM, and Bagshaw SM, "Association between e-alert implementation for detection of acute kidney injury and outcomes: a systematic review," Nephrology Dialysis Transplantation, vol. 32, no 2, pp. 265–272, 2017.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210869

Designing an Electronic Shared Care Plan to Enable Person-Centred, Team-Based Home Care

Susi WILKINSON[a,c1] and Elizabeth M. BORYCKI[b,c]
[a] *Interior Health Authority, Kelowna, British Columbia, Canada*
[b] *School of Health Information Science, University of Victoria, Canada*
[c] *Michael Smith Foundation for Health Research, Canada*

Abstract. Optimization of interdisciplinary team-based care is an important strategy to improve home health care for older adults. An electronic, shared care plan was designed with aim to support shifts toward greater client and family involvement and interdisciplinary team functioning. Content for the home health collaborative care plan (HHCCP) was determined through a review of existing care plans, engagement with client and family partners, and usability testing by home health clinicians using a paper prototype. Client and family consultation and usability testing led to the identification of new and unique requirements for the HHCCP. We provide an example of how a paper prototype was used to validate and determine additional requirements for electronic clinical documentation and conclude that end users should be involved in the design of electronic clinical documentation as early as possible.

Keywords. Electronic shared care plan, patient centred care, home care

1. Introduction

British Columbia's health system is progressing through a system-wide redesign to become more person-focused, accessible, sustainable, and achieve better population health outcomes. Important strategies for seniors care are to optimize functioning of home health interdisciplinary teams and linkages with clients' primary care providers and teams. Within Interior Health Authority (IHA), home health interdisciplinary teams currently complete electronic clinical documentation in discipline-specific electronic documents. IHA leaders determined the need to develop an electronic collaborative care plan (HHCCP) in order to enable a person-centred, interdisciplinary team-based approach to delivery of home health services.

A collaborative or shared care plan is both a patient-centred health record that supports communication amongst the health team and a tool for clients to become as involved in their health care as they wish to be [1]. A recent literature review of health care transformations and care plans for frail, multimorbid elderly populations found the care plans did not capture client-reported outcomes nor the monitoring of care interventions [2]. The majority of care plans did not capture client perspectives about

[1] Corresponding Author, Susi Wilkinson, Interior Health Authority, Penticton Health Centre, 740 Carmi Ave, Penticton, BC, Canada, V2A8P9; E-mail: susi.wilkinson@interiorhealth.ca.

their health needs or what mattered to them in their health care. This paper describes the methodology used to determine the content and other design requirements for the HHCCP. The methodology is novel as it engages client (i.e. patients) and captures information about patient perspectives and outcomes.

2. Methods

The initial requirements for HHCCP were to enable all involved home health clinicians to document their discipline-specific goals and interventions of care into a single care plan. The methodology consisted of three steps that were used to identify detailed requirements for the HHCCP: (1) a review of internal and external care plan documents; (2) a consultation with client and family partners; and (3) usability testing using a paper-based prototype.

2.1. Step 1: Review of Care Plans

The design team conducted a review of paper and electronic care plan documents used in IHA. Shared care plans provided by the Institute for Healthcare Improvement [3] and the Agency for Healthcare Research and Quality [1] websites were also examined to identify additional information that would be valuable to include in the HHCCP.

2.2. Step 2: Client and Family Consultation

The British Columbia Patient Voices Network assisted with identifying adults over the age of 65 years to participate in a virtual focus group that occurred via teleconference and optional videoconference. Focus group participants received a brief description of the project and a sample shared care plan prior to the focus group meeting. The focus group posed two questions:

- How might a shared care plan be helpful or useful to clients and/or families
- What information should be included in the shared care plan.

The focus group teleconference was recorded to allow for review and analysis.

2.3. Step 3: Paper Prototype Testing

Content identified in the review of care plans and input from the client/family focus group was used to create the paper prototype. The design team met with representative home health clinicians (test clinicians) from across the IHA region to test the paper prototype according to the following procedure:

- The design team first created two client scenarios to enable a simulation for future use of the electronic HHCCP. The fictional client scenarios described the health and social situation for two imaginary clients previously unknown to home health whose care would involve several different home health disciplines. The scenarios also specified the order of team member involvement, several client goals and related interventions of care.
- The prototype testing sessions were audio-recorded and began with a review of the project, the HHCCP prototype and the test scenario. Each test clinician

received a paper copy of the HHCCP prototype and a test client scenario to reference during the testing session.

- The first involved clinician described the steps they would take to complete the initial assessment and care planning session with the test client and to complete the HHCCP paper prototype. The design team directed the clinician to think aloud while completing the HHCCP paper prototype.
- The paper HHCCP prototype was then passed to each of the identified additional team members, each of whom were asked to think aloud while documenting their respective additional client goals and care interventions into the HHCCP prototype.
- The design team reviewed the test session audio recordings to document participant questions, areas of confusion and concerns expressed about the HHCCP and incorporated changes into the HHCCP paper prototype.

3. Results

All of the methods used to gather HHCCP requirements led to the identification of new and unique requirements for the HHCCP. Content of the HHCCP and the identification source are described in Table 1.

Table 1. Content requirements for the electronic home health collaborative care plan.

Requirement	Description	Source
Client identifying information	Such as name, address, phone, health number, account number, primary care provider or most responsible provider.	Care plan review
What Matters to the Client	Free text capturing client's expressions of what matters to them	Care plan review
Health Care Needs	Free text capturing client's response when clinician asks "How can I help?"	Care plan review
Client Background	Free text field capturing client's response to inquiry re: ethnic background, personal identify, or language	Care plan review
Culture and Spiritual Practice	Free text capturing client's response to inquiry re: cultural or spiritual practices or belief system	Prototype testing
Additional Information	Free text capturing client's response to "What else is important for your care team to know about you?"	Client & Family engagement
Formal/Informal Supports	Free text capturing client's response to inquiry regarding others who help them with their health.	Care plan review
Team Members	Name, discipline, phone number and service end date for each involved home health team member	Care plan review, Prototype testing
Health Conditions	List of active health conditions	Care plan review
Advance Care Plan	Yes/No/In Progress response and mandatory free text description if Yes or In Progress is selected	Prototype testing
InterRAI Clinical Assessment Protocols	For clients assessed using the InterRAI-HC assessment, there is requirement to align care planning with any identified clinical assessment protocols	Care plan review

Care Goals and Next Steps	Client goals and related interventions are documented in the appropriate category: Functional Mobility; Pain Management; Medication Management; Bowel/Bladder Management; Nutrition/Hydration/Swallowing; Cognitive Functioning; Mental Health and Substance Use; Social Determinants of Health; Other Health Challenges	Care plan review Prototype testing
Team Notes	Free text used for client-sensitive planning information which does not appear on client copy of the HHCCP	Paper prototype testing
Print Client Copy	Yes/No response indicating if client would like a copy of the HHCCP	Client & Family engagement
Name and Credentials		Care plan review

4. Discussion

While a review of existing care plans identified most of the content for the HHCCP, additional requirements were determined through consultation with end users of the HHCCP; namely client and family partners and clinicians. All of the client and family partners consulted in this project identified benefits of the HHCCP and provided examples of how they would use the HHCCP if/when it was to become available to them. Clients and families identified a new requirement to ask clients if there was anything else they wanted their care team to know about or consider in their health care and confirmed the importance of offering clients a copy of the care plan. During the paper prototype tests, home health clinicians identified requirements to document advance care plan information, create a section for client-sensitive information that would not appear on the client copy, and to separate the client background, identity, culture and spirituality question into two distinct questions. Clinicians also identified the need to include the service end date for each involved team member.

5. Conclusions

We conclude that clients and families should be engaged into the design of clinical documentation that is intended to be shared or used by them. We provided an example of how a paper prototype can be used to define and validate requirements for electronic clinical documentation and also recommend that clinician users become involved early in the design of electronic clinical documentation.

References

[1] Agency for Healthcare Research and Quality. *Develop a Shared Care Plan.* Available from: https://www.integrationacademy.ahrq.gov/products/playbooks/behavioral-health-and-primary-care/implementing-plan/develop-shared-care-plan [Accessed April 30, 2019].

[2] Berntsen G, Strisland F, Ing D, Malm-Nicolaisen K, Smaradottir B, Fensli R, Rohne M. The Evidence base for an ideal care pathway for frail multimorbid elderly: Combined scoping review and systematic intervention review. JMIR. 2019;21(4):e12517.

[3] Institute for Healthcare Improvement. My Shared Care Plan. Available from: https://www.ihi.org/resources/Pages/Tools/MySharedCarePlan.aspx [Accessed April 30, 2019].

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210870

Reviewing and Content Analysis of Persian Language Mobile Health Apps for COVID-19 Management

Leila ERFANNIA[a,b], Morteza AMRAEI[c], Goli ARJI[d], Azita YAZDANI[b,1], Maryam SABZEHGAR[a] and Leila YAGHOOBI[a]

[a] *Health information technology department, Paramedical school, Zahedan University of Medical Sciences, Zahedan, Iran*
[b] *Clinical education research center, health human resources research center, School of health management and information sciences, Shiraz University of Medical Sciences, Shiraz, Iran*
[c] *Department of health information technology, faculty of allied medical sciences, Lorestan university of medical sciences, Khorramabad, Iran*
[d] *Saveh university of medical sciences, school of nursing and midwifery, health information technology department, Saveh, Iran*

Abstract. The present study aimed to systematically search in app stores and intended to carry out content analysis of free Persian mobile health apps in the management of COVID-19 and, ultimately determine the relationship between the popularity and quality of these apps. According to a researcher-made checklist including five axes of ***ease of use***, ***privacy***, ***data sharing***, ***education***, and ***monitoring***, four app markets such as ***Myket***, ***Bazzar***, ***Google Play*** and ***App Store*** were searched from May 2021 up to now. The findings showed that all selected apps performed well in terms of ease of use and privacy but they needed to be improved in terms of education, monitoring, and data sharing. Also, there was no significant relationship between the popularity and quality of these apps. Owing to the high penetration rate of smartphones in Iran and the low popularity of COVID-19 apps, government, developers, and investors are required to improve the quality of apps and their marketing.

Keywords: Mobile health app, COVID-19, Persian app, Content analysis.

1. Introduction

Recently, outbreak of SARS-CoV-2 pandemic and the subsequent increase in mortality rate stressed the need for gaining information on various aspects of the virus [1]. Of course, The number of mobile health (m-health) apps has grown dramatically since the outbreak of the COVID-19 pandemic. COVID- related-m-health apps enjoy a number of other useful functions, including contact tracing and symptom control by health care

[1] Corresponding author, Azita YAZDANI Ph.D. of Medical Informatics, Assistant Professor, Clinical Education Research Center, Health Human Resources Research Center, School of Health Management and Information Sciences, Shiraz University of Medical Sciences, Shiraz, Iran. Email: a_yazdani@sums.ac.ir. ORCID ID: *0000-0002-5190-286X*

professionals and monitoring the number of infections [2].Hence, Evaluating the content and features of mobile apps is very important to guide users in choosing the right mobile app based on their needs. It should be noted that the studies conducted so far to evaluate the quality of COVID-related apps are mainly in English [3]. Also, A study that focused on Persian language considered all apps and systems and evaluated them based on ***MARS*** scale [4].Accordingly, the present study focuses solely on free Persian m-health apps in managing COVID-19 and uses a researcher-made checklist based on specific pandemic features to evaluate the apps.

2. Method

Based on PRISMA 2020 steps, to find Persian mobile apps, ***Myket***, ***App Store*** and ***Bazaar*** mobile app stores were selected due to the coverage of ***Persian*** m-Health apps and ***App Store*** and ***Google Play*** platforms and web search were also selected to get the desired results. Then, Persian keywords including "***کرونا***" and "***کووید-19***" from May 2021 were considered as our main search. Inclusion Criteria, here, included mobile apps that launched at the time of the epidemic**,** and focused on the management of COVID disease in terms of diagnosis, education, follow-up and counseling**.** 46apps selected for full review**.** At this stage, the assessors fully evaluated the remaining apps regarding capabilities and functions and excluded apps that focused only on one aspect and did not cover other related areas. Then 21apps entered the final evaluation stage (Figure1).

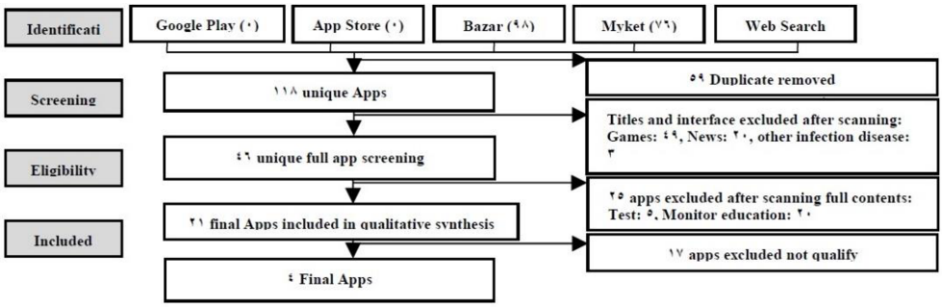


Figure1. The PRISMA flowchart of the screening process of the reviewed national Persian apps related to COVID

A researcher-made checklist (this is available upon request from the corresponding author) was designed based on m-Health programs evaluation criteria, literature review, and WHO guidelines on digital health interventions to classify the app functions [5-7]. It consisted of 37 questions in five axes (includes: ***ease of use***, ***privacy***, ***data sharing***, ***education*** and ***monitoring)*** scored on a two-point scale, one for ***Yes*** and zero for ***No***. Two assessors evaluated each app independently and every discrepancy was dealt with in a joint meeting with a third assessor. Subsequently, Apps with scores higher than the average score (18.5) were rated as acceptable. Accordingly, only did four apps receive acceptable scores, and the others were excluded from the study.

3. Results

All mobile apps in the qualitative step were evaluated according to each of the five axes. Five items were examined in ***ease of use*** axis namely as: flexibility and customization, understandable figures and icons, application assistance, data entry, and page forward and backward easiness). Overall, ***Mask***, ***AC19*** and ***Safirane Salamat*** apps got the most points from this axis (5 points each). Four items were examined in ***privacy*** axis -i-e- Legal commitment, username and password, username and passwords change, reregistration). ***Safirane Salamat*** and ***Corona (online diagnosis/statistics/education)*** apps obtained the highest scores in this axis (4 points each). Eleventh items were reviewed in ***data sharing*** axis-i-e- access to health history, volunteers assistance in crisis, search in app, send health reminders, link to educational pages, demographic data entry, data sharing from app, automatic data entry from user profile, online consultations, users feedback entry, users experience sharing and forums). The highest score in this axis belonged to ***Safirane Salamat***, ***Corona (online diagnosis/statistics/education)***, and ***AC19*** apps (9, 6, 6 points respectively). In ***education*** axis, nine items were examined as follows: education by images, health events information, patients and relatives educations, COVID death information, latest COVID cases according geographical parts, COVID statistics by tables and graphs, prediction COVID pattern in future, provide prevention video, send periodic massages for users). Overall, ***AC19*** app obtained the highest score in this axis (Score 9 out of 9). In ***monitoring*** axis, eight items were examined as follows: user tracking, patients monitoring from home, introducing labs and medical imaging centers, follow up patient's symptoms, evaluating possibility of COVID infected, monitored quarantine, introducing symptoms which need physician visit, providing health center address and map). Altogether, ***Safirane Salamat*** and ***Mask*** apps received the highest scores in this axis (6 points each). Having determined the final score of the apps, the download rate was extracted based on the download statistics of the app stores. Then, Pearson correlation test was used to test the relationship between the total score of the apps and the download rate and, of course, no significant correlation was observed between them ($p = 0.690$, $r = 0.310$).

Table1. Final national Persian COVID-19 mobile apps features.

App Name	Star rating/form	Download	Main Features	Platform	Size (Meg)
Safirane Salamati	4/169	+23000	Education	Android/iOS	25.3
AC19	4.1 /6651	+50000	Risk assessment/self-assessment	Android	5
Mask	4.1 /6719	+700000	Education/risk assessment	Android	27
Corona(online diagnosis/statistics/education)	3.5/11	+500	Education/ risk assessment	Android	14

Based on the total points, the ***Safirane Salamat*** app obtained the highest score (score of 29), followed by ***Ac19*** (score of 26) and ***Mask (***score of ***21)***. These selected apps, and **Corona (online diagnosis/statistics/education)**; (score of 19), obtained more than 50% of the quality review score. The features of final selected apps are demonstrated in Table1.

4. Discussion

Needless to say that Public-private partnerships can contribute to the applications success and involve in achieving health goals [8]. *Mask* app, the most popular Persian m-health app, was launched through the government/ non- government participation. Reportedly, the ease of using m-health apps and education seems to be one of the effective items for m-health apps. All four apps received acceptable scores for ease of use. User education is one of the capabilities of m-health apps in preventing the spread of disease. *AC19* app obtained the highest scores in this area and provided education content. The only Persian app which was capable of tracking users' calls from the beginning was the *Mask*. *Mask* and *Safirane Salamat* were able to provide strategies to support people in quarantine. According to the results of the correlation analysis, there was no significant relationship between the popularity of apps and their quality, which is consistent with the results of Wang's study [9].Hence, it can be concluded that the popularity of apps such as ads or search engine optimizations can be increased without necessarily being a higher quality than other apps. The final apps have been successful in *ease of use* and *privacy*. These apps need to be upgraded in terms of data sharing, *education*, and *monitoring*. Online consultation were not considered in the selected apps, though. As far as *education axis was concerned*, most apps provided educational information from the *Ministry of Health and Official Scientific* resources due to government affiliation, which were the strong points of these apps. In terms of monitoring, apps were only capable to assess the risk and make recommendations to visit the doctor, and were not able to monitor patients or suspected cases or people in quarantine. User tracing and quarantine monitoring were among the most important features of COVID- 19 apps, which were neglected in Persian m-health apps, though. The results of the present study showed that the most successful Persian m-health apps have been designed, launched, and even marketed through government support.

References

[1] Abd-Alrazaq A, et al. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. Journal of medical Internet research, 2020 March;22(4):e19016.
[2] Abeler J, et al. COVID-19 contact tracing and data protection can go together. JMIR mHealth and uHealth, 2020 April;8(4):e19359.
[3] Blacklow SO, et al. Usability, inclusivity, and content evaluation of COVID-19 contact tracing apps in the United States. Journal of the American Medical Informatics Association, 2021 June;28(9):1982-9.
[4] Abdolkarimi S, Mousavi B. Content Analysis of Health-Based Iranian Systems and Apps on Covid-19. Iranian Journal of War and Public Health, 2021 March;12(4):223-233.
[5] Ming LC, et al. Mobile health apps on COVID-19 launched in the early days of the pandemic: content analysis and review. JMIR mHealth and uHealth, 2020 Sep;8(9):e19796.
[6] Nouri R, et al. Criteria for assessing the quality of mHealth apps: a systematic review. Journal of the American Medical Informatics Association, 2018 Aug;25(8):1089-98.
[7] Guideline W. Recommendations on digital interventions for health system strengthening. World Health Organization, 2019:2020-10.
[8] Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, Smith AC. The role of telehealth in reducing the mental health burden from COVID19. Telemedicine and e-Health. 2020;26(4):377-9
[9] Wang X, Markert C, Sasangohar F. Investigating Popular Mental Health Mobile Application Downloads and Activity During the COVID-19 Pandemic. Human Factors. 2021 Mar;0018720821998110.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210871

Mobile-Based Self-Care Application for COVID-19: Development Process Using the ADDIE Model

Hamid Reza SAEIDNIA[a], Marcel AUSLOOS[b], Zahra MOHAMMADZADEH[c], Afshin BABAJANI[a] and Mohammad HASSANZADHH[a,1]
[a] *Department of Knowledge and Information Science, Tarbiat Modares University (TMU), Tehran, Iran*
[b] *Department of Statistics and Econometrics, Bucharest University of Economic Studies, Bucharest, Romania*
[c] *Department of Health Information Technology, Kashan University of Medical Sciences and Health Services, Kashan, Iran*

Abstract. During the COVID-19 era, technology-enhanced protection of this disease has saved lives in developed countries in which citizens have the privilege of accessing and using such technologies to fight Coronavirus. In the undeveloped countries, on the other hand, citizens have had no accession or ability to use digital technologies to prevent COVID-19. Having this in front, in the MyShield research project, we aim to address how to teach self-care skills in undeveloped countries in the era of COVID-19 using a mobile low-cost application effectively based on a standard educational model (ADDIE). This paper reports a framework that arises from the results of semi-structured interviews and online workshops conducted in the ADDIE design process for the self-care mobile application. The specialists contributed to indicte the appropriate content for teaching self-care skills while informants contributed to optimize the user experience flow.

Keywords. Self-care, COVID-19, Mobile Application, Undeveloped Countries.

1. Introduction

In the early days of the Coronavirus epidemic, developed and high-income countries used digital technologies (e.g. Contact tracing, IoT, Big-data analytics, and AI) to fight the COVID-19 pandemic [1]. Singapore used an AI Chatbot, Hong Kong and South Korea used contact tracing technology, Canada and Australia used (COVID Alert app) and (COVIDSafe app) respectively to prevent and inform citizens about COVID-19 [1-3]. Undeveloped countries, on the other hand, are quasi unable to create or use digital technology to prevent and fight the pandemic COVID-19. Moreover, undeveloped countries in the COVID-19 era face more problems and harm than developed countries due to poverty and corruption in the economy, a weak social health system, and an inability to meet self-care requirements [4]. In a 2020 study examining the number of mobile applications available to fight and prevent COVID-19, researchers found that

[1] Corresponding Author, Mohammad HASSANZADHH, Department of Knowledge and Information Science, Tarbiat Modares University, Tehran, Iran. E-mail: hasanzadeh@modares.ac.ir

over 25 pertinent applications have been developed in areas such as remote assistance, patients monitoring, current status, and COVID-19 prevention [5]. Nonetheless, none of these applications focuses on training self-care skills in undeveloped countries.
MyShield is a research project that aims to develop a low-cost and high-impact digital technology for undeveloped countries through a standalone mobile application based on the standard educational design method (ADDIE). Our paper presents the results from an early project phase, where the specialists contributed appropriate content for teaching self-care skills while informants contributed to the user experience flow.
In the study, our aim is to answer two questions:

RQ1: *In the era of COVID-19, what are the appropriate contents for teaching self-care skills?*

RQ2: *How should the features and user experience flow of a mobile application for teaching self-care skills be designed?*

2. Methods

The MyShield project employs the ADDIE design approach [6] for the mobile application development that is conducted in five phases: 1) Analysis, 2) Design, 3) Development, 4) Implementation, and 5) Evaluation. The project is in an early stage; this paper reports on the first phase, where 5 specialists and 14 informants helped us identify and collect the data required for the mobile application. For the response to the first question, semi-structured interviews were conducted with 5 specialists (3 infectious disease specialists and 2 nursing specialists) for achievement on the appropriate content for teaching self-care skills. To answer the second question, we held an online workshop with the presence of informants with the Zoom software. These were (are) experts in the fields of health information technology, medical informatics, user experience designers, and as well as Android mobile application developers. This cohort was enrolled using a call via social networks (Instagram and Facebook) published on 1-20 March 2021. Ultimately, 14 informants responded to our call: 6 females and 8 males. The average age of the informants was 38 years. At the beginning of the online workshop, we introduced the participants to the research and explained the objectives of the MyShield project. In the next step of our online workshop, we asked the informants to explain in detail the requirements, feature list, and user flow of their mobile application user experience. Data was collected through video recordings and comments during the online workshop, which lasted an average of three hours. Qualitative methods were used to analyze and categorize user needs, requirements application, and user experience for the self-care (MyShield) mobile application. In order to display the user experience flow, the Balsamiq Mockups software was used.

The study was conducted obtaining the approval of the Ethics Committee of Tarbiat Modares University, Tehran, Iran (IR.MODARES.REC.1399.142).

3. Results

Response to RQ1: As a result of the interviews with specialists, the following outcomes were reached: In the application, standard and scientifically accepted contents should be considered. As a result, the application must be easily updated so that users can obtain any useful and valuable content in the subject area of COVID-19 self-care without

interruption. Specialists agreed that the application would be better if covering the following topics: motivating yourself, daily life management, personal hygiene, healthy eating, and exercise. Self-motivation content needs to teach users how to think positively, motivate themselves, and utilize their intrinsic motivations. Even in these matters, spirituality motivation can be used. In matters related to daily life,, the teaching of strategies for stress, fear, and anger management must be considered. Education is also essential in other areas, such as managing life during personal quarantine and social activities. Personal hygiene can include education about masking and observing social distance, washing hands, and knowing how to employ disinfectants. Educating people about healthy eating should include these topics: useful supplements, beverages, fatty foods, proteins, dairy products, nutrition programs, and smoking. The topics that are covered in exercise education include exercising at home, conditions for exercising outdoors, professional exercise, intensity of exercise and, exercise in self-quarantine.

Response to RQ2: Informants agreed on the following flow of features of the mobile application (MyShield): start the application with a splash screen. After that, the registration screen should appear. Registration should be quick and easy, and the user should be asked for only a few pieces of information (i.e. E-mail, user name, and password). Following that, the intro slider should be displayed on the first launch of the application. It is better to introduce the application in at least 3 slides. After the introduction slider, the showcase feature should appear for users to become familiar with the application. Following these screens, the user should be presented with the main menu. The main menu can use two action bars. One at the top of the page displaying the search features of the entire application and the other at the bottom of the page containing buttons, messages, help, settings, and profiles. Additionally, the hamburger button can be made visible in the left-hand corner of the screen. Tapping on it should display the application settings. The settings screen should include features i.e. font setting, title search, clean checkmark, find, last read, dark them, notification, changing password, and about us. Educational content must be button shared and checked marked. Content should also be able to be liked or disliked by the user.

Figure 1 shows the user experience flow using Balsamiq Mockups software.

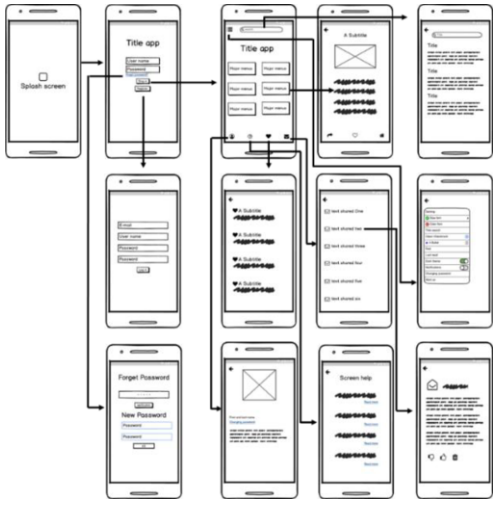


Figure 1. User experience mobile application self-care flow.

4. Discussion

Five specialists and fourteen informants participated in this study with the aim of responding to two research questions (RQs) which are parts of the first phase of the MyShield project (Analysis phase). On RQ1: According to the specialists, the main educational menus are five topics (motivating yourself, daily life management, personal hygiene, healthy eating, and exercise). The importance of managing stress and fear in the face of COVID-19 has been emphasized in other studies [7]. In addition, maintaining a healthy diet and regular exercise can greatly enhance the immune system's ability to fight against COVID-19 [8]. The content expressed by specialists is of great importance for use in the application and has been incorporated in most self-care research related to the COVID-19 era [9]. On RQ2: The informants in the online workshop helped us to draw the user experience flow and features of the application (MyShield). During the COVID-19 Pandemic, users will find m-Health applications of better acceptability if they are designed with an effective user interface and based on user experience [10]. Some limitation to this study is the small number of informants. Nevertheless, it was determined that the participants substantially represented the intended user group of the mobile application. Future research can address with a more number of informants.

5. Conclusion

We are pleased that interviews with specialists provided us with titles and content for the application, while the online workshop with informants led to a pertinent final version of the application user experience flow for MyShield.

References

[1] Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nature medicine. 2020;26(4):459-61.
[2] Gómez-Ramírez O, Medeiros P, Wainer R, Iyamu I. 13 Does the 'Canada COVID-19 alert'app stand up to critical scrutiny? A rapid qualitative assessment. British Medical Journal Publishing Group; 2021.
[3] Abbas R, Michael K. COVID-19 contact trace app deployments: learnings from Australia and Singapore. IEEE Consumer Electronics Magazine. 2020;9(5):65-70.
[4] Dahab M, Van Zandvoort K, Flasche S, Warsame A, Ratnayake R, Favas C, et al. COVID-19 control in low-income settings and displaced populations: what can realistically be done? Conflict and health. 2020;14(1):1-6.
[5] Islam MN, Islam I, Munim KM, Islam AN. A review on the mobile applications developed for COVID-19: an exploratory analysis. IEEE Access. 2020;8:145601-10.
[6] Branch RM. Instructional design: The ADDIE approach: Springer Science & Business Media; 2009.
[7] Hagger MS, Keech JJ, Hamilton K. Managing stress during the coronavirus disease 2019 pandemic and beyond: Reappraisal and mindset approaches. Stress and Health. 2020.
[8] Sidor A, Rzymski P. Dietary choices and habits during COVID-19 lockdown: experience from Poland. Nutrients. 2020;12(6):1657.
[9] Saeidnia H, Mohammadzadeh Z, Saeidnia M, Mahmoodzadeh A, Ghorbani N, Hasanzadeh M. Identifying Requirements of a Self-care System on smartphones for preventing coronavirus disease 2019 (COVID-19). Iranian Journal of Medical Microbiology. 2020;14(3):241-6.
[10] Saeidnia HR, Mohammadzadeh Z, Hassanzadeh M. Evaluation of Mobile Phone Healthcare Applications During the Covid-19 Pandemic. Studies in health technology and informatics. 2021;281:1100-1.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210872

Minimization of the Drug and Gene Interactions in Polypharmacy Therapies Augmented with COVID-19 Medications

Azra LAGUMDZIJA-KULENOVIC[a,1] and Adnan KULENOVIC[a]
[a]*Absolute Information Age, Inc. Toronto, Canada*

Abstract. Medications *Dexamethasone, Remdesivir* or *Colchicine,* used to treat *COVID-19* patients, have significant interactions with other medications and the human genome. The study presented in this paper investigates how to use the *Personalized Medicine Therapy Optimization Method (PM-TOM*) to minimize these interactions in polypharmacy therapies of *COVID-19* patients. We applied *PM-TOM* on the *EMR* database of *Harvard Personal Genome Project (PGP)*, drug database *DrugBank* and *Comprehensive Toxicogenomics Database (CTD)* to analyze polypharmacy therapies augmented with these medications. The main finding is that these *COVID-19* medications significantly increase the drug and gene interactions in partially optimized (or unoptimized) therapies, which is not the case in the fully optimized ones. For example, the test results show that in polypharmacy treatments for patients having between 3 and 8 conditions, the average number of drug and gene interactions in partially optimized therapies ranges from 3 to 18 after adding *Remdesivir*, 4.3 to 20 *Colchicine,* and 4.7 to 23 *Dexamethasone*. On the other hand, these interactions in fully optimized therapies range only 0.6 to 5.2, 1.2 to 7, and 2.7 to 11, respectively. These results suggest that polypharmacy therapies should be carefully examined before adding these medications. This recommendation applies to all other situations when polypharmacy patients may conduct new serious conditions, such as *COVID-19*, requiring additional medications with a high number of drug and gene interactions.

Keywords. *COVID-19*, *PM-TOM*, Clinical Decision Support Systems (*CDSS*), Dexamethasone, Remdesivir, Colchicine, Polypharmacy, Multimorbidity, Drug-Drug Interactions *(DDI)*, Drug-Gene Interactions *(DGI)*.

1. Introduction

The *Medical Guidelines for Treating COVID-19 Patients* [1] of *the US National Institutes of Health (NIH)* recommend drugs for various categories of these patients. For example, high-risk hospitalized patients who do not require supplemental oxygen should be treated with *Remdesivir*. *Dexamethasone* is recommended only for patients who need invasive mechanical ventilation or extracorporeal membrane oxygenation (*ECMO)*. *NIH* also reported several active or completed studies on the benefits of *Colchicine* for *COVID-19* patients. For example, study [2] found that the median hospitalization time was reduced from nine to seven days for the *COVID-19* patients treated with *Colchicine*.

[1] Corresponding Author, Azra Lagumdzija-Kulenovic, Absolute Information Age, Inc. Toronto, Canada; E-mail: azra.kulenovic@rogers.com.

While these medications could help the *COVID-19* patients to overcome this infection without serious outcomes, they may also pose a significant risk due to their numerous drug-drug interactions and drug-gene interactions (*DDIs*, *DGIs*). For example, the total number of *DDIs* for *Dexamethasone* is 1,211, *Colchicine* 1,297 and *Remdesivir* 442, according to *DrugBank* [3].

Several studies, for example [4], report a high occurrence of adverse drug and gene interactions in polypharmacy cases (66-87% of the patients had *DDIs* and 34-73% *DGIs*). In addition, study [5] found that adverse drug reactions and interactions were between the fourth and sixth causes of death in the US. So, the possible adverse drug and gene interactions of the *COVID-19* medications should be carefully considered before adding them to polypharmacy treatment of multimorbidity patients infected with *SARS-CoV-2.*

This study used the *Personalized Medicine - Therapy Optimization Method* (*PM-TOM*) [6, 7] to optimize *DDIs* and *DGIs* in polypharmacy treatments extended with the *COVID-19* medications. To the best of our knowledge, *PM-TOM* is the only published method for optimizing both the drug-drug and drug-gene interactions in polypharmacy treatments. The results suggest that these interactions can be significantly reduced, thus increasing the chances for multimorbidity patients to overcome *COVID-19* with minimal consequences.

2. PM-TOM Method

The ***PM-TOM* datasets** are a database of Electronic Medical Records (*EMR)*, Drug Repository *(DRG)*, and Gene-Drug Repository *(GDR)*. *PM-TOM* retrieves the patient's conditions, prescriptions and genome from *EMR* (only pharmacogenetic and pathogenic gene variations). The drug repository *(DRG)* offers data about the active medication ingredients (referred to as "drugs" in this paper), commercial drug products (referred to as "drug products") and drug-drug interactions. Gene-Drug Repository *(GDR)* keeps data about drug-gene interactions and their severities. Further information about the *PM-TOM* Data Model can be found in papers [6, 7].

***PM-TOM* inputs** are a patient's: conditions, genome and drug products considered by a physician, clinician or clinical pharmacist for treating these conditions. Several drug products can be provided for each condition, with an indication of the preferred one. The list of preferred drug products forms so-called *Initial Therapy (IT)*.

***PM-TOM* outputs** are *(i) Optimal therapy for considered drugs (OTCD)*, *(ii) Optimal therapy for all drugs (OTAD)* and *(iii)* a *Personalized Therapy Report (PTR)*. *PM-TOM* creates *OTCD* after examining drug and gene interactions in the considered drugs and *OTAD* after looking into all drugs in *EMR* prescribed for the same conditions by other physicians, clinicians or clinical pharmacists. In the rest of this paper, *OTCD* will also be referred to as *partially optimized therapy* and *OTAD* as *fully optimized therapy*. Lastly, *PTR* reports the patient's demographic data, conditions, considered drugs and drug products, partially and fully optimized therapies, along with their *Cumulative Adverse Drug Interactions Indicators* (*CADI*), explained below.

As the number of all combinations of potential medications for a patient's conditions could be rather high, *PM-TOM* implements an **iterative heuristic algorithm** [6] that examines a select set of candidate therapies rather than all possible ones. For finding an *OTCD,* these candidate therapies are derived from the initial therapy *(IT)* by replacing only one drug preferred for a condition with another drug considered for the

same condition. For creating the corresponding *OTAD,* the *PM-TOM* would repeat the same process, starting with *OTCD* rather than *IT* and examining all medications from *EMR* indicated for the same conditions.

For each candidate therapy, *PM-TOM* calculates indicator *CADI* as a weighted sum of its drug-drug and drug-gene interactions [6, 7]. The optimal therapy will be formed by selecting a drug for each condition whose candidate therapy has minimal *CADI* against the candidate therapies formed from other drugs considered for the same condition. In that way, the algorithm evaluates only a small subset of all candidate therapies, which significantly reduces its complexity.

3. Testing Results and Discussion

In this study, we implemented the *PM-TOM* method on 484 *EMRs* from the *PGP* database of the *Harvard Personal Genome Project* [8]. These records include the patient's conditions, drugs prescribed for the treatment of these conditions and genome. In addition, the drug database *DrugBank* [3] was used to find the drug-drug interactions (*DDIs)* and *the Comprehensive Toxicogenomics Database (CTD)* [9] to retrieve drug-gene interactions *(DGIs).*

We added condition *COVID-19* and then medications *Dexamethasone, Remdesivir* or *Colchicine* to the *PGP EMRs*. Then we used *PM-TOM* to calculate the ***CADIs*** of their ***OTCD*** and ***OTAD*** therapies before adding each of these medications and after. Table 1 summarizes information about the groups of examined *EMRs*: the number of conditions before the patient contracted *COVID-19* ***(Conditions#)***, the total number of patients in each group ***(Cases#)***, the average number of patients' pharmacogenetic and pathological genes ***(Genes#)*** and the average number of all drugs from *EMR* indicated for patient's conditions ***(Drugs#)***. It also shows the average ***CADIs*** of ***OTCDs*** and ***OTADs*** in the original therapies and therapies augmented with each *COVID-19* medication.

Table 1. Testing cases and their drug and gene interaction indicator *(CADI)* before and after applying *COVID-19* medications. *OTCD-Optimal Therapy for Considered Drugs, OTAD-Optimal Therapy for All Drugs.*

Conditions #	Cases #	Genes #	Drugs #	OTCD CADI before COVID	OTAD CADI before COVID	+Remdesivir OTCD CADI	+Remdesivir OTAD CADI	+Colchicine OTCD CADI	+Colchicine OTAD CADI	+Dexameth. OTCD CADI	+Dexameth. OTAD CADI
1	272	5.4	9.6	0.2	**0.03**	0.5	**0.06**	1.0	**0.33**	2.8	**2.2**
2	122	7.4	18.5	1.0	**0.3**	1.7	**0.4**	2.4	**0.7**	4.7	**3.2**
3	49	3.9	28.5	2.1	**0.4**	3.0	**0.6**	4.3	**1.2**	5.5	**2.7**
4	16	4.4	28.7	3.4	**1.1**	4.9	**1.6**	6.1	**2.7**	7. 6	**4.4**
5	13	4.5	41.8	4.9	**2.1**	6.6	**2.7**	8	**3.5**	10.2	**6.1**
6	3	0	49. 7	8.3	**5.3**	10.3	**5.7**	12. 7	**7**	12.3	**7**
7-8	5	6	61.8	15.4	**5**	18	**5.2**	20.4	**7**	23.2	**11**
>8	4	15.5	49.5	24.2	**19.5**	25.2	**21.2**	31.2	**29**	34	**28**

These results confirm previous findings that *PM-TOM* can find polypharmacy treatments with significantly reduced drug and gene interactions. For example, in the

group of three conditions, the average *OTCD's CADI* is 3.0 after adding *Remdesivir,* while the average *OTAD's CADI* is only 0.6. For *Colchicine,* these reductions are from 4.3 to 1.2 and *Dexamethasone* from 5.5 to 2.7. In the group with five conditions, these decreases are even more substantial: 6.6 to 2.7, 8 to 3.5 and 10.2 to 6.1.

The second observation is that these *COVID-19* medications significantly increase the drug and gene interactions in partially optimized therapies and, consequently, unoptimized ones, which is not the case in fully optimized therapies. For example, as shown in Table 1, in polypharmacy patients having 3 to 8 conditions, the *CADIs* of partially optimized therapies including *Remdesivir* would increase from 3 (in 3 conditions) to 18 (in 8 conditions), for *Colchicine* from 4.3 to 20 and *Dexamethasone* 4.7 to 23. On the other hand, the *CADI* of the fully optimized treatments ranges only from 0.6 to 5.2, 1.2 to 7, and 2.7 to 11, respectively.

4. Conclusion

These results further emphasize a need for the minimization of adverse drug and gene interactions in polypharmacy therapies. The above findings also show significant room for further fine-tuning of these therapies should a physician, clinician, or clinical pharmacist opt to modify *PM-TOM* recommendations due to the aspects not currently considered in this method, such as patient's age, gender, ethnicity, etc.

Proper selection of polypharmacy therapies is particularly important when the polypharmacy patients could acquire new severe conditions, such as *COVID-19*, which may need medications with a high number of drug and gene interactions, like *Dexamethasone* or *Remdesivir* or *Colchicine*.

References

[1] COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19). National Institutes of Health (NIH).

[2] Lopes MI, Bonjorno LP, Giannini MC, et al. Beneficial effects of Colchicine for moderate to severe COVID-19: A Randomized, Double-blinded, Placebo-controlled clinical trial. RMD Open 2021;7: e001455. doi:10.1136/rmdopen-2020-001455

[3] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic acids research. 2018 Jan 4;46(D1):D1074-82.

[4] Reynolds KK, Pierce DL, Weitendorf F, Linder MW. Avoidable drug-gene conflicts and polypharmacy interactions in patients participating in a personalized medicine program. Personalized Medicine. 2017;14:221-233.

[5] Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 1998;15:1200-1205.

[6] Kulenovic A, Lagumdzija-Kulenovic A. PM-TOM: A method for finding personalized polypharmacy therapies with minimal adverse drug-drug, drug-gene and drug-condition interactions. InDigital Personalized Health and Medicine 2020 (pp. 648-652). IOS Press.

[7] Kulenovic A, Lagumdzija-Kulenovic A. Using PM-TOM for the Minimization of Adverse Drug and Gene Interactions in Therapies for Common Multi-Diseases. Studies in Health Technology and Informatics. 2020 Jun 1;272:205-8.

[8] Ball MP, Bobe JR, et al. Harvard Personal Genome Project: lessons from participatory public research. Genome medicine. 2014 Feb;6(2):1-7.

[9] Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wiegers J, Wiegers TC, Mattingly CJ. The comparative toxicogenomics database: update 2019. Nucleic acids research. 2019 Jan 8;47(D1):D948-54.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210873

Introduction of the EU Digital COVID Certificate in Slovenia: Technological and Process Aspects

Dalibor STANIMIROVIC[a,1] and Lucija TEPEJ JOCIC[a]
[a] *National Institute of Public Health, 1000 Ljubljana, Slovenia*

Abstract. In early 2021, the European Commission presented a proposal to introduce an EU Digital COVID Certificate, which should enable safe border crossings for citizens within the EU during the COVID-19 pandemic. Subsequently, all EU Member States successfully introduced the EU Digital COVID Certificate by 1 July 2021. This article focuses on a review of the technological and process aspects identified in the introduction of the EU Digital COVID Certificate in Slovenia. The research applies a case study framework, including focus group discussions, as the primary data collection method. The research findings expose the technological and process complexities related to the dispersed data sources and fairly intricate and copious business rules used for the creation of the EU Digital COVID Certificate. Moreover, the study implies that the *ad hoc* introduction of such demanding and sensitive digital solutions in the future will not be possible without the establishment of effective national health information infrastructures across the EU.

Keywords. COVID-19 pandemic, EU Digital COVID Certificate, technological and process aspects, case study, Slovenia

1. Introduction

The EU Digital COVID Certificate (DCC) was introduced on 1 July 2021 in all EU Member States and it is based on the uniform EU framework for the issuance, verification, and acceptance of interoperable certificates of COVID-19 vaccinations, tests, and recovery, which are intended to facilitate free movement under the pandemic health conditions [1]. The DCC is recognized in all EU Member States, which generally enables the harmonized elimination of potential travel restrictions, and above all ensures the same approach to issuing and in particular verifying DCCs. All EU citizens and third-country nationals legally residing in the EU who hold a DCC should be exempt from restrictions on free movement while traveling in the same manner as nationals of the EU Member States visited. However, regardless of the uniform arrangement of and agreement on the DCC, EU countries will still be able to introduce measures independently to a certain extent [2], which means that the conditions for free movement may still differ between countries. The DCC contains essential information such as name and surname, date of birth, date of issue, relevant information on the vaccination against COVID-19, negative test results, and recovery from COVID-19, and a unique certificate identifier. The DCC

[1] Corresponding Author, Dalibor Stanimirovic, Centre for Healthcare Informatics, National Institute of Public Health, Trubarjeva Street 2, 1000 Ljubljana, Slovenia; E-mail: dalibor.stanimirovic@nijz.si.

is only issued for COVID-19 and the countries visited may not keep this information. Only the validity and authenticity of the DCC are verified, and the authority that issued and signed the DCC is checked. The health information remains the responsibility of the Member State that issued the DCC.

In Slovenia, the entire DCC system was established by the National Institute of Public Health (*Nacionalni Inštitut za Javno Zdravje*, hereinafter: NIJZ), which also manages the national eHealth solutions and is the only authorized DCC issuer in the country. The digital solution enabling the creation of the DCC is based on the records contained in the Central Registry of Patient Data (CRPD) and provides access to the DCC for patients, healthcare professionals, and authorized personnel at the NIJZ. The DCC can be stored in paper form or on a mobile device. Both versions are equipped with a QR code that contains the essential information and a digital stamp, which attests to the authenticity and validity of the DCC. This article focuses on a review of the technological and process aspects identified in the introduction of the DCC in Slovenia.

2. Methods

A case study framework, including focus group discussions as the primary data collection method, was applied to review the technological and process aspects identified in the introduction of the DCC in Slovenia. The choice of the research framework was based on the specifics of the research topic [3], and the approach used was deemed to be the most reasonable one.

Following a comprehensive examination of the related literature and EU recommendations and guidelines, an extensive analysis of the introduction of the DCC in Slovenia was carried out with the focus group participants. NIJZ experts in charge of the introduction of the DCC in Slovenia, representatives from the Ministry of Health, and experts from external IT providers participated in twenty focus group sessions conducted from April to July 2021 (each teleconference session lasted approximately two hours). The expertise and experience of the fifteen prominent experts in the field participating in the focus group discussions (excluding medical expert groups dealing with the epidemiological concerns) ensured the credibility of the research findings. The insights and opinions of the participating experts were supported by the data and reports from the statistical and business intelligence modules contained in the DCC Application and other eHealth solutions. The views and discussions from the focus group sessions were documented in writing. A conventional content analysis was used for the analysis of the data obtained [4], and, along with the literature and EU materials, provided a means to review the technological and process aspects identified in the introduction of the DCC in Slovenia.

3. Results

A new eHealth solution, i.e. the DCC Application, ensures the issuance of the DCC in accordance with uniform EU standards. It is based on the CRPD, which is the fundamental source of data and provides essential functionalities for the DCC Application. Data sources for the DCC are records of nucleic acid amplification test (NAAT) and rapid antigen test (RAT) results, as well as records of COVID-19 vaccinations. The DCC is available immediately upon the submission of the respective

test or vaccination record. A request for a DCC is triggered each time a common CRPD document query is performed, and the DCC is added to the patient's other healthcare documents. As a result, all healthcare providers in Slovenia have immediate access to the DCC for every patient via existing local hospital information systems (and there is no need to upgrade these systems at points of care). Alternatively, healthcare professionals can obtain the DCCs of patients via the zVEM+ web solution, which is aimed at those healthcare providers whose information systems are not yet connected to the CRPD (e.g. pharmacies). Patients can obtain their own DCC via the zVEM Patient Portal (https://zvem.ezdrav.si) or the zVEM Mobile Application. The digital solution is located among the core eHealth services and contains the building blocks and functional components depicted in Figure 1.

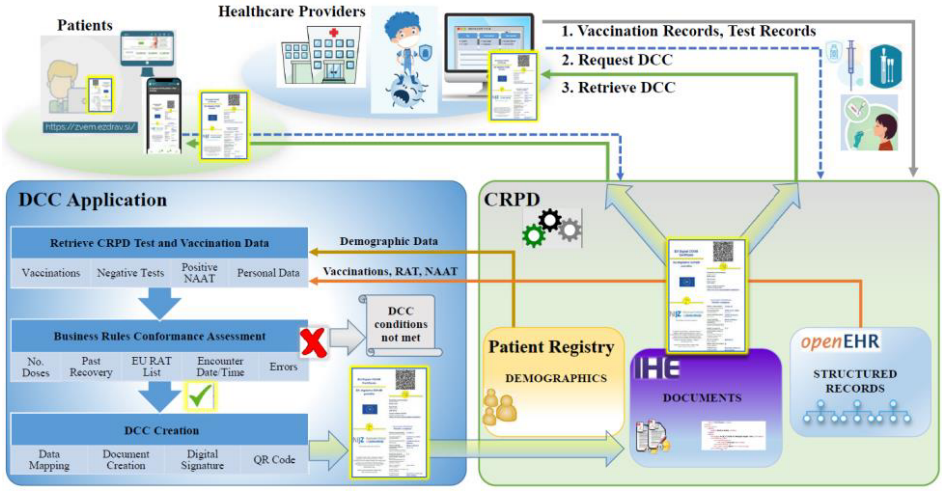


Figure 1. The DCC system architecture.

The DCC Application ensures the retrieval of data from the CRPD and its conversion in accordance with the EU standards, the verification of data conformance with the rules for issuing the DCC, and the creation of the DCC with a digital signature and QR code. High performance and dedicated data queries of vaccination and COVID-19 test result records extract the current data that are relevant for the issuance of the DCC. Queries are separate from the general CRPD services, which ensures sufficient responsiveness with a minimal load on the basic building blocks. The creation of a DCC is automatically triggered each time a CRPD query is made. For end users, the process of retrieving the DCC is exactly the same as for other documents. The DCC is associated with other documents available to patients and healthcare professionals, which is a great advantage. There are four types of DCC available in the CRPD: a DCC of a completed vaccination, a DCC of a negative NAAT, a DCC of a negative RAT, and a DCC of recovery from COVID-19. Only tests from the common list approved by the European Commission allow the issuance of a DCC. In the event of non-compliance with the required criteria, an error or failure to meet the DCC retrieval criteria is displayed on the document.

A prerequisite for the creation of a DCC is properly collected and recorded data. Healthcare providers must correctly submit records of tests and vaccinations to the CRPD. Providers of the RAT must properly report the testing results, and accredited

microbiological laboratories must provide NAAT results. As a rule, laboratories also provide the original microbiological report in pdf format. Records can be submitted through local information systems to the Patient Summary or the Electronic Vaccination Registry. The query for a DCC is triggered in the same way as for other documents in the CRPD. When a user of the zVEM Patient Portal or a local information system at a healthcare provider initiates a query for a DCC, the DCC Application is automatically called. The DCC Application obtains data from the CRPD and, in accordance with the rules, creates the DCC based on the data currently available (vaccinations, a DCC of the last negative test in the past 8 days, a DCC of a positive NAAT in the last 180 days). In accordance with the technical rules, the document is signed in the DCC Application with a valid system certificate issued by the NIJZ. An essential component of the DCC is the QR code, which enables authorized personnel to conduct automatic verification. The DCC Mobile Verification Application reads the data from the QR code, and verifies the authenticity and validity of the DCC. If all conditions are met and all business rules are followed, the DCC is available to the user; if not, an error or failure to meet the DCC retrieval criteria is displayed on the document.

4. Discussion

Slovenia successfully introduced the DCC even before the deadline of 1 July 2021. However, the whole project – which was directed by the European Commission, which provided technical guidelines and the central digital platform (EU Gateway) required for the issuance and verification of the DCC – was managed awkwardly and inconsistently in some parts. This is somewhat understandable, as it was necessary to find the lowest common denominator of the EU Member States, EU institutions, and national and international healthcare organizations. Despite the lengthy harmonization process, the EU Member States have retained some autonomy in setting national rules for crossing their state borders, which further complicates the establishment of uniform DCC solutions at the EU level and brings about considerable confusion among citizens, who are not adequately informed of the fast-changing rules.

The whole DCC concept is relatively complex due to the multiple and rather dispersed data sources, on one hand [5], and the quite complicated and numerous business rules, on the other. If we disregard the pressure from all sides and the frantic circumstances in which the project was implemented [6], it should be noted that the latter factors caused the most problems in the development of the DCC solution and its introduction into everyday use. The lack of qualified digital certificates required to retrieve the DCC via the zVEM Patient Portal, the poor quality of the captured data, missing data or incomplete reporting of data on vaccinations, recovery, and testing by healthcare providers, legal issues – including data security and protection, rapid changes in code lists (especially the valid RATs on the EU list), problems with citizens who have recovered from COVID-19 or have been vaccinated or tested abroad, the lack of relevant data on foreigners living in Slovenia, and dilemmas in formulating and complying with the business rules due to unpredicted combinations of vaccines, tests, and time frames, are some of the key issues that Slovenia encountered in introducing the DCC solution. As elsewhere, delayed actions due to the coordination of policymakers and other stakeholders (national and EU), the lack of awareness of citizens, inadequate and inconsistent communication, and the misleading information that has flooded the media space, have only exacerbated the above problems. Nonetheless, approximately 990,000

distinct DCCs have been issued in Slovenia (the vast majority sent by post), which is a rather high number in terms of the population (2.1 million) and the share of those fully vaccinated, which is 42%. The Slovenian DCC application enables easy modifications of business logics without the need to upgrade the core components. Moreover, connection to the national CRPD platform guarantees the highest level of both reliability and inclusion. The central DCC service is ensuring medical-grade quality, and the DCC is available to the entire population. Operational experience has proven that the Slovenian DCC architecture is providing the required resilience, flexibility, and responsiveness. According to the available data reported by the EU Member States [7], Slovenia is among the countries with the highest number of DCC issued per capita.

However, the effectiveness of limiting the spread of COVID-19 infections by means of the DCC and related approaches has yet to be evaluated [8], and the wider implications of similar measures is an issue that future research in this area should address.

5. Conclusions

The establishment of political consensus and the construction of central digital platforms at the EU level, with well-defined protocols and rules for the integration of Member States, could be a major advantage in the future development of cross-border eHealth services, which will hopefully not be associated only with pandemic-related challenges.

Nevertheless, as the experience with the DCC has revealed, all such EU projects require well-developed national infrastructure, an appropriate architecture of national healthcare digital solutions and databases, a sufficient number of competent experts, precise rules for processing data, and, of course, sufficient funding and a suitable legal framework. Without awareness thereof, such projects can quickly turn into unproductive improvisation that fails to address the needs of EU citizens and cannot contribute to achieving the declared public health objectives.

References

[1] Regulation (EU) 2021/953 of the European Parliament and of the Council of 14 June 2021 on a framework for the issuance, verification and acceptance of interoperable COVID-19 vaccination, test and recovery certificates (EU Digital COVID Certificate) to facilitate free movement during the COVID-19 pandemic. Brussels, 14.6.2021.
[2] Chen LH, Petersen E, Blumberg L, Piyaphanee W, Steffen R. COVID-19 health passes: Current status and prospects for a global approach. J Travel Med. 2021 Aug 3;taab118.
[3] Thomas G. How to do your case study. Thousand Oaks, CA: Sage Publications; 2021.
[4] Lindgren BM, Lundman B, Graneheim UH. Abstraction and interpretation during the qualitative content analysis process. Int J Nurs Stud. 2020 Aug;108:103632.
[5] Mbunge E, Dzinamarira T, Fashoto SG, Batani J. Emerging technologies and COVID-19 digital vaccination certificates and passports. Public Health Pract. 2021 Nov;2:100136.
[6] Cardenas NC. Advancing strategic policy on European Union digital COVID-19 certificate. J Public Health. 2021 Jun;fdab227.
[7] European Commission. eHealth Network. Statistics and reporting. Brussels, 24.9.2021.
[8] Mehtar S, Preiser W, Lakhe NA, Bousso A, TamFum JJM, Kallay O, Seydi M, Zumla A, Nachega JB. Limiting the spread of COVID-19 in Africa: one size mitigation strategies do not fit all countries. Lancet Glob Health. 2020 Jul;8(7):e881-e883.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210874

Latent COVID-19 Clusters in Patients with Opioid Misuse

Fatemeh SHAH-MOHAMMADI[a,1], Wanting CUI[a], Keren BACHI[a], Yasmin HURD[a] and Joseph FINKELSTEIN[a]
[a]*Icahn School of Medicine at Mount Sinai, New York, NY, USA*

Abstract. The goal of this paper is to apply unsupervised machine learning techniques in order to discover latent clusters in patients who have opioid misuse and also undergone COVID-19 testing. Target dataset has been constructed based on COVID-19 testing results at Mount Sinai Health System and opioid treatment program (OTP) information from New York State Office of Addiction Service and Support (OASAS). The dataset was preprocessed using factor analysis for mixed data (FAMD) method and then K-means algorithm along with elbow method were used to determine the number of optimal clusters. Four patient clusters were identified among which the fourth cluster constituted the maximum percentage of positive COVID-19 test results (20%). Compared to the other clusters, this cluster has the highest percentage of African Americans. This cluster has also the highest mortality rate (16.52%), hospitalization rate after receiving the COVID-19 test result (72.17%, use of ventilator (7.83%) and ICU admission rate (47.83%). In addition, this cluster has the highest percentage of patients with at least one chronic disease (99.13%) and age-adjusted comorbidity score more than 1 (83.48%). Longer participation in OTP was associated with the highest morbidity and mortality from COVID-19.

Keywords. Cluster analysis, COVID-19, Opioid Treatment Program

1. Introduction

In recent years, opioid use disorder has become a significant public health problem in the United States, leading to thousands of death. According to CDC, prescription opioid and heroin overdose deaths have been increasing since 1999 [1]. Opioid abusers are exposed to high risk of not only contracting infectious disease, but also development of mental illness [2]. In addition to health concerns, opioid crisis also creates serious financial costs. In 2009, the annual costs of prescription and illicit opioid abuse, including lost productivity and health care costs, were estimated to be over $55 billion [3]. Methadone and buprenorphine has been reported as effective treatments for opioid dependence, and their widespread use could mitigate the negative health and societal effects of opioid use disorder [4]. Opioid Treatment Programs (OTPs) are among the licensed providers of medication for opioid abusers, and usually require patients to take medication at a clinic.

Novel Coronavirus Disease 2019 (COVID-19) caused a worldwide pandemic outbreak and resulted in large number of infections and million deaths worldwide [5].

[1] Corresponding Author, Fatemeh Shah-Mohammadi, Icahn School of Medicine at Mount Sinai, 1770 Madison Ave, 2nd Fl, New York, NY, USA, 10035; E-mail: fatemeh.shah-mohammadi@mountsinai.org.

Recent studies were able to identify latent COVID-19 clusters in general population [7] as well as in people with chronic respiratory conditions [10]. Identification of latent COVID-19 subgroups can help prioritize COVID-19 screening and treatment [7, 10]. In our previous study [11] we also demonstrated that socio-demographic and OTP characteristics can be successfully used for OTP machine learning analytics. The aim of this study was to identify the latent clusters of patient characteristics who have enrolled in OTP and also tested for COVID-19 using an unsupervised machine learning approach. This approach is also critical for comprehensive understanding of the COVID-19 risk factors for patients with opioid misuse.

2. Methods

Two datasets (AIMS and a COVID-19) were merged to generate an analytical dataset. AIMS is a unique dataset that contains patients' demographic information, employment status, admission, transfer and discharge records, and the substance usage information from the MSHS opioid treatment programs of the New York State area with the size of 158,989 records for 22,044 unique patients. COVID-19 dataset was generated by querying electronic health records at Mount Sinai Health System (MSHS) in New York to identify all patients who underwent COVID-19 testing between March 2020 and July 2021. The data in AIMS was collected from the New York State Office of Addiction Service and Supports' (OASAS), OTP. We only considered the information for the patients who received treatments at MSHS. COVID-19 dataset contains 704,223 records of 388,981 unique patients. Some patients tested multiple times for COVID-19. These patients have been considered as tested negative only if all their records show negative results. From AIMS, we considered the variables "Sex", "Race", "Admission date", "Discharge date", "Date of Birth", "Employment status" and "Secondary abuse substance" which showing the gender, race, the date that the patient gets admitted to the OTP, discharge date from the program, birth date of the patient, whether the patient is employed and what the patient secondary abuse substance is, respectively. Since primary abuse substance for majority of patients (more than 97%) in all clusters was heroin, we did not consider it as a variable in our analysis. Discharge date and the admission date from this dataset was used to calculate new variable "Length of Enrollment in OTP". To calculate this new variable, the value for both dates needs to be valid. Since unknown values for discharge date conveys the fact that the patient is still enrolled in the OTP, this value has been replaced with the most recent admission date in the AIMS. For the patients who admitted in the program multiple times with different discharge dates, we calculated the time between the first time's admission date and the most recent discharge date. Variables in COVID-19 dataset are as follow: hospitalization date, admission to ICU, ICD-10 codes, use of ventilator, alive indicator, COVID-19 test result and the date that the patient received the test result. Patient's age was calculated based on subtracting COVID-19 test result delivery date and the birth date. In addition, the age-adjusted comorbidity score was calculated based on patients' medical history using ICD-10 codes and patient's age [12]. We also considered a variable named "Hospitalized after COVID-19 test" that shows whether the patient was hospitalized after receiving COVID-19 test result. This variable was calculated based on comparing the patient's hospitalization date and the date when patient received the result for COVID-19 test. The variable "Presence of a Chronic Condition", showing whether a patients has a history of chronic disease, was also added and identified based on the ICD10 codes [13]. After merging two datasets

(finding the patients who had records in both datasets) and eliminating patients with missing values, the number of unique patients was reduced to 866 patients. Among 866 patients, 105 patients tested positive for COVID-19.

We used factor analysis for mixed data (FAMD) method for preprocessing [8]. It should be used when we have mixed variables (categorical as well as numerical variables). To discover the latent clusters, we applied K-means clustering on preprocessed data (we only selected the first 2 principal components). Optimal number of clusters were found through cluster analysis. To be more exact, we ranged the number of clusters from 2 to 20, and for each number of clusters we performed K-means and calculated within cluster sum of squares (WCSS). We plotted the WCSS against the number of clusters and used the elbow method to determine optimal number of clusters. All analysis was performed in Anaconda Jupyter Notebook, using Python 3.9.0.

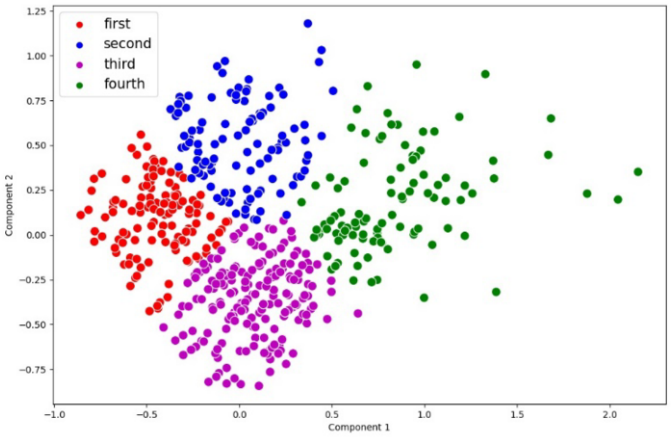


Figure 1. Visualization of the 4 clusters based on the first two principal components

3. Results

Four clusters were identified (Figure 1). According to Table. 1, there was a significantly larger amount of male patients compared to the female patients in all clusters. The largest percentage of positive COVID-19 result cases (20%) belongs to the cluster # 4, while the lowest (around 4%) comes from cluster # 2. More than 50% of patients in cluster # 2 have no comorbid conditions. Compared to the patients mostly in their adulthood in cluster # 2, patients in cluster # 4 were generally in their older adulthood with almost three times more "Length of Enrollment in OTP". More than 80% of patients in cluster # 4 have age-adjusted comorbidity score more than 1 while this value is around 7% for cluster # 2. All of the patients in both clusters have been hospitalized at least once among which around 73% of hospitalizations in cluster # 4 occurred after receiving the COVID-19 test result while this number is 40% for cluster # 2. The highest percentage of patients admitted in ICU belongs to the cluster # 4 (around 50%). This cluster has the highest percentage of patients with chronic disease (99.13%), the highest percentage of death

(16.52%) and the longest participation in OTP. Moreover, most of the patients in this cluster have African American and Hispanic race (around 70%) and are unemployed (46.15%).

4. Discussion

Among the selected variables for the analysis "Race", "Age-Adjusted Comorbidity Score", "Post-Testing Status", "COVID-19 Testing Result", "ICU Status", "Hospitalized after COVID-19 test" and "Age" were important factors in separation of the clusters. Since the highest percentage of mortality, ICU admission, hospitalization after COVID-19 test and the use of ventilator belongs to the cluster with the largest percentage of positive COVID test (i.e. cluster # 4), patients in this cluster need more rigorous testing and treatments. Percentage of patients with comorbidity score more than 1 is the highest for this cluster. On the other hand, more than 50% of patients in the cluster with the lowest percentage of positive COVID-19 cases (i.e. cluster # 2) have no comorbidities. These findings confirm that for OTP patients comorbidity condition can be considered as the COVID-19 risk factor. Moreover, the highest percentage of African Americans and Hispanics (70.44) are in cluster # 4 while the same races constitute the lowest percentage (28.24%) in cluster # 2. The cluster # 3&4 are also represented by the longest duration in OTP.

Our results are congruent with previous reports which used similar clustering techniques for patients with chronic respiratory conditions and pregnant women [7,9].

Table 1. Descriptive statistics o clusters

Clusters	**1**	**2**	**3**	**4**
Count	239	177	335	115
Age				
Young	20.08%	24.86%	0.00%	0.00%
Adult	77.41%	73.45%	13.73%	20.00%
Older adult	2.51%	1.69%	86.27%	80.00%
Sex				
Female	29.71%	20.34%	26.87%	21.74%
Male	70.29%	79.66%	73.13%	78.26%
Race				
Black / African American	7.98%	9.03%	38.51%	40.87%
White / Non-Hispanic	51.26%	54.24%	14.93%	15.65%
Hispanic	26.89%	19.21%	31.04%	29.57%
Other	13.87%	17.51%	15.52%	13.91%
Obesity				
Obese	17.99%	25.99%	18.51%	26.09%
No obese	82.01%	74.01%	81.49%	73.91%
Hospitalized (ever) - Yes	9.62%	100.00%	30.75%	100.00%
Presence of a Chronic Condition-Yes	68.20%	99.00%	82.39%	99.13%
Age Adjusted Comorbidity Score				
0	51.88%	53.67%	0.00%	0.00%
1	46.44%	39.55%	23.88%	16.52%
2	1.68%	6.78%	48.36%	37.45%
3	0.00%	0.00%	20.90%	38.34%
4	0.00%	0.00%	5.96%	6.82%
5	0.00%	0.00%	0.90%	0.87%
Length of Enrollment in OTP				
Mean number of days	981.23	612.91	2304.66	1826.76
Employment				
Employed	12.28%	12.41%	15.79%	17.31%
Not in labor force	13.16%	13.87%	29.47%	36.54%
Unemployed	74.56%	73.72%	54.74%	46.15%

Secondary Substance Use				
Alprazolam(Xanax) and Benzodiazepine	6.58%	6.57%	2.45%	1.92%
Opioid-based substance	12.73%	5.84%	5.26%	9.61%
None	71.05%	65.69%	84.21%	82.69%
COVID-19 Testing Result-Positive	14.17%	3.95%	14.03%	20.00%
Hospitalized after COVID-19 Test- Yes	0.00%	40.00%	0.60%	72.17%
On Ventilator - Yes	0.00%	0.00%	0.00%	7.83%
ICU Status -Yes	0.00%	6.00%	0.00%	47.83%
Post-Testing Vital Status				
Alive	100.00%	98.87%	100%	83.48%
Deceased	0.00%	1.13%	0.00%	16.52%

5. Conclusion

Four clusters have been identified in patients who participated in OTPs and underwent COVID-19 testing. K-means algorithm is used for cluster analysis in order to determine the number of optimal clusters. The largest percentage of COVID-19 positive cases belonged to the fourth cluster with patients mostly unemployed and in their older adulthood with the highest percentage of African American and Hispanic races. This conveys the fact that the COVID-19 pandemic's impact has more exposed race inequities. This cluster has the highest percentage of hospitalization rate after COVID test, death rate, ventilator usage, ICU admissions and comorbidity score more than 1. Due to the highest death rate, patients in this cluster need more rigorous treatments.

References

[1] Centers for Disease Control and Prevention, National Center for Health Statistics. Multiple Cause of Death 1999-2017 on CDC WONDER Online Database, 2018.
[2] Ehrich E, Turncliff R, Du Y, et. al., Evaluation of opioid modulation in major depressive disorder, Neuropsychopharmacology, 2015;40:1448-1455.
[3] Birnbaum, Howard G, et. al., Societal costs of prescription opioid abuse, dependence, and misuse in the United States, Pain medicine, 2011;12:657-667.
[4] Mattick RP, Kimber J, Breen C, Davoli M, Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence, Cochrane Database Syst Rev., 2008.
[5] WHO-China Joint Mission, Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), 2020.
[6] Centers for Disease Control and Prevention. Coronavirus disease 2019 (COVID-19): cases in US, Accessed April 24, 2020. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html
[7] Cui W, Robins D, Finkelstein J. Unsupervised Machine Learning for the Discovery of Latent Clusters in COVID-19 Patients Using Electronic Health Records. Stud Health Technol Inform. 2020;272:1-4.
[8] Mori Y., Kuroda M., Makino N. Multiple Correspondence Analysis. In: Nonlinear Principal Component Analysis and Its Applications. Springer Briefs in Statistics. Springer, Singapore. 2016.
[9] Qi M, Li X, Liu S, Li Y, Huang W., Impact of the COVID-19 epidemic on patterns of pregnant women's perception of threat and its relationship to mental state: A latent class analysis, PLoS One. 2020; 15.
[10] Cui, Wanting et al., Latent COVID-19 Clusters in Patients with Chronic Respiratory Conditions, Studies in health technology and informatics. 2020;275:32-36.
[11] Cui W, Bachi K, Hurd Y, Finkelstein J. Using Big Data to Predict Outcomes of Opioid Treatment Programs. Stud Health Technol Inform. 2020;272:366-369.
[12] Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. J Clin Epidemiol. 2004;57(12):1288-1294.
[13] Healthcare Cost and Utilization Project (HCUP). Chronic Condition Indicator for ICD-10-CM, v2021.1. Agency for Healthcare Research and Quality, October 2020.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210875

Smartphone-Based Healthcare Apps for Older Adults in the COVID-19 Era: Heuristic Evaluation

Hamid Reza SAEIDNIA[a,1], Ali GHORBI[b], Marcin KOZAK[c] and Claudiu HERTELIU[d]
[a] *Department of Knowledge and Information Science, Tarbiat Modares University, (TMU) Tehran, Iran*
[b] *Department of Knowledge and Information Science, Tehran University, Tehran, Iran*
[c] *Department of Media, Journalism and Social Communication, University of Information Technology and Management, Rzeszow, Poland*
[d] *Department of Statistics and Econometrics, Bucharest University of Economic Studies, 010552 Bucharest, Romania*

Abstract. Many national governments have attempted to prevent and combat COVID-19 using mobile health (mHealth) technologies during the epidemic. During this time, governments began developing smartphone-based apps for prevention, call tracking, and monitoring people with COVID-19. An important question is, does everyone benefit from these technologies equally and fairly? To answer this question, we evaluated the user interface of smartphone-based apps developed during the COVID-19 era by considering their design for older adults, in order to determine whether social justice has been considered in the development of these apps.

Keywords. smartphone, user interface, healthcare, older adults, COVID-19

1. Introduction

In order to combat the coronavirus pandemic, governments and health organizations used mobile health (m-Health) technology in the areas of self-care, contact tracing, education, and others. In the meantime, smartphone-based apps have become a more noticeable presence, and many countries unveiled their own official apps. As these apps are used by all segments of society (e.g., older adults, the elderly, young people, and the physically disabled), all requirements related to the user interface must be considered. The ability for all users to benefit from software (including apps) is so crucial that it is considered a social justice [1] and human rights issue, in that its importance goes beyond software into human experience [2]. Numerous articles have shown that poor user interface design repels people from using software services [2, 3].

Therefore, in this study, we seek the answer to the question of whether social justice, in the user interface design of smartphone-based healthcare apps during the coronavirus,

[1] Corresponding Author, Hamid Reza Saeidnia, Department of knowledge and information science Tarbiat Modares University, (TMU) Tehran, Iran; E-mail: hamidrezasaednia@gmail.com.

has been observed. For this purpose, we selected two apps from two countries considered to be developed and economically advantaged.

2. Method

This applied and descriptive research used the heuristic evaluation method. We examined the user interfaces of two official national apps in this study: COVID Alert and COVIDSafe. The former was developed by the Canadian government [4], and the COVIDSafe app by the Australian government [5]. We benefited from the experience of five IT experts to evaluate the user interfaces of these apps. These five experts have experience in app development and possess master's degrees or higher in health informatics, biomedical engineering, or health information technology disciplines (Table 1).

Table 1. Experts' qualifications.

	Expert 1 (pilot)	Expert 2	Expert 3	Expert 4	Expert 5
Sex	Male	Female	Male	Female	Male
Age	28	28	38	39	42
Education	MSc (Engineering, Information Management)	MSc (Health Information Management)	PhD (Health Informatics)	PhD (Biomedical Engineering)	PhD (Health Informatics)
Profession	PhD fellow	PhD fellow	Professor	Professor	Professor

Each of the five experts spent two hours testing the apps. We used the recommendations of Morris (1994) [6] as a tool to analyze their user interfaces for older adults, which are basic principles of user interface design [7]. The recommendations included training in visual, auditory, physical, memory, attitudinal, and app training areas [6]. We provided these recommendations as a questionnaire including 44 items in 6 sections and asked the experts to rate their observations based on a 5-point severity ranking (1=Not at all, 2=Partially observed, 3=Fair, 4=Very good, 5=Excellent). We calculated that the questionnaire was reliable (Cronbach's alpha of 0.87), and five professors of health informatics and information management evaluated and confirmed the validity of the instrument. Each of the recommendations was considered as items properly adapted for older adults, if that item obtained an average of 3 or greater. Finally, a one-hour debriefing was held after the heuristic evaluation, during which the experts further commented on the apps. The data analysis was conducted using SPSS 22 software.

3. Result

The results of the heuristic evaluation indicate that, for both apps, most recommendations related to user interface design for older adults achieved a mean score under 3. Table 2 shows the mean points for the recommendations.

Table 2. Average points of two selected apps

Recommendations	COVIDAlert Ave.	COVIDSafe Ave.	Recommendations	COVIDAlert Ave.	COVIDSafe Ave.
Visual			Memory		
Font size	3.2	3.2	Visibility of objects	2.4	2.2
Text characteristics	2.6	28	Navigational aids	2.2	2

Resolution	3.2	3.4	Interaction style	2.2	2.2
Lighting	2.8	3.2	Design for consistency	1.8	2
Glare	2.8	3	Instructions	2.2	
Color of text	3.4	3.4	Learning	1	1
Brightness	3.2	3.8	Note-taking	1	1
Object speed	3	3	Task structure	2	2
Scrolling speed	3.2	2.8	Information presentation	2.6	2.8
Multiple sensory modalities	1.8	2.4	User experience	3	3
Interaction style	2	2	Attitudinal		
Highly visually-impaired users	2	1.4	Designer assumptions	1	1
Auditory			System messages	2.2	2
Arrangement	3.2	3	Wording	2.4	2.2
Noise	2	3	Error-free design	2	2
Task	1.8	2.4	Notification of interaction	1	1
Frequencies	2.2	2.6	App training		
Synthesized	2.4	3.4	Written materials	1.8	2
Dimensions	2.6	4	Training session length	1	1
Highly hearing-impaired users	2	2	Self-pacing	1	1
Physical			Help	2	2
Menu structure	1.6	2	Instructional technique	1	1
Documentation	1.8	2	Training strategy	1	1
Experienced typists	2	1.8			

4. Discussion

In this study, we examined the user interface of two official government apps developed during the coronavirus. By this, we aimed to answer the question of whether the user interface design of these two apps considered the issue of accessibility for older adults.

According to the findings of the heuristic evaluation, these two apps did not apply any of the fundamental design principles well, except for a few items that should be considered as general, basic recommendations for older adults. For the items font size, resolution, text color, and brightness, both apps averaged 3 or above, which can be described as good. For a limited number of the other items, the COVIDSafe app averaged higher than the COVIDAlert app.

Both apps had the lowest average scores in most other items, such as interaction style, hearing-impaired access, learning, note-taking, designer assumptions, notification of interaction, training session length, self-pacing, help, instructional technique, and training strategy. Over half of the 44 items used to design user interfaces for older adults in both apps did not comply with the recommendations. However, older adults are among the most vulnerable people to the coronavirus [8], and so, certainly, these people should have been given more attention while developing these apps. Based on a report by the World Health Organization, during the next 40 years (until 2060), 29.7% of the worldwide population is projected to be older adults [9]. This illustrates the importance of addressing this huge and growing segment of the population. What is apparent at the moment is the lack of attention in app development for this group at the national level, and for design in general.

After the heuristic evaluation in the debriefing, the experts agreed that in addition to the complex user interface, other essential aspects of these apps, such as connecting to the Internet at least once a day and Bluetooth or Wi-Fi support, were difficult and often incomprehensible to older adults. This issue—that the technology required to help older adults often is not understood—is referred to as the "elderly paradox" [10]. Numerous studies have emphasized the importance of user interface design and user experience for all community members to benefit from healthcare apps in the coronavirus era [11, 12]. This is because these coronavirus apps must be downloaded by a majority of users to be effective [10].

Among the study's limitations, there was the absence of a standardized tool or questionnaire (e.g., SUS) for review of user interface design apps aimed at older adults. We compensated for this limitation by using fundamental recommendations for user interface design. Another limitation was the number of apps studied—only two. In order to evaluate the user interfaces designed for older adults during the coronavirus, more apps should be analyzed. Future research can address this issue.

5. Conclusions

In designing and developing healthcare apps that are implemented nationally, it is crucial to pay attention to the design of the user interface for older adults or groups with physical disabilities (e.g., hearing or vision). In these circumstances, either a separate version should be developed for these people, or features should be included in the app that allow easy use by older adults or people with physical disabilities. This work is partially supported by a grant from the Romanian National Authority for Scientific Research and Innovation, CNDS-UEFISCDI, project number PN-III-P4-ID-PCCF-2016-0084.

References

[1] Costanza-Chock S. Design Justice and User Interface Design. Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology; 2020.
[2] Harihareswara S. User experience is a social justice issue. Code4Lib Journal. 2015(28).
[3] Walls DM. User experience in social justice contexts. Proceedings of the 34th ACM International Conference on the Design of Communication; 2016.
[4] COVID Alert today 2020. Available at: https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19/covid-alert.html.
[5] COVIDSafe app 2020. Available at: https://www.health.gov.au/resources/apps-and-tools/covidsafe-app.
[6] Morris JM. User interface design for older adults. Interacting with Computers. 1994;6(4):373-93.
[7] Wagner N, Hassanein K, Head M. Computer use by older adults: A multi-disciplinary review. Computers in human behavior. 2010;26(5):870-82.
[8] Shahid Z, Kalayanamitra R, McClafferty B, Kepko D, Ramgobin D, Patel R, et al. COVID-19 and older adults: what we know. Journal of the American Geriatrics Society. 2020;68(5):926-9.
[9] Mattiuzzi C, Lippi G. Worldwide disease epidemiology in the older persons. European geriatric medicine. 2020;11(1):147-53.
[10] Rizzo E. COVID-19 contact tracing apps: the 'elderly paradox'. Public health. 2020;185:127.
[11] Saeidnia H, Mohammadzadeh Z, Saeidnia M, Mahmoodzadeh A, Ghorbani N, Hasanzadeh M. Identifying Requirements of a Self-care System on smartphones for preventing coronavirus disease 2019 (COVID-19). Iranian Journal of Medical Microbiology. 2020;14(3):241-6.
[12] Saeidnia HR, Mohammadzadeh Z, Hassanzadeh M. Evaluation of Mobile Phone Healthcare Applications During the Covid-19 Pandemic. Studies in health technology and informatics. 2021;281:1100-1.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210876

A-Test Method for Quantifying Structural Risk and Learning Capacity of Supervised Machine Learning Methods

Arash GHAREHBAGHI[a,1] and Ankica BABIC[a,b]
[a] *Department of Biomedical Engineering, Linköping University, Sweden*
[b] *Department of Information Science and Media Studies, University of Bergen, Norway*

Abstract. This paper presents an original method for studying the performance of the supervised Machine Learning (ML) methods, the A-Test method. The method offers the possibility of investigating the structural risk as well as the learning capacity of ML methods in a quantitating manner. A-Test provides a powerful validation method for the learning methods with small or medium size of the learning data, where overfitting is regarded as a common problem of learning. Such a condition can occur in many applications of bioinformatics and biomedical engineering in which access to a large dataset is a challengeable task. Performance of the A-Test method is explored by validation of two ML methods, using real datasets of heart sound signals. The datasets comprise of children cases with a normal heart condition as well as 4 pathological cases: aortic stenosis, ventricular septal defect, mitral regurgitation, and pulmonary stenosis. It is observed that the A-Test method provides further comprehensive and more realistic information about the performance of the classification methods as compared to the existing alternatives, the K-fold validation and repeated random sub-sampling.

Keywords. A-Test method, structural risk, learning capacity, heart sounds

1. Introduction

Artificial Intelligence (AI), as an advancing context, is creating a significant influence on different aspects of social sustainability including healthcare. AI-based tools are consistently becoming part of the healthcare system towards providing better healthcare for all the individuals of a society, where supervised machine learning methods serve as a central part for making the appropriate medical decision [1–5]. The performance of such machine learning methods is critically important, sometimes with vital value, as a mistaken error can lead to incorrect patient management. It is, therefore, crucial for any machine learning method to be properly trained, especially when it comes to medical applications [6]. In general, there are two main circumstances in the training of the machine learning methods, small-medium size and large size of the learning data. The main challenge for the large size data is to avoid biased training on any group. On the other hand, the small-medium size data is more likely to face overfitting, which affects the reproducibility of the results [6–7]. Small-medium size data is seen in many

[1] Corresponding Author, Arash Gharehbaghi. Department of Biomedical Engineering, Linköping University, Sweden: arash.gharehbaghi@liu.se

applications of biomedical engineering and bioinformatics, where the data acquisition is problematic, particularly, under the limitations recommended by the new adaptation of Good Clinical Practices and the codes of the General Data Protection Regulation (GDPR). Extensive attention has been paid by the AI community to improve learning methods, whereas validation has been by far less investigated [8–9]. Repeated random sub-sampling and K-Fold, are considered as the two existing validation methods to evaluate the accuracy of supervised classification methods. These methods have been commonly employed by the researchers for validation purposes. Both of these two methods lack the capability of proving realistic and pervasive evaluation. In this paper, we introduce novel capabilities of our original validation method, the A-Test method, in estimating the learning capacity of any supervised classification method [10]. A study is performed for two different classification methods trained for a demanding clinical application, and the results are illustrated and compared.

2. Materials

Heart sound signals were recorded from the referrals to the Children Medical Centre of Tehran, using an electronic stethoscope of WelchAllyn Meditron Analyzer in conjunction with a portable computer. All the referrals underwent echocardiography, and the study was approved by the appointed ethics committee and was conducted according to the Good Clinical Practice. All the referrals or their legal guardians gave their informed consent to participate in the study. The patient population is listed in Table 1.

Table 1. Patient population of the study.

Heart Condition	Number of Patients	Age Range (years)
Aortic Stenosis	15	1–8
Mitral Regurgitation	15	4–8
Normal without murmur	30	4–15
Pulmonary Stenosis	15	1–10
Ventricular Septal Defect	25	1–9

3. Methods

3.1. Structural Risk and Learning Capacity

Structural risk of a classification method is defined as instability of performance measure of the classification method when the method is tested by a dataset out of the training data. The learning capacity of a classification method is defined as the capability of improvement in the performance of a classification method when the method is trained by a broader set of training data.

3.2. The A-Test Method

The A-Test method is based on using k-fold validation method for different values of k. In k-fold validation, the validation dataset is divided into k partitions with almost equal length. One partition is used for testing and the rest for training the classification method. This procedure is repeated k times with one partition is used only once for testing. The

A-Test employs the k-fold validation with different values of *k (k=2,…,K_{max})*, and the classification error is calculated for each k-value:

$$\Gamma_M = 100\ \frac{\sum_{k=2}^{K_{max}} \Gamma_{M,k}}{K_{max}-1} \tag{1}$$

where $\Gamma_{M,k}$ is the classification error of the classification method M, and k is the fold value for validation, called validation index. K_{max} is less than the minimum group size of the validation data. For a classification method, the percentage of the relative span of the classification rate, $\psi_{M,k}$, is employed as an indication of the learning capacity.

$$P_M = 100\ \frac{\max_k \psi_{M,k} - \min_k \psi_{M,k}}{\max_k \psi_{M,k}} \tag{2}$$

3.3. The classification methods

The A-Test method is employed to compare the structural risk of two different learning methods for classifying PCG signals, a deep time growing neural network (DTGNN) and a hidden Markov model (HMM), whose technical details are found in [10] and [11–12], respectively. Both the DTGNN and HMM are trained to learn the pathological characteristics of the signal, caused by aortic stenosis.

4. Results

Figure 1 shows a variation of the classification error due to the k-value. The descriptive statistics of the classification error, as well as the learning capacities are listed in Table 2.

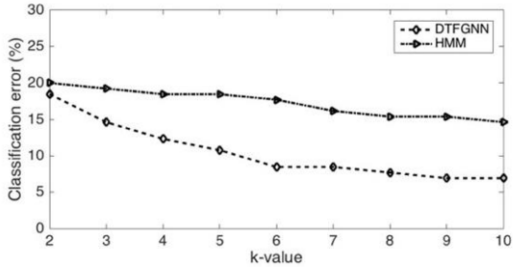


Figure 1. Variation of the classification error with respect to k-value for the two learning methods.

It is observed that the DTGNN provides a lower classification error, especially for higher k-values with a lower structural risk. Interestingly, for lower k-values, the two methods do not substantially differ in their performance.

Table 2. Descriptive statistics of the classification error for the two learning methods, the deep time growing neural network (DTGNN) and the hidden Markov mode (HMM).

Statistics	DTGNN	HMM
Average (%)	10.51	17.27
Minimum (%)	6.92	14.62
Maximum (%)	18.46	20.00
Median (%)	8.46	17.69
Learning Capacity (%)	14.2	6.7

5. Discussion

There are mainly two alternatives to the A-Test for exploring the structural risk of a classifier: the k-fold and the repeated random sub-sampling methods, both cannot provide an understanding of the learning capacity. Furthermore, it might lead to incorrect comparison for certain values of k, as was the case for the DTGNN and HMM with $k=2$. The learning capacity for the DTGNN and HMM is 14.2% and 6.7%, respectively, showing a higher capacity for the DTGNN because of the deep architecture.

6. Conclusions

This paper suggested the A-Test, as a powerful validation method. A-Test method provides means for validating not only the performance of a classifier but also the structural risk and the learning capacity of classification methods, by validating the classification methods' different ratios of training/test data and quantifying the results. This aspect cannot be seen in other alternatives, K-Fold validation and repeated random sub-sampling, which can potentially lead to an inappropriate validation result.

References

[1] Gharehbaghi A, Lindén M, Babic A. An Artificial Intelligent-Based Model for Detecting Systolic Pathological Patterns of Phonocardiogram based on Time-Growing Neural Network. Applied Soft Computing. 2019;83:105615.

[2] Gharehbaghi A, Ask P, Lindén M, Babic A. A novel model for screening aortic stenosis using phonocardiogram. In16th Nordic-Baltic Conference on Biomedical Engineering; 2015; Springer, Cham; p. 48-51.

[3] Gharehbaghi A, Sepehri AA, Linden M, Babic A. Intelligent Phonocardiography for Screening Ventricular Septal Defect Using Time Growing Neural Network. InICIMTH; 2017 Jul; p. 108-111.

[4] Gharehbaghi A, Lindén M, Babic A. A decision support system for cardiac disease diagnosis based on machine learning methods. Stud Health Technol Inform. 2017 Jan 1;235:43-7.

[5] Gharehbaghi A, Lindén M. An internet-based tool for pediatric cardiac disease diagnosis using intelligent phonocardiography. InInternational Internet of Things Summit. 2015 Oct 27; Springer. Cham; p. 443-447.

[6] Jain AK, et al. Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000;22(1):4-37.

[7] Gharehbaghi A, Babic A. Structural Risk Evaluation of a Deep Neural Network and a Markov Model in Extracting Medical Information from Phonocardiography. Studies in Health Technology and Informatics. 251:157-160.

[8] Gharehbaghi A, et al. An Intelligent Method for Discrimination between Aortic and Pulmonary Stenosis using Phonocardiogram, in World Congress on Medical Physics and Biomedical Engineering; 2015; Springer International Publishing: Cham. p. 1010-1013.

[9] Gharehbaghi A, et al. A Hybrid Machine Learning Method for Detecting Cardiac Ejection Murmurs. in EMBEC & NBC. Singapore; 2018; Springer Singapore; p. 787 – 790.

[10] Gharehbaghi A, Lindén M. A deep machine learning method for classifying cyclic time series of biological signals using time-growing neural network. IEEE transactions on neural networks and learning systems. 2017 Oct 12;29(9):4102-15.

[11] Gharehbaghi A, Ask P, Babic A. A pattern recognition framework for detecting dynamic changes on cyclic time series. Pattern Recognition. 2015;48(3):696-708.

[12] Gharehbaghi A, et al. A Hybrid Model for Diagnosing Sever Aortic Stenosis in Asymptomatic Patients using Phonocardiogram, in World Congress on Medical Physics and Biomedical Engineering; 2015; Springer International Publishing: Cham; p. 1006-1009.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210877

Factors for Individualization of Therapeutic Exercises for the Design of Health-Enabling Technologies

Lena ELGERT[a,1], Bianca STEINER[b], Birgit SAALFELD[a], Michael MARSCHOLLEK[a] and Klaus-Hendrik WOLF[a]
[a] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany*
[b] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Braunschweig, Germany*

Abstract. Designing health-enabling technologies (HETs) to support individualized physiotherapeutic exercises requires comprehensive knowledge of bio-psycho-social factors to be considered. Therefore, this review identified factors for individualization of therapeutic exercises in patients with musculoskeletal shoulder disorders in peer-reviewed articles searched in MEDLINE. The final full-text analysis included 16 of 335 search results and extracted nineteen main categories of individualization factors. The most frequently identified main categories include *progression of exercises*, *exercise framework*, and *assessment*. An iterative approach with constant reassessments represents the key principle for the process of individualization. Categories that are difficult to standardize were rarely mentioned, but should also be considered. The identified factors can improve the design-process of HETs by sensitizing developers, enable further formal modelling, and support communication between developers, physiotherapists, and patients.

Keywords. health-enabling technology, physical therapy, individualization, exercise

1. Introduction

Health-enabling technologies (HETs) are sensor-based information and communication technologies that promote a person's individual health, healthcare, and quality of life [1]. In physical therapy HETs can assist patients in performing exercises at home and support physiotherapists in assessments [2]. First studies indicate superiority of individualized compared to standardized non-individualized exercises in patients with musculoskeletal shoulder disorders [3]. HETs and treatment processes with HETs that assist patients and therapists with exercises should therefore promote individualized exercises and enable or support adaptation to the individual patient. Beyond this background, this literature review aimed to identify factors for individualization of therapeutic exercises in patients with musculoskeletal shoulder disorders to provide knowledge for designing HETs.

[1] Corresponding Author, Lena Elgert, Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany; E-mail: lena.elgert@plri.de.

2. Methods

The literature search and review process followed the PRISMA-ScR protocol [4]. Medical subject headings and text words related to the terms "upper extremity", "exercise and physical therapy", and "adaptation, progression, individualization, or personalization" constituted the search term in MEDLINE (via PubMed, on November 13, 2020). Two reviewers (LE and BSt) independently screened the 335 identified titles and abstracts against pre-defined inclusion and exclusion criteria. Reasons for exclusion were documented using the following main categories: 1. No focus on musculoskeletal shoulder disorders (n=184); 2. No exercise (n=24); 3. Exercises not individualized (n=82); 4. Other topics (e.g., prevalence of injuries during training) (n=13); 5. Language other than English or German (n=2). Full-text screening of the remaining 30 articles revealed five full-texts not reporting on individualized exercises. Nine full-texts were companion articles to the included main full-texts. The remaining 16 full-texts were incorporated in a systematic narrative synthesis based on a coding frame. The coding frame was built using a mix of concept- and data-driven-approaches [5] with some pre-defined main categories using the qualitative data analysis software MAXQDA (version 20.1.1; VERBI GmbH). The identified categories were illustrated in a code tree.

3. Results

Sixteen articles describing 15 projects were included [6-21]. One project was best represented by two complementing publications [19] and [20]. The 16 articles were published between 2007 and 2020. Treatment duration ranged from a single treatment to 5 months. Twelve weeks was the most common treatment period (n=6). Target group were patients with musculoskeletal shoulder disorders (e.g., subacromial impingement syndrome). The number of patients assessed ranged from one to 708 with a median of 90. Different study types were present ranging from the evaluation of a HET to case reports and qualitative interview studies to randomized controlled trials.

Different types of individualization were identified. Besides selecting, personalizing, and progressing exercises, creating exercises was mentioned as a type of individualization. However, a detailed description of the procedure for creating exercises was missing. Kromer et al. [17] reported to proceed in the sense of clinical reasoning and thus create exercises. Anton et al. [6] claimed that the interface of their HET supports the creation of new exercises, but did not elaborate or refer to expert knowledge.

The analysis yielded 19 main categories and 87 sub-categories for individualization of exercises. With 23 sub-categories occurring in 59 coded segments in 14 articles the *progression of exercises* was the most frequent main category. *Progression of exercises* was realized, for example, by *different performance levels* [10] with validated scales for perceived *exertion* or reported *pain* [14,15]. The second and the third most frequent main categories were *assessment* (34 coded segments, 10 articles) and *exercise framework* (22 coded segments, eight articles). *Validated assessment tools*, like the *Western Ontario Shoulder Index* [7], *range of motion* (ROM) [16], and *compliance measures* [7], were applied. Measurements, such as ROM, served both as decision criteria to decide whether a certain exercise can be started [7] and to evaluate whether an exercise is effective [6]. *Assessments* were performed with *re-tests* and thus served to adapt exercises to the patient both at the start of treatment and during the course of treatment [17,20,21]. The constant reassessment was inherent in the individualization process.

Factors maintaining or contributing to a patient's problem could be *leisure and sport activities* [20] or the *work place setting* [17]. In Anton et al. [6], the exercise individualization was based, among other things, on formalized knowledge in an ontology on biomechanical aspects. Figure 1 shows the main categories with frequency.

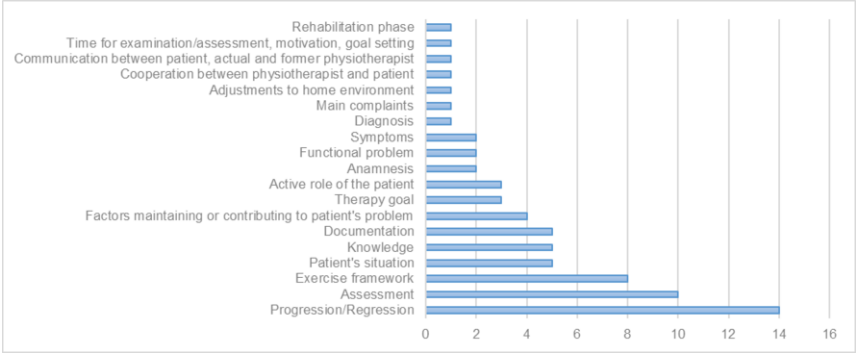


Figure 1. Main categories with frequency.

4. Discussion and Conclusions

Different individualization approaches of exercises are evident. Constant reassessment was identified as a key principle. Consequently, regardless of the selected approach, HETs should support or at least allow for reassessments. Since most rare individualization factors seem to be difficult to standardize, they may be underrepresented in this review of mostly highly standardized studies. However, that does not make them less important. Physiotherapy basically aims at sensomotoric self-determination of the individual and thus requires a meaningful frame of reference [22]. This means that exercises and therapy goals must be meaningful to the patient. Therefore, patients' everyday life, their environment, and their goals play a major role in the individualization of exercises. The rare as well as the frequent factors should be considered when designing HETs and physiotherapeutic treatment processes with HETs. Already in this limited literature search, the identified factors can provide insights, serve as a basis for further formal modeling and support the design-process of HETs. The results can sensitize developers and support communication between developers, physiotherapists, and patients. Further data collection and modelling, for example, in the form of an ontology, are planned to sufficiently support the design of HETs.

References

[1] Haux R, Koch S, Lovell NH, Marschollek M, Nakashima N, Wolf KH. Health-Enabling and Ambient Assistive Technologies: Past, Present, Future. Yearb Med Inform. 2016 Jun 30;Suppl 1(Suppl 1):S76-91.

[2] Elgert L, Steiner B, Saalfeld B, Marschollek M, Wolf KH. Health-Enabling Technologies to Assist Patients With Musculoskeletal Shoulder Disorders When Exercising at Home: Scoping Review. JMIR Rehabil Assist Technol. 2021 Feb 4;8(1):e21107.

[3] Roddy E, Ogollah RO, Oppong R, Zwierska I, Datta P, Hall A, Hay E, Jackson S, Jowett S, Lewis M, Shufflebotham J, Stevenson K, van der Windt DA, Young J, Foster NE. Optimising outcomes of exercise

and corticosteroid injection in patients with subacromial pain (impingement) syndrome: a factorial randomised trial. Br J Sports Med. 2021 Mar;55(5):262-271.
[4] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MDJ, Horsley T, Weeks L, Hempel S, Akl EA, Chang C, McGowan J, Stewart L, Hartling L, Aldcroft A, Wilson MG, Garritty C, Lewin S, Godfrey CM, Macdonald MT, Langlois EV, Soares-Weiser K, Moriarty J, Clifford T, Tunçalp Ö, Straus SE. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018 Oct 2;169(7):467-473.
[5] Schreier M. Qualitative content analysis in practice. London: SAGE; 2013. 272 p.
[6] Anton D, Berges I, Bermúdez J, Goñi A, Illarramendi A. A Telerehabilitation System for the Selection, Evaluation and Remote Management of Therapies. Sensors (Basel). 2018 May 8;18(5):1459.
[7] Bateman M, Smith BE, Osborne SE, Wilkes SR. Physiotherapy treatment for atraumatic recurrent shoulder instability: early results of a specific exercise protocol using pathology-specific outcome measures. Shoulder Elbow. 2015 Oct;7(4):282-8.
[8] Beaudreuil J, Lasbleiz S, Richette P, Seguin G, Rastel C, Aout M, Vicaut E, Cohen-Solal M, Lioté F, de Vernejoul MC, Bardin T, Orcel P. Assessment of dynamic humeral centering in shoulder pain with impingement syndrome: a randomised clinical trial. Ann Rheum Dis. 2011 Sep;70(9):1613-8.
[9] Bennell K, Coburn S, Wee E, Green S, Harris A, Forbes A, Buchbinder R. Efficacy and cost-effectiveness of a physiotherapy program for chronic rotator cuff pathology: a protocol for a randomised, double-blind, placebo-controlled trial. BMC Musculoskelet Disord. 2007 Aug 31;8:86.
[10] Christiansen DH, Frost P, Falla D, Haahr JP, Frich LH, Andrea LC, Svendsen SW. Effectiveness of Standardized Physical Therapy Exercises for Patients With Difficulty Returning to Usual Activities After Decompression Surgery for Subacromial Impingement Syndrome: Randomized Controlled Trial. Phys Ther. 2016 Jun;96(6):787-96.
[11] Crowell MS, Tragord BS. Orthopaedic manual physical therapy for shoulder pain and impaired movement in a patient with glenohumeral joint osteoarthritis: a case report. J Orthop Sports Phys Ther. 2015 Jun;45(6):453-61, A1-3.
[12] Eriksson L, Lindström B, Ekenberg L. Patients' experiences of telerehabilitation at home after shoulder joint replacement. J Telemed Telecare. 2011;17(1):25-30.
[13] Gleyze P, Georges T, Flurin PH, Laprelle E, Katz D, Clavert P, Charousset C, Lévigne C; French Arthroscopy Society. Comparison and critical evaluation of rehabilitation and home-based exercises for treating shoulder stiffness: prospective, multicenter study with 148 cases. Orthop Traumatol Surg Res. 2011 Dec;97(8 Suppl):S182-94.
[14] Ingwersen KG, Christensen R, Sørensen L, Jørgensen HR, Jensen SL, Rasmussen S, Søgaard K, Juul-Kristensen B. Progressive high-load strength training compared with general low-load exercises in patients with rotator cuff tendinopathy: study protocol for a randomised controlled trial. Trials. 2015 Jan 27;16:27.
[15] Keene DJ, Soutakbar H, Hopewell S, Heine P, Jaggi A, Littlewood C, Hansen Z, Barker K, Hamilton W, Carr AJ, Lamb SE. Development and implementation of the physiotherapy-led exercise interventions for the treatment of rotator cuff disorders for the 'Getting it Right: Addressing Shoulder Pain' (GRASP) trial. Physiotherapy. 2020 Jun;107:252-266.
[16] Kinsella R, Cowan SM, Watson L, Pizzari T. A comparison of isometric, isotonic concentric and isotonic eccentric exercises in the physiotherapy management of subacromial pain syndrome/rotator cuff tendinopathy: study protocol for a pilot randomised controlled trial. Pilot Feasibility Stud. 2017 Nov 14;3:45.
[17] Kromer TO, de Bie RA, Bastiaenen CH. Effectiveness of individualized physiotherapy on pain and functioning compared to a standard exercise protocol in patients presenting with clinical signs of subacromial impingement syndrome. A randomized controlled trial. BMC Musculoskelet Disord. 2010 Jun 9;11:114.
[18] Littlewood C, Ashton J, Mawson S, May S, Walters S. A mixed methods study to evaluate the clinical and cost-effectiveness of a self-managed exercise programme versus usual physiotherapy for chronic rotator cuff disorders: protocol for the SELF study. BMC Musculoskelet Disord. 2012 Apr 30;13:62.
[19] Roddy E, Zwierska I, Hay EM, Jowett S, Lewis M, Stevenson K, van der Windt D, Foster NE; SUPPORT trial team. Subacromial impingement syndrome and pain: protocol for a randomised controlled trial of exercise and corticosteroid injection (the SUPPORT trial). BMC Musculoskelet Disord. 2014 Mar 14;15:81.
[20] Stevenson K, Jackson S, Shufflebotham J, Roddy E, Foster NE. Development and delivery of a physiotherapist-led exercise intervention in a randomised controlled trial for subacromial impingement syndrome (the SUPPORT trial). Physiotherapy. 2017 Dec;103(4):379-386.
[21] Warby SA, Ford JJ, Hahne AJ, Watson L, Balster S, Lenssen R, Pizzari T. Effect of exercise-based management on multidirectional instability of the glenohumeral joint: a pilot randomised controlled trial protocol. BMJ Open. 2016 Sep 12;6(9):e013083.
[22] Probst A. Modell der menschlichen Bewegung. In: Höppner H, Richter R, editors. Theorie und Modelle der Physiotherapie, Ein Handbuch. Bern: Hogrefe; 2018. p. 99-113.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210878

Intellicor: Mobile Design for Monitoring Phonographic Signals

Malin WIST JAKOBSEN[a,1] and Ankica BABIC[b]
[a]*Department of Information Science and Media Studies, University of Bergen, Norway*
[b]*Department of Biomedical Engineering, Linköping University, Sweden*

Abstract. A mobile and web-based prototype was developed to explore utility of heart sound data in the context of patient self-monitoring. There are not many applications available despite measurement equipment that can be purchased. This research aimed at developing an application that could help patients understand and use phonocardiography. The resulting prototype Intellicor enables easy-to-use web and mobile solutions such as registering heart sound, review of previous heart signal recordings, summaries of terms related to patient condition, and medication taken. These functions can be utilized by both patients and physicians to create understanding of heart signals and build communication as a part of treatment. Three development iterations included several expert evaluators who gave positive feedback on the concept of the application. It was appreciated that patients could monitor heart signals and better understand the results. The medical experts would welcome such a system into their work domain if developed correctly and in accordance with the formal expectations, both legal and clinical. The prototype has shown the advantage of gathering data otherwise impossible to obtain. The Intellicor prototype presents the foundation that ought to be further developed in close cooperation of clinical and biomedical experts. The self-monitoring of this kind could benefit patients and the healthcare sector as demonstrated by the Intellicor prototype.

Keywords. Heart signals monitoring, Phonocardiography (PCG), application, eHealth system, expert evaluation

1. Introduction

Life changes due to complication of heart signals affect a large part of the Norwegian population, therefore indicating needs for monitoring of heart signals to a greater extend. There are medical indications, such as early symptoms of heart attacks, angina, and heart failure. As many as approximately 40 000 Norwegians annually receive specialist healthcare services related to heart attacks or angina, and 16 000 for heart failure [1]. People also monitor themselves remotely to track their behavior and workouts to get the best possible health benefits. Medical staff is also interested in patient self-monitoring due to the large number of patients they need to follow up, so they also recognize the advantage of remote self-monitoring without the intermediate help of health personnel.

Mobile and web applications can show signal activity to help users determine how and when to monitor their hearts and receive information about when to contact medical personnel. Many patients struggle with basic instructions and information from their

[1] Corresponding Author, Malin Jakobsen; E-mail: malin@j-j.no

health personnel due to limited health literacy and strive to know what to do and how to interpret findings when they self-monitor [2]. Smart phones already offer some applications such as Samsung health that contribute to physical fitness and health. Kardia is a mobile application for monitoring ECG signals [3]. There are not many applications for monitoring heart sound, even though measurement equipment is available such as electronic stethoscopes that can be combined with application to monitor heart signals.

2. Method

The design science research [4] paradigm was utilized to create an artifact through three design iterations. The design solutions ranged from low to high fidelity and were tested by relevant stakeholders and user groups. Stakeholders for this artifact are health personnel who could help patients monitor their health remotely and make sense of their measurements using mobile-web solutions. The second stakeholders are the patients who will monitor themselves using mobile web applications. One more potential stakeholder is a patient's family member who could assist the patient to self-monitor when necessary. Additional stakeholders can be found in the form of insurance agencies, health systems and society at large. Due to the corona virus pandemic at the time of the research these groups were not involved in the development but will be considered in the future.

The artifact's goal and context were to show proof of concept by building a prototype and testing it with today's technology which included development and evaluation tools.

The evaluation was conducted in two parts: system requirements for the prototype addressing both patient and physician needs using semi-structured interviews with field experts. Standardized evaluation methods such as Nielsen's heuristics, system usability scale and system usability testing were applied by usability experts.

3. Results

3.1. The Artifact

The artifact has two views, one enabling *patients* to remote monitor heart signals, and one for physicians who intend to utilize the data gathered from remote monitoring as a consultation tool (**Figure 1**, left side). The functionalities available to the patients are measuring heart signals with the use of PCG method, contact information to medical personnel, overview of earlier PCG measurements and medication overview to help patient manage and remember their medication. To share the data with medical personnel a generate QR code functionality was implemented. Similar functions are already seen in existing medical application available on the market, but currently are not used in Norway. As compared to earlier interpretations of PCG measurements, the patient will have a more understandable overview of their own heart signals, which helps elevate the significant problems surrounding health literacy that many adults face [5].

Based on the feedback gathered from the expert interviews new functionalities were added in the second iteration of the prototype. The result was a *physician's view* (**Figure 1**, right side) with the following functionalities: Registering QR code for transmitting patient data, overview of patients and their earlier recordings, encyclopedia with terms related to patients, and overall statistics over patients and their health history.

3.2. Evaluation of the Intellicor prototype

First understanding of expert expectations regarding the utility came from conducted interviews with a medical and a biomedical expert. This way, the Intellicor prototype development was informed about how medical personnel would use such a system and what they would expect to have for a consultation.

For PCG to be appropriately done, the article will propose the following key points to be considered when creating such an application for the medial domain:

- Good training for the individuals who are going to use the application,
- Good, but simple results view of heart sounds,
- Good feedback to the user regarding the quality of the signal,
- Statistical feedback regarding measurement. The results also should indicate the probability of having certain symptoms and an explanation about those.

A total of eight participants evaluated the third and final prototype using evaluation tools such as Nielsen's heuristic evaluation and System Usability Scale. In addition, system usability tasks were designed **(Figure 2)**. Six of the participants were usability experts and two were medical experts with little evaluating experience. The average resulting usability score was 90, which indicates high users' satisfaction, confirmed also through observations and comments.

Figure 1. Mobile view of the prototype, patient view (left) and physicians view (right)

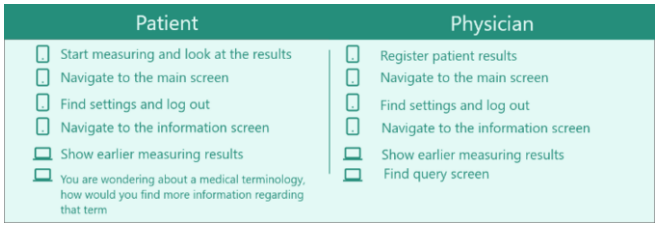


Figure 2. System Usability Tasks

4. Discussion

The main goal of creating the Intellicor prototype was to understand how feasible it would be for patients to monitor and record their heart sounds (PCG). Throughout all three development iterations efforts were made to secure a better understanding of medical content, especially following interaction design principles in the health-care sector. A key focus during evaluations and interviews was to address user needs, how potential users would want to have their health data presented. It is not often discussed during the patient consultation, but it could be assumed that patients are willing to self-monitor and share the data with healthcare personnel since there were already some commercial systems available for such purposes. However, Norwegian law sets clear and high demands regarding data sharing and privacy which was acknowledged in the development of the Intellicor prototype; no online consultations would be allowed, only collecting data, and transmitting it using QR code.

An assessment was conducted to better understand the patient's health situation and the requirements. The heuristic evaluation was instrumental in shaping different perspectives of two major user groups contributing to the iterative development. The interviews and the evaluation sessions with the participants resulted in positive feedback suggesting that an applied combination of medical informatics and user-centered design could tune into the patient needs. System usability was tested by performing a set of usability tasks returning a score of 90. Such a high usability score could be considered as a good indication to invest in further development by proper national authorities and other stakeholders.

5. Conclusion

Through research, interviews, and prototyping, it was possible to include user experience in designing of Intellicor prototype for patient to self-monitor heart signals. Intellicor prototype prospectively gathers data otherwise impossible to obtain and benefit from. Further development involves patients in their real environment.

Acknowledgements

The authors are thankful for the field experts; Dr. Henrik Holmstrøm, Dr. Arash Gharehbaghi, Dr Constance Weismann, Tuva Tronbøl for insight and feedback.

References

[1] Ariansen I, Selmer R, Graff-Iversen S, Egeland GM, Sakshaug S. Cardiovascular disease in Norway [cited 2020 Mar 23]; Available from: https://www.fhi.no/en/op/hin/health-disease/cardiovascular-disease-in-norway-

[2] Graham S, Brookey J. Do Patients Understand? Perm J. 2008;12(3):67.

[3] Hjem - ibeat.no [Internet]. [cited 2020 Apr 1]. Available from: https://ibeat.no/?v=c2f3f489a005

[4] Hevner AR, March ST, Park J, Ram S. Design Science in Information Systems Research. Management Information Systems Quarterly. 28. 75- 2004 [cited 2020 Sep 30].

[5] Wali H, Grindrod K. Don't assume the patient understands: Qualitative analysis of thechallenges low health literate patients face in the pharmacy. Res Soc Adm Pharm. 2016 Nov;12(6):885–92.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210879

Design and Implementation of a Computerized Management System for Adverse Drug Reactions at the Sourô SANOU University Hospital (CHUSS) in Bobo-Dioulasso, Burkina Faso

Arsene Sie Hamed TRAORE[a], Seydou Golo BARRO[a,b,1], Souleymane FOFANA[a], Issiaka SOMBIE[a,c], Rasmané SEMDE[d], Carine DE VRIESE[e], Dieudonne ILBOUDO[a] and Pascal STACCINI[b]

[a] *Nazi BONI University, Bobo Dioulasso, Burkina Faso*
[b] *RETINES Lab - Côte d'Azur University, Nice, France*
[c] *West African Health Organization, Bobo-Dioulasso, Burkina Faso*
[d] *Drug Development Laboratory, Joseph Ki-Zerbo University, Burkina Faso*
[e] *Pharmacology Department, Université Libre de Bruxelles (ULB), Belgique*

Abstract. Pharmacovigilance is the science and activities related to the detection, evaluation, understanding and prevention of adverse drug reactions or any other possible drug-related problems. In our tropics, this discipline is in an embryonic state. The availability of a management system capable of responding to pharmacovigilance activities is the main objective of our study. The coding was done on the DJANGO Framework. Signal detection was done using the ROR method. We designed three modules which are the notification module, the analysis module and the statistics module. This study has allowed us to launch the basis for a computerization of the pharmacovigilance information system and partly meets our objective. However, it could lead to the integration of the dictionary of adverse effects such as MedDRA as well as the International Classification of Medicines (ATC, EphMRA).

Keywords. Pharmacovigilance, Signal detection, Adverse reactions

1. Introduction

The development of pharmacovigilance in the world is quite recent and was accentuated following the occurrence of various health scandals in the 19th century, the most notable of which was the thalidomide scandal [1]. In 2002, the World Health Organization defined pharmacovigilance as the science and set of activities related to the detection, evaluation, understanding and prevention of adverse reactions or any other drug-related problem [2]. Indeed, all health products put on the market have benefits and risks, as clinical trials do not provide all the information on the safety of a drug at the time of marketing, as some adverse reactions (ARs) only appear during the use of the drugs in

[1] Corresponding Author, Seydou Golo BARRO; E-mail: seydou.golo@gmail.com.

the general population. This is the importance of pharmacovigilance, whose mission can be summarized around three actions: identification of AEs, which is based on reporting by health professionals, evaluation, which is based on imputability research, and prevention [3].

These three actions require the implementation of an efficient pharmacovigilance system. The first WHO collaborating center was created in Uppsala, Sweden (UMC) in 1978 [2]. Morocco and South Africa became the first African collaborating countries in 1992 [4]. The pharmacovigilance system in African countries is essentially characterized by spontaneous reporting, the inadequacy of local databases and the absence of tools for automatic generation of pharmacovigilance signals [4].

Burkina Faso is no exception to this observation, until 2013 the Sourô SANOU University Hospital (CHUSS) of Bobo-Dioulasso was the only hospital center to dedicate a service to pharmacovigilance [5]. This service is responsible for the detection, notification, analysis of adverse drug events and the use of this information for prevention purposes. In this service, spontaneous reporting is the main method of collection and individual analysis of imputability is done on a weekly basis.

A few isolated studies have consisted in evaluating pharmacovigilance signals in the department. However, the availability of a computerized management system for pharmacovigilance observations, which is able to carry out an imputability analysis and to generate signals automatically, is now essential. This has the advantage of improving the pharmacovigilance information system and making decisions that best meet patient safety in the use of medicines.

This is the general objective of our study, which consists of setting up a computerized management system for drug side effects at the Sourô SANOU University Hospital in Bobo-Dioulasso.

2. Materials and Methods

The study on the implementation of an automated management system of EIs was carried out at the CHUSS of Bobo Dioulasso which is a hospital of last resort in the health pyramid of Burkina Faso. At the server level, the programming language used is PYTHON 3.9. The HTML5, Javascripts and CSS3 languages were used to display, style and make dynamic the web elements of our pages. In order to avoid repetition of identical procedures during the creation of the web solution, we opted for the DJANGO Framework which is developed in python and responds to the DRY slogan " Don't Repeat Yourself ". This framework is open source and is dedicated to web development. Moreover, it uses an MVT[2] model which is inspired by the MVC[3] model. For the web server we used Apache 2.0. However, as almost all web servers do not communicate natively with python, we needed an interface to ensure this communication and DJANGO supports two interfaces which are WSGI[4] and the new standard ASGI[5]. We used as database management system, MySQL version 5.7. The analysis and design method used is 2TUP, the modeling language is UML. For the drug accountability analysis we used the WHO algorithmic method and the French method updated by

[2] Model View Template
[3] Model View Controller
[4] Web Server Gateway Interface
[5] Asynchronous Server Gateway Interface

Arimone et al. in 2011[6,7]. Signal detection was done using the Reporting odds Ratio [8]. Its formula is as follow:

$$\left(\frac{a}{b}\right) + \left(\frac{c}{d}\right) = ROR \tag{1}$$

$$ROR * e^{(\pm)1.96\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}} = IC \tag{2}$$

Table 1: ROR formula coefficients

Cases on drugs of interest	Report on events of interest	Report on others events
Cases on the drug of interest	a	b
Cases on other drugs	c	d
Total	a+c=N1	b+d=N2

3. Results

The designed solution is a web application which is then independent of operating systems and therefore accessible from all devices. Figure 1 displays the home interface of the solution.

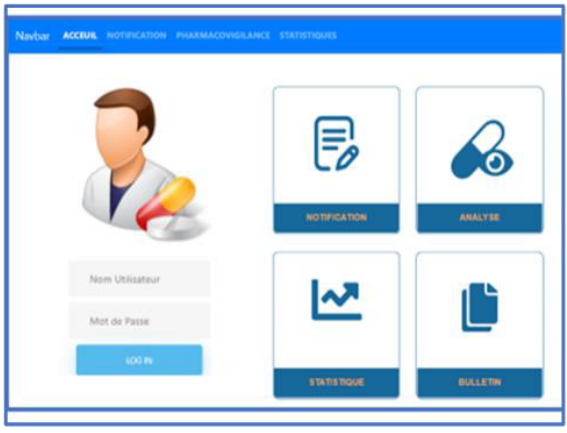


Figure 1. Welcome interface

Also the functionalities of our solution are divided into modules in order to facilitate the maintainability and to ensure an ergonomics live up to your expactations of all the users. The system includes essentially three modules which are:

- Notification management module that allows users to record information on the patient, on the drugs consumed, on the prescriber and on the adverse effects.
- Analysis module which allows the combination of each drug entered in the previous module with each of the reported adverse events. This module also allows an assignment of the imputability scores of the adverse effects.
- Statistics management module allowing to generate the signal for the drugs and the adverse reactions with the highest ORR scores. This module allows to elaborate dashboards for a decision support at the level of the

pharmacovigilance service. It also provides the possibility of a personalized search.

The access to the software is conditioned by authentication and the access to the different modules is conditioned by the authorization levels of the different actors and users of the system.

4. Discussion

Our objectives were to make available a database capable of responding in real time to the recording, analysis and restitution of data from the pharmacovigilance service. Our results show that these objectives have been achieved because we have developed an application that takes into account: the modules of recording, management and analysis of adverse drug reactions. The evaluation of our system is in progress through the compilation of the paper files already collected on the years spent in the pharmacovigilance service. The implementation of our system is a great innovation in the field of pharmacovigilance in our country and in the African sub-region, although it needs improvement, which are the next steps of this work. These are the integration of the dictionary of adverse reactions such as MedDRA, the International Classification of Medicines (ATC, EphMRA) etc. Another challenge remains the mobilization of ressources for deployment of the system on the system on a remote server to allow its use by other structures in Burkina Faso and even in the sub region.

References

[1] Lefrère J-J, Berche P. Les bébés du thalidomide. Presse Médicale. 2011;40:301–308.
[2] Weltgesundheitsorganisation, Collaborating Centre for International Drug Monitoring, editors. The importance of pharmacovigilance: safety monitoring of medicinal products. Geneva: WHO [u.a.]; 2002.
[3] The Importance of Pharmacovigilance - 2002 (WHO) - PAHO/WHO | Pan American Health Organization [Internet]. [cited 2020 May 15]. Available from: https://www.paho.org/en/documents/importance-pharmacovigilance-2002-who.
[4] Isah AO, Pal SN, Olsson S, et al. Specific features of medicines safety and pharmacovigilance in Africa. Ther Adv Drug Saf. 2012;3:25–34.
[5] Kabore L, Millet P, Fofana S, et al. Pharmacovigilance Systems in Developing Countries: An Evaluative Case Study in Burkina Faso. Drug Saf. 2013;36:349–358.
[6] Arimone Y, Bidault I, Dutertre J-P, et al. Réactualisation de la méthode française d'imputabilité des effets indésirables des médicaments. Therapies. 2011;66:517–525.
[7] Bannwarth B. Critères d'imputabilité des effets indésirables des médicaments. Rev Rhum Monogr. 2010;77:173–175.
[8] Faillie J-L. Les études cas–non cas : principe, méthodes, biais et interprétations. Thérapie. 2018;73:247–255.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210880

Implementation of a Monitoring System for Pregnancy by Tele-Ultrasound in an African Context (Burkina Faso)

Seydou Golo BARRO[a,1], Idrissa ZAMPALIGRE[b] and Pascal STACCINI[c]
[a] *Nazi BONI University, Bobo-Dioulasso, Burkina Faso*
[b] *Banfora's Sanitary District*
[c] *RETINES Lab - Cote-d'Azur University, Nice, France*

Abstract. Prenatal ultrasound is a radiological examination that allows optimal follow-up of pregnancies. However, its implementation remains limited in poor countries due to a lack of equipment and trained health workers, such as in Burkina Faso. The aim of this work is to set up an ultrasound tele-expertise system. To achieve this objective, we mobilized human, material and IT resources. The design of the tele-expertise application was based on a generic open source software called "MedShakeEHR" that we have adapted to our context. The application runs in a network on a Linux system. It enables ultrasound data exchange and sharing with a remote expert for interpretation using the DICOM protocol. This device thus offers the possibility to pregnant women to carry out their prenatal ultrasound locally. It also allows the constitution of prenatal ultrasound database according security and confidentiality standards.

Keywords. *Teleexpertise, Ultrasonography, Pregnancy*

1. Introduction

The realization of obstetric ultrasound in the follow-up of the pregnancy makes it possible to detect in time the risks in both the mother and the fetus. However, in poor countries, this examination is inaccessible to the vast majority of women, and this is mainly due to a lack of equipment and/or trained health workers. In Burkina Faso, especially in rural areas, pregnant women have to travel great distances for a follow-up ultrasound, sometimes not accessible financially. This work is part of this logic of improving the geographic and financial accessibility of obstetric ultrasound. It consists in setting up a technical and organisational system according national and international recommendations [1]:

- design and implementation of a tele-expertise web application from an open source one;
- organization of activities and distribution of tasks.

This device allows pregnant women to perform ultrasounds on the one hand and to interpret ultrasound results remotely. For the experiment of the device, we chose the rural

[1] Corresponding Author, Seydou Golo Baroo, University of Nazi BONI, Bobo-Dioulasso, Burkina Faso; E-mail: seydou.golo@gmail.com.

medical center of Niangoloko, Burkina Faso, as a pilot site for the deployment of the device in order to test his different functionalities.

2. Method

Analysis of the existing system shows this problem: pregnant women in rural areas don't have the opportunity to perform ultrasound on site due to lack of equipment and trained health workers. During the prenatal consultation, ultrasound is then prescribed as an external examination to be performed in a specialized medical center.

These centers are mostly located in other locations. For our example at Niangoloko, the nearest specialized medical center is the Banfora Regional Hospital Center located 45km away.

The system set up takes account of the existing situation and includes technical IT and organisational component. For the technical and computer aspects, we chose open-source software that has proven to be effective. These applications have been installed on the Linux operating system (Ubuntu 18.04.5 LTS). It concerns:

- MedShakeEHR, free medical software that we used for the design of the teleexpertise application [2].
- Orthanc, standalone DICOM server, for ultrasound data management [3].

We have made changes to the software to adapt it to our project. All actors have been trained to use it. On the organisational side, the midwife at the rural medical center is responsible for handling the ultrasound system for the examination of the woman and forwarding the ultrasound data to a physician expert for interpretation. This expert is then responsible for forwarding to the staff of the rural health center a written report, attached to the ultrasound images that are given to the patient by appointment.

3. Results

3.1. Application design and implementation

We used MedShakeEHR-base and MedShakeEHR-modGynObs modules of the MedShakeEHR software, in which we made modifications to adapt them to our project.

The following changes were made:

- implementation of a unique patient identification algorithm from the person identification profile (strict traits, extended traits, complementary traits) [4,5];
- integration of a database of localities in the country (village, sector,…);
- integration of a reference system for the country's professions;
- integration of a ICD-10 database for pathological diagnosis.

Fields on the base form that didn't meet our needs have been removed. The finalized application was installed on the CFRTM (Medical Technology Research and Training Center) server with remote access. All actors (assistants, midwives, specialists and supervisors) were trained in the use of tele-expertise application by their level of intervention. Midwives have received essential training, during six months from

specialist (gynecologist-obstetrician doctor), to perform an obstetric ultrasound: collection of ultrasound measurements and images, transmission of data.

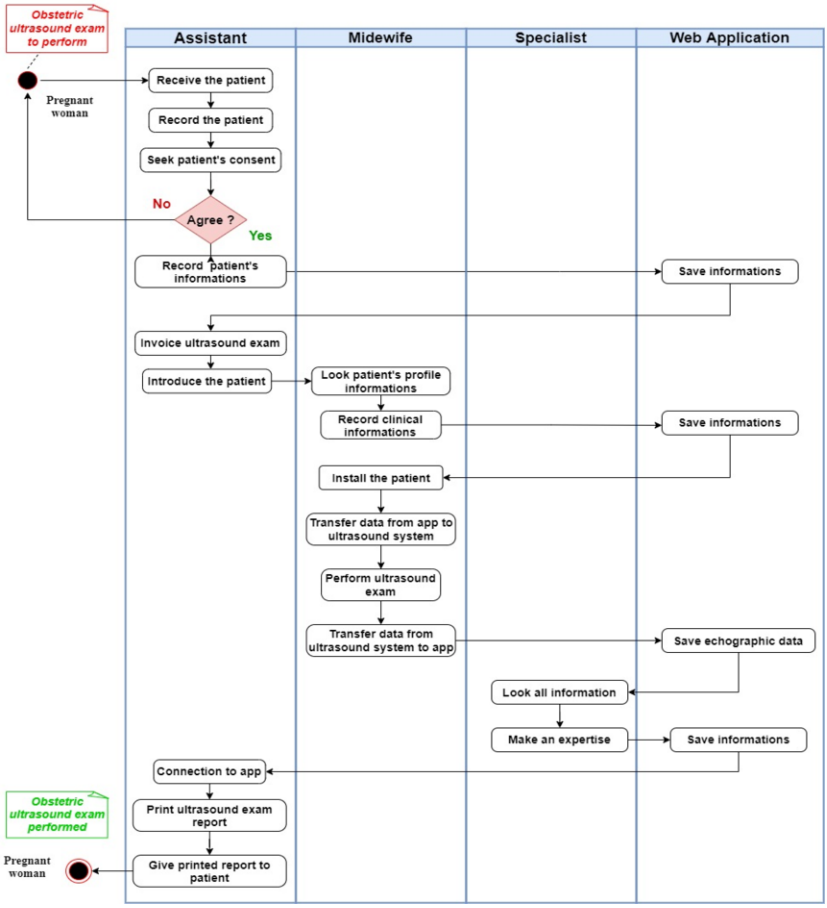


Figure 1. Device operation

3.2. Device operation

Ultrasound activities are carried out on three working days during a week: monday, wednesday and friday.

- ultrasound is performed in the morning from 08:00 to 12:00. After 12:00, a nominative list with the patient ID identifiers of the day are sent by email to the gynecologist-obstetrician doctor no later than 14:00.
- the gynecologist-obstetrician doctor connects to the application, consults the various records of the day and delivers his expertise according to his availability within 24 hours.
- The report of the prenatal ultrasound is therefore available; it will be printed and given to the patient on an appointment fixed beforehand in agreement with her within 48 hours.

At the decision of the steering committee, ultrasound examinations are invoiced at 5.000 CFA franc ($9) against 6000 CFA franc ($11) applied in public health

establishments excluding transport costs. The operation of the device, from patient reception to ultrasound interpretation is described by the following diagram.

3.3. Patient data security and confidentiality

Since we worked with health information, which is therefore highly confidential, it was crucial to ensure the security of health data as governed by Article 15 of Law 010-2004/AN on the protection of personal data [6].
Steps have been taken to ensure this confidentiality:

- a single administrator;
- encryption of the data storage hard drive on the server;
- an automatic backup system on external media;
- user authentication;
- a limitation of users' access rights according to their profile;
- a firewall to protect our server from remote attacks.

4. Discussion and conclusions

One of the major reasons for the low completion of ultrasound for pregnancy follow-up is the lack of availability of this examination in rural areas, which is essential for the smooth running of pregnancy. The main advantage of the ultrasound tele-expertise system we have set up is the accessibility of this examination, which is a guarantee of the motivation of rural populations. It is also a source of motivation for the staff of rural health centres who feel more involved in monitoring pregnant women.

Beyond this work objective, the scheme will eventually allow the creation of a database for monitoring pregnant women, which is necessary for the production of reliable statistics and research. The unique identification system implemented could open avenues for its extension to the HIS of our country.

References

[1] Conférence Nationale d'Echographie Obstétricale et Fœtale. L'échographie de dépistage prénatal. 2016. Available from: (http://www.cngof.fr/component/rsfiles/telechargement-fichier/fichiers?path=Clinique%252Finfo%2Bpatientes%252Fcngof_info_17-echographie_grossesse_.pdf&Itemid=756).

[2] Logiciel cabinet médical. MedShakeEHR : logiciel modulaire, universel, open source pour les praticiens santé (et son 1er module pour la gynécologie obstétrique). Available from: https://www.medshake.net/blog/a199/medshakeehr-logiciel-modulaire-universel-open-source-pour-les-praticiens-sante-et-son-1er-module-pour-la-gynecologie-obstetrique/. Date accesed: 25/05/2021.

[3] Orthanc-DICOM Server. About. https://www.orthanc-server.com/static.php?page=about. Date accesed: 25/05/2021

[4] ANAP. Identification du patient. Access date: 29/05/2021. https://www.anap.fr/ressources/publications/detail/actualites/identification-du-patient/.

[5] Programme Identifiant National calculé (INS-C) de Santé, Algorithme de calcul. ASIP Santé / PRAS ed. Paris: ASIP Santé, 2009;5-9. Available from: (https://esante.gouv.fr/securite/identifiant-national-de-sante/ins-c).

[6] CIL. Législation nationale. http://www.cil.bf/index.php/legislation/legislation-nationale. Date accesed: 31/05/2019.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210881

Mobile Monitoring of Diabetic Patients in the Active File of the Medical Center of Do (Burkina Faso)

Bry SYLLA[a], Seydou Golo BARRO[b,1], Jeoffray DIENDERE[c], Rachid G. KONFE[a], Ghislain BOUDA[d] and Pascal STACCINI[e]
[a] *Health Information System Department, Ministry of Health, Burkina Faso*
[b] *Nazi BONI University, Bobo-Dioulasso, Burkina Faso*
[c] *Health Sciences Research Institute (IRSS)*
[d] *Health District of Do, Burkina Faso*
[e] *RETINES Lab - Cote-d'Azur University, Nice, France*

Abstract: The care of diabetic patients in peripheral medical centers in Burkina Faso faces many difficulties. This work, which is a new experience, aimed to set up an information system for the care of diabetic patients in the context of Burkina Faso. The system thus conceived consists of a web application (MedshakeEHR), used by the doctor and a mobile application (Glucosio) for the patient. The system has advantages such as remote appointment scheduling, appointment reminder, patient information sharing. The device also makes it possible to store data for the production of statistics and for scientific research. This experience has enabled us to meet certain challenges related in particular to the problem of HIS such as organizational constraints, the creation of a unique identifier, the modeling of the main business processes, etc.

Keywords: HIS; ICT and health; Diabetes; Telemedicine

1. Introduction

Burkina Faso, like most African countries, is facing an epidemiological transition in health with the significant increase in chronic non-communicable diseases. These are cardiovascular diseases, diabetes, cancers requiring regular treatment throughout the patient's life. According to WHO, the number of people with diabetes worldwide increased from 108 million in 1980 to 422 million in 2014. In the African region the prevalence fell from 3.1% to 7.1% over the same period. New eating habits, sedentary lifestyle or even genetic predispositions are the major causes of the spread of this disease [1]. Unfortunately, our health systems have not adapted to these phenomena at the same speed, which naturally has consequences on the quality of life of patients and those around them. The solution to these difficulties lies in setting up an information system to effectively monitor patients and build a database for statistics and research on diabetes. The search for electronic experiments in this area and in our context has remained poor.

This is why we have undertaken to design and implement an electronic register of diabetic patients at the Medical Center with Surgical Antenna (CMA) of the health dis-

[1] Corresponding Author, Seydou Golo Baroo, University of Nazi BONI, Bobo-Dioulasso, Burkina Faso; E-mail: seydou.golo@gmail.com.

trict of DO, Bobo-Dioulasso, Burkina Faso. It is an application that integrates an appointment reminder system and patient access to their data via a mobile phone system. The use of telehealth has shown potential for the management of chronic pathologies and research [2]. The interest of our system is to improve the follow-up of diabetic patients with a view to their empowerment but also to contribute to improving the quality of data on this pathology.

2. Methods

2.1. System implementation site

The implementation site of our system is the Medical Center with Surgical Antenna (CMA) of the Do health district in Bobo Dioulasso (Burkian Faso). The health district of Do has, in addition to the CMA, twenty-seven (27) peripheral Health Centers.

2.2. Modeling and choice of tools

The analysis of the existing system and the new system to be implemented was carried out using UML (Unified Modeling Language). The offline version of draw.io and the power AMC software were used for the diagrams. Our system is an application made up of two parts: a web part used by doctors and a mobile part used by the patient. Open-source software called "MedshakeEHR" was used to design the web application. MedshakeEHR is a Hospital Information System (HIS). The mobile part was developed using Android software which is also open source "Glucosio". For our work, we used the Apache HTTP server version 2.4.3 which is free software respecting the client-server communication protocol. From this server, our application runs through a browser, regardless of the power of the computer or the operating system used. We preferred the Linux 18.04.2 operating system for this work. and the "MySQL" Database Management System. Mobile app users were equipped with a smartphone with Android OS version 4 and later operating system. To establish the connection between MedshakeEHR and Glucosio and facilitate data exchange, we have set up an API (Application Programming Interface). We choose the "REST" API (REpresentational State Transfer) because it meets the norms and standards in data exchange. This API retrieves data from MedshakeEHR and provides it to Glucosio in JSON (JavaScript Object Notation) format. For the safety and reliability of exchanges, we have integrated a unique patient identification system using the method proposed by the GMSIH (Group for the Modernization of the Hospital Information System). This method recommends the use of identity traits to generate the unique identifier by an automatic algorithmic calculation process; this process was used by ASIP Santé during its national health identification program [3] [4].

3. Results

3.1. System analysis and modeling

The analysis of the new system defines the role of each actor as well as the data provided and consulted. For the doctor, he connects to the "MedshakeEHR" application for recording the data of the patient's medical consultation; As for the patient, he uses the “Glucosio” mobile phone part to make his appointments and consult his personal information [figure 1].

To record a consultation, the doctor logs into the MedshakeEHR application after mandatory authentication. He then accesses the MedshakeEHR home interface and can finally record a medical consultation. The "appointment making" process involves the patient, the doctor and the two applications. The patient uses the mobile application for the appointment request which sends a notification to the web application that will be processed by the doctor.

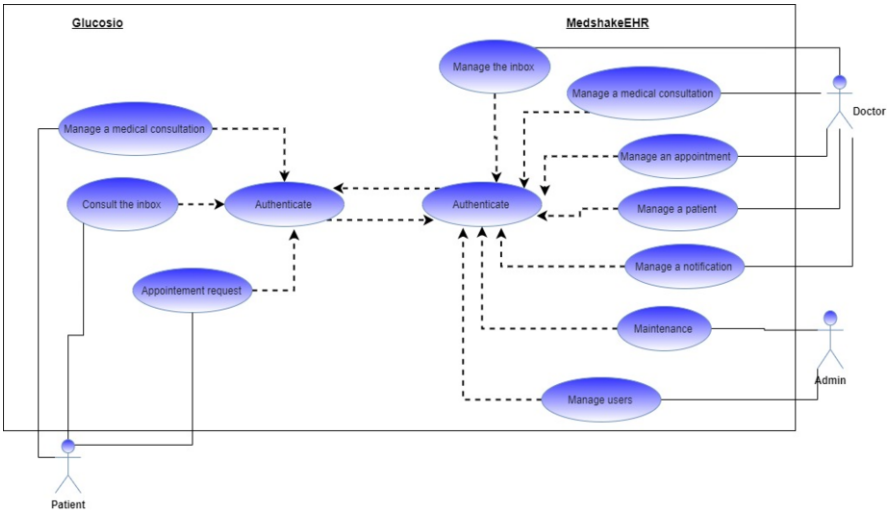


Figure 1. Use case Diagram

3.2. System implementation

The installation of the application was done on a desktop computer in the doctor's consulting room. Two user accounts have been created for the two physicians responsible for consulting diabetic patients with different levels of access. Security measures have been taken: an automatic backup system on external media, user authentication, limitation of user access rights according to their profile, etc. Physical measures were also taken regarding access to the consultation room and the technical room. For the protection of personal data, legal measures relating to digital health data, governed by Article 15 of Law 010-2004 / AN on the protection of personal data have been taken [5]. A system of data quality controls has been put in place during and after data entry. An interface for managing appointments and notifications has been created. For the use of the mobile phone part of the "Glucosio" application, a user account has been created allowing the patient to authenticate himself before accessing the content of the application. For security reasons, patient access has been limited to some information. The "REST" API, which was deployed on the same server as MedshakeEHR, allows the two applications to communicate with each other and facilitate the exchange of information between two or more doctors, but also between doctors and patients. When Glucosio is first installed, the patient's unique identifier will be stored on the application database. When requesting Glucosio, the unique identifier is used by the API to query MedshakeEHR and authorize secure exchanges [figure 2].

The system thus put in place has been operational since May 6, 2021. All users have been trained. We administered a system evaluation questionnaire to physicians responsible for

consulting diabetic patients with respect to the ergonomics of the application, functionality and handling. Another interview guide was used to assess patient satisfaction. The results of this assessment show that the application is generally satisfactory (70%). At the time of writing this article 36 patients were already registered in the system.

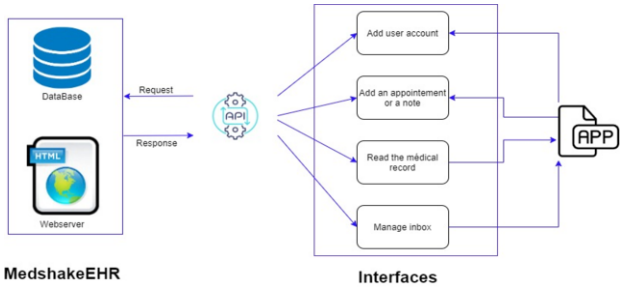


Figure 2. Application interface architecture

4. Conclusions

The electronic monitoring system for diabetic patients set up at the Do Medical Center is an innovation in the management of chronic diseases in the healthcare system of the country. It is a system with a dual objective (clinical and research) and a dual advantage (for the doctor and for the patient). Nevertheless, our project has its limits, in particular its restriction to a single medical center and a single pathology, the lack of perspective for a good evaluation, especially the clinical research component, etc. However, these limits remain as challenges to be taken up for future digital health projects in the care of patients. We set ourselves as a perspective the gradual integration of different pathologies to make this system a complete computerized patient record; the development of a billing module; the use of other features of MedshakeEHR in particular: Apycript mail, HPRIM, DICOM. It is also planned to install the system in the endocrinology department of the Souro Sanou University Hospital Center (CHUSS). Enabling the sending of SMS in addition to emails and notifications will greatly improve the management of appointment reminders in view of the mobile network coverage much higher than that of the Internet.

References

[1] Organisation Mondiale de la Santé (OMS). Rapport mondial sur le diabète. http://www.who.int/diabetes/global-report/fr/.

[2] Guo X, Gu X, Jiang J, Li H, Duan R, Zhang Y, Sun L, Bao Z, Shen J, Chen F. A Hospital-Community-Family-Based Telehealth Program for Patients With Chronic Heart Failure: Single-Arm, Prospective Feasibility Study. JMIR Mhealth Uhealth. 2019 Dec 13;7(12):e13229. doi: 10.2196/13229.

[3] ANAP. Identification du patient. https://www.anap.fr/ressources/publications/detail/actualites/identification-du-patient/.

[4] *Programme Identifiant National calculé (INS-C) de Santé, Algorithme de calcul*. ASIP Santé/ PRAS ed. Paris: ASIP Santé, 2009;5-9. Accessible à: (https://esante.gouv.fr/securite/identifiant-national-de-sante/ins-c).

[5] CIL. Législation nationale. http://www.cil.bf/index.php/legislation/legislation-nationale. Date d'accès : 20/05/2019.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210882

Length of Stay Prediction in Neurosurgery with Russian GPT-3 Language Model Compared to Human Expectations

Gleb DANILOV[a,1], Konstantin KOTIK[a], Elena SHEVCHENKO[a], Dmitriy USACHEV[a], Michael SHIFRIN[a], Yulia STRUNINA[a], Tatyana TSUKANOVA[a], Timur ISHANKULOV[a], Vasiliy LUKSHIN[a] and Alexander POTAPOV[a]
[a]*Laboratory of Biomedical Informatics and Artificial Intelligence, National Medical Research Center for Neurosurgery named after N.N. Burdenko, Moscow, Russian Federation*

Abstract. Patients, relatives, doctors, and healthcare providers anticipate the evidence-based length of stay (LOS) prediction in neurosurgery. This study aimed to assess the quality of LOS prediction with the GPT3 language model upon the narrative medical records in neurosurgery comparing to doctors' and patients' expectations. We found no significant difference ($p = 0.109$) between doctors', patients', and model's predictions with neurosurgeons tending to be more accurate in prognosis. The modern neural network language models demonstrate feasibility in LOS prediction.

Keywords. Length of stay, neurosurgery, prediction, neural networks, deep learning, natural language processing

1. Introduction

Patients and relatives anticipate the evidence-based risk assessment and length of stay (LOS) prediction in high-tech surgery [1,2]. LOS prognosis can also be utilized in clinical resource management. This paper continues a series of our publications on LOS predicting in neurosurgery based on unstructured textual data [5,6]. The current study aimed to assess the quality of LOS prediction with the GPT3 language model upon the narrative medical records in neurosurgery comparing to doctors' and patients' expectations.

2. Methods

Our study consisted of two components: 1) training a neural network language model to predict LOS based on unstructured text data collected retrospectively from electronic health records (EHR); 2) comparing the model predictions with the expectations of

[1] Corresponding Author, Gleb Danilov, N.N. Burdenko Neurosurgery Center, 4th Tverskaya-Yamskaya str. 16, Moscow 125047, Russian Federation; E-mail: glebda@yandex.ru.

neurosurgeons and patients on admission (before surgery) collected in prospective research.

The narrative textual data from the EHR of N.N. Burdenko National Medical Research Center of Neurosurgery was obtained for the period from 2000 to 2017 to shape the training corpus. The texts were extracted from the following source documents created upon patient admission: initial diagnosis, complaints, life history, history of present illness, somatic status, mental status, neurological examination, comorbidity, allergies, primary nursing examination, complications. Additionally, we added the narrative operative reports (for surgery commonly performed 1-2 days after admission) to the training set of texts. The omissions in all text records were replaced with blank lines. The texts from all document sources were concatenated through the ".\n" symbol into one field for each treatment case that served as the model's input. We replaced all contiguous zero-width spaces with an empty string, line-breaking spaces with a single newline, non-breaking spaces with a single space, then stripped any leading/trailing whitespace. Texts were converted to lower case. The target variable we used for machine learning was ln(1 + LOS), where LOS was the total number of days a patient stayed at the hospital since admission.

The underlying model (Generative Pre-trained Transformer 3 trained with large (600+ GB) corpus in the Russian language – ruGPT3) and tokenizer we utilized can be found here [3]. It was chosen as the promising state-of-the-art technology in Russian language modeling. The model vocabulary size was equal to 50 257 tokens. The length of the input text was limited to 2048 tokens. This, in turn, was also the maximum context length. The dimension of a token's vector representation was 768. To obtain the final vector representation of the token, the corresponding vectors were extracted from all 12 encoder layers and concatenated into one vector. To get a vector representation of the entire sentence, the final vector representations of its tokens were averaged into one vector dimensioned 768x12, which served the input to the fully connected layer (FCL) preceded by a dropout with a probability of 0.3.

The model fine-tuning was performed in two stages: 1) the top layer (FC) was trained during two epochs, the rest of the weights were "frozen"; 2) the whole model was trained during 20 epochs. The neural network training parameters were as follows: batch_size_top = 256 (training the top layer), learning_rate_top = 1e-5, batch_size_all = 16 (training the whole model), learning_rate_all = 1e-5, loss function: L1LOSS (https://pytorch.org/docs/stable/generated/torch.nn.L1Loss.html), optimizer: Adam (https://pytorch.org/docs/master/generated/torch.optim.Adam.html). The 25% random cases from 90 685 were used to evaluate prediction quality when fine-tuning the model. The inverse of the logarithm operation was applied to the model predictions to get the prognosis in days.

In a prospective study (carried out in the first half of May 2021), patients admitted to the N.N. Burdenko Neurosurgery Center and their treating neurosurgeons were asked to predict the total length of inpatient stay. Simultaneously, we registered neural model predictions.

Mean Absolute Error (MAE) in days was primarily used to measure the quality of model prediction and the correctness of doctors' and patients' expectations in a prospective study. Additionally, we calculated median absolute error (AE) and AE interquartile range (IQR) to better characterize AE distributions from different types of responders. We tested the difference in AE of prediction with the Kruskal-Wallis test. Spearman correlation coefficient was used to estimate the correlation between patients',

doctors', and model's predictions. A p-value < 0.05 was considered statistically significant.

3. Results

A total of 90 688 cases of neurosurgical treatment were identified retrospectively to provide unstructured medical texts for model training. A total of 111 patients with data collected prospectively were enrolled in the testing study. The study patients (adults with average age 52.8 ± 15.0 years; males / females = 49 / 62) underwent surgery for brain tumors (n = 78), vascular pathology (n = 12), hydrocephalus (n = 5) and other (n = 16). The length of the preoperative period in 86 (77.5%) of patients in the testing sample did not exceed 2 days (median = 1). The average LOS for the studied group was 7.5 days (median – 7 [6;9] days). The maximum reached 21 days in a complicated case. The LOS prognosis MAE and median AE with IQR obtained from doctors, patients, and the model are presented in Table 1. There was no statistically significant difference in the AE of prediction between the three types of responders (p = 0.109). However, doctors' prognoses tended to be more accurate.

Table 1. MAE – mean absolute error, AE – absolute error, IQR - interquartile range

Responders	MAE	Median AE [IQR]
Neurosurgeons	1.88	1.00 [1.00, 3.00]
Patients	2.37	2.00 [1.00, 3.00]
Prognostic model	2.37	2.00 [1.00, 3.00]

We observed a weak and statistically significant correlation between doctors' and patients' expectations (rho = 0.30, p = 0.002), patients' and model's prognosis (rho = 0.31, p = 0.001) and doctors' and model's prediction (rho = 0.29, p = 0.002). A boxplot of AE distributions for each type of responders with the estimation of statistically significant differences is shown in Figure 1.

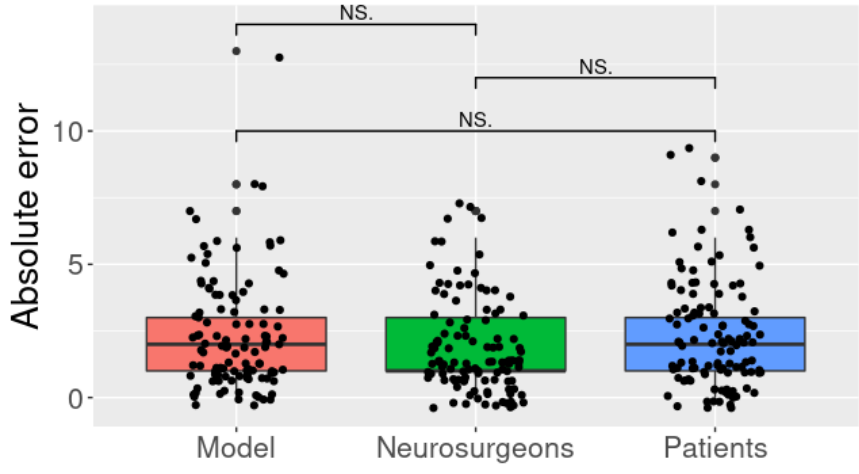


Figure 1. The distribution of prediction AE from three types of responders (the model, neurosurgeons and patients). NS. - non-significant.

4. Discussion

The length of stay is an important metric for healthcare providers and patients, and its prediction is sought to be automatized [4]. This study focused on utilizing narrative clinical text data, which seems informative and underlooked. We admit that our model had an advantage over the doctors and patients since it benefitted from the operative reports. Thus, the model, in fact, gave a prognosis a little later. However, given that surgery was performed 1-2 days after admission in most cases, the model's prognosis can still be considered timely. We also found that the neurosurgeons tend to predict LOS better than the patient or model. Nevertheless, the quality of the ruGPT3 model's prediction demonstrated in our research (even accounting for some unfairness of the comparison to humans) justifies considering such solutions as potentially applicable in automating LOS prediction. The testing sample size and its possible heterogeneity might be the limitations of the current study.

5. Conclusion

We found the medical texts sufficiently informative to solve complex tasks with modern neural network language models. The approach we justify should be considered in addition to predictive modeling based on structured feature space.

Acknowledgements

This project was supported by the Russian Foundation for Basic Research grants 18-29-01052 (information extraction) and 18-29-22085 (predictive modeling).

References

[1] Tanzer D, CRA KS, Tanzer M. Changing Patient Expectations Decreases Length of Stay in an Enhanced Recovery Program for THA. Clin. Orthop. Relat. Res. 2018;476(2):372-378. doi:10.1007/S11999.0000000000000043.

[2] Auer C, Laferton J, Shedden-Mora M, Salzmann S, Moosdorf R, Rief W. Optimizing preoperative expectations leads to a shorter length of hospital stay in CABG patients: Further results of the randomized controlled PSY-HEART trial. J. Psychosom. Res. 2017;97:82-89. doi:10.1016/J.JPSYCHORES.2017.04.008.

[3] sberbank-ai/rugpt3small_based_on_gpt2 Hugging Face, (n.d.). https://huggingface.co/sberbank-ai/rugpt3small_based_on_gpt2 (accessed August 29, 2021).

[4] Awad A, Bader-El-Den M, Mcnicholas J. Patient length of stay and mortality prediction: A survey. Heal. Serv. Manag. Res. 2017;30(2):105–120. doi:10.1177/0951484817696212.

[5] Danilov G, Kotik K, Shifrin M, Strunina U, Pronkina T, Potapov A. Prediction of Postoperative Hospital Stay with Deep Learning Based on 101 654 Operative Reports in Neurosurgery. Stud. Health Technol. Inform. 2019;258:125–129. http://www.ncbi.nlm.nih.gov/pubmed/30942728.

[6] Danilov G, Kotik K, Shifrin M, Strunina U, Pronkina T, Potapov A. Predicting Postoperative Hospital Stay in Neurosurgery with Recurrent Neural Networks Based on Operative Reports. Stud. Health Technol. Inform. 2020;270:382–386. doi:10.3233/SHTI200187.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210883

Drawing on the Doctor-Patient Relationship in e-Health Services

Rolf WYNN[a,1]
[a] *UiT The Arctic University of Norway, Tromsø, Norway*

Abstract. The doctor-patient relationship has been the cornerstone of medical practice. As e-health is coming to play an increasingly important role, it is necessary to consider how e-health can draw on, integrate and safeguard aspects of this relationship and thereby improve services and engage patients.

Keywords. E-health, doctor-patient relationship, engagement

1. Introduction

The doctor-patient relationship is the cornerstone of medical practice [1]. This relationship is sometimes, and especially within the field of counseling, referred to as 'the therapeutic alliance' [2].

During the two last decades, there has been a gradual and significant increase in the availability and use of different e-health services, including on the web, on apps, and on video [3,4]. In Norway, the Covid-19 pandemic has resulted in an increase in the use of e-health also in the public health services, as the need to push back against the pandemic has forced many health providers to reduce the availability of face-to-face services [5]. As e-health now moves into a period with an increased uptake, it becomes important to consider how to draw on, integrate and safeguard the elements of traditional face-to-face communication that have been such an important part of clinical medicine and the doctor-patient relationship, in order to further the uptake and success of e-health services.

In this paper, we discuss some central characteristics of the doctor-patient relationship and the advantages that might be gained by e-health providers by taking these into consideration.

2. Methods

Drawing on central concepts from doctor-patient interaction research, we present and discuss some clinically important phenomena that e-health providers may consider in optimizing their services.

[1] Corresponding Author, Rolf Wynn, Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, N-9037 Tromsø, Norway; E-mail: rolf.wynn@uit.no.

3. Results and Discussion

The relationship between the doctor and the patient is the cornerstone of traditional medical treatment. This relationship draws on the doctor's professionalism in terms of her knowledge and skills, but it also relies on other central elements including confidentiality and trust between the parties [1,2]. This relationship is a tool in itself that may help patients in their healing [1,2,6]. The lack of trust in technology and lacking technical skills have been two central factors in discouraging people's use of e-health [7].

While it may be difficult to isolate which elements that are most effective in the provider-patient relationship, one can hypothesize that e-health services that emulate characteristics of this relationship will be more effective in engaging and helping patients. Some e-health tools provide an opportunity to transfer central elements of the face-to-face relationship to a different medium, such as video consultations. However, this may be more challenging in other applications such as apps and self-help online programs. Drawing on this line of reasoning, some techniques that might be used to enhance engagement with e-health services could be video recordings or audio recordings of providers, the possibility to interact synchronously or asynchronously with providers through chat or email, etc. In this respect, interacting with providers, albeit in writing only, would probably be preferable to services where the user only provides or receives information, without any interactional element [8,9]. Regardless of the technology, it will remain central that the services inspire trust and require few technological skills.

4. Conclusions

As e-health is becoming an increasingly important part of the public health services, it is necessary to consider how e-health can draw on and integrate aspects of the traditional doctor-patient relationship in order to improve the services and engage the patients.

References

[1] Wynn R. Provider-patient interaction. Norwegian Academic Press, Oslo, 1999.
[2] Littauer H, Sexton H, Wynn R. Qualities clients wish for in their therapists. Scand J Caring Sci 2005;19**:**28-31.
[3] Wynn R, Oyeyemi SO, Budrionis A, Marco-Ruiz L, Yigzaw KY, Bellika JG. Electronic Health Use in a Representative Sample of 18,497 Respondents in Norway (The Seventh Tromsø Study - Part 1): Population-Based Questionnaire Study. JMIR Med Inform. 2020;8:e13106.
[4] Wynn R, Gabarron E, Johnsen JK, Traver V. Special Issue on E-Health Services. Int J Environ Res Public Health. 2020;17(8):2885.
[5] Wynn R. E-health in Norway before and during the initial phase of the Covid-19 pandemic. Stud Health Technol Inform. 2020;272:9-12.
[6] Bystad M, Bystad C, Wynn R. How can placebo effects best be applied in clinical practice? A Narrative Review. Psychol Res Behav Manag. 2015;8:41-5.
[7] Mata-Cervantes G, Clay CE, Baxter C. Maximising patient uptake and engagement with eHealth. NHS England, 2017.
[8] Del Hoyo JD, Nos P, Faubel R, et al. Adaptation of TECCU App Based on Patients´ Perceptions for the Telemonitoring of Inflammatory Bowel Disease: A Qualitative Study Using Focus Groups. Int J Environ Res Public Health. 2020;17:1871.
[9] Löhr H, Wynn R, Rosenvinge J. E-therapy as an adjunct to face-to-face therapy in the treatment of patients suffering from chronic psychiatric disorders. J Inform Technol Healthcare. 2007;5:67-79.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210884

Tracking Activities of Daily Living in the Home Office

Ju WANG[1], Marcel AGUIRRE MEHLHORN, Katharina DITLER, Benjamin EFFNER and Thomas M. DESERNO
Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School

Abstract. Due to the COVID-19 pandemic, home-office has turned to be a common practice in many companies to limit physical contact to reduce the rate of infections in the workplace. To quantify office workers' ADLs, this work demonstrates unobtrusive monitoring of activities of daily living (ADLs) of an office worker in a home-office environment with three low-cost sensors: an accelerometer and two light sensors. We extract four elementary events: distinct and fain chair movement, monitor, and fridge usage, from which we derived seven ADLs using predefined rules. This simple system can support the quantification of ADLs of home-office workers.

Keywords. Activities of Daily Living, home-office, COVID-19 pandemic, sensor, AAL

1. Introduction

The COVID-19 pandemic has changed our lives in many ways. With the availability of telecommunication platforms, it is feasible to conduct much work without being at the office. To limit physical contacts and reduce infection rate in the workplace, home-office has turned to be a common practice in many companies. However, home-office could negatively influence office workers. For instance, the work-life balance can be disturbed to change the office workers' physical and mental well-being [1].

The term activities of daily living (ADLs) is defined as the self-care activities that are important for health maintenance and independent living [2]. ADL is an indicator of health status as performing ADLs depends on cognitive (e.g., reasoning, planning), motor (e.g., balance, dexterity), and perceptual (including sensory) abilities [3]. With the advances in sensor technology, big data, and artificial intelligence, low-cost sensors can implement, when installed at proper locations, physiological monitoring and functional monitoring [4] for certain subjects, e.g., the elderly or patients with certain disabilities. For instance, with motion sensors and contact sensors, Kaye et al. recorded Alzheimer patients' daily life [5]. Similarly, our previous work profiled the hip fracture patients' behavior after discharge with multimodal Internet of things (IoT) sensors such as accelerometer, motion, vibration, and contact sensors [6][7]. The Massey University Student Enterprise (MUSE) group tracked the toilet use of an elderly person with a light

[1] Corresponding author, Peter L. Reichertz Institute for Medical Informatics, Muehlenpfordstr. 23, 38106 Braunschweig, Germany; E-mail: ju.wang@plri.de.

sensor in the rest room [8]. However, home-office has not yet been monitored by such simple systems.

Inspired by the related work of monitoring the ADLs of the elderly and the patients, this work will demonstrate unobtrusive ADL monitoring of an office worker with low cost and simple out-of-the-shelf sensors. We also showcase the information that can be extracted from the sensor data.

2. Methods

2.1. Sensor setup

We used a sensor kit (PLUX Biosignalsplux, PLUX wireless biosignals, S.A., Lisboa, Portugal) with sensors and hubs for data collection [9]. We used only three sensors (Fig. 1):

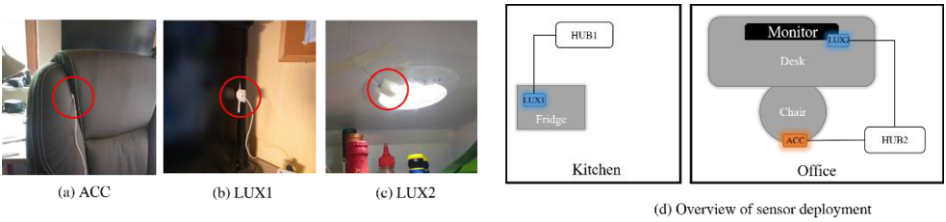


Figure 1. Hardware setup.

- An accelerometer (ACC) attached to the headrest of the office chair tracks the movement of the office worker (Fig. 1-a).
- A light sensor (LUX1) orthogonally attached to the monitor of the computer measures the display activity (Fig. 1-b).
- A light sensor LUX2 attached next to the refrigerator light checks if the refrigerator is open (Fig. 1-c).

We cabled ACC and LUX1 to a hub (HUB1) that then with Bluetooth connected to a smartphone (SF1), and LUX2 via another hub (HUB2) to another smartphone (SF2), which was placed near the fridge, as the computer was not within the Bluetooth reach from HUB2.

Our sensors recorded data over four weekdays. The system was started each day at about 07:00 and shut down at about 17:00. We configured the sampling rate for all sensors to 1 Hz.

2.2. ADL event extraction

We divided the ADL event extraction from raw sensor data into two stages: elementary event and ADL events. We defined and extracted four types of elementary events:

1. *Distinct chair movements* are caused by evident activities like standing up or sitting down in the office chair. We employed Butterworth and median filters to smooth the data and an empirical threshold to detect distinct movement.
2. *Faint chair movements* are the slight movements at the workplace while the person remains in the chair.

3. *Monitor on/off* event is derived by the light sensor, which illustrates clearly ascending and descending slopes when the monitor turns on and off, respectively. We applied an offset to the sensor data due to cope with interferences from the surroundings and ambient light.
4. *Fridge open/closed* event is detected directly from the signal as there is nearly no interfering light when the fridge is closed.

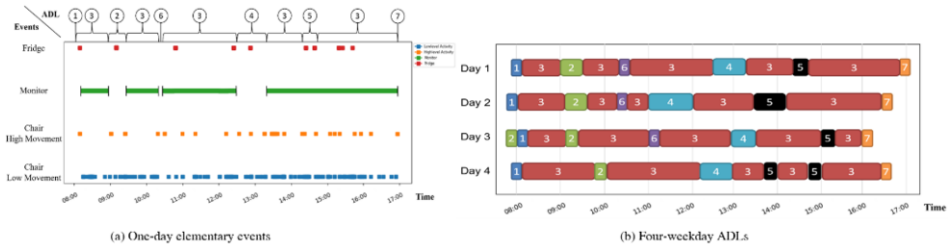
Based on the four types of elementary events, we extracted seven ADL events based on the social rhythm metrics (SRM) [10]. Table 1 lists the rules.

Table 1. Event extraction rules

ID	ADL	Rule	Timeslot
1	Begin of the working day	Distinct chair movement + monitor on	07:00 – 08:00
2	Breakfast	≥ 2 × distinct movement chair + monitor off + fridge open/closed	08:00 – 11:00
3	Concentration phase	Faint movement chair + monitor on	07:00 – 17:00
4	Lunch	≥ 2 × distinct chair movement + monitor off + fridge open/closed	11:00 – 14:00
5	Afternoon Snack	≥ 2 × distinct chair movement + fridge open/closed	14:00 – 17:00
6	General working break	≥ 2 × distinct chair movement + monitor off	07:00 – 17:00
7	End of the working day	Distinct chair movement + monitor off	16:00 – 17:00

3. Results

We obtained the elementary events. A sample set of events of a certain day is shown in (Fig. 2-(a)) with their starting- and ending-timestamps. Furthermore, after the event extraction, we can derive the target ADL events (Fig. 2-(b)) applying our rule set (Table 1). To quantify the office worker's ADL, we provide a statistical summary (Table 2).



(a) One-day elementary events (b) Four-weekday ADLs

Figure 2. Elementary and ADL events. The IDs of the ADL events are at the top of the graph.

Table 2. Summary of ADL events in four weekdays

ID	ADL	Average time	Average timeslot
1	Begin of the working day	-	07:45 – 08:15
2	Breakfast	ca. 20 min	09:00 – 10:00
3	Concentration phase	ca. 90 min	08:00 – 09:00
			09:30 – 11:00
			14:00 – 16:30
4	Lunch	ca. 45 min	11:00 – 13:30
5	Afternoon Snack	ca. 15 min	13:30 – 15:15
6	General working break	ca. 10 min	10:30 – 11:15
7	End of the working day	-	16:00 – 17:00

4. Discussion

Our experiment demonstrated that applying a limited number of low-cost and out-of-the-shelf sensors is feasible to track ADLs of workers in home-office. We designed the sensor deployment carefully to monitor the ADLs in an unobtrusive manner. Our data processing pipeline is easy to implement.

Our results suggest that the monitored office worker's lifestyle was relatively stable within the monitored duration. There was barely any movements during the workday besides standing up to get some food. The obtained information may offer tips for healthy lifestyle. For example, to keep in work-life balance, people need to break more frequently during the workday [11].

In this work, we defined the ADL events according to the SRM. On the one hand, we showed the potential of automatically collecting data for the SRM using sensor technology. On the other hand, the sensor data can be used to verify the precision of the SRM diary, which is usually collected by manually survey with the subjects.

The variety and the amount of sensors in the demonstrated system limit the ADL types that can be detected. The system cannot distinguish the work-relevant and non-relevant activities when the worker is in the chair. The system cannot detect the activies out of the sensing scope either, e.g., the phone calls with colleagues in a different room.

In future work, additional data from more test persons with more sensors can be derived to broaden our results over an extended period. In particular, we intend to include phone calls performed with wireless devices. Furthermore, a smartwatch with an accelerometer or motion sensor can be used to track a person's movements at home to verify the findings of the accelerometer on the office chair.

References

[1] Majumdar P, Biswas A, Sahu S. COVID-19 pandemic and lockdown: cause of sleep disruption, depression, somatic pain, and increased screen exposure of office workers and students of India. Chronobiol Int. 2020;37(8):1191–200.
[2] Troyer AK. Activities of daily living (ADL). Encycl Clin Neuropsychol. 2011;28–30.
[3] Mlinac ME, Feng MC. Assessment of activities of daily living, self-care, and independence. Arch Clin Neuropsychol. 2016;31(6):506–16.
[4] Wang J, Spicher N, Warnecke JM, Haghi M, Schwartze J, Deserno TM. Unobtrusive health monitoring in private spaces: the smart home. Sensors. 2021 Jan 28;21(3):864.
[5] Kaye JA, Maxwell SA, Mattek N, Hayes TL, Dodge H, Pavel M, et al. Intelligent systems for assessing aging changes: home-based, unobtrusive, and continuous assessment of aging. J Gerontol B Psychol Sci Soc Sci. 2011;66B:i180--i190.
[6] Marschollek M, Becker M, Bauer JM, Bente P, Dasenbrock L, Elbers K, et al. Multimodal activity monitoring for home rehabilitation of geriatric fracture patients-feasibility and acceptance of sensor systems in the GAL-NATARS study. Informatics Heal Soc Care. 2014;39(3–4):262–71.
[7] Wang J, Bauer J, Becker M, Bente P, Dasenbrock L, Elbers K, et al. A novel approach for discovering human behavior patterns using unsupervised methods. Z Gerontol Geriatr. 2014;47(8):648–60.
[8] Moretti G, Marsland S, Basu D, Sen Gupta G, The MUSE Group. Towards a monitoring smart home for the elderly: one experience in retrofitting a sensor network into an existing home. J Ambient Intell Smart Environ. 2013 Jan 1;5(6):639–56.
[9] biosignalsplux - PLUX Store [Internet]. [cited 2021 May 10]. Available from: https://plux.info/12-biosignalsplux
[10] Monk T, Flaherty J, Frank E, Hoskinson K, Kupfer D. The Social Rhythm Metric: an instrument to quantify the daily rhythms of life. J Nerv Ment Dis. 1990;178(2).
[11] Roche - 20 tips for maintaining a healthy work-life balance [Internet]. [cited 2021 Sep 8]. Available from: https://www.roche.com/careers/our-locations/asia/india/service/folder/20_tips_for_maintain.htm

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210885

Developing an Ontology for Documenting Adverse Events While Avoiding Pitfalls

Stefanie NEUENSCHWANDER[a], Patricia ROMAO[a], Jürgen HOLM[a] and Murat SARIYAR[a,1]
[a]*Bern University of Appl. Sciences, Department of Medical Informatics, Switzerland*

Abstract. Ontologies promise more benefits than terminologies in terms of data annotation and computer-assisted reasoning, by defining a hierarchy of terms and their relations within a domain. Here, we present central insights related to the development of an ontology for documenting events during interoperative neuromonitoring (IOM), for which we used the Basic Formal Ontology (BFO) as an upper-level ontology. This work has the following two goals: to describe the development of the IOM ontology and to guide the practice with respect to documenting of biomedical events, as available ontologies pose difficulties on certain issues. We address the following issues: (i) differentiate between the sets {documentation, identification, continuant} and {explanation, understanding, occurrent} as we had problems in applying the available ontology of adverse events, (ii) covering diseases and injuries in a consistent way, and (iii) deciding on which level to define relations.

Keywords. Adverse events, BFO, clinical documentation, intraoperative neuromonitoring, phenomenology

1. Introduction

Medical interventions related to the use of certain medical devices and products might cause unintended adverse events (AEs) [1]. Frequently, drug-related AEs are addressed in the literature [2], but AEs caused by surgery processes also generate significant burdens in terms of mortality and associated healthcare costs [3]. In order to capture the type of AEs, it is instrumental to document the results of associated health interventions in an unequivocal and standardized manner. For this purpose, standard terminologies [4] and ontologies [5] have been developed.

The main terms used in ontological hierarchies are universals, i.e., they represent what certain entities have in common [6]. It is frequently difficult to capture what such common characteristics are. For example, in the ontology of adverse events (OAE [1]), an AE is defined as "a pathological bodily process that occurs after a medical intervention", which also applies to processes after treatment of a chronic disease. Main motivation for this work is related to such ambiguous definitions and additional problems in the OAE, which we have encountered during the development of our ontology.

Our use case is the documentation of adverse events during interoperative neuromonitoring (IOM), for which we used the Basic Formal Ontology (BFO) as an

[1] Corresponding Author, Murat Sariyar, Bern University of Applied Sciences, Quellgasse 21, CH2502 Biel/Bienne, Switzerland; E-mail: murat.sariyar@bfh.ch.

upper-level ontology [7, 8]. IOM is the continuous measurements of neural electrical activity during neurosurgery to prevent postoperative deficits. BFO is a small upper-level ontology used by over 250 ontology-driven projects worldwide, especially for the biomedical domain. The fundamental distinction of BFO is between continuants, i.e., those entities that persist through time and are fully present at any point in time at which they exist at all (e.g., material entities such as syringes), and occurrents, i.e., those entities that unfold themselves in time and are present at any time only through their parts, e.g., my whole life with my childhood as one part at a certain time span.

Barry Smith has contributed central methods and notions to the scientific field of ontology. By consolidating his philosophical insights into the work on the BFO, he allowed many researchers to use sound concepts without having to understand the philosophies behind them. However, our work shows that it is sometimes crucial to understand the philosophical theories to make and justify certain decisions in practice. Without reference to philosophical works, it would have been difficult to detect the essential reason why the ontology of adverse events was not appropriate for us.

This work has the following two goals: to describe the development of the IOM ontology and to guide the practice with respect to documenting of biomedical events, as available ontologies pose difficulties on certain issues. We address the following issues: (i) differentiate between the sets {documentation, identification, continuant} and {explanation, understanding, occurrent}, (ii) covering diseases and injuries in a consistent way, and (iii) deciding on which level to define relations.

2. Methods

Guiding the practice required to adequately justify certain decisions during ontology development. For this purpose, the philosophical backbone of BFO were briefly analyzed and central implication for the practice was deduced from it.

To develop the IOM ontology, we first collected a list of relevant terms. In the second step, we defined a hierarchy between these terms through is_a (taxonomy) and part_of relations (mereology). In this step, the construction of the taxonomy and the reuse of existing ontologies was facilitated by referring to an upper-level ontology. Finally, further relations between two terms were defined (i.e., they are connected through a predicate), e.g., bearer_of (dependence) or is_connected_to (topology), and the terms themselves received further clarification by providing data properties, e.g., has_date or has_color. We used Webprotégé as development environment, Ontofox for reusing available ontologies, JavaFX for the frontend, and Apache Jena for the backend.

3. Results

3.1. Philosophical orientation in the IOM ontology

BFO has a philosophical orientation like other upper-level ontologies such as DOLCE and SUMO. It is strongly inspired by the phenomenologist Roman Ingarden, who advocated an immanent-realist view: there are mind-independent things in the spatiotemporal world and universals (kinds, properties, and relations), which capture them [9]. To put it in another way, the things are not just what we can sensually perceive, but also composed of categories, e.g., being a continuant. However, there is no direct

experience of things in terms of sensual and categorical properties. Experience is always intentionally directed towards things, through thoughts, images, desires, etc. A consequence is that we are always dealing with objects-as-intended (noematic structure), i.e., our intentional structure should be accounted for, when describing things.

During our ontology development it became crucial to consider the intentional structure for inferring following insight: continuants are associated with the intention to identify things, while occurrents are related to the intention to retrace something. Hence, if we want to identify a disease, we handle it like a continuant, if we want to retrace it (i.e., understand the course of the disease and its events), it functions as an occurrent. In other words, it is not useful to subsume events to the occurrence class if the goal of the final ontology is to facilitate documentation. These leads to the differentiation between the sets {documentation, identification, continuant} and {explanation, understanding, occurrent}.

3.2. Results and decisions in the IOM ontology

The IOM ontology is available via a user-friendly JavaFX frontend (access to the code at https://gitlab.ti.bfh.ch/neues4/IOMDO is granted upon request). It is not used for real-world documentation yet, even though some preliminary user tests were promising. Most of the terms in the IOM ontology were collected in a previous work, which lacked a taxonomy and relations [10]. For documenting the possible events during IOM with the help of an ontology, we relate those events to measurement data of evoked potentials and to induced postoperative deficits. With respect to the handling of events in our ontology, we initially relied on the OAE. However, in [1], it is stated that the ontology "allows the development and application of new analysis methods to better understanding the mechanisms of adverse events associated with or induced by different medical interventions", which means that the primary intention is to understand or retrace the processes that lead to adverse events [11].

What was the problem for the IOM ontology? In the adapted process-oriented view of the OAE, instances were of the form "Process P02 taken place on Sunday 10:00 pm leading to bleeding". Without further characterization of the processes in terms of dynamics and concrete components, such instances were useless for documenting events and associated data. In the gene ontology, for example, the process "hexose biosynthetic process" is integrated into a hierarchy of processes in order to understand the components of the metabolic processes. The aim is to retrace the sub-processes of the metabolic processes in general, not to identify concrete processes within an individual, which would be difficult anyway. In our initial approach, the difference of our perspective and that of OAE was overlooked. Even if one can combine both perspectives, the complexity of the resulting ontology would be very high. For documenting events, we finally used the generally dependent continuant term of BFO, as it has been used by the Ontology for General Medical Science (OGMS) for diagnoses [12]. Events are subsumed under the class 'information content', as we want to identify them, and for this goal, there is no benefit in referring to them additionally or alternatively as processes.

Regarding the classification of injuries, we relied again on the OGMS diagnosis and on the disposition concept for diseases. In OGMS, a disease "is a disposition to undergo pathological processes that exists in an organism because of one or more disorders in that organism". First, we decided that injuries belong to the category of disease, even though they are not dispositions per se. There are two reasons for this decision: (a) ICD10 has injuries as part of its categories (S00 – T88), (b) the effects of an injury depend on

the disposition of an individual. Second, we used injury diagnoses as entities with content, for the same reasons as for the classification of events in our ontology.

Considering relations, we could define all relations at the class level of our ontology or rely on other ontologies. The second approach means that we have to import the corresponding ontologies if we want to enable inferences based on reasoners. We decided to import the relation ontology because it provides a BFO-based collection of relations intended for standardization across different biomedical ontologies, e.g., we could use the "before", "has_phenotype", "has_participant" and other relations out-of-the-box [13].

4. Conclusion

Developing and understanding the distinction between different intentional structures (identification versus explanation) helped us to make informed decisions with respect to the classification of the terms "events" and "injuries" in an ontology destined to facilitate documentation. In contrast to the OAE, our ontology can be used off-the-shelf for documentation. Our insights have implications for several discussion related to the use of BFO. For example, uncertainties regarding the increase of complexity, when SNOMED CT is based on BFO can be reduced by referring to the distinction mentioned in the previous paragraph. For example: When diagnosing a patient with asthma in a healthcare setting, it is always an act of identifying the disease, not to retrace it. The related asthma attacks contribute to the characterization of asthma by being referred to them as 'information contents' rather than processes. The processes are relevant to the biomedical researcher who wants to better understand the disease by investigating certain dynamic patterns and who is, in most cases, not interested in individual cases.

References

[1] He Y, Sarntivijai S, Lin Y, et al. OAE: The Ontology of Adverse Events. J Biomed Semant 5 (2014), 29.
[2] Huang H-C, Wang C-H, Chen P-C, et al. Bibliometric Analysis of Medication Errors and Adverse Drug Events Studies. J Patient Saf 15 (2019), 128–134.
[3] Classen DC, Pestotnik SL, Evans RS, et al. Adverse Drug Events in Hospitalized Patients: Excess Length of Stay, Extra Costs, and Attributable Mortality. JAMA 277 (1997), 301–306.
[4] Breder CD. A Method for Retrieval of Adverse Event Terms in Clinical Trial Databases Using Standardised MedDRA Queries. Pharm Med 30 (2016), 103–108.
[5] Guo A, Racz R, Hur J, et al. Ontology-based collection, representation and analysis of drug-associated neuropathy adverse events. J Biomed Semant 7 (2016), 29.
[6] Armstrong DM. A Theory of Universals: Volume 2: Universals and Scientific Realism. CUP Archive, 1980.
[7] Arp R, Smith B, Spear AD. Building Ontologies with Basic Formal Ontology. MIT Press, 2015.
[8] Smith B. Ontology. In: Floridi L (ed) The Blackwell Guide to the Philosophy of Computing and Information. Blackwell Publishing Ltd, pp. 153–166.
[9] Johansson I. Proof of the Existence of Universals—and Roman Ingarden's Ontology. Metaphysica 10 (2009), 65–87.
[10] Zbinden C, Strickler M, Sariyar M, et al. A Protocol Entry Catalog for Intraoperative Neuromonitoring - Steps Towards an Ontology. Stud Health Technol Inform 272 (2020), 318–321.
[11] Yu H, Nysak S, Garg N, et al. ODAE: Ontology-based systematic representation and analysis of drug adverse events and its usage in study of adverse events given different patient age and disease conditions. BMC Bioinformatics 20 (2019), 199.
[12] El-Sappagh S, Franda F, Ali F, et al. SNOMED CT standard ontology based on the ontology for general medical science. BMC Med Inform Decis Mak 18 (2018), 76.
[13] Guardia GD, Vêncio RZ, de Farias CR. A UML profile for the OBO relation ontology. BMC Genomics 13 (2012), S3.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210886

Analyzing Topics and Sentiments from Twitter to Gain Insights to Refine Interventions for Family Caregivers of Persons with Alzheimer's Disease and Related Dementias (ADRD) During COVID-19 Pandemic

Sunmoo YOON[a,b,1], Peter BROADWELL[c], Carmela ALCANTARA[d], Nicole DAVIS[e], Haeyoung LEE[f], Amanda Bristol[a], Dante Tipiani[a], Joo Young Nho[g] and Mary MITTELMAN[g]
[a] *General Medicine, Department of Medicine, Columbia University, USA*
[b] *Data Science Institute, Columbia University, USA*
[c] *Center for Interdisciplinary Digital Research, Stanford University, USA*
[d] *School of Social Work, Columbia University, USA*
[e] *School of Nursing, Clemson University, USA*
[f] *Department of Nursing, Chung-Ang University, South Korea*
[g] *Department of Psychiatry, NYU Grossman School of Medicine, USA*

Abstract. We randomly extracted Tweets mentioning dementia/Alzheimer's caregiving-related terms (n= 58,094) from Aug 23, 2019, to Sep 14, 2020, via an API. We applied a clustering algorithm and natural language processing (NLP) to publicly available English Tweets to detect topics and sentiment. We compared emotional valence scores of Tweets from before (through the end of 2019) and after the beginning of the COVID-19 pandemic (2020-). Prevalence of topics related to caregiver emotional distress (e.g., depression, helplessness, stigma, loneliness, elder abuse) and caregiver coping (e.g., resilience, love, reading books) increased, and topics related to late-stage dementia caregiving (e.g., nursing home placement, hospice, palliative care) decreased during the pandemic. The mean emotional valence score significantly decreased from 1.18 (SD 1.57; range -7.1 to 7.9) to 0.86 (SD 1.57; range -5.5 to 6.85) after the advent of COVID-19 (difference -0.32 CI: -0.35, -0.29). The application of topic modeling and sentiment analysis to streaming social media provides a foundation for research insights regarding mental health needs for family caregivers of a person with ADRD during COVID-19 pandemic.

Keywords. dementia caregiving, online intervention, disparities, topic modeling

1. Introduction

Alzheimer's disease and related dementias (ADRD) is the sixth leading cause of death in the United States. More than 16 million family members or friends usually provide care for people with ADRD in their homes every year [1]. Family members of persons with

[1] Corresponding Author, Sunmoo Yoon, General Medicine, Department of Medicine, Columbia University Irving Medical Center, 630W 168 Street, PH105, New York, NY, 10032, USA; E-mail: sy2102@cumc.columbia.edu.

ADRD have been physically and psychologically affected by COVID-19 during the pandemic, including loss of support and increased loneliness, as well as likelihood of contracting COVID-19 [1,2]. News media have described patients with ADRD in nursing homes as COVID-19's hidden victims. Although recent qualitative studies from the U.K. and Canada have reported on the burden of family caregivers of persons with ADRD during the COVID-19 pandemic [1,2], it remains unclear which aspects of caregiving gained greater prominence among the general public during the pandemic, and to what degree they were already present in the public consciousness before COVID-19 emerged.

Topic modeling applies statistical analyses to counts of words or word groups that co-occur within text documents to reveal latent word groupings with potential semantic significance across a collection of documents. Compared to manual reading and keyword/keyphrase extraction algorithms (TF-IDF, TextRank), topic modeling offers the ability to find semantically meaningful groupings of multiple words across very large document collections, even in cases when all of the related words do not appear together within any single document or section. Further, sentiment analysis helps quantify the affective state of social media users as calculated via methods such as the Afinn lexicon, proposing a total emotional valence score [2]. Therefore, the purpose of this study was to detect changes in caregiving topics and sentiments to understand emotional distress before and during the pandemic as a foundation for developing future Twitter-based interventions for Hispanic and African American dementia caregivers.

2. Methods

We applied topic modeling and sentiment analysis for content mining of a corpus of publicly available English Tweets mentioning dementia/Alzheimer's caregiving-related terms (keywords: Alzheimer's/dementia caregiving/caregiver/care, n= 58,094) from Aug 23, 2019, to Sep 14, 2020. Any politicians' names were excluded upon data collection to exclude irrelevant Tweets during corpora creation. We used the NCapture and ORA software packages and other scripts run on a High-Performance Computing Cluster (https://cuit.columbia.edu/shared-research-computing-facility) to collect and analyze the data. First, 127 XML files were merged using Perl regular expression commands and shell scripting on the Terremoto cluster. Second, we used NLP techniques to remove n-grams occurring fewer than three times, and applied the Newman clustering algorithm to group associated Tweets in the corpus (Newman modularity: pre-COVID-19: 0.420, during COVID-19: 0.523). Next, we compared emotional valence scores of Tweets from before (through the end of 2019) and after the beginning of the COVID-19 pandemic (2020-). Lastly, experts in dementia caregiving studies reviewed detected topics. The larger study was approved by the Institutional Review Board (IRB). Resources including analytic Python codes and de-identified data are available on GitHub and OSF.io (https://osf.io/qruf3).

3. Results

A total of 58 topics were detected in the English Tweet corpus (n= 58,094), including 27 topics from 2019, prior to the emergence of COVID-19, and 31 in 2020 during the COVID-19 pandemic. The prevalence of topics related to caregiver emotional distress

(e.g., depression, helplessness, stigma, loneliness, elder abuse) and caregiver coping (e.g., resilience, love, reading books as in bibliotherapy) increased (Figure 1). Conversely, the prevalence of common topics related to late-stage dementia caregiving, such as nursing home placement, hospice, or palliative care, decreased during the pandemic (figure 1). The mean emotional valence score significantly decreased from 1.18 (SD 1.57) to 0.86 (SD 1.57) after the beginning of the COVID-19 pandemic in 2020 (difference -0.32 CI: -0.35, -0.29) (figure 2).

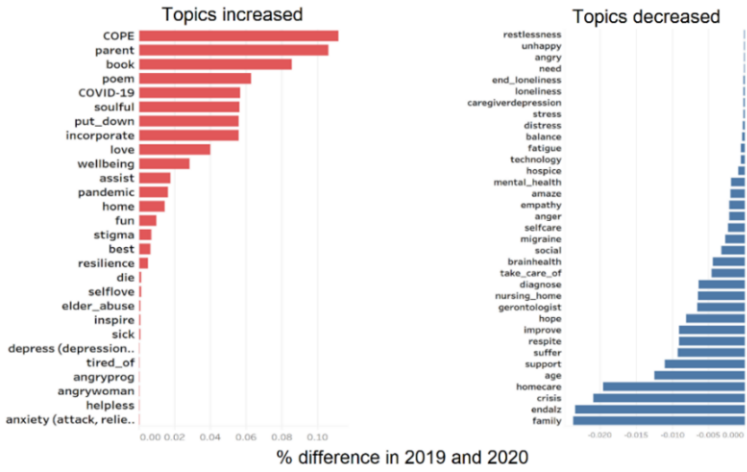


Figure 1. Changes in topic prevalence in English Tweets mentioning dementia caregiving in 2019 (before COVID-19) and 2020 (during the COVID-19 pandemic).

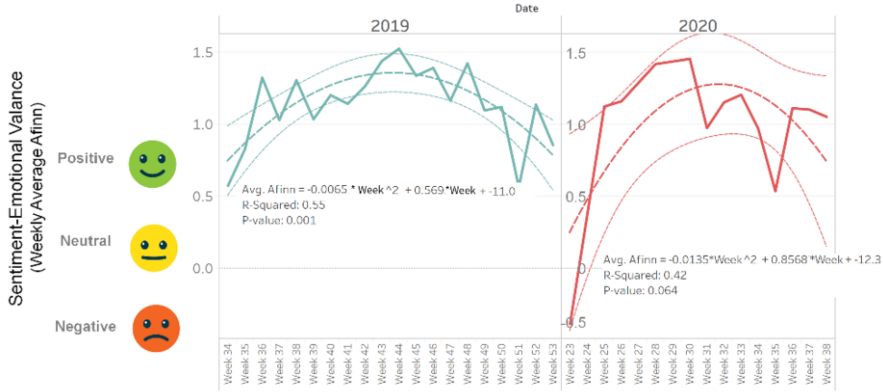


Figure 2. Weekly average emotional valence score of dementia caregiving English Tweets in 2019 (before COVID-19) and 2020 (during the COVID-19 pandemic).

4. Discussion and Conclusions

This study explored the changes in topics and emotional valence scores of English Tweets mentioning dementia caregiving before and during the COVID-19 pandemic. Consistent with the findings from other qualitative studies [1,2] regarding caregiving burdens among family caregivers in the U.K. and Canada, we found increased prevalence of family caregiver emotional distress such as depression, helplessness, and loneliness in

Tweets mentioning dementia caregiving during COVID-19 pandemic. In the meantime, we also found an increased prevalence of coping strategies and resilience topics in Tweets mentioning dementia caregiving. Historically it is not new to observe mentions of the specific coping strategy of reading books containing caregiving-related stories and poems to alleviate emotional distress, conflicts, and tensions in Tweets that circulate among dementia caregivers. Nevertheless, it was new to observe the actual terms of "resilience" and "cope," rather than their proxy terms (e.g., reading caregiving stories) in Tweets mentioning dementia caregiving during the pandemic. Consistent with a recent qualitative study emphasizing the role of external encouragement and social support such as from the health system (e.g., a nurse or a social worker in the health system communicates with caregivers and provides support) during the stress of the pandemic [1], this study suggests that communities and organizations on Twitter have been spontaneously taking an active role to promote resilience and coping among dementia caregivers. Moreover, it is remarkable to find new topics on positive coping strategies and resilience (e.g., encouraging love, well-being, and improving spiritual health) in Tweets mentioning dementia caregiving. As considerable evidence has continuously shown the effectiveness of off-line spiritual support for family caregivers for dementia patients since 1985 [4], our finding adds to the knowledge that online social media has the potential to be a platform to promote positive coping strategies and resilience (e.g., love, spiritual support, well-being) at the societal and system level, rather than providing individual-level decision support for late-stage dementia caregiving, such as nursing home placement, hospice, or palliative care, as we found their prevalence decreased over the period considered in this study. In conclusion, the application of topic modeling and sentiment analysis to streaming social media provides caregiving topics and sentiments to understand emotional distress before and during the COVID-19 pandemic as a foundation for developing future Twitter-based interventions for Hispanic and African American dementia caregivers.

Acknowledgments

U.S. federal grant TweetS2 R01AG060929 (P.I.: Yoon).

References

[1] Hanna K, Giebel C, Butchard S, Tetlow H, Ward K, Shenton J, Cannon J, Komuravelli A, Gaughan A, Eley R, Rogers C. Resilience and supporting people living with dementia during the time of COVID-19; A qualitative study. Dementia. 2021 Aug 4:14713012211036601.

[2] Brown EE, Kumar S, Rajji TK, Pollock BG, Mulsant BH. Anticipating and mitigating the impact of the COVID-19 pandemic on Alzheimer's disease and related dementias. The American Journal of Geriatric Psychiatry. 2020 Jul 1;28(7):712-21.

[3] Finn Årup Nielsen A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages 718 in CEUR Workshop Proceedings 93-98. 2011 May. http://arxiv.org/abs/1103.2903.

[4] Wright SD, Pratt CC, Schmall VL. Spiritual support for caregivers of dementia patients. Journal of Religion and Health. 1985 Mar;24(1):31-8.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210887

Identification of COVID-19 Vaccines Concerns in Health-Related French Web Forums : A Topic Modelling Approach

Pierre KARAPETIANTZ[a,1], Bissan AUDEH[a] and Cédric BOUSQUET[a]
[a] Inserm, Sorbonne Université, université Paris 13, Laboratoire d'informatique médicale et d'ingénierie des connaissances en e-santé, LIMICS, F-75006 Paris, *France*

Abstract. Since December 2019 and the first reported cases of COVID-19 in Wuhan, China, there have been 199,466,211 confirmed cases of COVID-19 in the World. The WHO defined vaccination hesitancy as one of the top ten threats to global health in 2019. Our objective was thus to identify topics and trends about COVID-19 vaccines from French web forums to understand the perception of the French population on these vaccines before the vaccination campaign started. We performed a topic model analysis on 485 web forums' posts. 10 topics were found. We reviewed 120 posts from 6 of these 10 topics. One topic was about vaccine hesitancy, refusal, and mistrust, and two topics were related to what the users think about the government, the political and economic choices made towards this epidemic.

Keywords. Pharmacovigilance, Social Media, Internet, Forum, Side effects, COVID-19 vaccines

1. Introduction

Since December 2019 and the first reported cases of COVID-19 in Wuhan, China, there have been 199,466,211 confirmed cases of COVID-19 in the World including 4,244,541 deaths according to the World Health Organization (WHO) [1]. Vaccines have been developed, and now a total of 3,984,374,918 vaccine doses have been administrated [1]. However, vaccine hesitancy is a hot topic that should be taken into consideration by vaccination campaigns, especially in France [2]. The WHO defined vaccination hesitancy as one of the top ten threats to global health in 2019 [3].

Social media are potentially interesting to investigate such hesitancy, and any information may be useful to health authorities to promote vaccine use, adherence, and confidence. Moreover, as the web 2.0 can affect the decision about getting vaccinated [4], exploring personal experiences and opinions from social media users is desirable to measure how such media may modify the perception of the population. Our objective was thus to identify topics and trends about COVID-19 vaccines from French web forums to understand the perception of the French population on these vaccines before the vaccination campaign started.

[1] Corresponding Author. Pierre Karapetiantz Limics, 15 rue de l'Ecole de Médecine, 75005 Paris, France; E-mail: pierre.karapetiantz@gmail.com.

2. Material and Methods

2.1. Material

We used the open-source tool Vigi4Med Scraper [5] to extract posts from web forums. The study period was from 1st February 2020 to 31st January 2021, before the vaccination campaign in France which started in February 2021.

2.2. Method

The preliminary data processing to reduce noise and incoherence [6] and the identification of COVID-19 vaccine related posts were done in 7 steps:

- Conversion to lower case: the text of all posts was converted to lower case text as R software (The R Project for Statistical Computing, Vienna) discriminates between lowercase and uppercase words.
- Identification of COVID vaccines related posts: These posts were selected according to the following decision rule:
    - The post contained the words coronavirus or COVID
    - The post contained the word vaccine, or any word related to COVID vaccines such as BioNTech, Cominarty, Pfizer, Johnson & Johnson, Moderna, AstraZeneca, Vaxzevria or Covishield. Spelling variations were tolerated.
- Punctuation and stop words removal
- Removal of certain identified words specific to web forums (e.g., users names)
- Removal of posts dealing about influenza
- Removal of overrepresented tokens that appear in more than 95% of the posts and in half of the posts or more and rare tokens that are present in less than 5% of the posts, less than 2 times and less than 5 posts. Words used to build the corpus (eg, covid) were thus removed as they were systematically present in all posts and did not carry any useful information.
- Multiple whitespace characters were collapsed to a single blank.

We generated *document-term matrix* (DTM) from processed users' posts. This matrix shows the frequency of terms that occur in the collection of posts: rows correspond to posts, and columns to terms. If a term occurs in a particular post, then the matrix entry is 1, if not it is 0. We decided to keep unigrams and bigrams. This made it possible to retain frequent contiguous sequences of two items, such as adverse effects (AEs).

We applied topic modeling with Latent Dirichlet Allocation (LDA) algorithm developed by Blei et al with the Variational Expectation-Maximization (VEM) algorithm [7, 8]. A 10-fold cross-validation was performed, testing the models for each number of topics from 2 to 10 topics. First, the data set was split into 10 test data sets (size of 49 posts) with the remaining data as training data. The number of topics for the final model was chosen according to the minimum perplexity in mean obtained by the 10 testing models and we initialized the parameter alpha of the final model considering the mean of the alphas generated for this number of topics [9]. For topics we found possibly relevant, 20 posts with less than 500 words were sampled for a manual review.

The motivation of the use of this method is that it answers our objective of identifying topics from users' posts. Indeed, when users write a post, they refer to a certain number of topics using words with a certain probability from the set of terms that correspond to that topic. Thus, each message contains several topics among all the identified topics, and the probability distribution shows how prominent the identified topics are in this message.

All statistical analyses were performed with the R language and environment for statistical computing and graphics [10] using the tm, textmineR and topicmodels packages [9, 11, 12].

3. Results

Among the 66 176 posts from 1st February 2020 to 31st January 2021, 704 were identified as concerning COVID-19 vaccines. 211 were excluded as they also referred to influenza, and 8 were excluded after data-management (removal of all the terms in the post). A total of 485 posts were included in the study.

After generating the DTM, we performed the 10-fold cross-validation. The perplexity we obtained is described in Figure 1.

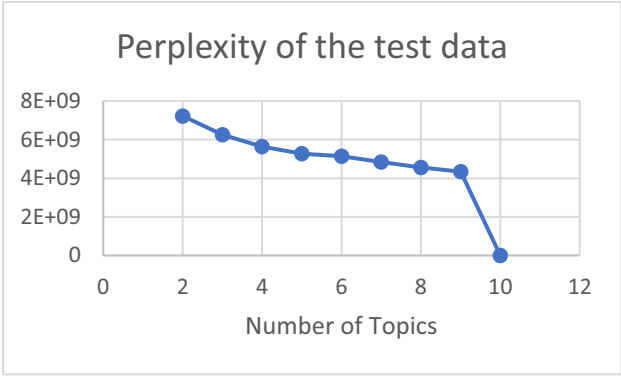


Figure 1. Perplexity according to the number of topics used

For the final model, we determined the number of topics to 10, and alpha was estimated as the mean of the 10 VEM test models for 10 topics.

After analyzing the most related words to each topic, we decided to manually review posts from topics 3, 4, 5 and 6 which seemed to be the most meaningful for our research. Table 1 gives the summary of this analysis.

Table 1. Main themes found for each of the studied topics generated

Topic	Main themes
3	Government-big pharma connivance; vaccination campaign and when it will start; children vaccination; health pass; chloroquine
4	Epidemic in Europe (beginning); vaccine research; vaccine refusal, hesitation and mistrust; civic responsibility
5	Chloroquine; hope for a vaccine, vaccine; Remdesivir; bored with politicians, Chloroquine AE; economic issues
6	Childbirth and COVID; how long before a vaccine?; COVID and children; collective immunity; conspiracy

4. Discussion

We identified 10 main topics. We reviewed 120 posts from 6 of these 10 topics. In future analysis, posts from topic 4 should be the object of a deeper manual analysis to learn more about vaccine hesitancy, refusal, and mistrust. Topics 3 and 5 could be explored to know more about what the users think about the government, the political and economic choices made towards this epidemic.

Our study had some limitations. As social media users may invent neologisms, we plan to reduce noise by selecting only words that are part of the French language, but there is a risk of losing potentially interesting information. Stemming was tried during this study, but results were difficult to interpret. Thus, lemmatization should be privileged. Finally, Dynamic Topic Models should also be explored with this kind of data to evaluate how topics may evolve.

References

[1] WHO, Coronavirus (COVID-19) Dashboard [Internet] 4 August 2021 [cited 4 August 2021] Available from: https://covid19.who.int/.
[2] Le Monde, Les Français sont les plus sceptiques face aux vaccins, selon une enquête mondiale. [Internet] 19 June 2019 [cited 4 August 2021] Available from: https://www.lemonde.fr/societe/article/2019/06/19/les-francais-sont-les-plus-sceptiques-face-aux-vaccins-selon-une-enquete-mondiale_5478259_3224.html.
[3] ANMJ, WHO's top ten threats to global health in 2019 [Internet] 3 May 2019 [cited 4 August 2021] Available from: https://anmj.org.au/whos-top-10-threats-to-global-health-in-2019/.
[4] Betsch C, Sachse K. Dr. Jekyll or Mr. Hyde? (How) the Internet influences vaccination decisions: recent evidence and tentative guidelines for online vaccine communication. Vaccine 2012;30(25):3723-6.
[5] Audeh B, Beigbeder M, Zimmermann A, Jaillon P, Bousquet C. Vigi4Med scraper: a framework for web forum structured data extraction and semantic representation. PloS one. 2017 Jan 25;12(1):e0169658.
[6] Iavindrasana J, Cohen G, Depeursinge A, et al. Clinical data mining: a review. Yearbook of Medical Informatics. 2009:121-133. PMID: 19855885.
[7] Blei DM, Lafferty JD. Topic Models. Srivastava AN, Sahami M, editors. Text Mining: Classification, Clustering, and Applications. CRC Press; 2009. p. 23.
[8] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res [Internet]. 2003;3:993-1022. Available from: http://portal.acm.org/citation.cfm?id=944937.
[9] Grün B, Hornik K. topicmodels: An R Package for Fitting Topic Models. Journal of Statistical Software [Internet]. 2011;40(13):1--30. Available from: http://www.jstatsoft.org/v40/i13/.
[10] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016.
[11] Feinerer I, Hornik K. tm: Text Mining Package [Internet]. 2012. Available from: http://CRAN.R-project.org/package=tm.
[12] Tommy J. TextmineR: Functions for Text Mining and Topic Modeling. R package version 3.0.5 [Internet] 2021. Available from: https://CRAN.R-project.org/package=textmineR.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210888

MDA Framework for FAIR Principles

Zahra MOHAMMADZADEH[a1], Hamid Reza SAEIDNIA[b], Marcin KOZAK[c] and Ali GHORBI[d]
[a] *Department of Health Information Technology, Kashan University of Medical Sciences and Health Services, Kashan, Iran*
[b] *Department of Knowledge and Information Science, Tarbiat Modares University, (TMU) Tehran, Iran*
[c] *Department of Media, Journalism and Social Communication, University of Information Technology and Management, Rzeszow, Poland*
[d] *Department of Knowledge and Information Science, Tehran University, Tehran, Iran*

Abstract. This paper shows that the MDA framework can be helpful for designing and implementing FAIR principles. We reached this conclusion based on a focus group interview with six experts, during which we focused on the three MDA components: mechanics, dynamics and aesthetics.

Keywords. Design, Implementation

1. Introduction

FAIR (Findability, Accessibility, Interoperability, and Reusability) principles were published in 2016 [1] and have been used ever since. In this paper, we explore whether the MDA (Mechanics, Dynamics, and Aesthetics) framework, a tool used to analyze games during the design phase [2], can be used to design and implement FAIR principles.

2. Method

We conducted a focus group interview with three experts in database development and information architecture, and three professors of health information technology. During online semi-structured interviews, we asked one question: *Is it possible to use MDA components to implement FAIR principles?* The interview lasted four hours. Then we analyzed the expert comments using narrative analyses.

3. Result & Discussion

The MDA framework consists of three design components: mechanics, dynamics, and aesthetics. From the interviews, it follows that *mechanics* can be used to determine features and rules that can be applied in certain situations when implementing FAIR.

[1] Kashan University of Medical Sciences and Health Services, Isfahan Province, Kashan, Iran; E-mail: Mohammadzadeh.z@umsu.ac.ir

Dynamics can be viewed in terms of how users navigate to a certain location; *dynamics* can also be extended by looking at users' behavior during their activities. *Aesthetics* refer to user interface design and user experience; in implementing FAIR, aesthetics are related to visual appearance. Visual design comprises all aspects relevant to the design theme, such as font face and size, color combinations, layout, images, charts, graphics, animation and videos.

The MDA framework formalizes the use of games by breaking them down into their distinct components and establishing their design counterparts. Due to the MDA components and their use in the design and implementation of FAIR, FAIR platforms can be more applicable and accessible. Widely used [1], FAIR principles have been tested in a variety of fields [3]. The MDA framework has been introduced in numerous studies as an appealing approach to website implementation [4], design and development of mobile health applications [5], and software development [6]. This shows that, because of the flexibility of the MDA framework, it can be used in various scenarios. Originally, the framework was used to design games; since games are considered as more attractive to users than other media [7], MDA can increase the level of user engagement in implementing FAIR principles.

A limitation of the study was the small number of specialists interviewed. However, in addition to their expertise, they offered a significant amount of time and involvement during the interviews. We believe that our results offer a starting point for further research on the practical aspects of applying MDA to design and implementation of FAIR principles; such research should include more experts and provide more detailed questions.

4. Conclusions

The study shows that the MDA framework can assist in designing and implementing FAIR principles.

References

[1] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data. 2016 Mar 15;3(1):1-9.

[2] Hunicke R, LeBlanc M, Zubek R. MDA: A formal approach to game design and game research. Proceedings of the AAAI Workshop on Challenges in Game AI; 2004: San Jose, CA.

[3] El Emam K, Sibbald D, Barnes R. Implementing the FAIR Data Sharing Principles: Technology and Experiences. 2021.

[4] Santosa P. Creating Websites Using MDA Framework: A Proposed Mapping 2015.

[5] De Schutter B. Gerontoludic design: extending the MDA framework to facilitate meaningful play for older adults. International Journal of Gaming and Computer-Mediated Simulations (IJGCMS). 2017;9(1):45-60.

[6] Nikiforova O, Nikulsins V, Sukovskis U. Integration of MDA framework into the model of traditional software development. Databases and Information Systems V: IOS Press; 2009:229-39.

[7] Kusuma GP, Wigati EK, Utomo Y, Putera Suryapranata LK. Analysis of Gamification Models in Education Using MDA Framework. Procedia Computer Science. 2018;135:385-92. doi: https://doi.org/10.1016/j.procs.2018.08.187.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210889

Students' Satisfaction and e-Learning Courses in Covid-19 Pandemic Era: A Case Study

Leila ERFANNIA[a,b], Roxana SHARIFIAN[c], Azita YAZDANI[b,c], Aida SARSARSHAHI[d,1], Rezvaneh RAHATI[e] and Saeedeh JAHANGIRI[e]
[a] *Health Information Technology Department, Paramedical School, Zahedan University of Medical Sciences, Zahedan, Iran*
[b] *Clinical Education Research Center, Shiraz University of Medical Sciences, Shiraz, Iran*
[c] *School of Health Management and Information sciences, Health Human Resources Research Center, Shiraz University of Medical Sciences, Shiraz, Iran*
[d] *Medical Equipment Department, Urmia University of Medical Sciences, Urmia, Iran*
[e] *Scientific research center, Zahedan University of Medical Sciences, Zahedan, Iran*

Abstract. The present study is conducted to determine the status of e-learning, student satisfaction and the relationship between these two variables in Zahedan University of Medical Sciences (ZAUMS). According to a descriptive study, there was just a significant difference between the mean score of e-Learning experience and student satisfaction, and a positive correlation between the education level and student satisfaction. Also, there was a positive correlation between all variables of e-learning and student satisfaction The findings showed that the more capable learners were outcome of better educational content, stronger e-learning infrastructure, better support and assessment of e-learning quality, which ,in turn, resulted in the greater the students' satisfaction. As a result, the experiences from the evaluation of e-learning in the Covid-19 pandemic period may be regarded a good guide in improving the course during the Covid-19 pandemic, and also it can be considered a key factor in providing educations in the post-Covid-19 period.

Keywords. E-Learning, Satisfaction, Student, Covid-19 pandemic,

1. Introduction

With the outbreak of the cCovid-19 pandemic and the resulting restrictive measures such as quarantine and social distance, many people worldwide were forced to change their lifestyles and use online platforms for education, entertainment and work [1]. Before the Covid-19 pandemic outbreak, e-learning had been growing at a rate of 15.4% per year in educational institutions around the world without pressure from institutions or learners, followed with a dramatic boom with the occurrence of the Covid-19 pandemic [2].

Allegedly, Student attitude evaluation is one of the important methods of evaluation in the educational system [3]. A survey of students' views around the world about virtual

[1]Corresponding Author, Aida Sarsarshahi, Medical Equipment Department, Urmia University of Medical Sciences, Orjhans Street, Resalat Blvd, Urmia, IR Iran; E-mail: sarsarshahi81@gmail.com.

education offered during the Covid-19 pandemic reveals two general views: The first group of students who have a positive attitude refer to easy access to educational resources at any time and any place, maximum coverage of education for a wide range of audience and more freedom to communicate with professors [3, 4]. However, those who had a negative attitude towards the virtual education point to their unwillingness to use, insufficient infrastructure and technology constraints, managerial and technical support, access problems such as Internet connection, mobile phone, and laptop ownership and usage costs as their most important challenges in using e-learning [5-7].

Hence, owing to the complexity of e-learning structure and evaluating e-learning, there is a need to consider students' perspectives and other components related to e-learning structure at Zahedan University of Medical Sciences where all educational courses were offered through virtual education. Therefore, the purpose of this study was to determine the status (according 5 variable in present research) of virtual education, student satisfaction, and the relationship between these two variables.

2. Methods

The present study was a descriptive cross-sectional study in nature conducted at Zahedan University of Medical Sciences in 2021. Based on a simple random sampling method, 325 people were selected (Total number=2150) of whom 300 participated in the study. The research tool was a mixed questionnaire adapted from previous research and included three general sections [8]. The first part covered the demographic information of the respondents. The second part consisted of 33 questions about the status of e-learning and included 5 dimensions of individual characteristics of the learner, educational content, infrastructure, support, and assessment. The third section encompassed 19 questions related to students' satisfaction with e-learning and consisted of five Likert scales from strongly agree to strongly disagree. The questionnaire was designed online and shared on educational platforms for students.

Descriptive and analytical statistical methods were used to analyze the data. Therefore, ANOVA and Pearson correlation tests were used to analyze the mean score and correlation between demographic variables and satisfaction, .Also, Pearson correlation test was used to analyze the correlation between the variables of e-learning status and students' satisfaction.

3. Results

Based on the findings, most of the participants were women (87.7%). The highest mean age was 25 to 35 years. Also, 11.3% of the participants were students of health information technology. 74.7% of the participants had used the virtual education system for less than a year. Most people had a bachelor's degree (81.3%). (Table 1). According to the comparison tests of the means, there was just a significant difference between the mean of e-Learning experience and student satisfaction (P-value = 0.027, T-score=2.56).

Table1. Frequency of demographic variables and statistical.

Variable	Variable Category	Frequency (%)	Test	P-Value (T-score)
Education Field	Health Information Technology	34(11.3)	ANOVA	0.950(0.93)
	Radiology	32(10.7)		
	Nursing	27(9)		
	Midwifery	18(6)		
	Anesthesia	23(7.7)		
	Medicine	22(7.3)		
	Environmental Health	27(9)		
	Nutrition	10(3.3)		
	Physiotherapy	15(5)		
	Laboratory Sciences	26(8.7)		
	Professional Health	14(4.7)		
	Audiology	3(1)		
	Optometry	9(3)		
	Speech Therapy	7(2.3)		
E-Learning Experience	<1	224(74.7)	ANOVA	0.027(2.56)
	1-3	62(20.7)		
	>3	14(4.7)		
Education Level	B.S	252(84)	ANOVA	0.267(1.01)
	M.S	26(8.7)		
	M.D	22(7.3)		

Based on the results of Pearson correlation test; There was just a positive correlation between the education level and student satisfaction ($p = 0.049$, $N = 300$, $r = 0.398$), implying that the higher the level of education of students, the more satisfaction they show with virtual education. To evaluate the status of e-learning, five variables of learner individual characteristics, educational content, infrastructure, support, and assessment were used. To determine the relationship between these variables and student satisfaction, the Pearson correlation test was used. The results showed that there was a positive correlation between all variables and student satisfaction, the more capable learners, better educational content, stronger e-learning infrastructure, better support and assessment of e-learning quality, the greater the students' satisfaction. (Table 2).

Table2. Correlation between variables and students' satisfaction.

Variable	Students Satisfaction		
	N	Pearson correlation	Sig.(2-tailed)
Learner Individual Characteristics	300	0.508	0.000
Educational Content	300	0.445	0.000
Infrastructure	300	0.509	0.000
Support	300	0.517	0.000
Evaluation	300	0.569	0.000

4. Discussion

The results of the present study showed that there was no significant difference between the mean of variables such as level of education, and level of student satisfaction. Diab's and Shahzad's study showed that a significant difference between the gender variable and the attitude towards virtual education during the Covid-19 pandemic [5, 9]. In the present study, there was a significant difference between the average experience of using the e-learning system and the level of satisfaction. However, these two variables did not

correlate and in contrast, there was a direct and positive correlation between education level and satisfaction, which was consistent with the results of Diab`s research [5]. Students at ZAUMS were satisfied with the virtual education (Mean=99.10, Std.Dev=16.54) which was offered during the Covid-19 pandemic. They also considered virtual education as favorable and satisfactory. Two factors-i-e- ease of use and usefulness of working with the system, - were effective in students' intention to use virtual education [9]. Therefore, the development of e-learning based on easy and useful tools can be considered as one of the effective strategies of e-learning planners at universities [10]. In conclusion, the results of the present study showed that a set of factors involved in students' satisfaction such as learner ability, infrastructure, support, educational content, .As a result ,evaluation can assess the status of virtual education and can serve as a model in future research. Accordingly, more attention should be paid to educational content and infrastructure improvement in software and hardware, support the virtual program by top management, e-learning programs evaluation. Using the results of this study and considering the effective factors, health informatics specialists will have a valuable consulting role in designing better e-courses as well as tools related to content production and evaluation. To sum up, experiences from the evaluation of e-learning in the Covid-19 pandemic period, may act as a good guide in improving the course during the Covid-19 pandemic, and in the post-Covid-19 period.

Acknowledgments

The present study is adapted from a scholarly student research that was approved in scientific research center in Zahedan University of Medical Sciences with IR.ZAUMS.REC.1399.450 ethical code.

References

[1] Favale T, Soro F, Trevisan M, Drago I, Mellia M. Campus traffic and e-learning during covid-19 pandemic. Computer networks. 2020;176:107290.
[2] Alqahtani AY, Rajkhan AA. E-learning critical success factors during the covid-19 pandemic: A comprehensive analysis of e-learning managerial perspectives. Education sciences. 2020;10(9):216.
[3] Edelhauser E, Lupu-Dima L. Is romania prepared for elearning during the covid-19 pandemic? Sustainability. 2020;12(13):5438
[4] Khan MA, Nabi MK, Khojah M, Tahir M. Students' perception towards e-learning during covid-19 pandemic in india: An empirical study. Sustainability. 2021;13(1):57.
[5] Diab G, Elgahsh NF. E-learning during covid-19 pandemic: Obstacles faced nursing students and its effect on their attitudes while applying it. American Journal of Nursing. 2020;9(4):300-14.
[6] Aboagye E, Yawson JA, Appiah KN. Covid-19 and e-learning: The challenges of students in tertiary institutions. Social Education Research. 2021:1-8.
[7] Olum R, Atulinda L, Kigozi E, Nassozi DR, Mulekwa A, Bongomin F, et al. Medical education and e-learning during covid-19 pandemic: Awareness, attitudes, preferences, and barriers among undergraduate medicine and nursing students at makerere university, uganda. Journal of Medical Education and Curricular Development. 2020;7:2382120520973212.
[8] Ismaili H, Rahmani SH, Kazemi A, Ali Ahmadi M. Evaluation of E-Learning of the virtual learning program from the student's point of view. Journal of Public Management Researches. 2016;9:203-22.
[9] Shahzad A, Hassan R, Aremu AY, Hussain A, Lodhi RN. Effects of covid-19 in e-learning on higher education institution students: The group comparison between male and female. Quality & quantity. 2021;55(3):805-26.
[10] Vuorela M, Nummenmaa L. How undergraduate students meet a new learning environment? Computers in Human Behavior. 2004;20(6):763-77.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210890

Cultural Adaptation and Piloting of iSupport Dementia in Greece

Areti EFTHYMIOU[a,b,1], Nomiki KARPATHIOU[b], Eleni DIMAKOPOULOU[b], Panagiota ZOI[b], Chryssa KARAGIANNI[b], Michalis LAVDAS[b], Anastasios MASTROYIANNAKIS[c], Eirini SIOTI[c], Ilias ZAMPETAKIS[c] and Paraskevi SAKKA[b]

[a] *Quality-of-Life Lab, Department of Social Work, Hellenic Mediterranean University, Heraklion, Crete*

[b] *Athens Association of Alzheimer's Disease and Related Disorders, Athens, Greece*

[c] *CMT Prooptiki, Athens, Greece*

Abstract. The COVID-19 pandemic brought into the spotlight the technological needs of carers together with accessibility and availability of disease-related web-based services. Athens Alzheimer Association undertook the cultural adaptation of the platform iSupport in Greece. The process included multiple methodological steps based on Ecological Validity Model, and the platform was pilot tested to 15 carers of people with dementia. The duration of this project lasted for one year (April 2020-March 2021). Today more than 160 carers in Greece have registered on the platform. All 23 lessons are easily accessible by the isupportdementia-greece website.

Keywords. carers, dementia, eHealth, eLearning

1. Introduction

Carers make multiple decisions every day for themselves and the person with Dementia (PwD). The complexity of carers' needs depends on the type, stage of the disease, carers' knowledge, skills and attitudes towards caregiving and available health services and social network. Carers' reported needs according to the Nuffield Council of Bioethics were financial and social support, access to counselling, preservation of identity, establishing a partnership with the patient and health professionals and the access to confidential information for the cared-for person [1].

The COVID-19 pandemic has strongly influenced the needs of carers and the PwD. In a survey of 204 carers living in Greece during the lockdown in Spring 2020, the family burden has increased due to limited health service provision and access [2]. Web-based services improved carers' well-being, quality of life, alleviate the burden, increase social networking, and are cost-saving [3,4]. A higher level of eHealth literacy among carers is associated with higher self-efficacy in managing behavioural disorders [5]. In Greece, there is a lack of eLearning platforms for carers of people with dementia. The present study aimed to culturally adapt the World Health Organization iSUPPORT

[1] Corresponding Author, Areti Efthymiou, Quality-of-Life Lab, Social Work Department, Hellenic Mediterranean University, Estavromenos, 71410 Heraklion, Crete; E-mail: aefthymiou@hmu.gr.

platform in the Greek language to support the carers with the provision of a tailor-made eLearning course during the period of lockdown in Greece. In this paper, we present the adaptation methodology and focus on the pilot-test of the platform.

2. Methods

2.1. The iSUPPORT platform and the cultural adaptation phases

The WHO developed an eLearning platform for carers of people with dementia in 2019 which was updated during the pandemic. The platform included five modules with 23 lessons in total: module 1-Introduction of dementia (lesson one), module 2-Being a caregiver (four lessons), module 3-Caring for me (three lessons), module 4-Providing everyday care (five lessons), module 5- Dealing with cognitive, psychological, and behavioural challenges (ten lessons). Athens Alzheimer association was granted by the WHO the access to adapt the platform in the Greek language culturally.
The methodology followed was based on the Ecological Validity Model [6], focusing on the adaptation of the content (language used, persons, context, names, concepts) and on the cultural adaptation process used by Teles et al [7] for the Portuguese adaptation.

A survey was organized to identify the training needs of carers in Greece. The modules were officially translated (May to June 2020), the research team checked the final translated deliverable and made minor changes. The contents were also evaluated by an expert panel of six experts working in dementia and a small group of four carers. The six experts worked in parallel and met in person to reach a consensus. After integrating experts' and carers comments, all changes were reported to WHO. At the same time, the application programming interface (API) was being developed, and the platform was customized by the contracted organization, using AngularJS, TypeScript and MongoDB. The IT representatives participated as part of the expert panel in all meetings (Figure 1).

The cultural adaptation lasted for 12 months (from April 2020 to March 2021).

1.	Descriptive study of carers' web-based need
2.	Official translation of the iSupport Modules
3.	Translation check by the research team
4.	Evaluation by an expert panel
5.	Evaluation by a group of carers
6.	Fidelity check by the organisation
7.	Final adaptation of the contents
8.	API development & platform customization
9.	Pilot test of the iSupport platform
10.	Expert meeting to discuss results and feedback by carers
11.	Update of the iSupport Platform

Figure 1. Methodological process of the cultural adaptation.

2.2. Piloting methodology

The piloting methodology followed a qualitative design with four phases: 1) the platform was presented by the research team to a group of 15 carers (pre-assessment

phase) 2) the carers registered in the platform and used the platform according to their needs for one month (January 2021), 3) Two carers' online focus groups were organized after the end of one month (post-assessment phase). The online focus group meetings were recorded, and two researchers transcribed and analysed the discussions. All carers who participated in the focus groups received a questionnaire guide in the pre-assessment phase, 4) a consensus meeting among the experts was held to finalize the platform.

2.3. Pilot-test participants

A convenience sample of 15 carers participated in the pre-assessment phase; 94% (N=14/15) women, with a mean age of 50 years old, 60% (N=9/15) children caring for their parents most of them highly educated (tertiary education 67%, N= 10/15). They had attended services of the dementia center previously.

Two online focus groups were organized for the post-assessment. Of the 16 carers, 8 (50%) agreed to participate in the discussions. The reason for drop-out for most carers was the lack of time due to the Holiday season (Christmas Period).

2.4. Ethics

The Scientific Board approved the study of the Athens Alzheimer Association. Carers were informed about the aims of the study and provided their written consent. Researchers supported carers during the pilot-phase by email and telephone if carers had queries with regards to the platform.

3. Results

3.1. Experts and carers comments on the original translation

During the evaluation of the contents by the expert panel and the group of carers, the core issues raised concerned: words and expressions, the clarity and the precision of the concepts and the scenarios, titles in Greek, available resources, and cultural adaptation of habits within the scenarios.

The group of carers mostly commented on the scenarios. In total, there were 60 scenarios in all 23 sections. Carers required more Greek related information on the resources, more in-depth information for the severe stages of dementia and more scenarios in some behavioural disorders.

3.2. Focus group results

Four core themes emerged from the discussion content analysis:

- **Face to face interaction and sensory problems**. Carers admitted that they did not always prefer online reading, and they preferred face-to-face interaction. They felt tired reading on the screen. They would prefer a downloadable option

- **Interactive aspects of the platform**. They preferred an interactive version of the platform with videos and a forum where they would receive feedback from health professionals.
- **Navigation issues**. Carers suggested navigation changes to improve the user-friendliness of the platform.
- **Cultural adapted content**. New content was suggested to be included; a section for the communication with the paid carer, health services in Greece, legal and financial issues, how to treat severe dementia

4. Discussion

The iSupport Dementia for Greece was officially launched in March 2021 on the Greek Carers' Day (https://isupportdementia-greece.gr). Experts, carers and IT developers participated from the beginning of the adaptation process. Most of the proposed modifications have been considered providing a user-friendly platform easily accessible by a basic internet user. Before the pandemic one in two carers used eLearning services even if there was a lack of dementia-specific courses [8]. During the pandemic, iSupport eLearning course is considered a valuable tool for carers and healthcare professionals. Future research could assess the impact of the iSupport platform on caregiving variables, digital and eHealth literacy skills.

References

[1] Nuffield Council of Bioethics. The needs of carers. In: Dementia: ethical issues. 2009. p. 113–25.
[2] Tsapanou A, Papatriantafyllou JD, Yiannopoulou K, Sali D, Kalligerou F, Ntanasi E, et al. The impact of COVID-19 pandemic on people with mild cognitive impairment/dementia and on their caregivers. International Journal of Geriatric Psychiatry. 2021 Apr 1;36(4):583–7.
[3] . Dam AEH, de Vugt ME, Klinkenberg IPM, Verhey FRJ, van Boxtel MPJ. Review article: A systematic review of social support interventions for caregivers of people with dementia: Are they doing what they promise? [Internet]. Vols. 85 OP-I. 2016. p. 117.
[4] Hopwood J, Walker N, McDonagh L, Rait G, Walters K, Iliffe S, et al. Internet-based interventions aimed at supporting family caregivers of people with dementia: Systematic review. Journal of Medical Internet Research. 2018;20(6).
[5] Efthymiou A, Middleton N, Charalambous A, Papastavrou E. Health literacy and eHealth literacy and their association with other caring concepts among carers of people with dementia: A descriptive correlational study. Health and Social Care in the Community. 2021;00:1–11.
[6] Bernal G, Bonilla J, Bellido C. Ecological validity and cultural sensitivity for outcome research: Issues for the cultural adaptation and development of psychosocial treatments with Hispanics. Journal of Abnormal Child Psychology. 1995;23(1):67–82.
[7] Teles S, Napolskij MS, Paúl C, Ferreira A, Seeher K. Training and support for caregivers of people with dementia: The process of culturally adapting the World Health Organization iSupport programme to Portugal. Dementia. 2021 Feb 1;20(2):672–97.
[8] Efthymiou A, Papastavrou E, Middleton N, Markatou A, Sakka P. How caregivers of people with dementia search for dementia-specific information on the internet: Survey study. JMIR Aging 2020;3(1):e15480

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210891

Continuing Education of Nurses in Patient Handovers: Development and Evaluation of a Digitally Enabled Problem-Based Learning Course

Nicole EGBERT[a,1], Birgit BABITSCH[b] and Ursula HÜBNER[a]
[a] *Health Informatics Research Group, Osnabrück University AS, Germany*
[b] *New Public Health, Osnabrück University, Germany*

Abstract. Communication deficits belong to the most frequent errors in patient handovers calling upon specialized training approaches to be implemented. This study aims to harness problem-based learning (PBL) methods in handover education and evaluated the learning process. A digitally enabled PBL course was developed and implemented at Klinikum Osnabrück from which eight nurses participated in the course. They agreed on the stimulating effect of the setting regarding self-directed learning and on the potential to translate the new knowledge and skills into the daily clinical practice. In conclusion, the findings are promising that a digitally enabled PBL course is a suitable learning format for handover education.

Keywords. Education, nursing, handover, problem-based learning, online course

1. Introduction

Patient handovers and handoffs require a timely and proper transmission of nursing and medical information from the out-going to the incoming shift. Without such information transfer, patient safety is jeopardized [1]. Communication deficits belong to the most frequent errors that account to up to 80% of the critical incidents [2,3]. While it is known that handover trainings [4] can mitigate this situation, there is still a lack of appropriate opportunities during undergraduate, postgraduate and continuing education [5]. Students benefit from learning handovers that includes active elements such as role playing, simulation [5], real-world scenarios [6] or case-based learning [7].

One of the most well-known pedagogical methods for activating the learners is problem-based learning (PBL). Employing realistic case descriptions, so called case vignettes, it is a method that highlights the access to acquiring practically relevant knowledge and skills [8]. In the ideal case, PBL should enable the learners to take over the responsibility for the learning process [9]. PBL has been augmented towards its application in electronic learning environments in the last decade [10].

Despite its activating character, the realistic scenarios it relates to and its dedicated focus on working in teams, PBL has not been applied for teaching handovers or hand-

[1] Corresponding Author, Nicole Egbert, Health Informatics Research Group, Osnabrück University of Applied Sciences, Albrechtstr. 30, 49076 Osnabrück, Germany; E-mail: n.egbert@hs-osnabrueck.de.

offs so far. It therefore seems worthwhile to harness PBL in handover trainings in particular for continuing education and evaluate the learning process. As training adults often is supported by digital elements or is entirely performed online, digitally enabled PBL should be the method of choice to be tested. The aim of this study, therefore, was to describe the development and the evaluation of a handover course in continuing education using the digitally enabled PBL approach. The research questions are as follows: 1) To what extent was this course suitable to foster active learning? 2) Did the course elements refer to clinically relevant situations and to what extent did the course help to transfer the new knowledge into practice?

2. Methods

The course development including the one of the case vignette was based on prior own work that helped to identify communication errors as the primary critical safety issue in patient handover via a CIRS analysis [11]. The case vignette thus reflected a face-to-face patient handover between two nursing shifts with typical communication deficits.

The overall learning objective of the course was that, upon completion of the course, participants will have knowledge and communication skills that will enable them to ensure information continuity in patient handovers. Hereby, the aim is to improve patient safety proactively identifying and avoiding potential harm. The course was structured into three phases following the 7 steps of PBL (Fig. 1).

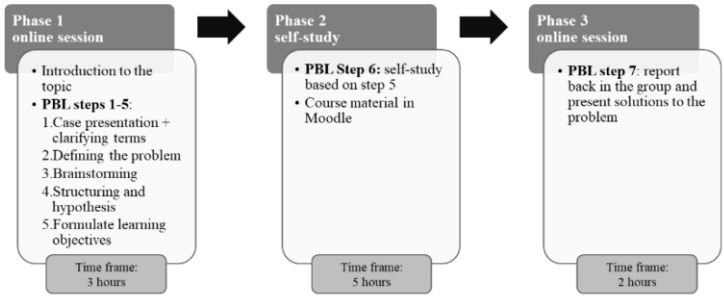


Figure 1. Course timeline

All in all, it took 10 hours and included two synchronous online sessions together with two tutors (phase 1 and 3) and one asynchronous self-study phase (phase 2). The online sessions were realized via Zoom and were recorded, the course content was published via the learning management platform Moodle. During the first online phase the students had to work on the definition of the problems in the case vignette, the collection and systematic compilation of ideas and solutions, and the identification of the learning objectives. The results of the group were documented on the digital pin wall Padlet®. The course material for the self-study phase (phase 2) consisted of texts, graphics, videos, tests and quizzes as well as assignments to reflect the topics patient safety, high reliability organisations, patient handovers, communication, structuring the handover information and electronic systems to support handovers. During this phase, students worked on their own. The second online session was used to report back to the group and present solutions to the problems.

The course was held at Klinikum Osnabrück, a 768 beds municipal hospital, involving 8 registered nurses in leading positions (6 females, 2 males, one person was

younger than 30 years, three between 30 and 44 years and three persons older than 44). It took place between August 26th and September 16th, 2020 allowing three weeks for the self-study phase. The learning process was evaluated in a pre-post design, whereby the pre-questionnaire addressed amongst others previous experience, and prior knowledge. The post-questionnaire consisted of questions concerning the learning experience in the PBL course and the translation of the knowledge into practice. Some of the questions were identical to allow for a before-after comparison.

3. Results

The analysis of the pre-questionnaire indicated that there was a wide variety of prior experience regarding eLearning, case vignettes and PBL from "not at all" to "a large extent". The course participants had either none or experience "to a medium extent" with learning management platforms. The analysis of the post-questionnaire revealed that the participants agreed either "rather" or "fully" to statements concerning the relevance of the case vignette, the stimulating effect of the learning setting regarding self-study and obtaining multiple perspectives (Fig. 2).

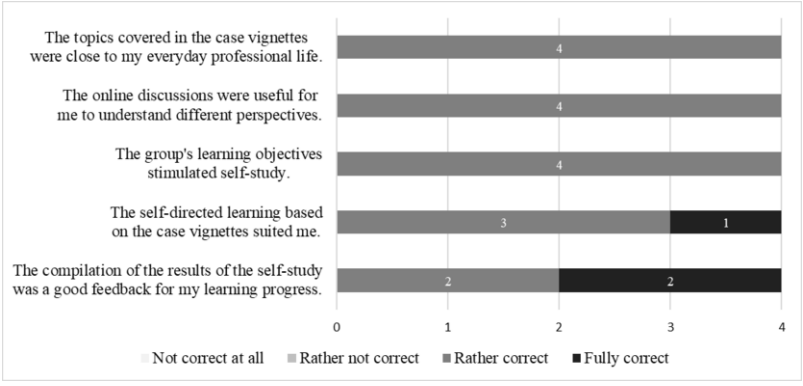


Figure 2. Evaluation of the problem-based learning setting (post-questionnaire) (n=4)

Alike, the participants concurred with the statements about the translation of the new knowledge gained in this course into the daily clinical practice (Fig. 3). Due to organizational problems not all course participants submitted the questionnaires. However, all participants gave feedback stating that the use of the learning platform was easy and helpful as was expressed by one of the persons "The course structure was very intuitive. It was very useful for me to watch the recordings of the online sessions." Also, all participants agreed that taking this course was compatible with their work.

4. Discussion

The present study demonstrated the feasibility and usefulness of a digitally enabled PBL course in continuing education. Despite the heterogeneity of the prior experience in eLearning, it was generally agreed that using the electronic learning environment was easy and provided added value. In accordance with the literature [8], the PBL course activated the students to engage in self-directed learning. Due to the perceived practical

relevance of the case vignette, the students expected the new knowledge and skills to be sustainable and transferable into practice.

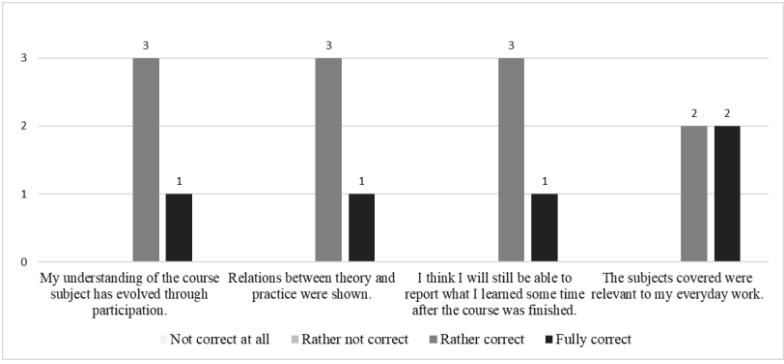


Figure 3. Evaluation of translation of knowledge into practice (post-questionnaire) (n=4)

The interpretation of the findings suffers from the fact that not all participants took part in the post-questionnaire. However, the final feedback including all participants yielded a positive impression. In conclusion, the findings are promising that a digitally enabled PBL course is a suitable learning format in continuing education and for training handovers particularly. Additional implementations of the course would be useful for further improvement and development of the course.

Acknowledgement

This study was supported by a grant from the VolkswagenStiftung Hannover, Germany (grant no. ZN3123). We wish to thank Klinikum Osnabrück for their interest and support.

References

[1] Australian Council for Safety and Quality in Health Care. Clinical Handover and Patient Safety Literature Review Report. March 2005.
[2] Flemming D, Hübner U. How to improve change of shift handovers and collaborative grounding and what role does the electronic patient record system play? Results of a systematic literature review. Int J Med Inform. 2013 Jul;82(7):580-592.
[3] Galatzan BJ, Carrington JM. Exploring the State of the Science of the Nursing Hand-off Communication. Comput Inform Nurs. 2018 Oct;36(10):484-493.
[4] Halm MA. Nursing handoffs: ensuring safe passage for patients. Am J Crit Care. 2013;22(2):158-162.
[5] Gordon M, Hill E, Stojan JN, Daniel M. Educational Interventions to Improve Handover in Health Care: An Updated Systematic Review. Acad Med. 2018 Aug;93(8):1234-1244.
[6] Riesenberg LA, Leitzsch J, Cunningham JM. Nursing Handoffs: A Systematic Review of the Literature. Am J Nurs. 2010 Apr;110(4):24-34.
[7] Klaber RE, Macdougall CF. Maximising learning opportunities in handover. Arch Dis Child Educ Pract Ed. 2009 Aug;94(4):118-122.
[8] Bligh J. Problem-based learning in medicine: an introduction. Postgrad Med J. 1995;71:323–326.
[9] de Grave W, Moust J, Hommes J. The role of the tutor in a problem based learning curriculum. Maastricht: Maastricht University; 2003.
[10] Saqr M, Nouri J, Vartiainen H, Malmberg J. What makes an online problem-based group successful? A learning analytics study using social network analysis. BMC Med Educ. 2020 Mar 18;20(1):80.
[11] Egbert N, Babitsch B, Hübner U. Learning from CIRS to optimise patient safety in handovers. GMS Med Inform Biom Epidemiol. 2020;16(3):Doc10. DOI: 10.3205/mibe000213.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210892

Multilingual Approach to COVID-19 Online Learning Response on OpenWHO.org

Heini UTUNEN[a,1], Ngouille NDIAYE[a], Melissa ATTIAS[a], Lama MATTAR[a], Anna TOKAR[a] and Gaya GAMHEWAGE[a]
[a] *World Health Organization, Geneva, Switzerland*

Abstract. In pursuit of equitable access to emergency-related knowledge, the World Health Organization (WHO) translates COVID-19 and other infectious disease courses into multiple languages on its open-access online learning platform OpenWHO.org. Languages spoken by vulnerable or underserved populations in low- and middle-income countries and in outbreak-prone and affected areas are prioritized. Accessing learning in preferred languages enhances uptake and comprehension. In this study, we assess and compare the initial enrollment levels and global reach of these multilingual courses. On average, OpenWHO's 38 COVID-19 courses have each been translated into 4.8 languages. The platform hosts courses in 55 different languages with 10.4 million words translated. The findings identify which available languages were most utilized for COVID-19 learning to inform course production and outreach strategies. Languages were used differently across geographic regions, calling for localized learning offerings. A streamlined multilingual publishing scheme, ensuring quick and effective delivery of diverse languages, is critical to achieving greater equity of access to knowledge.

Keywords. OpenWHO, multilingual, online learning, COVID-19 pandemic

1. Introduction

OpenWHO.org is an open-access, low-bandwidth online learning platform created by the World Health Organization's Health Emergencies Programme (WHE) in 2017 to democratize global access to trusted public health knowledge. The platform is designed to support preparedness and response to outbreaks [1]. OpenWHO has delivered critical learning content during the pandemic and ensured broader equity through courses in 55 languages, including the official languages of every WHO region, the 15 most-spoken languages worldwide, 14 African languages and the official languages of 42 out of 46 of the least-developed countries.

WHE has been producing materials for emergencies since the West Africa Ebola outbreak. The solutions and services put into practice include local language translation capacity, unlimited online dissemination to the front lines of emergencies through a dedicated low-bandwidth platform, and a process to turn evidence-based, emergency guidelines into knowledge resources for responders. The World Health Assembly (2019) was convinced of the importance of respecting the diversity of cultures and the plurality

[1] Corresponding Author, Heini Utunen; E-mail: utunenh@who.int.

of international languages. This was echoed in the United Nations Joint Inspection Unit report on multilingualism (2020) [2].

Comprehension can be greatly improved when using the appropriate languages and formats [3]. Learning in one's preferred language enhances learning effectiveness [4]. This is particularly true for fast-changing epidemic and outbreak contexts, as well as in humanitarian settings with multiple health concerns for affected populations [5].

This paper presents the multilingual approach pioneered by WHO's OpenWHO.org learning platform for health emergencies and the COVID-19 pandemic. It examines which languages are most utilized for COVID-19 learning and language use patterns by geographic region to inform and strengthen the platform's approach.

2. Methods

Data were extracted from the built-in OpenWHO.org reporting system and a descriptive analysis was conducted to assess early course traction by language and geographic spread from 26 January 2020 to 26 August 2021. OpenWHO user metrics included self-reported data on location, gender, language, age and affiliation of users, which may be utilized for research under the platform terms of use. As this study was deemed 'non-invasive', ethical clearance was not required.

Platform and course-based data were integrated to depict global and local use of COVID-19 courses by language. Course registrations were segregated by language used to access the course and by geographic location mapped according to WHO region. The first four weeks of course availability were targeted to establish a comparable baseline, as course topics and language versions are published at different times. To account for variations in language availability per course topic, each language's total enrollment count was divided by the number of topics available in that language.

3. Results

Multilingualism in OpenWHO's large-scale online production strategy has resulted in more than 5.6 million course enrollments driven largely by 38 COVID-19 courses, with 2.9 million certificates awarded for course completion or achievement. Courses are available in 55 languages, with a total of 10.4 million words translated thus far, hosted across 113 course topics. On average, OpenWHO's COVID-19 courses have each been translated into 4.8 languages, with the Introduction to COVID-19 course available in 44 languages and the Infection Prevention and Control course in 24 languages.

The top languages by enrollment are English (76.2%), Spanish (13.0%), French (3.1%), Arabic (1.5%), Portuguese (1.5%), Indian sign language (1.0%), Hindi (0.7%), Indonesian (0.6%), Russian (0.5%) and Italian (0.4%). All other languages combined constitute 1.5% of use.

English language course use varies significantly by WHO region. The WHO Region of the Americas is characterized by Spanish (48.0%), English (46.7%) and Portuguese (2.8%) use. Enrollments in the English learning resources for other WHO regions are as follows: European (69.2%), Eastern Mediterranean (84.1%), South-East Asia (88.2%), African (90.6%) and Western Pacific (94.6%). Spanish is the second most popular language among OpenWHO learners, with 14.5% of all enrollments in Spanish-language courses.

Average enrollments per course by language in the first four weeks following course launch are shown in Table 1.

Table 1. Top languages by first four-week average enrollments per COVID-19 course.

Languages	Enrollments	Languages	Enrollments
English	26327	Chinese	783
Spanish	12628	Bengali	661
French	7510	Turkish	503
Portuguese	3878	Serbian	417
Arabic	2052	Japanese	380
Indian Sign	1877	Persian	359
Hindi	1862	Urdu	352
Indonesian	1667	Hungarian	328
Russian	1151	Oriya	277
Italian	878	German	269

4. Discussion

The OpenWHO platform ramped up learning production for the pandemic, deploying record-setting crowdsourcing power in its efforts to deliver life-saving knowledge with an emphasis on multilingualism to cater to the demands of a massive global audience. Examining the initial enrollment surge by language in the first four weeks after course launch helps account for the distribution of courses across languages – courses are published more often in some languages than others – and variations in course publication date, as courses published earlier have more overall enrollments. It also highlights the traction of languages such as Indian sign language with notable levels of popularity. The importance of national and local language resources far exceeds the use percentage in terms of providing equitable access to learning for the most vulnerable and contributing to co-ownership.

The results inform OpenWHO's multilingual approach by suggesting which languages are in highest demand for COVID-19 learning for future course production, as well as which languages should be targeted for outreach and promotion to reach additional learners. Geographically, the study shows that English courses are used in a much lower proportion in the Americas and Europe and at a much higher rate in Africa and the Western Pacific This suggests that research into the drivers of these regional variations, whether unmet language needs or limited platform reach, could further strengthen the multilingual production strategy. Overall, the importance of Spanish among learners has implications for learning production and confirms the urgency of timely Spanish translation. The popularity of Indian sign language, the seventh most popular language on the platform, illustrates the demand for learning resources by special needs audiences.

The scalability of OpenWHO's response was achieved through a fast-paced production system with strengthened commitment to equity of access. This was bolstered through local and co-ownership, as WHO country offices, health institutes, individuals and other volunteers translated the materials to meet their needs and further adapted them for field use in plain language and other localized formats. A multiplier effect has occurred with national language materials in particular, as the translated and adapted learning resources have been used extensively outside of the platform context.

One limitation of the data is that the English version of a course is launched first, likely contributing to the higher uptake in the first four weeks as it may be the only language version available. This might have led to an unequal distribution of enrollments across languages. Nevertheless, the findings testify to the importance of multilingualism and inform the translation planning process to support global preparedness and response with wider equity in access to trusted public health knowledge.

5. Conclusion

WHO's health emergency online learning platform supports global COVID-19 preparedness and response while seeking enhanced health literacy through multilingualism. The translation production system has successfully scaled up to meet the global demand for learning during the pandemic and can be further refined based on usage patterns. Additional work is required to study comprehension and assess communication preferences across affected and at-risk populations. This area continues to evolve to support more efficient and equitable access to a life-saving knowledge platform.

References

[1] Utunen H, Attias M, George R, Ndiaye N, Piroux C, Farzi MR, Gamhewage G. Global Access to OpenWHO's Online Learning Resources for COVID-19. Stud Health Technol Inform. 2020 Jun;272:304-305.DOI:10.3233/shti200555.

[2] Lozinskiy N. Report of the Joint Inspection Unit. Multilingualism in the United Nations system. United Nations. 2020 Jun; Geneva, Switzerland: The United Nation.

[3] Nwokediuko AA. Advancing e-learning in African native communities: The language factor. International Conference on Education and e-Learning Innovations. 2012 Jul;1-4. DOI: 10.1109/ICEELI.2012.6360594.

[4] Watkins PG, Razee H, Richters J. (2012). 'I'm Telling You … The Language Barrier is the Most, the Biggest Challenge': Barriers to Education among Karen Refugee Women in Australia'. Australian Journal of Education. 2012 Aug;56(2).

[5] Finset A, Bosworth H, Butow P, Gulbrandsen P, Hulsman RL, Pieterse AH, Street R. Tschoetschel R, van Weert J. Effective health communication – a key factor in fighting the COVID-19 pandemic. Patient Education and Counseling. 2020 May;103(5):873–876.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210893

Could Linguistic Complexity Be Automatically Evaluated? A Multilingual Study on WHO's Emergency Learning Platform

Giuseppe SAMO[a,1], Yu ZHAO[b], Maria Teresa GUASTI[c], Heini UTUNEN[b], Oliver STUCKE[b] and Gaya GAMHEWAGE[b]
[a] *Beijing Language and Culture University*
[b] *World Health Organization*
[c] *University of Milano-Bicocca*

Abstract. The ability of assessing any type of linguistic complexity of any given contents could potentially improve knowledge reproduction, especially tacit knowledge which can be expensive during a pandemic. In this paper, we develop a simple and crosslinguistic model of complexity which considers formal accounts on the study of linguistic systems, but can be easily implemented by non-linguists' groups, e.g., communication experts and policymakers. To test our model, we conduct a study on a corpus extracted from the World Health Organization (WHO)'s emergency learning platform in 6 languages. Data extracted from open-access encyclopaedic entries act as control groups. The results show that the measurements adopted signal a trend for a minimization of complexity and can be exploited as features for (automatic) text classification.

Keywords. Digital learning; Linguistic complexity; Digital Health; Emergency training; Knowledge Reproduction

1. Introduction

Digital technologies revolutionized the way we retrieve information and acquire knowledge [1]. The recent COVID-19 pandemic has increased the demand for reliable information to help frontline health personnel respond to the outbreak, communities better protect themselves, and health policymakers draft new policies to accommodate emerging needs [2]. Being capable to apply distance learning became instrumental for health workers, who demonstrate different degrees of accessibility to content due to multilingual background, especially for non-native speakers. Under such circumstances, content providers are required to facilitate teaching and/or training activities in the context of emergencies, possibly in a variety of languages [3], which could be fairly time-consuming and labour intensive. Having the ability to assess the linguistic

[1] Corresponding Author, Giuseppe Samo, Department of Linguistics, Beijing Language and Culture University, Mailbox 82, Xue Yuan Road 15, 100083 Beijing, People's Republic of China; E-mail: samo@blcu.edu.cn.

complexity of any given contents in advance could potentially optimize resources and enabling the learning contents to be more readable/accessible [4].

Although indices of readability do exist, they are mainly language-specific (see [5] for an overview) and the language-specificity is not efficient to measure such indices across multiple languages. Moreover, many indices are also very coarse, generally on mean length of words and sentences.

Is there a way for us to quantify linguistic complexity from a crosslinguistic perspective? How can one observe and interpret trends of complexity, even without technical background in formal accounts of linguistics?

In this paper, we aim to discuss a simplified measurement of linguistic complexity that takes into account a simple intuition from the study of language in theoretical linguistics, psycholinguistics, computational linguistics, yet simultaneously appear to be easily implemented by non-linguists' group e.g., communication experts and policymakers.

The primary objective of this paper intends to operate on small-sized corpora, which are often under-investigated by the linguistic community, yet reveals critical value for under-resourced languages. Iterated from previous studies, we propose a model, an improved version from origin discussed and developed in [3] and [4], to minimize manual investigation for assessing complexity in health-related context.

We then continue discussing complexity measures in Section 2, and present an observational study in section 3, by comparing an emergency corpus from OpenWHO.org, World Health Organization (WHO)'s emergency learning platform with similar content from open-access encyclopaedic entries in six languages. In Section 4 we discuss and explore directions for further in-depth research, followed by a conclusion.

2. A simple model to measure linguistic complexity.

One of the objectives in theoretical and experimental linguistics is to investigate complexity in language comprehension and production, in adult and development grammars, and in language pathology (see [6] for an introduction and overview).

In experimental linguistics, different linguistic tasks are created to evaluate whether specific linguistic architectures (e.g., specific word orders, specific grammatical structures) are harder than others. Various types of measures such as reading (parsing) times, neuronal activity or, in specific populations, performances in specific tests (e.g., picture/scenario matching, see [7]) are defined for evaluation.

The evaluation results from experimental linguistics confirm a very basic idea originated from formal models of language: there are structures (e.g., a specific configuration of a sentence which opens to grammatical re-orderings of syntactic constituents) that are more complex to be parsed by human mind than others. Many of these syntactic re-orderings (e.g., relative clauses, topicalizations, cleft structures), which are also complicated for machine learning architectures in deep learning (see [8] for an overview), are cued by functional elements.

In Zipf's distributions [9, 10], functional elements have the tendency to be more frequent in corpora than lexical entries. The distribution of functional elements can be then detected by measures such as Gini's coefficient and Shannon *H* entropy (see 10 [for] a detailed discussion). We will not discuss elements Type/Token ratio (TTR), as content-specific text (e.g., medical information) require also the occurrences of technical terms. The role of measurements such as TTR, however, could be investigated in future studies.

Our main hypothesis is that in less complex texts, functional elements should be minimized, leading to an expected lower Gini's coefficient and higher Shannon entropy correlates if a text of the same size and similar TTR is compared.

3. A crosslinguistic study of learning contents from OpenWHO.org

We test our hypothesis on an emergency learning corpus, in which the materials have been well studied (see [3] and [4]) and labelled as less complex to other corpora (encyclopaedic entries, legal, news and social media) *via* a manual linguistic analysis on a typology of functional elements. All the materials (languages and size of the corpora are to be found in Table 1) are extracted from OpenWHO.org [11]. The dataset from the test group is extracted from the course *Infection Prevention and Control (IPC) for COVID-19 Virus* (downloaded in May 2020) in six languages: English, French, Italian, Polish, Portuguese and Spanish. For the control group, due to the transferring-knowledge nature of the dataset (see the discussion in [3]), we collected content from open-encyclopaedic entries (Wikipedia) on Covid-19 in the relevant six languages.[2] To avoid bias on the measures (see [10]), we extracted similar number of tokens, guaranteeing a readability of the content.

We compiled all the relevant textual contents in a *.txt* file and uploaded them to Zipfsexplorer (https://zipfexplorer.herokuapp.com/zipfexplorer, [10]). The online tool provides relevant measures discussed in Table 1.

The results in Table 1 support our hypothesis. All the Shannon's entropy measures are higher, while all the Gini's coefficients are equal or lower, as predicted by the hypothesis.

Table 1. Languages, Set (Wikipedia, OpenWHO.org), size, TTR (Type/Token ratio), Gini's coefficient and Shannon *H* entropy for every corpus under investigation (https://zipfexplorer.herokuapp.com/zipfexplorer. 08/2021). Results in bold confirm our hypothesis.

Language	Set	Size (Tokens)	TTR	Gini	*H*
English	Wiki	5,621	0.14	0.70	3.34
English	OpenWHO	5,087	0.15	**0.69**	**3.36**
French	Wiki	6,845	0.12	0.76	3.15
French	OpenWHO	6,886	0.11	**0.74**	**3.48**
Italian	Wiki	6,713	0.14	0.73	3.18
Italian	OpenWHO	6,646	0.12	**0.71**	**3.56**
Polish	Wiki	6,076	0.19	0.65	3.08
Polish	OpenWHO	5,994	0.18	0.65	**3.13**
Portuguese	Wiki	6,582	0.12	0.73	3.44
Portuguese	OpenWHO	6,591	0.12	0.73	**3.46**
Spanish	Wiki	7,960	0.10	0.78	3.36
Spanish	OpenWHO	8,042	0.10	0.78	**3.44**

[2] The control group data are extracted from the relevant URLs (last access, August 2021):
English: https://en.wikipedia.org/wiki/COVID-19;
French: https://fr.wikipedia.org/wiki/Maladie_%C3%A0_coronavirus_2019;
Italian: https://it.wikipedia.org/wiki/COVID-19; Polish: https://pl.wikipedia.org/wiki/COVID-19;
Portuguese: https://pt.wikipedia.org/wiki/COVID-19; Spanish https://es.wikipedia.org/wiki/COVID-19.

4. Conclusions

To conclude, we attempt to evaluate linguistic complexity of digital learning contents from OpenWHO.org, WHO's health emergency learning platform, across multiple languages.

We decide to apply an improved model to largely reduce manual investigation and provide possibility for non-linguistic experts to independently measure the complexity of health information prior to translation and information retrieval for training or monitoring purpose.

Finally, findings in this paper reveal potential possibility of further studies in different language families to test the reliability of such hypothesis. If adequately demonstrated, such model might be applied and developed as an additional feature for automatic text classification or text detection, contributing to a better machine learning performance in information retrieval and machine translation.

References

[1] Foray D. Economics of knowledge. Cambridge, MA: MIT press; 2004. 287 p.
[2] Gamhewage G, Utunen H, Attias M, George R. Fast-tracking WHO's COVID-19 technical guidance to training for the frontline. Weekly Epidemiological Record. 2020, Jun;95: 257-64.
[3] Samo G, Zhao Y, Gamhewage G. Syntactic Complexity of Learning Content in Italian for COVID-19 Frontline Responders: A Study on WHO's Emergency Learning Platform. Verbum. 2020, Dec; 11:1-4.
[4] Zhao Y, Samo G, Utunen H, Stucke O, Gamhewage G. Evaluating Complexity of Digital Learning in a Multilingual Context: A Cross-Linguistic Study on WHO's Emergency Learning Platform. Stud Health Technol Inform. 2021, May; 281: 516-7.
[5] Dell'Orletta F, Montemagni S, Venturi G, READ-IT: assessing readability of Italian texts with a view to text simplification. In: Alm N, editor. SLPAT '11 Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies, 2011 Jul 30; Edinburgh, UK. Stroudsbourg, PA, USA. p. 73-83.
[6] Guasti MT. Language acquisition: The growth of grammar. Cambridge, MA: MIT press; 2017. 672 p.
[7] Friedmann N, Belletti A, Rizzi L. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. Lingua. 2009, Jan; 119: 67-88.
[8] Linzen T, Baroni M. Syntactic structure from deep learning. Annual Review of Linguistics. 2021, Jan; 7: 195-212.
[9] Zipf GK. Human Behavior and the Principle of Least Effort. Cambridge, MA: Addison-Wesley Press; 1949. pp. 573
[10] Coats S. Comparing word frequencies and lexical diversity with the ZipfExplorer tool. In; Reinsone S, Skadina I, Baklāne A, Daugavieti J, editors. Proceedings of the 5th Digital Humanities in the Nordic Countries Conference; 2020 Oct 21-23; Riga, Latvia. Aachen, Germany: CEUR, c2020. p. 219-25.
[11] Rohloff T, Utunen H, Renz J, Zhao Y, Gamhewage G, Meinel C. OpenWHO: Integrating Online Knowledge Transfer into Health Emergency Response. In: Dimitrova V, Praharaj S, Fominykh M, Drachsler H, editors. EC-TEL Practitioner Proceedings 2018: 13th European Conference on Technology Enhanced Learning; 2018 Sep 3-6; Leeds, UK, CEUR-WS, c2018. p. 1-14.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210894

User-Based Evaluation of a Data-Driven Medical Education Platform

Dimitrios ZIKOS[a,1] and Cassie E. SPRAGUE[a]
[a] *Central Michigan University*

Abstract. TrainCoMorb is an online data-driven training platform for medical students and residents who can practice recognizing comorbidities and their attributable risk for negative hospital outcomes. This is a subsequent cross-sectional evaluation study designed to examine four dimensions of the platform (navigation, usefulness, validity, features) and their association with external factors (age, experience in simulation systems, opinion about data-driven education). Eighteen medical residents participated in a scenario-based evaluation session and completed an online survey. The residents evaluated the four composite dimensions with scores between 3.77 and 4.15 (5-scale) and thought highly of data-driven medical education. Those more familiar with clinical simulation systems, and more positive about data-driven education, evaluated the "usefulness", "validity", and "features" dimensions with higher scores. TrainComorb is intended to be a supplementary tool for the education of future physicians, and this user-based evaluation study provided positive feedback that it could serve its intended scope.

Keywords. evaluation, health informatics, medical education, data-driven

1. Introduction

TrainCoMorb (Training with Comorbidities) is an online data-driven training platform for medical students and residents who can practice recognizing comorbidities and their attributable risk for negative outcomes. A comorbidity is the presence of one or more diseases co-occurring with a principal diagnosis (D_x) [1]. The authors of this study led the development of *TrainCoMorb,* which was completed in the year 2020 as a joint work between the College of Health Sciences and College of Medicine, at the Central Michigan University (CMU). With the use of *TrainCoMorb*, trainees (medical students and residents) can create scenarios of comorbidities in a computer assisted manner, to have at their disposal dynamically updated realistic snapshots of patient safety outcomes (hospital mortality, hospital acquired conditions) and the length of stay. The backend of the system is a relational database with 11 million inpatient records. The link to the app is: https://floating-caverns-75251.herokuapp.com. The system design and its clinical significance are explained in [2].

This present paper describes a subsequent study that was designed to evaluate the *TrainComorb* platform by the intended target audience (medical residents) and to examine several external factors associated with their perceptions about the platform.

[1] Corresponding Author, Dimitrios Zikos, Central Michigan University, Rowe Hall 208c, Mt. Pleasant MI 48858, United States; E-mail: zikos1d@cmich.edu.

2. Methods

2.1. Survey design and deployment

The survey included 17 questions which explored the following dimensions: (i) System Navigation, (ii) Usefulness, (iii) Validity of information, (iv) System features. These are typical evaluation components that are examined in health informatics research [3] and in software evaluation [4]. The residents were also asked to provide their opinion about data-driven medical education. The questions used a 5-category Likert scale (Strongly Agree…Strongly Disagree). The survey also asked about the experience with use of computers, exposure with medical simulation systems, and their demographics.

After the study design protocol received approval by the Institutional Review Board, the director of Internal Medicine, Central Michigan University, made a formal invitation and eighteen internal medicine residents agreed to participate. Each resident was compensated with the amount $100 upon participation. The principal investigator of the study was not disclosed to the participants to avoid authority bias. After a 15-minute introduction of the platform, the residents were asked to complete four scenarios where they would create, step-by-step, patient comorbidity profiles (Table 1). Upon completion of the four scenarios, the participants were asked to follow a SurveyMonkey link and complete the evaluation survey.

2.2. Internal Validity and Statistical Analysis

To examine the internal validity of the four dimensions, Cronbach's Alpha statistic was used (Table 2). The separate questions that were designed to examine each dimension, were found to have an alpha score between 0.704 (navigation dimension) and 0.908 (features dimension). Finally, four questions were used to calculate the composite dimension 'Opinion about data driven medical education' (alpha = 0.84). The composite score of each new dimension was calculated as the mean value of the individual items. Since the sample size was relatively small, data analysis was conducted at the 90% of statistical significance. The analysis was completed with SPSS version 26 [5].

Table 1. Scenarios that the residents completed during the evaluation session

Scenario 1: Male patient, 75-79yrs, with a Principal D_x of Acute cerebrovascular disease [Submit]
Chronic Conditions: Essential Hypertension, Congestive Heart Failure [Submit]
Comorbidities: Diagnosed in hospital: None, Non-hospital acquired: Aspiration Pneumonitis [End]
Scenario 2: Female patient, 80-84yrs, with a Principal D_x of Multiple Myeloma [Submit]
Chronic Conditions: Chronic Kidney Disease [Submit]
Comorbidities: Diagnosed in hospital: Deficiency & other anemia, non-hospital acquired: Acute & unspecified renal failure [End]
Scenario 3: Female patient, <65yrs, with a Principal D_x of Cancer of Pancreas [Submit button]
Chronic Conditions: Essential Hypertension [Submit button]
Comorbidities: Diagnosed in hospital: Secondary malignancies non-hospital acquired: None [End]
Scenario 4: Male patient, 65-69yrs, with a Principal D_x of Septicemia [Submit]
Chronic Conditions: Diabetes mellitus with complications [Submit]
Comorbidities: Diagnosed in hospital: CKD, Non-hospital acquired: Fluid & electrolyte disorders [End]

3. Results

3.1. Univariate Analysis

The participants had an equal gender distribution (9 male, 9 female). Their mean age was 28.2 years (std. dev. = 1.7 years). Half of the participants answered that they are very familiar with medical simulation systems. When asked about their perceived expertise with computer use, seven participants (38.9%) said that they had average expertise, nine (50%) above average, and two participants (11.1%) thought that they have a far above average expertise. Finally, seven participants (38.9%) were somewhat familiar, and only two participants (11.1%) not as familiar with medical simulation systems. Of the four evaluation dimensions, the one with the highest score was the 'system features' (4.15); the dimension with the lower score was the 'information validity' (3.77). The composite variable about the participants' opinion on data-driven education was found to be 4.27.

Table 2. Individual item responses, Internal Validity and Composite Scores

	Mean (s.d)	Cronbach Alpha	Composite Mean (s.d)
Dimension 1: System Navigation			
The software that I used today was easy to navigate	3.89 (.83)		
I could navigate the software without any problems	3.89 (.96)		
I could understand the available choices that I was provided	4.17 (.51)	.70	3.96 (.47)
The navigation buttons were placed where they should be	4.17 (.51)		
It was easy to select options from drop-down lists	3.72 (.57)		
Dimension 2: System Usefulness			
Provided relevant information about comorbidities	3.94 (.63)		
Provided useful insights of the comorbidity effect on outcomes	3.89 (.83)		
The information of the system was interesting	4.11 (.47)	.87	3.98 (.47)
The system can help learn about the impact of comorbidities	4.06 (.53)		
Provides feedback & information relevant in actual clinical cases	3.94 (.41)		
Dimension 3: Validity of information			
The information that the system provided was correct	3.83 (.51)		
The information that the system provided was accurate	3.78 (.64)	.82	3.77 (.54)
The information that the system provided was unambiguous	3.72 (.75)		
Dimension 4: System Features			
Separating chronic vs hospital acquired conditions was important	4.17 (.51)		
The step-by-step system approach to develop my scenario was helpful	4.17 (.51)	.90	4.15 (.46)
The summary of the scenarios after completion is a useful feature	4.00 (.48)		
The updated risk estimations output after each step is a useful feature	4.28 (.57)		
Opinion about data-driven medical education			
Data-driven simulations systems are needed in medical curricula	4.22 (.54)		
Big data provide opportunities to examine clinical case combinations	4.39 (.50)	.84	4.27 (.40)
Comorb. with varying effect on outcomes should be easy to navigate	4.28 (.46)		
Familiarizing with data-driven estimations help proactive decisions	4.22 (.42)		

3.2. External Factors Associated with Perceptions about TrainComorb

Bivariate analysis was conducted to examine the association between the participants' age, their self-reported computer experience, and the simulation system familiarity, with each of the four composite evaluation dimensions. A negative and statistically significant correlation was observed between the participants' age and (i) their responses on the 'information validity' dimension ($r=-0.44$, $p<0.10$) and (ii) their opinion regarding data driven education ($r=-0.52$, $p<0.05$). The older the residents were, the more skeptical they

appeared to be regarding computer-generated clinical information, and they had a less positive opinion about the potential in data driven education.

A positive statistical association was observed between the familiarity with simulation systems and the 'information validity' dimension ($r=0.49$, $p<0.05$): Those exposed more in medical simulation systems appeared to be more confident about the information that the system was generating. Similarly, a positive association was observed between the residents' opinion about data-driven education and the 'system usefulness' ($r=0.44$, $p<0.1$) and the 'system features' ($r=0.62$, $p<0.01$) dimensions. No correlation was observed between the reported computer experience and any of the examined dimensions (Table 3).

Table 3. Association of External Factors with Perceptions about TrainComorb

	System Navigation	System Usefulness	Information Validity	System Features	Opinion Data-Driven educ.
Age	0.03 (p=.99)	-0.34 (p=.17)	-0.44 (p=.07)	-0.36 (p=.15)	-0.52 (p=.03)
Computer Experience	-0.21 (p=.39)	-0.19 (p=.44)	0.14 (p=.57)	-0.09 (p=.71)	-0.02 (p=.92)
Familiar with sim systems	-0.10 (p=.69)	-0.05 (p=.82)	0.49 (p=.03)	-0.01 (p=.96)	-0.09 (p=.71)
Opinion Data-Driven educ.	0.34 (p=.16)	0.44 (p=.06)	0.16 (p=.51)	0.62 (p=.01)	

4. Conclusions

The medical residents evaluated the four dimensions with average scores between 3.77 and 4.15 (5-scale) and thought highly of data-driven medical education. The finding supports a recently acknowledged need for data-driven physicians [6]. Although the residents' evaluation was not associated with their computer experience, those more familiar with simulation systems, and more positive about data-driven education, evaluated the dimensions 'usefulness', 'validity', and 'features' with higher scores: Apparently, the role of experience in user perception has been discussed in the literature, as in [7]. The new generation of medical students are gradually becoming familiar with alternative modes of medical training, and they appear to value evidence based and data driven modalities as useful supplementary learning approaches. This is supported by the external factor associations that were observed in the present study.

TrainComorb has the potential to become a useful supplementary tool for the education of future medical professionals. The results of this user-based evaluation study and the system prototype itself, are being communicated with the upper administration of a College of Medicine and its Simulation Center in the region of central Michigan, to identify effective venues of integration with the formal medical curricula.

References

[1] Van Weel C, Schellevis FG, Comorbidity and guidelines: Conflicting interests, Lancet 18 (2006), 550-1.
[2] Zikos D, Ragina N, Strong O, Stud Health Technol Inform 272 (2020), 83-86.
[3] Nykänena P, Brender J, Talmonc J, et al., Guideline for good evaluation practice in health informatics (GEP-HI). Int. J. Med. Inf. 80 (2011), 815-27.
[4] National Career Development Association, Software evaluation criteria, https://tinyurl.com/2725a9je
[5] IBM Corp, IBM SPSS Statistics for Windows (2019), Version 26.0. Armonk, NY: IBM Corp.
[6] Stanford Medicine 2020 Health Trends Report, The Rise of the Data-Driven Physician, (2020), Available at: https://med.stanford.edu/dean/healthtrends.html
[7] Ribeiro R, The Role of Experience in Perception. Human Studies 37 (2014), 559–81.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210895

Sharing the IT Educational Experience of Developing 3D Applications for Medical Students Training

Stelian NICOLA[a,1] and Lacramioara STOICU-TIVADAR[a]
[a] *Department of Automation and Applied Informatics*
Politehnica University Timisoara, Romania

Abstract. The importance of using the new technologies to develop educational and training applications for medical students is given by the times we live in and also by the continued development of the IT industry. New concepts used in 3D applications such as gamification bring added value to the use of learning applications. The introduction of technologies-based on virtual and augmented reality contributes to increasing of the interactivity. This paper presents an overview of the students' experience in developing a complex 3D application for medical students. The application development methodology is explained by splitting the design and development of the 3D learning application into steps. The main users' benefits provided by such applications are increased interactivity and learning benefits, and the continuous availability of the application wherein virtual laboratories within a medical clinic are simulated.

Keywords. gamification, virtual reality, 3D application, medical education

1. Introduction

Currently, the IT industry growth is also due to the emergence of new technologies based on virtual reality (VR), augmented reality (AR), and mixed reality (MR). With the aid of such technologies, the IT industry has become involved in the vast majority of fields such as: education, medicine, healthcare, engineering. Many applications have been developed using such technologies which are meant to help the users learn or understand better certain systems and processes. Thus, new ideas have emerged, which are based on games forteaching and explaining certain procedures. Among the new concepts we focus on gamification. Gamification introduces the use of certain concepts which are used in applications or in games such as: granting bonuses (e.g. stars or experience points), timekeeping, displaying progress and certain suggestions meant to help the user to use the application [1]. Serious games, which, besides entertainment, are also aimed at educating or helping the users to understand certain concepts in various fields, are also based on gamification.

Numerous applications used in education or training may be found in the literature. Paper [2] presents an application supporting medical, bioinformatics, computer science and engineering students to train in a virtual environment, built for intensive care (ICU).

[1] Corresponding Author, Stelian NICOLA, Politehnica University Timisoara, P-ta Victoriei no. 2, Timisoara, Romania; E-mail: stelian.nicola@aut.upt.ro.

Hardware-wise, the manner of operation of the devices as well as the processing of the resulting data are learned by the IT students. Medical students learn how to use the devices, but also how they should interact and react in certain situations with an ICU patient. The use of a 2D and 3D-based multimedia tool, of animations and video clips on certain systems of the human body (e.g. the circulatory system) is discussed in [3]. The benefit of such applications is given by the fact the students may learn without making errors or being afraid to make errors in real situations (e.g. dissecting a virtual human body). VR and AR-based applications are proposed for learning certain parts of the body or anatomical systems of the human body. Papers [4] and [5] discuss applications for learning the brain structure and, respectively, the skeletal system. AR is used to learn the heart structure and functions [6]. A 3D marker of the human body is proposed, that students may scan with the assistance of the application and then they may learn about the functions of the heart. VR is also used in the nursing field [7]. Based on VR applications, nurses may learn the steps they need to follow to change the bandage on a patient with a flesh wound. Visualizing a 3D virtual model, following each step to change the bandage as well as the provision of feedback and of instructions on how to change it introduces the concept of gamification in this application.

This paper will discuss the experience of the students enrolled in the Healthcare Information Systems (HIS) Master's degree of our university, in developing a complex 3D application based on VR. 20 students participated in the project. The application development methodology throughout a university semester (14 weeks) is also described. The motivation to develop such an application relies in the idea of providing learning and training applications to medical students who, in times like these (COVID 19) are prevented to physically attend laboratory classes.

2. Methods and tools

The application is divided into 5 modules and its purpose is to provide information on certain laboratories within a hospital or laboratory classes on human anatomical systems. The Unity editor together with several 3D models from the Unity asset store (e.g. the clinic model, the virtual nurse) was used for the application development, while the C# language was used for the implementation of the functionalities. The application was developed for desktop, and it is based on VR technologies and gamification concepts.

The methodology employed to develop the HIS Master's projects is presented in the followings.

2.1. Methodology

Step 1: presentation of the project themes and the technologies employed within the project by the teaching staff. They presented several project themes and what they entail to the students. Among the themes that were approached there were: a radiology laboratory, an MRI laboratory, a blood collection laboratory, a laboratory for cell and tissue visualization and a laboratory for learning the locomotor (musculoskeletal) system.

Step 2: division of the students based on knowledge and preferences. Each student had to choose, based on their own knowledge and preferences, one of the following cases: C# language programming knowledge, knowledge of use of the Unity 3D software package, minimal knowledge of the medical field (where minimal means medical knowledge of one of the themes that the teaching staff presented in step 1).

Step 3: making groups of 4 students, where each member must have knowledge of one of the cases described in step 2. Thus, each group was built in such a way as to cover all the knowledge needed to realize the project theme.

Step 4: establishing the project theme for each team. Each team chose one of the 5 project themes presented in step 1: the radiology laboratory, the blood collection laboratory, the laboratory for the exploration and visualization of cells and tissue, the muscular system - learning laboratory and the skeletal system - learning laboratory.

Step 5: elaboration of a scenario by each team for the chosen project theme. The elaboration of the scenario consisted of the main functionalities that the application had perform. In addition, the stages that the application had to respect were also defined.

Step 6: identifying and researching the software functionalities of the application, but also a medical research on the methodology that must be respected when performing procedures within the laboratory.

Step 7: developing the application from a software and visual standpoint. Writing C# scripts for the implementing the functionalities defined in steps 5 and 6. Among the developed functionalities are: zooming in/out, moving, rotating 3D models, controlling the character from the keyboard, opening/closing laboratory doors, informing or helping the user via a help button, turning devices on and off. Finding, developing and using 3D models in the application scene. The students made searches and selected free 3D models of the characters (e.g. nurses, physicians, patients) and of devices or pieces of furniture used in a laboratory (e.g. microscope, X-ray machine, test tubes, computer, monitors). All 3D models were arranged in the laboratory based on its kind.

Step 8: testing the applications. This was performed by testing the software functionalities realized by each team. This was made by exchanging the applications. For instance, the team that developed the radiology laboratory tested the application of the team that developed the laboratory for the learning of the muscular system and vice versa. The testing was performed by the students using a test scenario. The main tests that were done: software function tests and generally user evaluation.

Step 9: software integration. All the laboratories developed by the team were integrated in a single 3D desktop application. Thus, a complex application was realized, wherein all the users may learn or explore various laboratories based on their needs. The entire application has a defined scenario that allows the users to control a character who may explore several laboratories within a hospital. The 3D application starts with the user visualizing a clinic, the clinic entry and the possibility to enter one of the laboratories developed by one of the teams and then explore them. A color was chosen for each team (red, blue, green, pink and yellow) for a better demarcation and visualization of the laboratories developed by each team. Thus, the door of each laboratory within the clinic is colored based on each team's color.

Figure 1 shows several images from the laboratories developed by the students. Therefore, figure 1 contains. (a) - radiology laboratory, figure (b) - muscular system-learning laboratory, (c) – skeletal system-learning laboratory, (d) blood collection laboratory and (e) – cell and tissue-visualization laboratory.

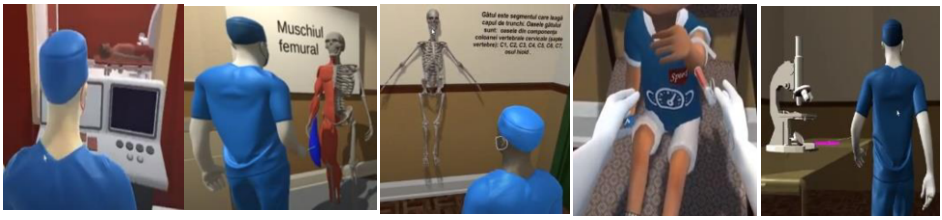


Figure 1. 3D Virtual Laboratories.

3. Conclusions

This paper highlighted the students' and staff experience in developing a complex 3D application meant to help medical students to learn certain laboratories more easily. Using a step or task-based methodology was efficient from the standpoint of delivering a functional application. The complexity of the application is given by the multitude of functionalities provided, by the 3D virtual environment, but also by the use of the virtual reality and gamification concepts.

In the future, we intend to test the application with medical students using new test methods by automatically identifying the user activity by means of a software service: Unity Analytics.

Acknowledgment

The authors would like to thank the Healthcare Information System Master's degree students from the Automation and Computers Faculty within the Politehnica University Timisoara, class of 2021, for their involvement and participation in the development of the complex 3D application.

References

[1] Lavoué E, Monterrat B, Desmarais M and George S, Adaptive Gamification for Learning Environments, IEEE Transactions on Learning Technologies. Jan.-March 2019;12(1):16-28.
[2] Forjan M, David V, Wagner M, Dolesch L, Lechner M and Sauermann S, Conceptualization of an ICU Infrastructure for Simulation Based Education in Medical Engineering & eHealth, 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).
[3] Majerník J, Szerdiová L,Preparation of Medical Students for Cadaveric Anatomy using Multimedia Education Tools, 2017 International Conference on Information and Digital Technologies (IDT).
[4] Wirza R, Afera CNS, Hod RB, Mobile Augmented Reality Application with Virtual Buttons as User Interaction and Stereoscopic View for Medical Learning, 2019 2nd International Conference on Applied Engineering (ICAE).
[5] Nicola S, Virag I, Stoicu-Tivadar L, VR Medical Gamification for Training and Education, 11th Annual Conference on Health Informatics Meets eHealth (eHealth), Austria, May 2017.
[6] Touel S, Mekkadem M, Kenoui M, Benbelkacem S, Collocated learning experience within collaborative augmented environment (Anatomy course), 2017 5th International Conference on Electrical Engineering - Boumerdes (ICEE-B).
[7] Choi KS, Virtual Reality Wound Care Training for Clinical Nursing Education: An Initial User Study, 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210896

Artificial Intelligence in Primary Care: An Overview

Darius TURCIAN[a,1] and Vasile STOICU-TIVADAR[a]
[a] *Politehnica University Timişoara, Department of Automation and Applied Informatics, Timişoara, Romania*

Abstract. With the increase in computing power and the development of numerous technological devices that facilitate remote work, the involvement of artificial intelligence in medicine has seen a significant increase to help the doctor make decisions and intervene in the medical process and telemedicine. In this paper, we gave an overview of the practical involvement of artificial intelligence through different support systems used in primary medicine or telemedicine and also identified the possibilities and opportunities for the development of new support systems for family medicine. Thus, we identified systems used for primary diagnosis, diagnosis of hypertension, early detection of heart abnormalities, detection of diabetes, support in the prescription process, helping clinicians in the daily workflow by providing certain answers to questions, treatment guidance, determining patient priority for treatment for SARS-CoV-2 infection or early detection of disease, support of artificial ventilation in medical emergency centers, remote support for treatments and medication.

Keywords. Artificial intelligence, primary care, family medicine

1. Introduction

One of the most important and emerging things today, in terms of technology is Artificial Intelligence (AI) that began to be used in a variety of domains. In all areas, the question is whether AI can be involved to improve and perfect the functioning of systems or help humans in taking decisions. In medicine, AI is seen as a great opportunity to increase or replace the limited processing power, to improve the identification accuracy of the diagnosis, to improve the efficiency in the doctor's workflow as well as patient scheduling, to increase productivity, to reduce medical errors, and to facilitate better patient monitoring. At the same time, there is a concern regarding what exactly AI can provide and also undermining trust in physicians. [1–4]

From the perspective of AI position as software with decision-making ability how it is presented in [3], AI has evolved from a rule-centric approach which is based on *if-then* rules to a data-centric approach which is based on capabilities to learn from data [1,2].

Family medicine which is a clinical specialty oriented towards primary care is a field where AI can be used too to help humans in making decisions. The aims of this paper are to expose the current state of AI used in primary care and family medicine.

[1] Corresponding Author, Darius Turcian, Politehnica University Timişoara, Department of Automation and Applied Informatics, B019, Vasile Pârvan 2, Timişoara, Romania; E-mail: darius.turcian@upt.ro.

2. Methods

To identify, evaluate and interpret the current state of AI used in family medicine and primary care, we performed a systematic search for specialized papers published starting with 2018, from the field of engineering and medicine. The main explorer was the IEEE Xplore database, and also National Center for Biotechnology Information (NCBI) including PMC and PubMed databases, Elsevier databases. The search strategy included terms related to AI and primary care such as artificial intelligence, machine learning, deep learning, family medicine, primary care, prediction, diagnosis, etc.

3. Results

3.1. Searches

Table 1 presents the search results based on the methods presented above.

Table 1. Search results

Keywords	Results description	Selections
Artificial intelligence in primary care	From the first 50 results, 17 matched the subject and 3 were selected	[5], [6], [7]
Machine learning in primary care	From the first 50 results, 12 matched the subject and 3 were selected	[8], [9], [10]
Decision support system in primary care	From the first 20 results, 5 matched the subject and 2 were selected	[11], [12]
Primary diagnosis	From the first 20 results, 8 matched the subject and 2 were selected	[13], [14]
COVID-19	From the first 15 results, 4 matched the subject and 2 were selected	[15], [16]

After the first selection for those papers which matched the subject based on the abstract study, we selected a subset of the found papers, based on the date of publication and the importance of systems for primary care.

3.2. Practical implications

In table 2 are presented AI systems and their techniques / methods used and the accuracy for each of them.

Table 2. AI systems developed for primary care and telemedicine

System	Technique/ Algorithm	Accuracy	Reference
ML model to predict hypertension	Three-layer Artificial Neural Network (ANN)	82%	[8]
Clinical Decision Support System (CDSS) for generating alerts during medication prescribing	Oracle Data Warehouse	Unknown	[11]
ML system for primary diagnosis from chief complaints	Naïve Bayes and Random Forest classifiers	66.1%-77.7%	[13]
Q&A system (called AI-Q)	DL with Natural Language Processing (NLP) operations	Low	[6]
CDSS used for guidance the patient's treatment	Rules-based	Unknown	[12]
ML system used to determine the priority of patients for treatment for Covid-19	Convolutional Neural Network (CNN) model	96.5%	[15]
CDSS used for an early detection of the coronavirus disease	DL model	92.5%	[16]

CDSS system which use IoT devices in ICU centers to provide alerts regarding ventilation therapy	ML algorithms	97.1%-97.93%	[9]
Portable AI system used for early detections of heart abnormalities	CNN algorithm	75.2%-97%	[14]
Mobile application which offers on-line information regarding treatments and drugs	ML and Data Mining	Unknown	[7]
ML system that predicts diabetes	ML algorithms based on Big Data analysis	58%-82%	[10]

4. Discussion

From the analysis of the presented systems, we noticed a great potential for improvement comparing with human decision only, both in terms of execution time and the limitation of human errors. In most cases the methods used are focused on the "data-centric approach" using ML and DL algorithms with ANN and CNN methods.

AI is increasingly used in primary medicine and implicitly this will revolutionize family medicine. In this sense, an important step is represented by CDSS that have the ability to predict a certain diagnosis. For diseases that require CT scans, we noticed that ML and DL analyzes can be used in a very precise way on X-Ray images to highlight a diagnosis and its severity. We also noticed that for monitoring patients, ML algorithms can be used to generate different alerts depending on the patient's health status. Then another important factor is the ability to create smart mobile or web applications based on artificial intelligence, to help the family doctor make decisions and prescribe drugs or give priority to patients with more serious conditions. Another step is to identify the possibilities of creating AI-based CDSS to diagnose and provide treatment for most conditions often present in a family doctor's office.

The main concern of AI systems is still an accurate prediction. On the other hand, there is a need for external validation, integration into clinical workflow, approaching special cases, or facing the clinicians' reticence. Then the ethical aspects should be considered for the possible decisions errors or medical malpractice. An ethical problem is that AI has the potential to change the dynamic relationship between patient-clinician by introducing a new factor of trust – AI.

5. Conclusions

In this paper, we have created an overview of the systems based on artificial intelligence, that are used in primary care and not only, but also that offer viable solutions for telemedicine and the possibility of working remotely with the family doctor with his patients.

A further study will be necessary to evaluate and identify a better technique that can be used as part of the CDSS in family medicine together with NLP in order to create intelligent systems. Also, a good way to improve these types of systems is to implement a global place in Cloud to store all the public EMR of patients as a dataset that can be used for training the ML systems, to become a more efficient training model.

The system used in Thailand for predicting diagnosis based on a text input is a great inspiration and together with AI-Q system, the prioritization system leads us to think about the development of an intelligent CDSS used in family medicine, on the one hand,

to give priority to patients and provide them with a solution to communicate with the system that proposes a diagnosis, validated by the doctor, proposes a treatment, validated by the doctor and on the other hand, solves many tasks in the doctor's office.

Future research will follow these aspects and will look for solutions not only to obtain greater accuracy but also to reduce the limitations and concerns presented in this paper regarding AI.

References

[1] Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. Gastrointestinal Endoscopy. 2020 Oct;92(4):807–12.

[2] Liyanage H, Liaw S-T, Jonnagaddala J, Schreiber R, Kuziemsky C, Terry AL, et al. Artificial Intelligence in Primary Health Care: Perceptions, Issues, and Challenges: Primary Health Care Informatics Working Group Contribution to the Yearbook of Medical Informatics 2019. Yearb Med Inform. 2019 Aug;28(01):041–6.

[3] Mistry P. Artificial intelligence in primary care. Br J Gen Pract. 2019 Sep;69(686):422–3.

[4] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019 Jan;25(1):44–56.

[5] Tsoi KKF. Application of Artificial Intelligence on a Symptom Diagnostic Platform for Telemedicine A Pilot Case Study. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). Bari, Italy: IEEE; 2019. p. 806–13.

[6] Hanada E, Wada K, Oda K, Nishi K, Kawazoe K. Practical use of Artificial Intelligence for Clinical Staff Other than Physicians. In: 2018 IEEE 8th International Conference on Consumer Electronics - Berlin (ICCE-Berlin). Berlin: IEEE; 2018. p. 1–4.

[7] Liu W, Yuhang L, Lingling P, Na C. Design and Implementation of Family Doctor App on Android Platform. In: 2018 9th International Conference on Information Technology in Medicine and Education (ITME). Hangzhou: IEEE; 2018. p. 128–32.

[8] LaFreniere D, Zulkernine F, Barber D, Martin K. Using machine learning to predict hypertension from a clinical dataset. In: 2016 IEEE Symposium Series on Computational Intelligence (SSCI). Athens, Greece: IEEE; 2016. p. 1–7.

[9] Rehm GB, Woo SH, Chen XL, Kuhn BT, Cortes-Puch I, Anderson NR, et al. Leveraging IoTs and Machine Learning for Patient Diagnosis and Ventilation Management in the Intensive Care Unit. IEEE Pervasive Comput. 2020 Jul 1;19(3):68–78.

[10] Ahmed H, Younis EMG, Ali AA. Predicting Diabetes using Distributed Machine Learning based on Apache Spark*. In: 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE). Aswan, Egypt: IEEE; 2020. p. 44–9.

[11] Albahly MS, Seliaman ME. Evaluation of the impact of Clinical Decision Support Systems: Descriptive Analytics. In: 2020 2nd International Conference on Computer and Information Sciences (ICCIS). Sakaka, Saudi Arabia: IEEE; 2020. p. 1–5.

[12] Duarte de Almeida JR, Oliveira JL. GenericCDSS - A Generic Clinical Decision Support System. In: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS). Cordoba, Spain: IEEE; 2019. p. 186–91.

[13] Pettakorn S, Sutantayawalee V, Thammasudjarit R. Primary Diagnosis Prediction from Chief Complaints. In: 2020 17th International Joint Conference on Computer Science and Software Engineering (JCSSE). Bangkok, Thailand: IEEE; 2020. p. 145–9.

[14] Singh KK, Singh SS. An Artificial Intelligence based mobile solution for early detection of valvular heart diseases. In: 2019 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT). Bangalore, India: IEEE; 2019. p. 1–5.

[15] Rahim A, Kurniawan M, Kusrini. Machine Learning Based Decision Support System for Determining the Priority of Covid–19 Patients. In: 2020 3rd International Conference on Information and Communications Technology (ICOIACT). Yogyakarta, Indonesia: IEEE; 2020. p. 319–24.

[16] Qjidaa M, Ben-fares A, Mechbal Y, Amakdouf H, Maaroufi M, Alami B, et al. Development of a clinical decision support system for the early detection of COVID-19 using deep learning based on chest radiographic images. In: 2020 International Conference on Intelligent Systems and Computer Vision (ISCV). Fez, Morocco: IEEE; 2020. p. 1–6.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210897

Development and Evaluation of a Bayesian Risk Stratification Method for Major Amputations in Patients with Diabetic Foot Ulcers

Jens HÜSERS[a,1], Guido HAFER[b], Jan HEGGEMANN[b], Stefan WIEMEYER[b], Swen Malte JOHN[c] and Ursula HÜBNER[a]
[a] *Health Informatics Research Group, Osnabrück University AS, Germany*
[b] *Christliches Klinikum Melle, Niels Stensen Kliniken, Germany*
[c] *Department Dermatology, Environmental Medicine, Health Theory, University of Osnabrück, Germany*

Abstract. The diabetic foot ulcer, which 2% - 6% of diabetes patients experience, is a severe health threat. It is closely linked to the risk of lower extremity amputation (LEA). When a DFU is present, the chief imperative is to initiate tertiary preventive actions to avoid amputation. In this light, clinical decision support systems (CDSS) can guide clinicians to identify DFU patients early. In this study, the PEDIS classification and a Bayesian logistic regression model are utilised to develop and evaluate a decision method for patient stratification. Therefore, we conducted a Bayesian cutpoint analysis. The CDSS revealed an optimal cutpoint for the amputation risk of 0.28. Sensitivity and specificity were 0.83 and 0.66. These results show that although the specificity is low, the decision method includes most actual patients at risk, which is a desirable feature in monitoring patients at risk for major amputation. This study shows that the PEDIS classification promises to provide a valid basis for a DFU risk stratification in CDSS.

Keywords: Clinical Decision Support System, Logistic Regression, Health Information Technology, Diabetic Foot Ulcer, Amputation, Bayesian statistics

1. Introduction

The diabetic foot ulcer (DFU), which 2% - 6% of diabetes patients experience, is a severe health threat that is characterized as a chronic lesion of the foot tissue, which is mainly caused by diabetes-related peripheral vascular impairment and neuropathic conditions [1]. The prevention of diabetes-related foot problems is the primary imperative, as they are linked with the risk of lower extremity amputation (LEA), which is associated with high economic and health burden [2]. This is especially true for amputations above the ankle, also defined as major amputations. When a DFU is present, it is desirable to identify major-amputation risk patients to initiate preventive actions early on.

[1] Corresponding Author, Jens Hüsers, Osnabrück University of AS, Health Informatics Research Group, PO Box 1940, 49009 Osnabrück, Germany; E-Mail: j.huesers@hs-osnabrueck.de.

The PEDIS classification, which reliably describes DFU through five risk factors (**P**erfusion status, ulcer **E**xtent and **D**epth, **I**nfection status, and foot **S**ensation) [3,4], is an international standard for DFU classification consented by the International Working Group of the Diabetic Foot. It is the aim of this study to develop a model based stratification scheme that allows for the classification of patients with and without a risk for a major amputations using the PEDIS system.

Accordingly, our research questions concerned the development (first question) and the evaluation (second question): 1.) Which risk (probability) constitutes the optimal cutpoint for major-amputation risk stratification? 2.) What is the corresponding diagnostic accuracy in terms of AUC, sensitivity and specificity?

2. Methods

We conducted a prospective cohort study at the Wound Care Centre of Christliches Klinikum Melle in Germany, which started in June 2013 and ended in June 2019. All subjects were screened before inclusion: Diabetes mellitus (type I and type II) patients with a DFU were included; other wounds, e.g., venouse or arteric leg ulcers, were excluded. Then, eligible subjects underwent baseline assessment where age and gender were documented, as well as the ulcer classified according to the PEDIS system. Six months after baseline assessment, the major-amputation status, defined as amputation above the ankle was determined and recorded. This prospective data collection was repeated, when no major-amputation was conducted, and the treatment was continued for a patient.

We aimed to stratify patients for the target condition, i.e., major amputation. Therefore, we utilised a Bayesian logistic regression model [5], in which major-amputation served as the criterion, each of the five PEDIS risk factors as a predictor and gender as well as age as covariates. We selected a logistic regression model as it allowes meaningful interpretation of the model parameters, i.e., odds ratio and therefore provides transparent insights for clinicians. Furthermore, the bayesian approach allows the incorporation of prior knowledge. In this study the bayesian model made use of information from the multicentre EURODIALE study [6].

The Bayesian posterior distributions of the model coefficients were the starting point for the cutpoint analysis, which was based on maximising the Youden-Index [7]. To analyse the diagnostic validity of the cutpoint we computed the AUC as well as sensitivity and specificity and their 95% high-density intervals (HDI) to quantify uncertainty [8].

3. Results

The data used to fit the model was based on a sample of 237 DFU patients. Of all patients, 12.2% (n=29) underwent major-amputation procedure. The mean age of the overall sample was 65.9 years (± 12.3). For the non-amputees and amputees, the average age was 65.6 years (± 12.5) and 65.5 years (± 12.4), respectively. The overall proportion of female and male patients was 16.5% (n=39) and 83.5% (n=198).

According to the Youden-Index, the posterior median estimate of the optimal cutpoint was 0.28, which means that patients with a predicted amputation probability above this threshold are classified as risk patients. Quantifying the uncertainty associated

with this estimate, the 95% HDI ranged from 0.16 to 0.41. The posterior median of the sensitivity was 0.83 (95% HDI 0.71 - 0.93), and that of the specificity was 0.66 (95% HDI 0.53 - 0.77) (Figure 1).

Table 1. Optimal cutpoint, AUC and the screening test metrics with corresponding 95% highest density intervals.

Metric	Posterior Median	95% HDI
Optimal Cutpoint	0.28	[0.16 - 0.41]
AUC	0.80	[0.78 - 0.80]
Sensitivity	0.83	[0.71 - 0.93]
Specificity	0.66	[0.53 - 0.77]

4. Discussion

In this study, we developed a PEDIS based method for binary risk stratification for DFU patients facing potential major-amputation. The optimal cutpoint for the predicted six-month amputation risk was determined by drawing on a Bayesian logistic regression model with an acceptable AUC value and its stratification performance was scrutinised using diagnostic metrics.

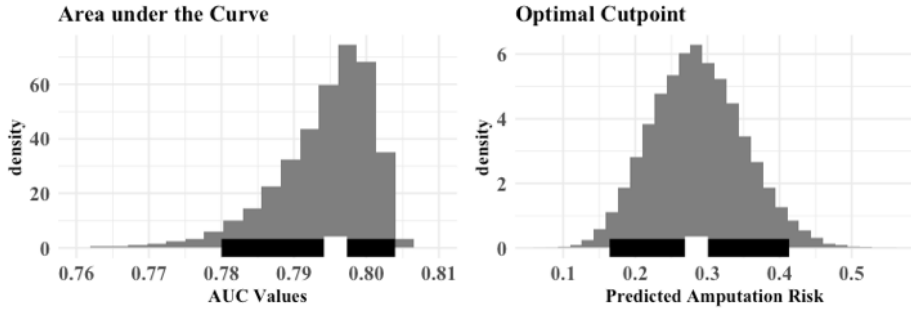


Figure 1. Histogram of posterior distribution of AUC (left) and cutpoint (right), horizontal black bar indicates range of 95% HDI and white box within the bar indicates the median. Left: AUC median = 0.80 (95% HDI 0.78 - 0.80) Right: Optimal cutpoint median = 0.28 (95% HDI 0.16 to 0.41).

Using the Youden index for cutpoint determination, the optimal cutpoint was 0.28. Patients with higher predicted risks are considered as patients at risk. The associated HDI ranges from 0.16 to 0.41; thus, indicating some uncertainty about the optimal cutpoint.

The sensivity and specificity was 0.83 and 0.66, respectively. These results reflect a typical situation where a tradeoff between sensitivity and specificity is present. This trade-offs have to be discussed critically: In this case, the rather low specificity is tolerable for the context of amputation because higher sensitivity is a desirable feature from the clinical perspective of monitoring these patients more closely: Through this test, a clinician can detect most of the actual risk patients at the expense of including non-risk patients in the risk strata.

As this method requires computational ressources, we recommend to implement it in routine record-keeping systems such as wound documentations as a clinical decision

support tool. Thereby, the clinician receives valuable additional information about the risk status of a DFU patient when completing the PEDIS classification procedure.

A limitation of this study is its single-centre approach. Thus, this study may represent a specific subgroup seen in specialised, multidisciplinary wound care centres rather than patients that nurses and general practitioners see in daily routine care. However, by incorporating external scientific knowledge in the Bayesian model, we tried to mitigate this effect. Furthermore, more data is required. Again, the Bayesian approach allows an update of the model when further data is available.

In summary, this study leads to two main conclusions. First, Bayesian statistics is a valuable method for designing and evaluating clinical decision models. Second, given further validation through more data, the PEDIS classification promises to be a valid and clinically applicable method for risk stratification for patients with a DFU.

Funding

The state of Lower Saxony funded this work, project ROSE - The Learning Health Care System in the region Osnabrück and Emsland (ZN3103)

References

[1] Zhang P, Lu J, Jing Y, Tang S, Zhu D, Bi Y. Global epidemiology of diabetic foot ulceration: a systematic review and meta-analysis. Ann Med. 2017 Feb 17;49(2):106–16.

[2] Bakker K, Apelqvist J, Lipsky BA, Van Netten JJ, Schaper NC, on behalf of the International Working Group on the Diabetic Foot (IWGDF). The 2015 IWGDF guidance documents on prevention and management of foot problems in diabetes: development of an evidence-based global consensus: Development of the 2015 IWGDF Guidance. Diabetes Metab Res Rev. 2016 Jan;32:2–6.

[3] Chuan F, Tang K, Jiang P, Zhou B, He X. Reliability and Validity of the Perfusion, Extent, Depth, Infection and Sensation (PEDIS) Classification System and Score in Patients with Diabetic Foot Ulcer. Santanelli, di Pompeo d'Illasi F, editor. PLOS ONE. 2015 Apr 13;10(4):e0124739.

[4] Schaper NC. Diabetic foot ulcer classification system for research purposes: a progress report on criteria for including patients in research studies. Diabetes Metab Res Rev. 2004 May;20(S1):S90–5.

[5] Hüsers J, Hafer G, Heggemann J, Wiemeyer S, John SM, Hübner U. Predicting the amputation risk for patients with diabetic foot ulceration – a Bayesian decision support tool. BMC Med Inform Decis Mak. 2020 Aug 24;20(1):200.

[6] Pickwell K, Siersma V, Kars M, Apelqvist J, Bakker K, Edmonds M, et al. Predictors of lower-extremity amputation in patients with an infected diabetic foot ulcer. Diabetes Care. 2015 May;38(5):852–7.

[7] Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and Optimal Cut-Point Estimated from Observations Affected by a Lower Limit of Detection. Biom J Biom Z. 2008 Jun;50(3):419–30.

[8] Kruschke JK. Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison. Perspect Psychol Sci. 2011;6(3):299–312.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210898

Effectively Detecting Left Bundle Branch Block False Defects in Myocardial Perfusion Imaging (MPI) with a Convolutional Neural Network (CNN)

Mohamed E. H. ABDI[a,1] Qais NAILI[a], Mahdi HABBACHE[a], Bilhadj SAID[a], Abdelhak BOUMENIR[a], Taouba DOUIBI[a], Dahlia DJERMANE[b] and Sid-Ahmed BERRANI[c]
[a] *Centre d'imagerie scintigraphique, Blida, Algeria*
[b] *Cardio A2, CHU Mustapha Bacha, Algiers, Algeria*
[c] *National Polytechnic School, Algiers, Algeria*

Abstract. Left bundle branch block (LBBB) is a frequent source of false positive MPI reports, in patients evaluated for coronary artery disease. *Purpose:* In this work, we evaluated the ability of a CNN-based solution, using transfer learning, to produce an expert-like judgment in recognizing LBBB false defects. *Methods:* We collected retrospectively, MPI polar maps, of patients having small to large fixed anteroseptal perfusion defect. Images were divided into two groups. The LBBB group included patients where this defect was judged as false defect by two experts. The LAD group included patients where this defect was judged as a true defect by two experts. We used a transfer learning approach on a CNN (ResNet50V2) to classify the images into two groups. *Results:* After 60 iterations, the reached accuracy plateau was 0.98, and the loss was 0.19 (the validation accuracy and loss were 0.91 and 0.25, respectively). A first test set of 23 images was used (11 LBBB, and 12 LAD). The empiric ROC (Receiver operating characteristic) Area was estimated at 0.98. A second test set (18x2 images) was collected after the final results. The ROC area was estimated again at 0.98. *Conclusion:* Artificial intelligence, using CNN and transfer learning, could reproduce an expert-like judgment in differentiating between LBBB false defects, and LAD real defects.

Keywords. left bundle branch block, obstructive coronary artery disease, GATED SPECT myocardial perfusion imaging, convolutional neural network

1. Introduction

Left bundle branch block (LBBB) is a frequent source of false positive reports in myocardial perfusion imaging (MPI). This has been reported in previous studies that have evaluated the relation between myocardial perfusion and LBBB [1, 2].

The method of interpretation used in MPI in patients having LBBB, may influence the sensitivity and specificity of the exam. In this paper by Higgins et al. [1], authors concluded that the use of certain features of the MPI scan can aid the clinician in differentiating true perfusion defects, distinguishing underlying ischemia from false

[1] Corresponding Author, M. ABDI, CIS Blida (scintigraphies.com); E-mail: dr.abdi.m.h@gmail.com

defects. We try to evaluate the usefulness of artificial intelligence, to reproduce this method of interpretation, offering a perspective to develop a diagnostic aid tool.

Some previous studies have evaluated the value of deep learning in the diagnosis of coronary artery disease (CAD) in MPI [3, 4, 5]. Betancur J et al. [3] concluded that deep learning improves automatic prediction of obstructive coronary artery disease from MPI, as compared to the current standard quantitative method.

However, such studies did not include some clinical information when training the model, such as for example, the existence of a LBBB.

2. Purpose

In this work, we evaluated the ability of a CNN based solution, using transfer learning, to produce an expert-like judgment in differentiating LBBB false defect, from left anterior descending artery (LAD) real perfusion defect. The study was conducted considering a small dataset, because collecting a larger dataset needs a proof of utility.

3. Materials and Methods

3.1. Study population

The study covered two groups of MPI polar map images, collected retrospectively from our department, with small, to large fixed antéroseptal perfusion defect (small: 1 segment, moderate: 2 segments, large: 3 or more segments, based on the 17 segments model [6]).

- The LBBB group included patients where the perfusion defect was judged as false defect by two experts (based on clinical assessment, and GATED-SPECT [1]). Expert judgment was reinforced by a flow up for 3 years (no cardiovascular event). All patients in this group had a LBBB.
- The LAD group included patients where the perfusion defect was judged as a true positive by two experts. Expert judgment was confirmed by angiography; >70% narrowing of LAD artery (patients with more than one vessel disease, or LBBB on ECG, were excluded from this group).

Study population baseline characteristics, are illustrated in table1

Table 1. Baseline Characteristics of Studied Population

Characteristic	Overall n=63	LBBB n=33 (52.38%)	LAD n=30 (47.62%)	*P*
Age	63.08 ± 2.46	64.18 ± 3.44	61.82 ± 3.46	0.0087
Male	26 (41.27%)	5 (15.15%)	21 (70.00%)	< 0.0001
Female	37 (58.73%)	28 (84.84%)	9 (30.00%)	< 0.0001
Hypertension	35 (55.55%)	21 (63.63%)	14 (46.66%)	0.1793
Diabetes	21 (33.33%)	11 (33.33%)	10 (33.33%)	1.0000
Dyslipidemia	6 (9.52%)	4 (12.12%)	2 (6.66%)	0.4644
Smoking	7 (11.11%)	0 (0%)	7 (23.33%)	0.0035
Chest pain	30 (47.61%)	20 (60.60%)	10 (33.33%)	0.0318
Dyspnea	4 (6.34%)	4 (12.12%)	0 (0%)	0.0506

3.2. Image Acquisition

A conventional single head, Gamma Camera was used for all patients, using Tc99m-Sestamibi radiotracer. Patients had various stress protocols, such as treadmill, pharmacological, or a mixed protocol. Stress and rest exams were performed either the same day, or on two different days.

3.3. Images Dataset

The dataset was composed of 107 perfusion polar maps (42 images in each class for training, with 29% for validation). Stress and rest images were used.

3.4. Deep learning

Several CNN were tested, and ResNet50V2 was chosen for achieving the best results.

Only the classification part of the network was re-trained, following a transfer learning approach (training a fully connected layer, with two neurons).

4. Results

After 60 iterations, the reached accuracy plateau was 0.98, and the loss was 0.19 (the validation accuracy and loss were 0.91 and 0.25, respectively).

A first test set of 23 images was used (11 LBBB, and 12 LAD). The empiric Receiver operating characteristic (ROC) Area was estimated at 0.98, with 95.7% accuracy.

A second test set (18 images in each group) was collected after the final results (but without the 3 years follow-up for the LBBB group). The ROC area for the model, was estimated again at 0.98.

An example of a man of 70 years old, smoking, having a rest angina, and a LBBB on ECG, is illustrated (Figure 1). He had a positive stress, with reduced LVEF at 35%. Images of this patient were not used nor in training, nor in validation. Stress image was predicted by our model as an LAD real perfusion defect. The rest image was predicted as a LBBB false defect, which means that according to our model, this patient is having an ischemia in the LAD territory (Figure 1). This patient had an angiography, confirming a severe narrowing of proximal LAD. He had a revascularization, resulting in an improvement of his LVEF (from 35% to 45%).

This example illustrates the capacity of the model, differentiating real from false defect, in a same patient having CAD in the LAD artery, along with LBBB.

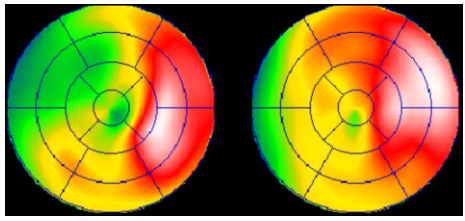


Figure 1. Example case. Left is the Stress polar map, predicted as a real LAD perfusion defect, and right is Rest image predicted as LBBB false defect.

5. Discussion

5.1. Study population

It is worth clarifying that the two groups used in our study are not of the same cardiovascular risk level, as long as we are looking for false defects in the LBBB group, and real defects in the LAD group. We believe that such contrast is mandatory to train the model on distinct features from each group.

5.2. Limitations

Even if these results are encouraging, this study is still a retrospective one, done on a small number of patients, from a single department. Also, it has to be said that our model was trained on a specific color map used in our department, so the evaluation of the model on other patients from other departments needs to convert images to this specific color map.

The aim of this study was to evaluate the ability of deep learning to reproduce an expert like judgment for this problem. The next step could be a multicentric study (different gamma cameras), with coronary angiography as ground truth.

6. Conclusion

Artificial intelligence, using CNN and transfer learning -even on a very small training dataset- could reproduce an expert-like judgment in differentiating between LBBB false defect and LAD real perfusion defect. These results are motivating for a multicenter prospective study, to develop a diagnostic aid tool for clinicians, offering an expert like lecture. Such tool, could probably reduce false positive MPI reports, and by the way, reduce the number of unnecessary invasive angiography.

References

[1] Higgins JP, Williams G, Nagel JS, Higgins JA. Left bundle-branch block artifact on single photon emission computed tomography with technetium Tc 99m (Tc-99m) agents: mechanisms and a method to decrease false-positive interpretations. Am Heart J. 2006;152(4):619-626.

[2] Gupta K, Bajaj NS, Hage FG, Bhambhvani P. Myocardial perfusion artifacts in left bundle branch block: A diagnostic challenge [published online ahead of print, 2019 May 2]. J Nucl Cardiol. 2019;10.1007/s12350-019-01717-9.

[3] Betancur J, Commandeur F, Motlagh M, et al. Deep Learning for Prediction of Obstructive Disease From Fast Myocardial Perfusion SPECT: A Multicenter Study. JACC Cardiovasc Imaging. 2018;11(11):1654-1663.

[4] Betancur J, Hu LH, Commandeur F, et al. Deep Learning Analysis of Upright-Supine High-Efficiency SPECT Myocardial Perfusion Imaging for Prediction of Obstructive Coronary Artery Disease: A Multicenter Study. J Nucl Med. 2019;60(5):664-670.

[5] Apostolopoulos ID, Papathanasiou ND, Spyridonidis T, Apostolopoulos DJ. Automatic characterization of myocardial perfusion imaging polar maps employing deep learning and data augmentation. Hell J Nucl Med. 2020;23(2):125-132.

[6] Cerqueira MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK, Pennell DJ, Rumberger JA, Ryan T, Verani MS; American Heart Association Writing Group on Myocardial Segmentation and Registration for Cardiac Imaging. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. Circulation. 2002 Jan 29;105(4):539-42.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210899

Usability of Venous Thromboembolism Prophylaxis Recommender System: A Pilot Study

Mehrdad KARAJIZADEH[a], Farid ZAND[b], Afsaneh VAZIN[c], Roxana SHARIFIAN[a,1] and Najmeh BAYATI[b]
[a] *Health Human Resources Research Center, School of Management & Medical Information Sciences, Department of Health Information Management, Shiraz University of Medical Sciences, Shiraz, Iran*
[b] *Anesthesiology and Critical Care Research Center, Shiraz University of Medical Sciences, Shiraz, Iran*
[c] *Department of Clinical Pharmacy, Faculty of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran*

Abstract. Poor usability of clinical decision support system impact negative on healthcare professionals, decrease usage and quality of clinical decision support system and result in a negative effect on patient outcome. Therefore, the objective of this study was the usability evaluation of the venous thromboembolism prophylaxis recommendation system. This study design is a pilot study. Totally seven individuals participate in the study that 4 out of 7 were ICU attending and 3 out of 7 were Residents in ICUs setting. System Usability Scale (SUS) was used to assess the usability of the clinical decision support system (venous thromboembolism prophylaxis recommendation system) integrated into the medication order entry system in the ICU setting. This study has shown that the mean System Usability Scale (SUS) score was 74.64. Summing up the results, it can be concluded that the usability quality of the venous thromboembolism prophylaxis recommendations system is good. Further research requires to evaluate the usability of the venous thromboembolism prophylaxis recommendation system by quantitative and qualitative methods in large scale.

Keywords. Usability, SUS, CDSS, VTE, Prophylaxis, Recommender System.

1. Introduction

Poor usability of clinical information systems was a global challenge in the healthcare information technology industry [1]. Poor usability of clinical information systems has a negative effect on healthcare professionals that due to burnout, low productivity, and alert fatigue [2]. The low productivity of health care professionals has negative effects on patient care. Furthermore, low-quality clinical information systems increase morbidity and mortality [3].

Previous studies have shown that the lack of involving the end-user in the clinical information system design process leads to poor usability [4]. Therefore, user-centered

[1] Corresponding Author, Roxana Sharifian, Health Human Resources Research Center, Almas Building, Alley 29, Qasrodasht Ave, Shiraz, Iran; Email: sharifianr@sums.ac.ir.

design is essential to a well-designed clinical information system [5, 6]. Still, there is a large gap between the current clinical information system and physician tasks [7]. ISO defines usability as the" extent to which a system can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [8]. To the authors' knowledge, no research has been investigated the usability of the venous thromboembolism prophylaxis recommendation system from the point of view of healthcare professionals in the ICU setting of Nemazee hospital. Therefore, in this paper, the usability level of the venous thromboembolism prophylaxis recommendation system was investigated. This study results could help to identify usability problems of the venous thromboembolism prophylaxis recommendation system and solve them.

2. Methods

This study design is a pilot study. The usability of the Venous Thromboembolism (VTE) prophylaxis recommendations system was evaluated in the ICUs setting of Nemazee Hospital in July 2021. Seven healthcare professionals were a participant in the study. 4 out of 7 were ICU attending and 3 of 7 were residents in ICUs. Venous thromboembolism prophylaxis recommendation system has been implemented in three ICUs (general ICU, central ICU, and emergency ICU) of Nemazee hospital. The usability of the VTE prophylaxis recommendation system has been assessed by System Usability Scale (SUS) questionnaire [9]. The data collection form had two parts: the first part included the SUS questionnaire items and the second part included a non-structured question about the usability of this system. The data collection form was attached as supplementary file 1. System Usability Scale (SUS) questionnaire consists of 10 items with one of five responses that range from Strongly Agree to Strongly disagree. The usability score was calculated based on the Brooke study [9]. SUS Score above 68 is considered excellent usability based on the general guideline on interpretation SUS score. The sampling method was convenient because all physicians in the three ICUs settings were included. Each questionnaire item was described by mean and standard deviation. All statistical analyses were performed by SPSS 24.

The study was approved by the local Institutional Review Board and ethics committee of Shiraz University of Medical Science (Approval ID: IR.SUMS.REC.1398.1046).

3. Results

These results have shown that use VTE prophylaxis recommendation system frequently, various functions in the VTE prophylaxis recommendations system were well integrated, and imagine that most people would learn to use the VTE prophylaxis recommendations system very quickly have a high mean (mean over 4.14) from respondent perspective. The mean System Usability Scale (SUS) Score was 74.64. Overall measurement results are summarized in Table 1. Only one participant had commented about the usability of this system (“It was difficult to move the form page up and down when using the system”). Other participants had no comment.

Table 1. Descriptive Statistics Each Item of System Usability Scale (SUS) Questionnaire for VTE prophylaxis recommendation system

System Usability Scale (SUS)	Mean	S.D
I think I would like to use the VTE prophylaxis recommendations system frequently.	4.29	0.76
I found the VTE prophylaxis recommendations system unnecessarily complex.	2.00	0.58
I thought the VTE prophylaxis recommendations system was easy to use.	4.29	0.49
I think that I would need the support of a technical person to be able to use the VTE prophylaxis recommendations system.	1.57	0.79
I found the various functions in the VTE prophylaxis recommendations system were well integrated.	4.14	0.69
I thought there was too much inconsistency in the VTE prophylaxis recommendations system.	1.71	1.11
I would imagine that most people would learn to use the VTE prophylaxis recommendations system very quickly.	4.14	0.90
I found the VTE prophylaxis recommendations system very cumbersome to use.	2.29	0.95
I felt very confident using the VTE prophylaxis recommendations system.	3.43	1.72
I needed to learn a lot of things before I could get going with the VTE prophylaxis recommendations system.	2.00	1.15
System Usability Scale (SUS) Score	74.64	12.86

4. Discussion

This study indicated that the usability quality of the VTE Prophylaxis Recommendations System was good. To our best knowledge, this is the first study to investigate the usability of a homegrown venous thromboembolism prophylaxis recommendation system in the ICU setting of Nemazee hospital. These results concur with the other studies which have shown that clinical decision support system usability quality was good [10-12]. In contrast to some reports in the literature, there was usability quality of clinical information systems was low [7, 13]. This study indirectly implies that end-user involvement in the design of clinical decision support systems and homegrown clinical information systems improves the success and adoption rate of these systems. Also, multi-disciplinary collaboration increases the usability of clinical information systems.

The strength of our survey is that intensive care specialists assess the usability of the VTE Prophylaxis Recommendation System. The major limitation of this study is that the analysis does not enable to determine the usability of the VTE prophylaxis recommendation system by quantitate and qualitative methods in large scale, because this system implemented in ICUs of single center. Thus, the limitation of this study was the sample size. A further research requires to evaluate the usability of the VTE prophylaxis recommendations system by mixed methods (quantitate and qualitative methods) in large scale.

Summing up the results, it can be concluded that the usability quality of the VTE Prophylaxis recommendation system is good. Further research is required to evaluate the usability of the VTE Prophylaxis recommendation system by quantity and qualitative methods in on a large. In addition, iterative usability redesigns the process needs to improve the usability of the VTE Prophylaxis Recommendation System.

5. Acknowledgment

This article was extracted from a doctoral thesis (research ID: 97-01-07-19299) supported by the Research Vice-Chancellor of Shiraz University of Medical Sciences. We are profoundly thanks the healthcare professionals at ICUs of Nemazee Hospital in Shiraz, Iran because this paper would not have been possible without their friendly collaboration.

References

[1] Staggers N, Xiao Y, Chapman L. Debunking health IT usability myths. Appl Clin Inform. 2013; 4(2):241-50. doi: 10.4338/ACI-2013-03-IE-0016. PubMed PMID: 23874361.

[2] Guo J, Iribarren S, Kapsandoy S, Perri S, Staggers N. Usability Evaluation of An Electronic Medication Administration Record (eMAR) Application. Appl Clin Inform. 2011;2(2):202-24. doi: 10.4338/ACI-2011-01-RA-0004. PubMed PMID: 23616871.

[3] Schulte F, Fry E. Death by 1,000 clicks: Where electronic health records went wrong. Kaiser Health News. 2019; 18.

[4] Rizvi RF, Marquard JL, Hultman GM, Adam TJ, Harder KA, Melton GB. Usability Evaluation of Electronic Health Record System around Clinical Notes Usage-An Ethnographic Study. Appl Clin Inform. 2017;8(4):1095-105. Epub 12/14. doi: 10.4338/ACI-2017-04-RA-0067. PubMed PMID: 29241247.

[5] Chou D. Health IT and patient safety: building safer systems for better care. Jama. 2012;308(21):2282-.

[6] Carayon P, Salwei ME. Moving toward a sociotechnical systems approach to continuous health information technology design: the path forward for improving electronic health record usability and reducing clinician burnout. Journal of the American Medical Informatics Association. 2021; 28(5):1026-8.

[7] Viitanen J, Hyppönen H, Lääveri T, Vänskä J, Reponen J, Winblad I. National questionnaire study on clinical ICT systems proofs: physicians suffer from poor usability. International journal of medical informatics. 2011;80(10):708-25.

[8] Available from: https://www.iso.org/obp/ui#iso:std:iso:21902:ed-1:v1:en:term:3.19.

[9] Brooke J. SUS-A quick and dirty usability scale. Usability evaluation in industry. 1996;189(194):4-7.

[10] Genes N, Kim MS, Thum FL, Rivera L, Beato R, Song C, et al. Usability Evaluation of a Clinical Decision Support System for Geriatric ED Pain Treatment. Appl Clin Inform. 2016; 7(1):128-42. doi: 10.4338/ACI-2015-08-RA-0108. PubMed PMID: 27081412.

[11] Hudson D, Kushniruk A, Borycki E, Zuege DJ. Physician satisfaction with a critical care clinical information system using a multimethod evaluation of usability. International Journal of Medical Informatics. 2018;112:131-6. doi: https://doi.org/10.1016/j.ijmedinf.2018.01.010.

[12] von Dincklage F, Suchodolski K, Lichtner G, Friesdorf W, Podtschaske B, Ragaller M. Investigation of the usability of computerized critical care information systems in Germany. Journal of intensive care medicine. 2019;34(3):227-37.

[13] Bloom BM, Pott J, Thomas S, Gaunt DR, Hughes TC. Usability of electronic health record systems in UK EDs. Emergency Medicine Journal. 2021;38(6):410-5. doi: 10.1136/emermed-2020-210401.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210900

Development of a Clinical Decision Support System for Smart Algorithms in Emergency Medicine

Stefan VOGEL[a,1], Andreas REISWICH[b,] Zully RITTER[a], Michael SCHMUCKER[b,] Angela FUCHS[d], Kerstin PISCHEK-KOCH[a], Stefanie WACHE[a], Katrin ESSLINGER[d], Michael DIETRICH[c], Tibor KESZTYÜS[a], Dagmar KREFTING[a], Martin HAAG[b] and Sabine BLASCHKE[d]
[a] *Department of Medical Informatics, University Medical Center Göttingen, Germany*
[b] *GECKO Institute, Heilbronn University of Applied Sciences, Germany*
[c] *German Research Center for Artificial Intelligence Berlin, Germany*
[d] *Emergency Department, University Medical Center Göttingen, Germany*

Abstract. The development of clinical decision support systems (CDSS) is complex and requires user-centered planning of assistive interventions. Especially in the setting of emergency care requiring time-critical decisions and interventions, it is important to adapt a CDSS to the needs of the user in terms of acceptance, usability and utility. In the so-called ENSURE project, a user-centered approach was applied to develop the CDSS intervention. In the context of this paper, we present a path to the first mockup development for a CDSS interface by addressing Campbell's Five Rights within the CDSS workflow.

Keywords. Clinical Decision Support System, Emergency Medicine, Artificial Intelligence, Clinical Process

1. Introduction

Preclinical and clinical emergency medicine has developed into a complex professional field in recent years [1, 2]. One of the major problems represents the increase in the number of emergency patients presenting with a broad spectrum of illnesses and injuries ranging from outpatient care to intensive care [3]. Due to this problem, time-critical clinical decisions based on interdisciplinary medical knowledge represent the most important challenge in terms of patient safety and emergency care process efficiency. Clinical decision support systems (CDSS) are one approach to relieve clinicians, to contribute to patient safety and to improve clinical outcomes [4, 5]. Similar to driver assistance in automobile clinical decision support systems can provide clinically relevant indications at the point of care to support the decision-making process of emergency physicians. The type of support can take many forms ranging from alerts in case of deterioration of vital signs to predictions of possible diagnoses and appropriate therapies.

[1] Corresponding Author, Stefan Vogel, University Medical Center Göttingen, Department of Medical Informatics, D-37075 Göttingen, Germany; E-mail: stefan.vogel@med.uni-goettingen.de.

There are various technical options for implementing this support such as artificial intelligence (AI) approaches.

In the ENSURE project a prototypical CDSS to support timely and targeted diagnostics and initial therapy in emergency care will be developed and tested. It was the aim of the present work to design this CDSS intervention and develop initial software mock-ups for the future prototype.

2. Methods

The development of a CDSS is complex, costly and interdisciplinary [6]. For the development of the ENSURE prototype the user-centered design approach was applied [7]. To add value for the end users the Five Rights according to Campbell [8] were included in the planning of the CDSS intervention. Therefore, the following questions had to be clarified for the CDSS intervention: what is the right information, who is the right person, what is the right intervention format, what is the right channel and when is the right time in the workflow?

An essential part of the project is that two different information-processing methods will be implemented in the prototype. On the one hand a pattern-based approach based on the SOP book [10] and on the other hand a machine learning (ML) approach based on emergency data are implemented in order to support the diagnosis and initial therapy strategy. To evaluate the prototype, it is planned to randomly provide the user with one of the two approaches in a blind selection manner. In this way, it will be investigated whether one of the two approaches is superior in terms of performance and acceptance.

3. Results

In order to address the Five Rights and the user-centered design approach an experienced emergency physician took part in the conception phase. Based on the top 20 reasons for presentation in the Emergency Department (ED) [9] accounting for more than 80% of relevant emergencies, 30 emergency care algorithms resp. standard operating procedures (SOPs) as described by Blaschke and Walcher et al. [10] were identified.

Based on the SOPs 153 leading symptoms (in sum) associated to these 30 leading diagnoses were identified. The 153 leading symptoms were reduced to 96 in three iterations. Each iteration was based on the questions which leading symptoms are relevant to the 30 leading diagnoses, which information can be aggregated and whether there are symptom intersections between the leading diagnoses.

During emergency care this information is gathered and used for differential diagnosis. These aspects are of great importance for CDSS conception due to the fact that the intended diagnostic support has to process this gathered information.

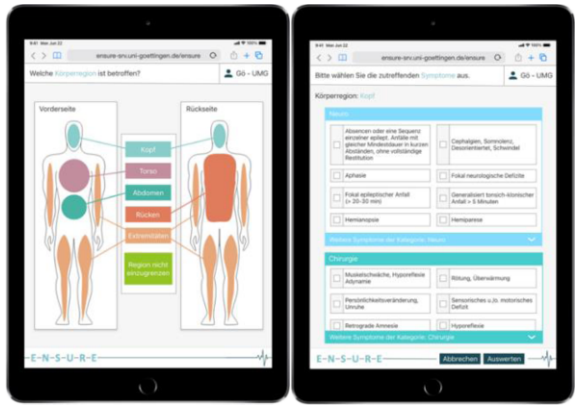


Figure 1. Mock-ups information aggregation of the CDSS prototype

Based on this preliminary work the first mock-ups (Fig. 1) were developed to allow users to collect information in a structured way using the future prototype. The 96 leading symptoms were assigned to five distinctive body regions as categories (head, thorax, abdomen, back, extremities). Another category "other" was introduced for symptoms that cannot be assigned to a specific body region. Furthermore, the symptoms were assigned to medical specialties to make the assessment intuitive and supportive in each emergency case situation.

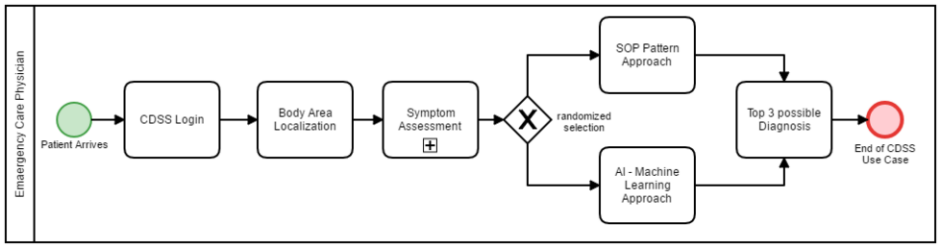


Figure 2. ENSURE CDSS prototype program flow created as BPMN with Camunda Modeler

Guided by the Five Rights and the user-centered design we were able to develop the prototypical program flow shown in Figure 2 and the first mock-ups in Figure 1.

4. Discussion

Emergency care is characterized by time-critical clinical decisions [11] to prevent possible long-term damage to patients or -in the worst case -even death. These conditions lead to a heavy burden on emergency physicians. Therefore, it is desirable to relieve this burden by providing a useful clinical supporting tool. Other research has shown that CDSS can have a positive impact in the emergency department [12]. To develop such a CDSS prototype, early involvement of experienced emergency physicians allows for a well-structured user-centered design.

In our research work, an SOP book with defined emergency care algorithms [10] was used for the development of a knowledge-based approach. The knowledge gained from this development will also be used in the development of an AI-based approach with ML algorithms. It is planned to develop different models for the ML approach and finally to use the model that has the best performance. Random forest models, support

vector machines and artificial neural networks of different types are addressed in the development. The ML approach is developed in parallel on the basis of real emergency case data. Both approaches will be used in the prototype evaluation.

5. Conclusion

In this project, we develop a CDSS prototype for clinical emergency care. The conceptualized program flow of the prototype as well as the first mock-ups were created. The finalization of the ML approach and the linkage of the individual components to a complete prototype will follow.

The future evaluation of the prototype, especially with respect to the comparison of the knowledge-based and ML approaches, is a promising endeavor to gradually transition CDSS into routine emergency care.

Acknowledgments

The ENSURE project was funded by the Federal Ministry of Health Germany (BMG) (Grant No. 2520DAT803).

References

[1] Gemeinsamer Bundesausschuss 19.4.2018 „Regelungen zu einem gestuften System von Notfallstrukturen in Krankenhäusern gemäß §136c Absatz 4 SGB V“, https://www.gba.de/beschluesse/3301/, accessed 25.06.2020

[2] BMG, Reform der Notfallversorgung – schnellere Hilfe im Notfall“ 22.7.2019, https://www.bundesgesundheitsministerium.de/notfallversorgung.html?fbclid=IwAR3k0u_DH7W4UEdzlgjamdFU-K3t4E0Ufyt3tcL9wPndVo6GMUYzMTETe9w, accessed 25.06.2020

[3] Schöpke T, Dodt C, Brachmann M, Schnieder W, Petersen PF, Böer J. Statusbericht aus deutschen Notaufnahmen. Notfall Rettungsmed 2014;17:660-670

[4] Sutton RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. npj Digit. Med. 2020;3:17

[5] Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support [published correction appears in J Am Med Inform Assoc. 2007 May-Jun;14(3):389]. J Am Med Inform Assoc. 2007;14(2):141-145. doi:10.1197/jamia.M2334

[6] Richter J, Vogel S. Illustration of Clinical Decision Support System Development Complexity. Stud Health Technol Inform. 2020 Jun 26;272:261-264. doi: 10.3233/SHTI200544. PMID: 32604651.

[7] Kashfi H. Applying a user centered design methodology in a clinical context. Stud Health Technol Inform. 2010;160(2):927-31.

[8] Campbell RJ. The five rights of clinical decision support: CDS tools helpful for meeting meaningful use. Journal of AHIMA. 2013 Oct;84(10):42-7.

[9] Greiner F, Brammen D, et al.: Standardisierte Erhebung von Vorstellungsgründen in der Notaufnahme. Med Klin Intensivmed Notfmed 2018;113:115-123

[10] Blaschke S, Walcher F (Hrsg). SOP Handbuch Interdisziplinäre Notaufnahme. Medizinisch Wissenschaftliche Verlagsgesellschaft (MWV), Berlin, 1. Auflage 03/2015, 2. Auflage 2021 (in press)

[11] Al-Azri NH. How to think like an emergency care provider: a conceptual mental model for decision making in emergency care. Int J Emerg Med 2020;13:17.

[12] Bennett P, Hardiker NR. The use of computerized clinical decision support systems in emergency care: a substantive review of the literature. J Am Med Inform Assoc. 2017 May 1;24(3):655-668. doi: 10.1093/jamia/ocw151. PMID: 28031285; PMCID: PMC7651902.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210901

Developing a Data Driven Approach for Early Detection of SIRS in Pediatric Intensive Care Using Automatically Labeled Training Data

Marcel MAST[a,1], Michael MARSCHOLLEK[a], ELISE STUDY GROUP[*], Thomas JACK[b] and Antje WULFF[a]
[a] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Karl-Wiechert-Allee 3, 30625 Hannover, Germany*
[b] *Department of Pediatric Cardiology and Intensive Care Medicine, Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany*
**ELISE STUDY GROUP: Louisa Bode[a]; Marcel Mast[a]; Antje Wulff[a]; Michael Marschollek[a]; Sven Schamer[b]; Henning Rathert[b]; Thomas Jack[b]; Philipp Beerbaum[b]; Nicole Rübsamen[c]; Julia Böhnke[c]; André Karch[c]; Pronaya Prosun[d]; Lena Wiese[d]; Christian Groszewski-Anders[e]; Andreas Haller[e], Mathias Slawik[e]; Torsten Frank[e]*
[a] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany*
[b] *Department of Pediatric Cardiology and Intensive Care Medicine, Hannover Medical School, Hannover, Germany*
[c] *Institute of Epidemiology and Social Medicine, University of Muenster, Muenster, Germany*
[d] *Research Group Bioinformatics, Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany*
[e] *medisite GmbH, Hannover, Germany*

Abstract. Critical care can benefit from analyzing data by machine learning approaches for supporting clinical routine and guiding clinical decision-making. Developing data-driven approaches for an early detection of systemic inflammatory response syndrome (SIRS) in patients of pediatric intensive care and exploring the possibility of an approach using training data sets labeled automatically beforehand by knowledge-based approaches rather than clinical experts. Using naïve Bayes classifier and an artificial neuronal network (ANN), trained with real data labeled by (1) domain experts ad (2) a knowledge-based decision support system (CDSS). Accuracies were evaluated by the data set labeled by domain experts using a 10-fold cross validation. The ANN approach trained with data labeled by domain experts yielded a specificity of 0.9139 and sensitivity of 0.8979, whereas the approach trained with a data set labeled by a knowledge-based CDSS achieves a specificity of 0.9220 and a sensitivity of 0.8887. ANN yielded promising results for data-driven detection of pediatric SIRS with real data. Our comparison shows the feasibility of using training data labeled automatically by knowledge-based approaches rather than manually allocated by experts.

[1] Corresponding Author: Marcel Mast, Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany. Email:marcel.mast@plri.de

Keywords. Pediatrics, SIRS, data-driven, machine learning

1. Introduction

Medical practice, especially within the pediatric intensive care, is a highly challenging and time pressuring domain. Long-term success of treatment and patient's outcome depend on the timely detection of underlying diseases and an early initiation of a therapy.[1] Nowadays, detection of serious diseases can be supported by analyzing routine clinical data though data-driven applications. Together with the increasing availability of medical data such approaches gain in importance for clinical settings.[2] Hence, data-driven approaches are an a recent considered topic in the field of research.[3, 4] One of these highly challenging diseases within the setting of the pediatric intensive care is the systemic inflammatory response syndrome (SIRS). In 2017, worldwide 2.9 million deaths associated with SIRS were reported for the group of children under five years.[5] To accommodate the differences between adults and children the International Pediatric Sepsis Consensus Conference (IPSCC) has modified the existing adult diagnostic criteria into six age-depending criteria groups for diagnosing SIRS in pediatric patients.[6] According to these criteria SIRS in children manifests when two of the following four criteria, from which one have to be an abnormal temperature or leukocyte count, are met: I) hyper- or hypothermia, II) tachycardia, III) tachy- or bradypnoea, IV) leukocyctosis or leukopenia.[6] The appropriate antimicrobial therapy should be started within an hour of the onset of first symptoms to avoid a two to five times higher mortality.[7] The use of clinical decision-support systems (CDSS) and data-driven approaches can support early detection of SIRS. However, such approaches are rarely used in the medical routine to support clinical decision making.[2–4] In the context of this work, we strive for developing a data-driven approach for the detection of SIRS within the pediatric intensive care. To achieve this goal we aim at implementing two approaches based on (I) a naïve Bayes classifier and (II) an artificial neuronal network (ANN). Both are trained using a ground truth created by domain experts. The manually creation of labeled data as used in this work is a highly time-consuming and challenging task. Therefore, the second objective of this work is evaluating the feasibility of using trainings data labeled automatically by a knowledge-based CDSS rather than using perfect ground truth label allocated manually.[1]

2. Methods

In this work we use an existing data set of routine data from the pediatric intensive care unit of the Hannover Medical School. This pseudonymized data set originates from a previously published study.[8] The data set consists of various vital parameters like temperature and blood pressure values, heart and respirations rates as well as laboratory test results and data from medical devices such as cooling blankets, ventilation and pacemaker for each of the included 168 pediatric patients. Laboratory test results include leukocyte, platelet and neutrophil counts as well as INR values derived from the prothrombin time. Every value comes with a specific timestamp documenting its measurement time by which a temporal sequence is ensured. Furthermore, the age of the respective patients is given, which is decisive for a correct pediatric SIRS diagnosis. In addition to the routine data, there is a ground truth describing whether patients suffered

from SIRS. For the generation of this ground truth two experienced pediatric intensive care physicians manually assessed the patients according to SIRS diagnostic rules defined by the IPSCC.[6] The resulting labeled data is used for training the data-driven algorithms. To improve the quality of the available data the data set has been cleaned and formatted. Furthermore different strategies for missing values were applied.[9] For ANN, missing values were replaced with previous occurred values within a parameter-dependent time window. Values that could not be replaced with that imputation strategy were replaced with NULL values. For the naïve Bayes classifier, missing values were ignored. The evaluation of the algorithms comprised splitting the available data set into trainings and test data set and running multiple 10-fold cross validation. Furthermore, the ANN were trained using data automatically labeled by a knowledge-based CDSS and evaluated with the ground truth allocated manually by domain experts.

3. Results

The trained algorithms were evaluated individually with the ground truth. The results of multiple 10-fold cross validations for every approach can be found in Table 1. The reported values represent the calculated average from all validation rounds. Folds were generated per patient to avoid splitting related data. The approach based on the naïve Bayes classifier achieves an accuracy from 0.5425 to 0.5683. Specificity ranges from 0.5848 to 0.6132 and sensitivity from 0.3510 to 0.4230. The ANN trained on data labeled by clinical experts achieves an accuracy of 0.8975, ranging from 0.8816 to 0.9174. The ANN using trainings data labelled by a knowledge-based CDSS achieves an accuracy of 0.8963, ranging from 0.8693 to 0.9263. This ANN shows a slightly higher specificity, but a lower sensitivity than the ANN trained with the data labeled by domain experts.

Table 1. Overview of aggregated results of multiple 10-fold cross validations

	Accuracy	Specificity	Sensitivity
Naïve Bayes Classifier	0.5512 [0.5425-0.5683]	0.6047 [0.5848-0.6132]	0.3894 [0.3510-0.4230]
ANN trained with data labeled by domain experts	0.8975 [0.8816-0.9174]	0.9139 [0.8558-0.9684]	0.8979 [0.8650-0.9231]
ANN trained with data labeled by a CDSS	0.8963 [0.8693-0.9263]	0.9220 [0.8357-0.9757]	0.8887 [0.8713-0.9163]

The architecture of the used ANN is shown in Table 2.

Table 2. Architecture of the used artificial neuronal network

Layer	Type	Number of neurons	Activation function
1	Input	8 x 1	
2	Hidden	6	tanh
3	Output	2	softmax

4. Discussion

With our work, we aimed at examining data-driven approaches for SIRS detection for patients of the pediatric intensive care. Furthermore, we strived to explore the feasibility of using a training data labeled by a knowledge-based CDSS instead of laborious labeled data by clinical experts. Limitations within this work are present in regards of transferability of the approach to other areas and the development and evaluation of the

used CDSS. Although the naïve Bayes classifier is well known as a solid baseline approach for starting data-driven research [9] our approach based on the naïve Bayes classifier reached a poor accuracy of 0.5512 for a binary classification problem. Even a comparably simpler classifier as ZeroR [10] that just predicts on basis of the majority class would have yielded a higher accuracy of approximately 0.74 for this use case. In contrast, the ANN achieved a promising accuracy. Moreover, literature review revealed that the developed ANN algorithm yielded better results than previously published data-driven approaches for SIRS detection in pediatric patients.[3, 4] The ANN using a data set labeled by a knowledge-based CDSS rather than clinical experts achieved relatively comparable results. This shows the feasibility of using such approaches more often in the future to relieve clinical experts and avoid laborious manual labeling of data sets. However, the development of such CDSS suited for providing label for this use case is a time-consuming task itself that could lead to vast amounts of training data.[1]

References

[1] Wulff A, Haarbrandt B, Tute E, Marschollek M, Beerbaum P, Jack T. An interoperable clinical decision-support system for early detection of SIRS in pediatric intensive care using openEHR. Artificial Intelligence in Medicine. 2018; 89:10–23.

[2] Shortliffe EH, Cimino JJ. Biomedical informatics: Computer applications in health care and biomedicine. 3rd edition. New York, NY: Springer. 2006.

[3] Islam MM, Nasrin T, Walther BA, Wu C-C, Yang H-C, Li Y-C. Prediction of sepsis patients using machine learning approach: A meta-analysis. Computer Methods and Programs in Biomedicine. 2019; 170:1–9.

[4] Wulff A, Montag S, Marschollek M, Jack T. Clinical Decision-Support Systems for Detection of Systemic Inflammatory Response Syndrome, Sepsis, and Septic Shock in Critically Ill Patients: A Systematic Review. Methods of Information in Medicine. 2019.

[5] World Health Organization. Sepsis; 2021 [cited 2021 Aug 24]. Available from: URL: https://www.who.int/news-room/fact-sheets/detail/sepsis.

[6] Goldstein B, Giroir B, Randolph A. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. Pediatric critical care medicine: a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies. 2005; 6(1):2–8.

[7] Scott L. Weiss, Mark J. Peters, Waleed Alhazzani, Michael S. D. Agus, Heidi R. Flori, David P. Inwald et al. Surviving sepsis campaign international guidelines for the management of septic shock and sepsis-associated organ dysfunction in children. Intensive Care Med, Springer. 2020; 46(1):10–67.

[8] Antje Wulff, M. Marschollek. Learning Healthcare Systems in Pediatrics: Cross-Institutional and Data-Driven Decision-Support for Intensive Care Environments (CADDIE). Studies in health technology and informatics. 2018; 251: 109-112.

[9] Berthold M, Borgelt C, Höppner F, Klawonn F. Guide to intelligent data analysis: How to intelligently make sense of real data. London: Springer; 2010.

[10] Witten IH, Frank E, Hall MA. Data mining: Practical machine learning tools and techniques. Third edition. Amsterdam: Morgan Kaufmann; 2011.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210902

Applying Social Network Analysis to Compare Dementia Caregiving Networks on Twitter in Hispanic and Black Communities

Sunmoo YOON[a,1], Carmela ALCANTARA[b], Nicole DAVIS[c], Peter BROADWELL[d], Haeyoung LEE[e], Amanda Bristol[a], Dante Tipiani[a], Joo Young Nho[f] and Mary MITTELMAN[f]
[a] *General Medicine, Department of Medicine, Columbia University, US*
[b] *School of Social Work, Columbia University, USA*
[c] *School of Nursing, Clemson University, USA*
[d] *Center for Interdisciplinary Digital Research, Stanford University, USA*
[e] *Department of Nursing, Chung-Ang University, South Korea*
[f] *Department of Psychiatry, NYU Grossman School of Medicine, USA*

Abstract. We applied social network analysis (SNA) on Tweets to compare Hispanic and Black dementia caregiving networks. We randomly extracted Tweets mentioning dementia caregiving and related terms from corpora collected daily via the Twitter API from September 1 to December 31, 2019 (initial corpus: n= 2,742,539 Tweets, random sample n= 549,380 English Tweets, n= 185,684 Spanish Tweets). After removing bot-generated Tweets, we first applied a lexicon-based demographic inference algorithm to automatically identify Tweets likely authored by Black and Hispanic individuals using Python (n= 114,511 English, n = 1,185 Spanish). Then, using ORA, we computed network measures at macro, meso, and micro levels and applied the Louvain clustering algorithm to detect groups within each Hispanic and Black caregiving network. Both networks contained a similar proportion of dyads and triads (Hispanic 88.2%, Black 88.9%), while the Black caregiving network included a slightly larger proportion of isolates (Hispanic 0.8%, Black 4.0%). This study provides useful baseline information on the composition of existing large groups and small groups. In addition, this work provides useful guidance for future recruitment strategies and the design of social support interventions regarding emotional needs for Hispanic and Black dementia caregivers.

Keywords. network analysis, social support, dementia, caregivers, social media

1. Introduction

The prevalence of Alzheimer's disease and related dementia (ADRD) is higher among Hispanics and Blacks than non-Hispanic Whites in the United States [1]. Among the proportion of people with dementia who live in the community rather than in assisted living facilities (70%-81%), most are cared for by family or friend caregivers [1]. A systematic review by Petosa and Smith found that peer social support is an effective contributor to meeting one's health goals [2]. Experimental evidence in behavioral science specifically highlights the critical role of dyad or triad friendship relations in

[1] Corresponding Author, Sunmoo Yoon, General Medicine, Department of Medicine, Columbia University Irving Medical Center, 630W 168 Street, PH105, New York, NY, 10032, USA; E-mail: sy2102@cumc.columbia.edu.

social support [3]. Although social scientists have reported a strong correlation between beneficial behavior and group cohesion in triad friendships since the mid 19th century, triad friendships and their associated benefits were rarely observed in previous social media-based studies of minority dementia caregivers [4].

Social network analysis is commonly used to assess social structures such as the composition of groups (e.g., large groups, small groups, triads, and dyads) and dynamics (e.g., relationships, links, and temporal interactions) using network and graph theory [4]. Social network analysis of online communities, particularly Twitter, may provide insights to aid the design of culturally sensitive and customized social support interventions for Hispanic and Black family caregivers for persons with dementia [4, 5]. This study aims to apply social network analysis to a large corpus of Tweets to gain insights about Hispanic and Black dementia caregiving networks.

2. Methods

We applied network analysis techniques to publicly available Twitter messages (Tweets) mentioning dementia (keywords/hashtags: #dementia, #demencia, #Alzheimers). We used the Twitter Application Programming Interface (API), Python, and a High-Performance Computing (HPC) Cluster (https://osf.io/qruf3/) to extract these messages from a random sample of publicly available Tweets collected daily using via the Twitter API from September 1 to December 31, 2019 (n= 2,742,539 Tweets). Our team's behavioral domain expert and clinician first extracted and categorized U.S.-based Tweets from the initial corpus (n= 549,380 English Tweets, n= 185,684 Spanish Tweets). Second, we applied a Twitter bot detection algorithm to remove bot-generated Tweets using the Botometer Python API (https://github.com/IUNetSci/botometer-python). Third, we applied a lexicon-based demographic inference algorithm to automatically identify Tweets likely authored by Black or Hispanic individuals (n= 114,511 English, n = 1,185 Spanish, https://osf.io/qruf3/). Fourth, we applied Social Network Analysis to the winnowed Tweet corpus to automatically detect similar communities and to identify stakeholders with a high degree of network centrality. Fifth, to avoid algorithm dependency, we applied and compared the results of five clustering algorithms: Leiden, Principal Component Analysis, Louvain, CONCOR, and Newman. Sixth, network measures at the macro, meso, and micro levels were calculated to compare Black and Hispanic dementia caregiving networks within Twitter using the ORA meta-network analysis software. Seventh, we visualized Black and Hispanic dementia caregiving networks within Twitter at macro, meso, and micro levels. Lastly, behavioral science experts and a clinician evaluated the results according to their clinical meaningfulness for the design and recruitment strategy of a social support intervention for Hispanic and Black dementia caregivers on Twitter. The larger study was approved by the Institutional Review Board (IRB).

3. Results

Dementia caregiving-related Tweets potentially created by 5,321 Black users were disseminated to 280,042,783 users, with retweets outnumbering original messages by a factor of 11. In contrast, the dementia caregiving-related Tweets in Spanish created by 409 users were disseminated to 6,141,424 users at a ratio of 11,798 retweets to each

original message. The top 10 stakeholders who disseminated dementia caregiving-related messages in English were identified as media outlets, a lay person, a psychologist, a social worker, and an association; among Spanish-language tweets, the top stakeholders were an individual, a media outlet, a lawyer, a community organization, and an economist. 123 groups (17.0%, Louvain modularity value: 0.802) were found in the Black dementia caregiving network, whereas 14 distinct groups (11.0%, Louvain modularity value: 0.801) were detected in the Hispanic caregiving network.

Both networks contained a similar proportion of dyads or triads (Hispanics: 88.2% dyads or triads, Blacks: 88.9%), while the Black caregiving network included a slightly larger proportion of isolates (Hispanics: 0.8%, Blacks: 4.0%) (Figure 1).

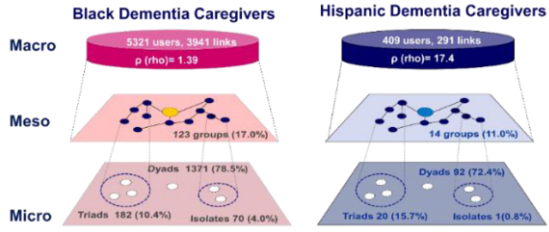


Figure 1. Characteristics of macro, meso, and micro level Twitter networks between Hispanic and Black dementia caregivers

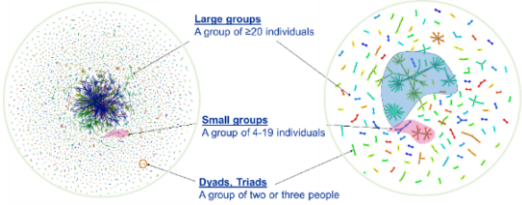


Figure 2. Social network structure of dyads, triads, and communities of Black and Hispanic dementia caregiving networks from September to December 2021

4. Discussion and Conclusions

This study examined the characteristics of macro, meso, and micro-level social network structures among Hispanic and Black dementia caregiving networks on Twitter. In the macro-level social networking structure, a substantially larger number of Black dementia caregiving network users were engaging in the conversation than in the Hispanic dementia caregiving network in the U.S. (5321 users, 3941 retweet activities vs. 409 users, 291 retweet activities) over three months. According to the U.S census in 2020, the total percentage of the population identifying as Black is smaller than the total number of Hispanics (13.4% vs. 18.5%).

Consistent with the existing literature on the strong presence and cohesiveness among Black communities in Twitter in the U.S., our findings confirms that cohesive communities among Black networks on Twitter are present in the context of dementia caregiving. Indeed, more community groups were observed in the Black dementia caregiving network than in the Hispanic dementia caregiving network in the U.S. (123 groups 17.0% vs. 14 groups, 11.0%) in the meso level social networking structure. Social cohesiveness, a determinant for population health (Healthy People 2030), has been referred to as the way in which individuals within a society are bound together by

attitudes and behaviors. Social cohesion and social support has been associated with decreased depression [6], stress and emotional burden among primary family caregivers of older adults [7].

A novel finding of this study is that the Hispanic dementia caregiving network included a very small proportion (0.8%) of isolates who express themselves without quoting others' messages. Not only is this number smaller than in the Black caregiving network (4.0%), but it is unusual in the context of social networks that develop around diseases such as HIV or COVID-19 on Twitter. Further, the social science-based concept of the optimal number of community members, which in these cases range from two to a hundred people, has not been explored in the context of providing social support for dementia caregivers in the Twitter space. Studies with an interventional design using Twitter are needed to investigate each community's optimal number, which then will enable interventions to deliver social support messages meaningfully and cost-effectively to racial/ethnic minority dementia caregivers.

The Louvain social network clustering algorithm was selected after comparing the utility of the five clustering algorithms listed above. The performance of several of the alternative clustering algorithms on our dataset was suboptimal for the following reasons: 1) Leiden: degree of clustering performance is not available; 2) Principal Component Analysis: results included one large dominant cluster, which is clinically less meaningful; 3) CONCOR: prone to subjective bias; 4) Newman: clustering modularity score was slightly lower than that computed via Louvain (0.801 vs. 0.802).

In conclusion, this study provides useful baseline information on the composition of existing large groups, small groups, and isolates among Hispanic and Black dementia caregivers on Twitter, which will help guide the design and future recruitment strategies of social support interventions involving community engagement and partnership in these minoritized communities.

Acknowledgments

U.S. federal grant TweetS2 R01AG060929 (PI: Yoon).

References

[1] Alzheimer's Ass. Race, Ethnicity, and Alzheimer's in America. Alzheimer's & Dementia. 2021;17(3).
[2] Petosa R, Smith LH. Peer mentoring for health behavior change: A systematic review. American Journal of Health Education. 2014;45(6):351-7.
[3] Simmel G. The sociology of georg simmel. Wolff KH, editor: The Free Press; 1964.
[4] Odlum M, Hwayoung CH, Broadwell P, Davis, Patrao M, Schauer D, Bales ME, Alcantara C, Yoon S. Application of Topic Modeling to Tweets to Learn Insights on the African American Lived Experience of COVID-19. Studies in health technology and informatics. 2020;272:24.
[5] Yoon S, Lucero R, Mittelman MS, Luchsinger JA, Bakken S. Mining Twitter to Inform the Design of Online Interventions for Hispanic Alzheimer's Disease and Related Dementias Caregivers. Hispanic Health Care International. 2020;18(3):138-43.
[6] Mittelman MS, Roth DL, Coon DW, Haley WE. Sustained benefit of supportive intervention for depressive symptoms in Alzheimer's caregivers. Am J Psychiatry. 2004 May;161(5):850-6.PMID: 15121650.
[7] Naganuma Y, Yamaoka K, Takahashi K. Relationship between social cohesion and the care burden of primary family caregivers in central Tokyo, Japan. Health Sci Rep. 2021;4(1):e238.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210903

The Biopsychosocial Actor in the Contexts of Health and Human Services Informatics

Jaakko HAKULA[a,1]
[a] *Department of Health and Social Management, University of Eastern Finland, Finland*

Abstract. The paradigm for health and human services informatics (HHSI) was developed by Finnish researchers. The four entities of the HHSI paradigm and their interrelations form the basics for informatics research and education in the University of Eastern Finland. The focus of the essay is on the entities of actors and action related to different conceptions of agency. The entities of data and technology are the backbones of digitalization. The further aim of the study is to modernize the Holistic Concept of Man (HCM) metaphor to take the form of the biopsychosocial (BPS) actor. The HCM metaphor with its Husserlian-Heideggerian backgrounds is renovated towards a more realistic model of an individual actor or decision-maker described by the BSP model. As the BSP actor is embedded in the contexts of the HHSI paradigm, the notion of the BPS-D actor or decision-maker emerges. The BPS-D actor is a hybrid agent, who has cognitive, emotional, informational, and action-oriented connections to other possible agencies and artificial systems in the digitalized encounter. The very context of future research is the HHSI neo-paradigm.

Keywords. HHSI paradigm, agency, HCM metaphor, BPS model, BPS-D actor

1. Introduction

The term informatics encompasses people, information, and technology. The paradigm for health and human services informatics (HHSI) is related to health and biomedical informatics. The HHSI paradigm was developed along the utilization of information technology in the health and social sciences in the University of Eastern Finland, where it is widely used in research and education. It is composed of four main entities – actors, action, data, and technology [1]. Health and human services informatics confronts many practical challenges in the integration of multifaceted ontologies in and between healthcare and social services [2]. Especially the role of artificial intelligence (AI) in future health technology assessment (HTA) is becoming increasingly important [3].

Real-world research practices define the order of preferences between the main entities of the HHSI paradigm. Depending on the emphasis of argumentation between the paradigmatic entities and/or the links, the contents of research and education vary. Research on information management in service processes is determined by the combination of data, as the value chain of information, and action. Research on action and technology produces evaluation and development research on information and communication technology. The combination of technology and actors generates

[1] Corresponding Author, Jaakko Hakula, Department of Health and Social Management, University of Eastern Finland, 70211 Kuopio, Finland; E-mail: jaakko.hakula@dnainternet.net.

research on knowledge management. Scientific and developmental work on the body of knowledge originate in the interrelationship between actors and data as information [1].

In 1977 Engel presented the biopsychosocial (BPS) model to broaden restrictions of the dominating biomedical model [4]. The BPS model inherits its holistic, multilevel ontology, and epistemological and methodological essence from general systems theory [5]. The abstraction of the nested organization of micro, meso, and macro levels of natural systems is the basis for scientific research of the BPS model. The holistic structure of the BPS model with functionalistic properties is of use also in making every day encounters of cure and care more humane [5-6].

Agency as human intentional action, and as possible, non-human artificial property, are not strictly separated in informatics research. Whether artificial technical systems fulfill the definition of agency is debatable. Human actors (as agents) acting in contexts of digitalized information systems communicate and cooperate individually sharing the agentic environment with another individual actor, or collectively in larger coalitions [7].

The three-partite Holistic Concept of Man (HCM) represents a philosophic metaphor of agency, based on Husserlian-Heideggerian schools of thought. The HCM metaphor originates in works of the late Lauri Rauhala, a Finnish philosopher, and of Pekka Pihlanto, a professor emeritus of management accounting. The HCM metaphor consists of consciousness, corporeality, and situationality. It emphasizes the role of individual actors in everyday contexts of life. It has been applied mainly by Finnish researchers, e.g. in areas of management, information sciences, and rehabilitation [8]. The aim of the study is to sketch a novel model of agency in the context of the HHSI paradigm.

2. Materials and methods

Materials and methods are closely connected to each other. The HHSI paradigm with its entities and their functions, the HCM metaphor, and the BPS model are the materials of the study that are conceptually analyzed and resynthesized. The "end-product" is but an instantiation of metamorphosis towards something new to be applied in future studies, e.g. in areas of management, information sciences, and rehabilitation.

The methods could be broadly called philosophical as the conceptions in relation to the language and drawings used reflect various historical schools of scientific thinking. Informatics, management and technology in healthcare are an interdisciplinary endeavour. Materials and methods are supported by general systemic principles.

3. Results

Figure 1 depicts the BPS actor reformulated from the HCM metaphor, which stands for the initial actor in the digitalized context steered by the HHSI paradigm. The HHSI paradigm as a whole is presented divergent from the original pictorial expression but the ontologies and interrelations of the paradigmatic entities are considered to remain as earlier defined [1]. Human agency as the BPS actor is positioned in the center of responsible action. The BSP actor inherits the concepts of mind and body, and activities towards objects from the HCM metaphor. However, the main tenets of the HCM metaphor − consciousness, corporeality and situationality − are replaced by those of psychological, biological and social determinants of the BPS model [4-6].

The HHSI neo-paradigm in Figure 1 reflects the three basic onto-epistemological and ethical properties of holistic conceptions of health and disease from an individual perspective. Cognitive and emotional factors intermingle with general and special knowledge and information [9]. Motor action is closely linked to cognition and emotion [10] (omitted in the contexts of the HCM metaphor). The BPS model can be expanded to digitalized encounters: the biopsychosocial-digital (BPS-D) actor as a hybrid agent is imagined to have human, intentional properties, and possible artificial elements.

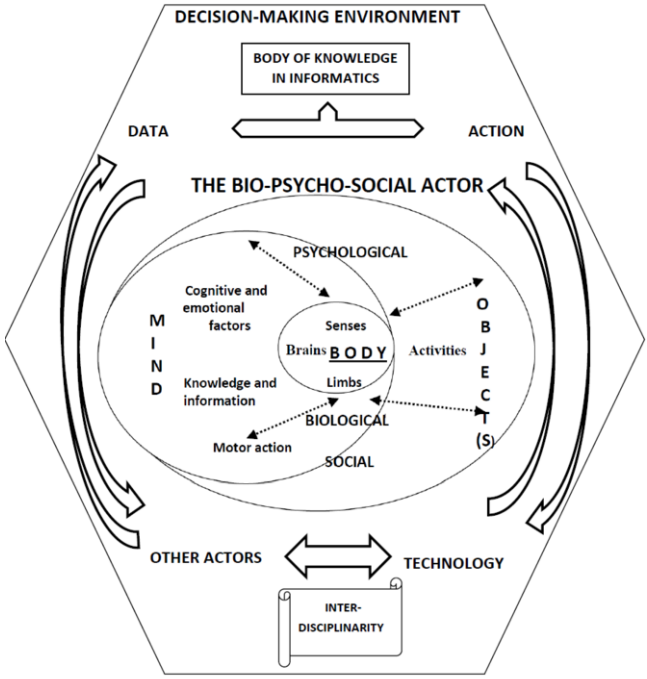


Figure 1. The BPS-actor and the HHSI neo-paradigmatic expansion

The BPS actor may represent a human expert or a layman but not an artificial system. Artificial systems are embedded in all of the entities of the HHSI (neo)-paradigm, with their interconnections, although they might be called artificial agents metaphorically. The BPS-D actor evolves in the systems of the neo-paradigmatic environment.

4. Discussion

The notion of the BPS-D actor, with its hybridity, amplifies the HHSI neo-paradigm towards more individually inclined research to be made both in information management of healthcare and social services, patient care, and rehabilitation. The problem of systemic, holistic models is their complexity: everything has an impact on everything. Quality and/or quantity – validity and repeatability – where is the relevance?

The HHSI (neo)-paradigm actualizes as a conglomerate of multiple paradigms that support transdisciplinary research and education – and indirectly everyday work

activities in health and human services informatics. Anomalies of the HHSI paradigm have been scarcely studied, which might justify a deeper thought on the Kuhnian symbolism. One relevant topic in this respect could be different scopes of agency.

References

[1] Kuusisto-Niemi S, Saranto K, Rissanen S. Informatics in social science: research, developments, and outcomes from the Finnish perspectives. In: Tavana M, Ghapanchi AH, Talaei-Khoei A, editors. Healthcare informatics and analytics: emerging issues and trends [Internet]. Hershey (PA): IGI Global, 2015 [cited 2021 Jan 04]. Chapter 2. Available from: https://www.igi-global.com/chapter/informatics-in-social-services/115105 DOI: 10.4018/978-1-4666-6316-9.ch002.

[2] Fichtenberg C, Delva J, Minyard K, Gottlieb LM. Health and human services integration: generating sustained health and equity improvement. Health Affairs [Internet]. 2020 Apr 20 [cited 2021 Jan 12];39(4):567-573. Available from: https://www.healthaffairs.org/doi/10.1377/hlthaff.2019.01594 DOI:10.1377/hlthaff.2019.01594.

[3] Alami H, Lehoux P, Auclair Y, de Guise M, Gagnon MP, Shaw J, Roy D, Fleet R, Ahmed MA, Fortin JP. Artificial intelligence and health technology assessment: anticipating a new level of complexity. Journal of medical Internet research. 2020;22(7):e17707.

[4] Bolton D. The biopsychosocial model and the new medical humanism. Archives de Philosophie [Internet].2020 Oct 28 [cited 2021 Sep 01];83(4):13-40. Available from: https://www.cairn-int.info/article.php?ID_ARTICLE=E_APHI_834_0013 DOI:10.3917/aphi.834.0013.

[5] Karunamuni N, Imayama I, Goonetilleke D. Pathways to well-being: Untangling the causal relationships among biopsychosocial variables. Social Science & Medicine [Internet]. 2020 Feb 10 [cited 2021 Sep 01];12 p. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0277953620300654?via%3Dihub DOI:10.1016/j.socscimed.2020.112846.

[6] Ahmadvand A, Gatchel R, Brownstein J, Nissen L. The Biopsychosocial-Digital Approach to Health and Disease: Call for a Paradigm Expansion. J Med Internet Res [Internet]. 2018 May [cited 2021 Aug 20];20(5):e189. Available from: https://www.jmir.org/2018/5/e189/ DOI: 10.2196/jmir.9732.

[7] Schlosser M. Agency. [Internet]. The Stanford Encyclopedia of Philosophy (Winter 2019 Edition). Edward N. Zalta, editor. Stanford (CA): Stanford University, Center for Study of Language and Information, the Metaphysics Research Lab; 2019 Oct 28 [cited 2021 Aug 30]. Available from: https://stanford.library.sydney.edu.au/archives/win2019/entries/agency/.

[8] Markopoulos E, Vanharanta H.Delphic Maxims Based Applied Philosophy for Business and Governance Management. In: Kantola J, Barath T, Nazir S, Andre T, editors. Advances in human factors, business management, training and education. Advances in Intelligent Systems and Computing, vol 498 trends [Internet]. Cham:Springer;2017 [cited 2021 Aug 28]. Available from: https://link.springer.com/chapter/10.1007%2F978-3-319-42070-7_4 DOI:10.1007/978-3-319-42070-7_4.

[9] Dionne SD, Gooty J, Yammarino FJ, Sayama H. Decision making in crisis: A multilevel model of the interplay between cognitions and emotions. Organizational Psychology Review [Internet]. 2018 Oct 28 [cited 2021 Sep 01];8(2-3):95-124. Available from: https://journals.sagepub.com/doi/10.1177/2041386618756063 DOI: 10.1177/2041386618756063.

[10] Glenberg AM, Witt JK, Metcalfe J. From the Revolution to Embodiment: 25 Years of Cognitive Psychology. Perspect Psychol Sci [Internet]. 2013 Sep; [cited 2021 Sep 08];8(5):573-85. Available from: https://pubmed.ncbi.nlm.nih.gov/26173215/ DOI:10.1177/1745691613498098.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210904

Towards Harmonized Data Quality in the Medical Informatics Initiative – Current State and Future Directions

Matthias LÖBE[a,1], Gaetan KAMDJE-WABO[b], Adriana Carina SINZA[c], Helmut SPENGLER[d], Marcus STROBEL[e] and Erik TUTE[f]
[a] *Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Germany*
[b] *Center for Preventive Medicine and Digital Health, Department of Biomedical Informatics, Medical Faculty Mannheim of the University of Heidelberg, Germany*
[c] *Data Integration Center, University Hospital Aachen, Germany*
[d] *Institute for Artificial Intelligence and Informatics in Medicine, Technical University of Munich, Germany*
[e] *Data Integration Center, University Hospital Leipzig, Germany*
[f] *Peter L. Reichertz Institute for Medical Informatics (PLRI), Hannover Medical School, Germany*

Abstract. Health data from hospital information systems are valuable sources for medical research but have known issues in terms of data quality. In a nationwide data integration project in Germany, health care data from all participating university hospitals are being pooled and refined in local centers. As there is currently no overarching agreement on how to deal with errors and implausibilities, meetings were held to discuss the current status and the need to develop consensual measures at the organizational and technical levels. This paper analyzes the discovered similarities and differences. The result shows that although data quality checks are carried out at all sites, there is a lack of both centrally coordinated data quality indicators and a formalization of plausibility rules as well as a repository for automatic querying of the rules, for example in ETL processes.

Keywords. Data Quality, Electronic health record, Medical Informatics Initiative

1. Introduction

Electronic health record (EHR) data from health care information systems have particular, well-recognized weaknesses in the area of data quality due to background of their collection (treatment, not completeness, as the primary purpose, limited human resources for documentation in hospitals, complex real-life processes with divergences between medical treatment and technical recording) [1]. As a result, some clinical researchers, as well as biometricians, have reservations about using EHR data for

[1] Corresponding Author, Matthias Löbe, Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany; E-mail: matthias.loebe@imise.uni-leipzig.de.

research and do not see them as equivalent to prospectively collected data backed by data management processes as in clinical trials.

In the German medical informatics initiative (MII) [2], data integration centers (DIC) have been established since 2018 at currently 29 university hospitals, which aim to make healthcare data available for research. The DICs are each assigned to one of four consortia [2], whereby each consortium is based on different technical information architectures. To ensure a comparable data corpus in all DICs, a central core data set was agreed upon, which contains modules from different areas (demographics, encounter, diagnoses, procedures, laboratory data, medications, etc.). The core data set is specified in HL7 FHIR format.

In the last years, the DICs have focused primarily on connecting sources to assemble a comprehensive data pool. Aspects of data quality have not been a focus of the work. However, for the broad and partially automated sharing of the data envisaged for the future, measures have to be developed that ensure the data quality required for each use case. This was exemplified in the so-called MII demonstrator study [3].

2. Methods

Within a joint working group, a series of questions was developed and sent to designated representatives of the consortia. These questions queried aspects of six complexes: 1) the organizational structures set up locally, 2) the technical infrastructure, 3) the handling of errors in the process of extracting, transforming, and loading from the primary information systems, 4) the curation processes in the research database, 5) the plan for developing further functionalities, and 6) the envisaged need for central conventions and consensual data quality rules. Responses were presented and discussed in a workshop.

3. Results

The topic ‘data quality’ was recognized in its importance by all consortia and addressed in a surprisingly comparable depth. There are local working groups that deal with conceptual, implementation-related and organizational aspects. These groups are also networked with other groups within the consortia (e.g., those responsible for data extractions from primary systems), but the data quality assurance procedures outlined are not yet universally used in routine operations.

However, the implementation status of data quality measures varies considerably *within* individual sites of a consortium. For the most part, pilot sites exist that also have lead responsibility for developing the concepts and tools. It was noticeable that beside the long scientific history, existing data quality approaches were only used to a limited extent. New technical developments are used in the majority of cases. Available information systems, e.g., from the area of clinical trials or cohorts, which address data quality problems technically [4], e.g., through dedicated query management or data curation boards, are not used. This may be due in part to the fact that the size of the project required the use of many staff from fields other than medical informatics. As the ETL pipelines mature and are more widely used, a deeper understanding of the structure and limitations of the local data bodies is now currently emerging.

[2] https://www.medizininformatik-initiative.de/en/consortia/data-integration-centres

Finally, it should be noted that the researchers' view of "errors" must also be questioned. Not every case of complaint are real errors, because in reality, the hospital data are recorded at a very detailed level, which has to serve different purposes of use. This includes provisional values, cancellations, error corrections, recoding, and similar operations that are not present in the smoothed view that is usually presented to researchers.

Differences between the consortia naturally concern the technical implementation. Since the underlying grant program was competitively bid, the information architecture is the same within consortium sites, but different between the consortia. The technical architecture of the ETL pipeline in DIFUTURE integrates comprehensive event logging through which data quality is viewed in a structured and detailed manner [5]. An audit service allows flexibly configurable quality analyses, which are executed on a SQL-based data mart but are independent of concrete schemas.

In HiGHmed, there is a special focus on data governance. Comprehensive organizational structures have been established including data stewards for modeling clinical concepts of a domain, responsible parties for each source system, and a data reviewing board for overall, regular analysis of data sets. The in-house development openCQA [6] makes commonly governed compilations (e.g. for reports or dashboards) of various data quality indicators applicable on HiGHmed's technical architecture.

MIRACUM stores important data quality indicators and rules into a metadata repository (MDR) [7]. The self-developed software DQAstats generates detailed, cross-site standardized error reports in PDF format on a quarterly basis to ensure, monitor and document the measures taken [8]. These reports are published anonymously in the consortium for comparison and self-assessment.

The SMITH consortium has designed a five-stage data quality assurance concept, which, starting with manual tests based on a coordinated catalog of data quality indicators and at defined intervals, continuing with automated procedures (in development), also plans for the use of central terminology services, the connection of a metadata repository and natural language processing of free text annotations.

4. Discussion and Future Work

In summary, promising approaches have been developed that now need to be rolled out across the range of consortia sites, put into operation, and feedback incorporated. Nevertheless, objective evidence of the qualitative suitability of the extracted data for the variety of potential research projects is still lacking. This will require the involvement of a broader community of domain experts from other areas of biomedical research such as biometrics and epidemiology, as they have already addressed a variety of similar problems and developed strategies to solve them. Another need is seen in the training and further qualification of staff, which will lead to a more effective involvement of specific and harmonized Data Stewards abilities across the different consortia. Furthermore, it is considered to construct a plausible clinical question to query real data and to test known typical error constellations on the shared data in a cross-consortium 'projectathon'. Those involved in the workshop agreed to use the framework of Kahn et. al. [9] as a taxonomy of error types.

In the longer term, common solutions are to be developed in three sub-areas. First, this concerns the development of a system of harmonized data quality indicators and rules for their operationalization. This approach should separate the conceptual

specification from the syntax for execution in order to be able to support different target platforms. It seems worthwhile to define common data quality indicators for certain common data elements of the MII core data set (e.g., LOINC Top 300 most common laboratory values) and also to define metrics or thresholds for quality assurance in data quality assessments.

Secondly, a repository as a storage and access location for the harmonized rules would be desirable. In all consortia, a Metadata Repository is already in productive use or at least announced. In principle, these systems could also be used to manage the data quality indicators and rules as outlined by MIRACUM, if suitable programming interfaces for mutual access or syndication become available. Further workarounds, such as the making use of some GIT-based repositories could also for this purpose be investigated, so that a distributed and location-wide access is ensured towards a continuous MDR-based assessment of data quality indicators.

The third issue relates to the availability of reference data sets for data validation. Many advanced data quality issues cannot be identified by looking at locally available data alone. Comparison to a gold standard is needed. This gold standard can vary widely in nature and scope; as an example, a comparison of the frequency of distribution of certain numbers of cases or answer categories between the local site and a larger (and thus more statistically robust) number of cases would be useful in identifying systemic biases. With this in mind, combined, i.e., aggregated reference data from all MII sites would be beneficial.

Acknowledgement

The work is supported by the corresponding BMBF/MII grants (01ZZ1803A, 01ZZ1801E, 01ZZ1803B, 01ZZ1804A, 01ZZ1803D, 01ZZ1802C, 01ZZ1910A).

References

[1] Kohane IS, Aronow BJ, Avillach P, Beaulieu-Jones BK, Bellazzi R, Bradford RL et al. What Every Reader Should Know About Studies Using Electronic Health Record Data but May Be Afraid to Ask. J Med Internet Res 2021;23(3):e22219. doi: 10.2196/22219.

[2] Semler S, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med 2018;57(S 01):e50-e56. doi: 10.3414/ME18-03-0003.

[3] Kamdje-Wabo G, Gradinger T, Löbe M, Lodahl R, Seuchter SA, Sax U et al. Towards Structured Data Quality Assessment in the German Medical Informatics Initiative: Initial Approach in the MII Demonstrator Study. Stud Health Technol Inform 2019;264:1508–9. doi: 10.3233/SHTI190508.

[4] Löbe M, Meineke F, Winter A. Scenarios for Using OpenClinica in Academic Clinical Trials. Stud Health Technol Inform 2019;258:211–5. doi: 10.3233/978-1-61499-959-1-211.

[5] Spengler H, Lang C, Mahapatra T, Gatz I, Kuhn KA, Prasser F. Enabling Agile Clinical and Translational Data Warehousing: Platform Development and Evaluation. JMIR Med Inform 2020;8(7):e15918

[6] Tute E, Scheffner I, Marschollek M. A method for interoperable knowledge-based data quality assessment. BMC Med Inform Decis Mak 2021;21(1):93. doi: 10.1186/s12911-021-01458-1.

[7] Kadioglu D, Breil B, Knell C, Lablans M, Mate S, Schlue D et al. Samply.MDR - A Metadata Repository and Its Application in Various Research Networks. Stud Health Technol Inform 2018;253:50–4.

[8] Kapsner LA, Kampf MO, Seuchter SA, Kamdje-Wabo G, Gradinger T, Ganslandt T et al. Moving Towards an EHR Data Quality Framework: The MIRACUM Approach. Stud Health Technol Inform 2019;267:247–53. doi: 10.3233/SHTI190834.

[9] Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC) 2016;4(1):1244. doi: 10.13063/2327-9214.1244.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210905

The Linkage Between Bone Densitometry and Cardiovascular Disease

Mahmoud A. REFAEE[a,b], Hamada R.H. AL-ABSI[a], Mohammad Tariqul ISLAM[c], Mowafa HOUSEH[a], Zubair SHAH[a], M.Sohel RAHMAN[d] and Tanvir ALAM[a,1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*
[b] *Hamad Medical Corporation, Doha, Qatar*
[c] *Computer Science Department, Southern Connecticut State University, New Haven, CT 06515, USA*
[d] *Department of Computer Science & Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh*

Abstract. Dual-energy X-ray absorptiometry (DXA) has been traditionally used to assess body composition covering bone, fat and muscle content. Cardiovascular disease (CVD) has deleterious effects on bone health and fat composition. Therefore, early detection of bone health, fat and muscle composition would help to anticipate a proper diagnosis and treatment plan for CVD patients. In this study, we leveraged machine learning (ML)-based models to predict CVD using DXA, demonstrating that it can be considered an innovative approach for early detection of CVD. We leveraged state-of-the-art ML models to classify the CVD group from non-CVD group. The proposed logistic regression-based model achieved nearly 80% accuracy. Overall, the bone mineral density, fat content, muscle mass and bone surface area measurements were elevated in the CVD group compared to non-CVD group. Ablation study revealed a more successful discriminatory power of fat content and bone mineral density than muscle mass and bone areas. To the best of our knowledge, this work is the first ML model to reveal the association between DXA measurements and CVD in the Qatari population. We believe this study will open new avenues of introducing DXA in creating the diagnosis and treatment plan of cardiovascular diseases.

Keywords. Cardiovascular disease, Dual-energy X-ray absorptiometry (DXA), Bone densitometry, Qatar Biobank (QBB)

1. Introduction

Dual X-ray Absorptiometry (DXA) is a fast and non-invasive approach to measure the body composition in terms of mass, fat and bone composition. Using DXA, the entire body can be scanned to calculate the composition of bone mass and soft tissue. DXA machine can also reveal results of fat content and lean muscle of the body when a full-body scanning is performed [1]. DXA is a standard method for evaluating bone mass and body composition for several reasons. First, the two X-ray attenuations from the DXA machine that pass through the body can measure the mass using the simple algebra and physical properties of materials. It can also measure the regional body compositions by

[1] Corresponding Author, Tanvir Alam, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar; E-mail: talam@hbku.edu.qa.

subdividing the whole body into well-defined body cut lines. Overall, it is a precise and stable measurement technique that has been used in clinics for several years. By using DXA Phantoms, it is quite simple to also verify measurement stability (nearly 0.5% variation) for body composition [1]. Recently, Hologic Inc. has started to use DXA for the prediction of cardiovascular diseases, including coronary heart disease, stroke, etc. This process is already approved by the FDA [2]. Lastly, this method can facilitate the discovery of comorbidities and mortality in cohorts from multiple ethnicities, as well.

According to the WHO, cardiovascular disease (CVD) is considered as the primary cause of fatality worldwide. In its 2018 report, The World Health Organization (WHO) stated that in 2016, non-communicable diseases (NCD) account for more than 70% of the global death. Out of this, CVDs contributed to over 44% of total NCD-related mortality. In the Gulf Cooperation Council Countries, CVD is the leading cause of fatality [3]. The Ministry of Public Health, Qatar has reported CVD as the leading cause of fatality in Qatar, with CVD now being considered as an economic burden in Qatar's healthcare sector. Many studies exist focusing on the early detection of CVD using machine learning (ML)-based approaches [4]. To our knowledge, no study has been published focusing on the early detection of CVD using DXA measurements in Qatar based on ML techniques. The objective of this study is to leverage DXA measurements to determine if they are effective in the diagnosis of CVD. We formulated this problem in the machine learning classification framework to classify the CVD group from the non-CVD group. Apart from diagnosing CVD, the proposed ML model will also help us to associate different metrics from body compositions to CVD in the Qatari population.

2. Materials and Methods

2.1. Data Collection from QBB

We collected a deidentified dataset from Qatar Biobank (QBB) for a CVD cohort of 250 participants. We also had a control (non-CVD) group of 250 cases that were free from CVD and other related diseases like diabetes, obesity etc. In the CVD group, most participants were diagnosed with hypertension (180 out of 250, nearly 72%). Other participants from the CVD group had abnormal heartbeats, angina, and revascularization.

2.2. Data Description and Preprocessing

There were 182 different measurements from DXA in the dataset. DXA machines primarily capture four types of measurements from different body parts: namely, (i) bone mineral density (BMD), (ii) lean mass, (iii) fat content, and (iv) bone area measurements. A few metrics from each category are described in Table 1.

Table 1: List of some measurements from the DXA dataset

Data Type	Description
BMD	Age-matched Z-score in the body area, Young adult T-Score in total body, BMD (g/cm^2) in total body, spine, etc.
Lean mass	Tissue lean mass (g) in leg, android, trunk, gynoid, etc.
Fat content	Tissue fat mass (g) in the trunk, arm, android, gynoid, total body etc.
Bone Area	Bone area (cm^2) in total body, legs, lumbar spine, etc.

Features with over 30% missing values were removed. For the remaining features, the missing values were replaced with the corresponding feature-mean. Data from both the CVD group and control group were normalized applying min-max normalization.

2.3. Machine Learning Model Development and Evaluation

We developed multiple ML models to distinguish the CVD group from the non-CVD group. We used four different ML algorithms: Multi-layer Perceptron (MLP), Decision Tree (DT), Random Forest (RF), and Logistic Regression (LR). We applied five-fold cross validation to evaluate the model. We considered multiple performance evaluation metrics such as: (i) Accuracy, (ii) Sensitivity, (iii) Specificity, and (iv) Matthews' Correlation Coefficient (MCC) to evaluate the performance of the proposed ML models.

3. Results and Discussions

In the CVD group, 54.80% participants were male and 45.2% were female, whereas in the control group had an even distribution of 50% between male and female participants). All the participants included in this study were adult Qatari nationals. The weight (72.2:64.2 in CVD:control) and BMI (26.12:23.2 in CVD:control) were relatively greater in the CVD group compared to the non-CVD group. In addition to considering the full DXA measurements in the ML model as features, an ablation study was conducted to measure the effectiveness of different types of DXA measurements by feeding them into ML models as different feature groups. Table 2 compares the performance of the ML models. Evidently, the LR-based model achieved the accuracy of 74% for BMD, 70% for lean mass, 76% for fat contents and 70.8% for bone area measurements. Considering all DXA measurements (182 features), the LR-based model achieved the highest accuracy of 77.4%.

Table 2 Machine Learning models performance based on ablation study

Feature	Model	Accuracy	Sensitivity	Specificity	MCC
BMD	DT	0.718	0.648	0.798	0.455
(85 features)	MLP	0.688	0.746	0.642	0.390
	RF	0.690	0.626	0.756	0.386
	LR	**0.740**	0.709	0.775	0.487
Lean mass	DT	0.662	0.643	0.679	0.326
(7 features)	MLP	0.592	0.437	0.730	0.171
	RF	0.672	0.654	0.689	0.344
	LR	**0.700**	0.642	0.757	0.402
Fat content	DT	0.722	0.655	0.787	0.449
(15 features)	MLP	0.724	0.739	0.715	0.458
	RF	0.744	0.679	0.801	0.486
	LR	**0.764**	0.722	0.806	0.551
Bone area	DT	0.646	0.600	0.702	0.301
measurements	MLP	0.636	0.723	0.572	0.296
(75 features)	RF	0.642	0.626	0.657	0.281
	LR	**0.708**	0.670	0.756	0.424
All features	DT	0.700	0.625	0.779	0.413
(182)	MLP	0.678	0.677	0.681	0.363
	RF	0.740	0.686	0.796	0.491
	LR	**0.774**	0.757	0.797	0.556

Our results demonstrated better bone health for the CVD group. The young adult T-score for lumbar spines (L1, L2, L4) were greater in the CVD group compared to non-CVD group. The Z-score for L2-L4 was -0.282: -0.57 in CVD vs. control group. Additionally, bone mass total was 2,510.45 in the CVD group vs. 2,416.17 in the control group. We also observed higher total lean mass (43,643.57: 40,614.88 in CVD:control) in CVD compared to the control group. Another prospective study based on the Greek general population recruited 3,042 adults without pre-existing CVD and the baseline skeletal muscle mass revealed an inverse association with ten-year CVD incidence [5]. Our study also observed higher fat content in the body composition of the CVD group (total fat mass=25908.57: 20962.12 in CVD: control). As mentioned earlier, the studied CVD group had higher BMI levels than the non-CVD group, and a higher weight elevated the fat level in the CVD group. In summary, most of the bone mineral density, fat content, muscle mass and bone surface area related measurements were greater in the studied CVD group compared to non-CVD group. The details of the DXA measurements can be found at: https://github.com/tanviralambd/CVD_DXA.

This study has some limitations worth mentioning. We worked with a relatively smaller sized cohort and we did not consider lifestyle or socio-demographic information of the participants in the model. Integrating such valuable information would add more insights into our study.

4. Conclusion

Although the FDA approved the use of DXA for diagnosing CVD recently, which measurements and what formula should be used in predicting and expanding the usage of DXA to CVD conditions still needs to be determined. The current study observes that machine learning can help in the selection of DXA measurements, which can potentially be included in the final report for more convenient clinical practice. In the future, we will evaluate our proposed approach in the CVD group with more advanced stages like ischemic heart attack and revascularization.

References

[1] Shepherd JA, Ng BK, Sommer MJ, Heymsfield SB. Body composition by DXA. Bone. 2017;104:101–5.

[2] Horizon DXA System [Internet]. [cited 2021 Aug 27]. Available from: https://www.hologic.com/hologic-products/skeletal-solutions/horizon-dxa-system.

[3] Al-Absi HRH, Refaee MA, Nazeemudeen A, Househ M, Shah Z, Alam T. Cardiovascular Diseases in Qatar: Smoking, Food Habits and Physical Activities Perspectives. Stud Health Technol Inform. 2020;272:465–9.

[4] Al-Absi HRH, Refaee MA, Rehman AU, Islam MT, Belhaouari SB, Alam T. Risk Factors and Comorbidities Associated to Cardiovascular Disease in Qatar: A Machine Learning Based Case-Control Study [Internet]. IEEE Access. 2021. p. 29929–41. Available from: http://dx.doi.org/10.1109/access.2021.3059469

[5] Tyrovolas S, Panagiotakos D, Georgousopoulou E, Chrysohoou C, Tousoulis D, Haro JM, et al. Skeletal muscle mass in relation to 10 year cardiovascular disease incidence among middle aged and older adults: the ATTICA study. J Epidemiol Community Health. 2020;74:26–31.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210906

Radiotherapy Information System in Cancer Treatment Improvement: A Case Study at the Brazilian National Cancer Institute

Antônio Augusto Gonçalves [a,b] , Cezar Cheng [a]
Sandro Luís Freire de Castro Silva [a1] , Emmanuel Tenorio Cavalcante [a] ,
Jose Geraldo Pereira Barbosa [b] , Francisco Santos Sabbadini [c]

[a] *Instituto Nacional de Câncer - COAE Tecnologia da Informação, Rua do Resende 195, Rio de Janeiro, 20230-092, Brazil*
[b] *Universidade Estácio de Sá - MADE, Avenida Presidente Vargas 642, Rio de Janeiro, 200071-001, Brazil*
[c] *Universidade do Estado do Rio de Janeiro - UERJ Rod Pres Dutra km 298, Resende, 27537-000, Brazil*

Abstract. Radiotherapy is one of the main means of treating cancer patients. Its application has grown worldwide. Around 50% of all cancer patients should receive radiation. Brazil faces a shortage of radiotherapy treatment because of a lack of enough treatment units, equipment availability, well-trained staff, and fair reimbursement. The Radiotherapy Information System (RIS) implementation to manage information about patient scheduling is vital to improve the efficiency of care and reduce the waiting time to start cancer treatment. The information system deployed can be indicated as a disruptive innovation in the Brazilian public health system, considering the radical improvement in the cancer treatment process at the Brazilian National Cancer Institute.

Keywords. Radiotherapy Information System, RIS, Cancer Treatment

1. Introduction

Cancer treatment involves a complex and multifaceted decision-making process. Brazilian citizens have faced long waiting times to start oncology treatment. Lengthy waiting periods increase the risk of cancer recurrence and decrease patients' survival rates [1].

Radiotherapy is one of the main means of treating cancer patients. Its application has grown worldwide. Around 50% of all cancer patients should receive radiation. Nowadays, the Brazilian public health system cannot attend to the radiotherapy demand, and many patients requiring radiotherapy are not to access this treatment. Brazil faces a

[1] Corresponding Author, Sandro Luís Freire de Castro Silva, Instituto Nacional de Câncer - Tecnologia da Informação, Rua do Resende 195, Rio de Janeiro, 20230-092, Brazil; Email: sandrofreire@gmail.com.

shortage of radiotherapy treatment because of insufficient treatment units, equipment availability, well-trained staff, and fair reimbursement [2].

Growing demand for radiotherapy enforces healthcare service managers to face the challenge of delivering in-time radiation therapy to cancer patients for the lowest possible costs. Long waiting times can raise patients' anxiety and tumor growth, negatively impacting clinical outcomes. The high costs are related to Linear Accelerators. Therefore, the focus is on scheduling many patients on a low number of available machines [3].

The availability of accurate information critically influences the quality of cancer care and its efficient management. Radiotherapy Information System (RIS) plays a vital role in cancer treatment by ensuring proper and safe delivery of radiation therapy. Efficient patient scheduling within oncology clinics ensures the delivery of adequate treatment at the exact time. The main objective is to provide the effective use of scarce medical resources [4,5].

The implementation of the RIS to manage information about patient scheduling is essential to improve the efficiency of care and reduce the waiting time to start cancer treatment. Scheduling rules are defined to determine when treatment can be made. The objective is to ensure the effective use of the medical resources and deliver the proper treatment to the patient at the right time.

2. Methods

The Brazilian National Cancer Institute (INCA) has five hospital units that share the same information systems based on an integrated patient database. The focus on interoperability makes the most of patient data accessible across devices and platforms.

The RIS infrastructure was developed using INCA's intranet network that provides safe access to applications designed to improve cancer care. This information architecture consists of a collaborative environment between physicians and technicians, simplifying the physician work and empowering technicians involved in the decision-making processes. The legacy systems feed the radiotherapy data repository, which is the basis to build the decision support system, as shown in Figure 1.

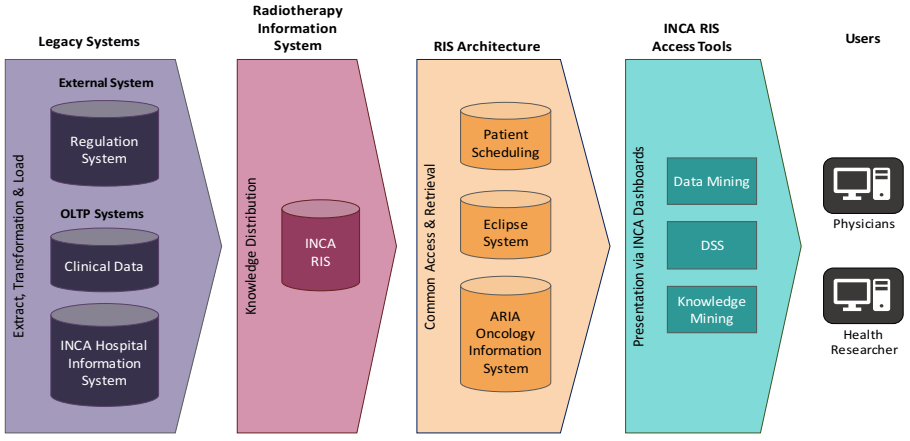


Figure 1. The INCA Radiotherapy Information Architecture

Radiotherapy is a time-consuming process requiring a significant volume and variety of data. Developing a web-based information system is intended to provide a tool that allows the exchange of therapeutic knowledge and patients' clinical data between INCA physicians. Treatment plans and best practices need to be shared to be made accessible to the whole team of clinicians.

Scheduling radiotherapy treatment appointments is a complex problem due to various medical procedures and scheduling constraints, such as machine availability, patient category, physicians' rota, waiting time targets, number of machines, and significant demand for patients.

Different functionalities had to be implemented, including reducing machines idle time, reducing the usage of overtime slots, and reducing the number of patients who do not meet their waiting time targets. The Radiotherapy Information System was developed to schedule cancer patients and solve severe scheduling problems.

The developed system runs on a web application with a modular and layered architecture searching for usability, interoperability, and ease of maintenance. Generally, communication and collaboration with other physicians are necessary. The web-based information system was implemented to improve patient scheduling for radiotherapy treatment, as shown in Figure 2.

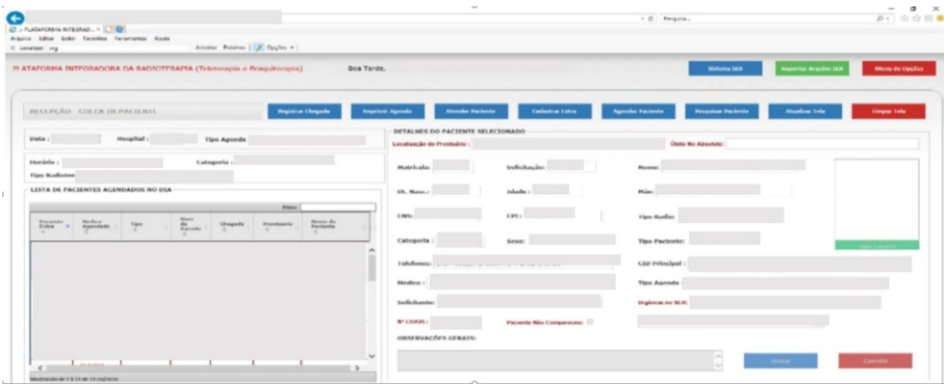


Figure 2. INCA Radiotherapy Information System (RIS)

3. Radiotherapy Information System (RIS)

The information architecture described in this paper has allowed the implementation of an Internet-based information system for radiotherapy. The system was deployed for patient scheduling, treatment decision-making, investigation, and results analysis.

The study describes the implementation of a Radiotherapy Information System (RIS) at INCA hospital units. The information architecture contains clinical data and therapeutic information accessible to INCA physicians. All these components can interact together.

The RIS is integrated with other information systems that support the radiotherapy decision-making process by supplying therapeutic information and sample plans, which can be improved to specific patient cases. Since treatment planning is a critical optimization process, complex plans are created not by a single specialist but by a nuclear medicine team, as shown in Figure 3.

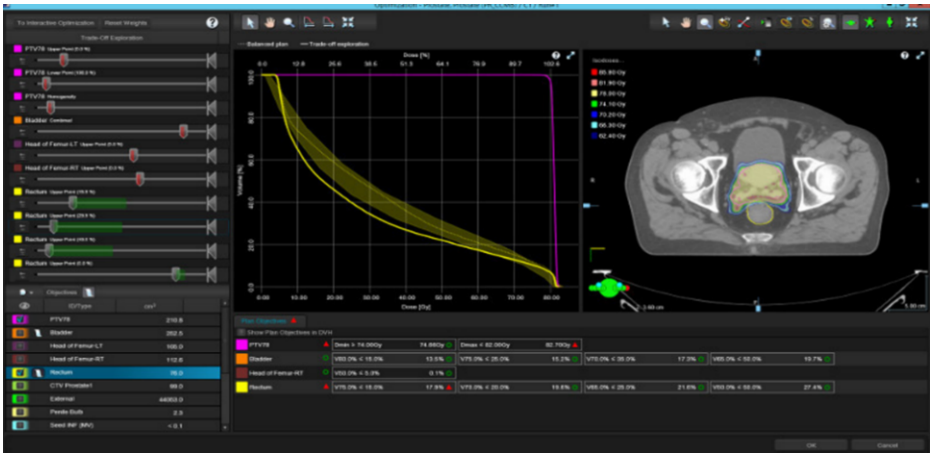


Figure 3. Eclypse System

4. Conclusions

The efficient radiotherapy patient scheduling within INCA clinics plays a crucial role in ensuring the delivery of the proper treatment at the right time. In this scenario, generating a high-quality solution is a challenging task since different goals such as reducing patient waiting time to start cancer treatment could be achieved, and a set of constraints, as the number of available machines, should be considered.

An integrated Radiotherapy Information System is a critical success factor in cancer treatment. The traditional decision-making process cannot deal with the vast amount of data that must be processed for modern medicine. Web-based information systems need to be developed to support complex radiation therapy treatments. This technology connects radiotherapy best practices, specialists, and patient information.

There were many challenges involved in organizing and communicating data in INCA radiation therapy treatment processes. The focus on interoperability has integrated the most patient data across devices and platforms that simplified to share patients' information between cancer hospital units. The system deployed can be indicated as a disruptive innovation in the Brazilian public health system, considering the radical improvement in the cancer treatment process at INCA.

References

[1] Gonçalves, A.A., De Castro Silva, S.L.F., Martins, C.H.F., Cheng, C., Barbosa, J.G.P., De Oliveira, S.B. Decision Support Systems in Cancer Treatment: A Case Study at Brazilian National Cancer Institute (2018); Studies in Health Technology and Informatics, 251, pp. 199-202.

[2] Moraes FY, Marta G, Hanna SA, Leite ET, Ferrigno R, da Silva JL, et al. Brazil's challenges and opportunities. Int J Radiat Oncol Biol Phys 2015; 92(4): 707-712.

[3] Chen Z, King W, Pearcey R, Kerba M, Mackillop WJ (2008) The relationship between waiting time for radiotherapy and clinical outcomes: a systematic review of the literature. Radiother Oncol 87(1):3–16

[4] Mandal A, Asthana AK, Aggarwal LM. Development of an electronic radiation oncology patient information management system. J Cancer Res Ther (2008) 4:178. doi:10.4103/0973-1482.43342.

[5] S. Silva, A.A. Goncalves, C. Cheng, C. Martins, (2018) Mobile technology applications in cancer palliative care, Studies in Health Technology and Informatics, vol 249, pp. 199–202.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210907

eHealth Literacy Among Health Sciences Students in Greece

Trisevgeni TRANTALI[a], Christina ATHANASOPOULOU[b], Areti LAGIOU[c,d] and Evanthia SAKELLARI[c,d,1]
[a] *Master's Programme "Health Promotion and Health Education", National and Kapodistrian University of Athens & University of West Attica, Athens, Greece*
[b] *Department of Occupational Therapy, School of Health and Care Sciences, University of West Attica, Athens, Greece*
[c] *Department of Public and Community Health, School of Public Health, University of West Attica, Athens, Greece*
[d] *Laboratory of Hygiene and Epidemiology, University of West Attica, Athens, Greece*

Abstract. eHealth literacy is a necessary skill to find and make good use of online health information. However, the general public lacks this skill; it is essential for future health professionals to be able to guide and facilitate the public. The study aimed to examine the perceived eHealth literacy level of health sciences students in Greece. A cross-sectional, online survey was conducted (N=113 students). The questionnaire included socio-demographic data and the eHealth Literacy Scale (eHEALS). Participants' mean eHEALS score was 31.9 with medicine and dentistry students having the highest score (33.7) and other health and caring sciences students the lowest (29.8). There was no statistically significant difference at eHEALS score among participants of different academic year. However, there was a statistically significant difference at eHEALS score among University Departments (p=0.009). Further research in representative samples is required to assess specific needs and improve current educational curricula.

Keywords. eHealth literacy, eHEALS, health information, health sciences students

1. Introduction

The internet is a major source of health information [1]. So, "the ability to find, understand, and apply online health information in order to make appropriate health decisions", known as eHealth literacy, is a necessary skill [2]. However, previous studies have shown that not only the general public lacks of this skill [3] but also the health professionals [4]. Investigating eHealth literacy among future health professionals plays an important role as they help in guiding and facilitating the general public to find, understand and apply online health information. Especially in Greece, there is limited literature about eHealth literacy among health sciences students. Thus,

[1] Corresponding Author, Evanthia Sakellari, Department of Public and Community Health, Laboratory of Hygiene and Epidemiology, School of Public Health, University of West Attica, 196 Alexandras Av., 11521, Athens, Greece; E-mail: sakellari@uniwa.gr.

the aim of this study was to examine the perceived eHealth literacy level among undergraduate health sciences students in Greece.

2. Methods

A cross-sectional, online survey was conducted between June 4th and July 13th, 2021. A convenient sample of health sciences students from Greek Universities was recruited using a passive recruitment strategy, through invitation via Facebook groups. The study was granted permission from the Ethics Committee of the University of West Attica (41141-25/05/2021).

The eligibility criteria were: 1) internet access 2) 18 years of age or older, 3) undergraduate health sciences student at Greek University 4) understanding Greek language 5) accepting to complete the questionnaire voluntarily and anonymously.

Data collection was made using a structured, self-administered questionnaire through Google Forms, which consisted of 16 items: 1) socio-demographic characteristics (8 items), 2) eHealth Literacy Scale-eHEALS (8 items) [5,6], measured participants' combined knowledge, confidence and perceived skills in finding, evaluating and applying eHealth information to health problems. Each item was scored on a 5-point Likert scale, from 1 "Strongly disagree" to 5 "Strongly agree" (min:8; max:40). Higher score represented higher perceived eHealth literacy level. In this study the eHEALS had an internal consistency of Cronbach a=0.89.

Data analysis was made using SPSS Version 26. Descriptive statistics were used to identify participants' socio-demographic characteristics and eHealth literacy scores. Moreover, One way Analysis of Variance on ranks (Kruskal-Wallis Test) was used to compare the eHEALS scores among participants. The statistical significance level was set at 0.05 [7].

3. Results

The total study sample was 113. Median age was 22 years, the majority were female (n=92, 81.4%), single (n= 65, 57.5%), students at universities of Attica region (n=90, 79.6%), at 2nd academic year (n=34, 30%), who studied health and caring sciences (midwifery, occupational therapy, physiotherapy or nursing) (n=37, 32.7%) (Table 1).

Table 1. Socio-demographic characteristics of participants.

Socio-demographic characteristics	Total sample N=113
Age: Median (min, max)	22 (18, 53)
Gender: N (%)	
Male	19 (16.8)
Female	92 (81.4)
Other	2 (1.8)
Marital status: N (%)	
Single	65 (57.5)
Partnership/Married	42 (37.2)
Other	6 (5.3)
University: N (%)	
Attica region	90 (79.6)
Outside Attica region	23 (20.4)

Department	
Medicine & Dentistry	32 (28.4)
Pharmacy	11 (9.7)
Public & Community Health	33 (29.2)
Other Health & Caring Sciences	37 (32.7)
Academic year: N (%)	
1st	9 (8)
2nd	34 (30)
3rd	16 (14.2)
4th	25 (22.1)
≥ 5th	29 (25.7)
Employment status: N (%)	
Working	31 (27.4)
Not working	82 (72.6)
Living situation: N (%)	
Own household (with partner/family)	22 (19.5)
Own household (alone)	24 (21.2)
Rented apartment complex/University campus	6 (5.3)
Parents' household	60 (53.1)
Other	1 (0.9)

The mean total eHEALS score was 31.9 (SD: 4.9, min: 18; max: 40). Table 2 below, shows the mean scores of individual items of the eHEALS scale (min:1; max:5). The lowest scoring item was 3.4, indicating the weakest eHealth literacy item while the highest scoring item was 4.2, indicating the strongest eHealth literacy item.

Table 2. Participants' score (Mean & SD) on the eHEALS items.

eHEALS Items	Mean (SD)
I know what health resources are available on the Internet	3.9 (0.7)
I know where to find helpful health resources on the Internet	4.1 (0.8)
I know how to find helpful health resources on the Internet	4.2 (0.7)
I know how to use the Internet to answer my questions about health	4 (0.8)
I know how to use the health information I find on the Internet to help me	4 (0.8)
I have the skills I need to evaluate the health resources I find on the Internet	4.1 (0.7)
I can tell high-quality health resources from low-quality health resources on the Interne	4.1 (0.8)
I feel confident in using information from the Internet to make health decisions	3.4 (1.1)

No significant difference in terms of total eHEALS scores was observed among participants of different academic years. However, there was a statistically significant difference of total eHEALS score among the different departments represented in the study (Table 3). Further statistical analysis using Bonferroni correction indicated that the mean eHEALS score of medicine and dentistry students (mean: 33.7) was higher than the score of other health and caring sciences students (mean: 29.8) (Bonferroni Post Hoc Test, p=0.004).

Table 3. Comparison of the total eHEALS scores between departments and academic years.

Variable		Total eHEALS score			
		Mean	Median	Min, Max	p-value*
Department	Medicine & Dentistry	33.7	33.5	23, 40	0.009
	Pharmacy	31.8	31	26, 40	
	Public & Community Health	32.6	32	22, 40	
	Other Health & Caring Sciences	29.8	30	18, 40	
Academic	1st	31.3	32	18, 40	
	2nd	30.3	31	21, 40	

year	3rd	32.7	32	26, 40	0.173
	4th	32.9	32	22, 40	
	≥ 5th	32.8	32	27, 40	

* Kruskal-Wallis rank test

4. Discussion

This is the first study examining the perceived eHealth literacy among health sciences students in Greece. The current study found a higher mean eHEALS score (31.9) among health sciences students compared to the score found in previous similar studies examining eHEALS score among medical and health sciences students (mean: 28.21) [8] and nursing students (mean: 24.52) [9]. Due to the small number of participants, the study results cannot be generalized. Further research is needed in order to assess the characteristics and specific needs. Moreover, the examination of additional information such as more socio-economic characteristics, internet usage and devices, primary connectivity in relation to eHEALS score among health sciences students could provide more specific data in this field. In conclusion, based on this important insight, departments could develop curricula that emphasize on practical eHealth literacy skills and critical thinking, in order to educate the future health sciences professionals.

References

[1] Eurostat. One in two EU citizens look for health information online. 2021. Available at: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20210406-1.
[2] Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res. 2006 June;8(2):e9.
[3] Park H, Cormier E, Gordon G, Baeg JA. Identifying Health Consumers' eHealth Literacy to Decrease Disparities in Accessing eHealth Information. Comput Inform Nurs. 2016 Feb;34(2).
[4] Stellefson M, Hanik B, Chaney B, Chaney D, Tennant B, Chavarria EA. eHealth Literacy Among College Students: A Systematic Review With Implications for eHealth Education. J Med Internet Res. 2011 Dec;13(4):e102.
[5] Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res. 2006 Nov;8(4):e27.
[6] Athanasopoulou C, Valimaki M, Koutra K, Lottyniemi E, Bertsias A, Basta M, Vgontza A, Lionis C. Internet use, eHealth literacy and attitudes toward computer/internet among people with schizophrenia spectrum disorders: a cross-sectional study in two distant European regions. BMC Med Inform Decis Mak. 2017 Sept;17(123).
[7] Crosby R, DiClemente R, Salazar L. Research methods in health promotion. San Francisco: Jossey-Bass; 2006. 317-346 p.
[8] Dashti S, Peyman N, Tajfard M, Esmaeeli H. E-Health literacy of medical and health sciences university students in Mashhad, Iran in 2016: a pilot study. Electron Physician. 2017 Mar; 9(3):3966–3973.
[9] Tanaka J, Kuroda H, Igawa N, Sakurai T, Ohnishi M. Perceived eHealth Literacy and Learning Experiences Among Japanese Undergraduate Nursing Students: A Cross-sectional Study. Comput Inform Nurs. 2020 Jan-Mar;4(1):e2.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210908

Trust in E-Health System and Willingness to Share Personal Health Data

Dasa BOSANAC[a,1] and Aleksandra STEVANOVIC[bc]
[a]*University North, Croatia*
[b]*University of Rijeka Faculty of Medicine, Croatia*
[c]*University of Rijeka Faculty of Health Studies, Croatia*

Abstract. Electronic personal health records (ePHR) are web-based tools that enable patients to access their personal health data. Since the data in PHR are systematized, they can be used in scientific research with the patient's consent. Despite the potential benefits of using ePHR, their adoption in Croatia remains low. Multiple factors are influencing the use of PHR and willingness to share personal health data (PHD). The purpose of this paper was to determine familiarity with the e-health system among Croatian citizens, trustfulness in the system itself, and willingness to share PHD with physicians or researchers. Results showed that 34% of respondents use ePHR, and have less confidence in the electronic system than doctors' confidentiality. However, health professionals have lover trust in doctors' confidentiality compared to non-health workers. Respondents rated mental health data and STD data as most sensitive but are overall willing to share that data with doctors and researchers.

Keywords. E-health, personal health records, data sharing, trust, privacy

1. Introduction

E-health contributes to improving the efficiency of health services, enables the communication between the patient and the doctor, and ensures the continuity of medical care. Electronic personal health records (ePHR) enable data systematization for easier scientific research [1,2]. However, to provide appropriate services and conduct research through health applications and e-health services, users' consent is required [3].

Previous studies examined factors that influence the frequency of use of personal health data (PHD) and willingness to share them for scientific research. Generally, people are more willing to share their ePHD if they perceive them useful for public health research, not sensitive, and if they trust the anonymity of research [3,4]. The intention to use an ePHD application is strongly affected by the social impact [5], by the ease of usage of the system [6,7], while age, gender, education, and profession showed different influences on the frequency of e-health use and trust in the system itself [1,6,7].

In Croatia, there is an e-Health Portal since 2016 and a Mobile e-Health Portal since 2017, which in addition to basic services, enables citizens to allow or prohibit

[1]Corresponding Author: Daša Bosanac, University North, Trg dr.Zarka Dolinara 1, 48000 Koprivnica, Croatia; E-mail: dabosanac@unin.hr.

family doctors, dentists, gynecologists, and pediatricians to access a specific PHD. It also provides insight into who and when viewed which data. Considering previous research, we were interested in how much people in Croatia are familiar with the e-Health system, how much they trust such a system, how sensitive they consider specific medical data, and depending on that, how much of the content of the ePHD are they willing to share with doctors or for research purposes.

2. Methods

We conducted the research using an anonymous online survey on the KwikSurveys platform. Data were collected in July 2021. We asked respondents if they use the e-Health system and how much they trust e-data and family doctor confidentiality on a scale from 0 to 10. Furthermore, we asked participants to assess the perceived sensitivity of 13 different health data selected for the study and their willingness to share them, both on a scale from 0 to 10. Finally, we collected demographic data (age, gender, occupation, self-assessment of digital literacy).

Data are presented as mean and standard deviation or number and percentage. Differences were assessed by chi-squares or t-test for independent groups. Dependent on the type of data, correlations were assessed by either Pearson or Spearman correlation coefficients.

3. Results

Out of a total of 198 questionnaires started, 102 were sufficiently completed for analyses (62 F, 35 M, 5 other), average age M=43.04 (range 25-75), SD=10.89. In the sample, 92% of the participants were employed, of which 37% were health professionals. 88% of the sample was highly educated. On a linear scale from 1-10, respondents rated their digital literacy as M=7.79, SD=1.63.

A total of 73% of participants reported they were aware of the e-Health Portal's existence in Croatia, yet only 34% were using it, and 38.2% were familiar with its possibilities. Confidence in e-Health data protection averaged 6.23, SD=2.43, and confidence in the family doctor's confidentiality averaged 7.49, SD=2.65. Participants have significantly greater confidence in doctor's confidentiality than online data protection ($t=4.603$, $p<0.001$). Also, 61% of participants reported that family doctors should have complete insight into all patients' medical data. Participants rated their digital literacy as M=7.8 (SD=1.63, range 1-10). Digital literacy significantly correlated with trust in e-Health data protection ($r=0.27$, $p=0.007$) but not with doctor's confidentiality ($r=0.09$, $p=0.395$).

Employment status, educational level, gender, and age were not significantly related to the e-Health Portal awareness, use, or knowledge of its possibilities.

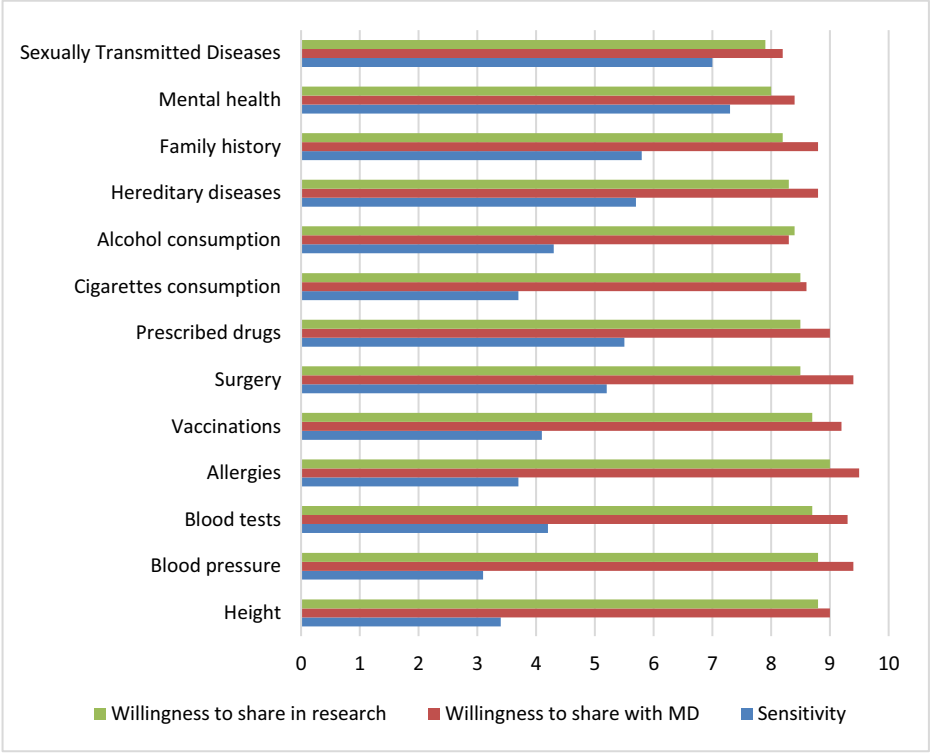


Figure 1. Sensitivity of health-related data and willingness to share

As presented in Figure 1., participants rated mental health and STD as the most sensitive health-related data followed by family history and hereditary diseases. Data with the lowest sensitivity was blood pressure, followed by height, allergies, and cigarettes consumption. Overall, participants were highly willing to share their health data with their physicians and research, with scores ranging from 7.9 for STD to 9 for allergies. They rated their trust in the anonymity of data in research on an average of 7.21, SD=2.67.

There was a low yet significant correlation between trust in research anonymity and computer skills level (r=0.206) and a moderate correlation between trust in research anonymity and physician confidentiality (r=0.298). Health professionals (M=6.86, SD=2.82) had significantly lower trust in doctors confidentiality compared to others (M=8.1, SD=2.35) (t=-2.324, p=0.022).

4. Discussion

Our data showed low use of ePHR in Croatia despite awareness of its existence. In Croatia, there are several problems with the e-Health system. The first concerns the elderly population, which is not predominantly digitally literate, and the second is the lack of infrastructure due to which some households do not have Internet access. The same issues arose as the limitation of this online study. Respondents of lower

educational status, older respondents and those from rural areas did not participate in the study. The third concern is trust in the e-Health system itself.

Our data showed that respondents believe significantly more in doctors' confidentiality than in data protection in the e-Health system. Although distrust of institutions is characteristic of post-socialist countries, it is interesting that health professionals have less trust in doctors' confidentiality than non-health workers in our sample. It raises the question of what health system employees know about the system itself that others do not. As expected, higher digital literacy was related to higher confidence in online data protection, suggesting that promoting computer skills and digital literacy could be beneficial for the broader implementation of ePHD.

As in previous study [4], the results showed that not all PHD were perceived as equally sensitive, yet the respondents were willing to share them with physicians and researchers. The interpretation of data-sharing attitudes is a complex process, related to and influenced by various factors, and involves medical ethics issues [8]. In our highly educated sample, a high willingness to share PHD might be influenced by an interest in personal health improvement or health care promotion, albeit some reserves toward trust in the e-Health system.

5. Conclusions

The use of the e-Health Portal in Croatia remains low, although there is a high level of trust in confidentiality, anonymity, and willingness to share data for research. Future research should involve a more diverse sample. Efforts in promoting e-Health tools could be beneficial for both the users and researchers.

References

[1] Karampela M, Ouhbi S, Isomursu M. Connected Health User Willingness to Share Personal Health Data: Questionnaire Study. JMed Internet Res. 2019 Nov; 21(11):e14537.doi:10.2196/14537

[2] Whetstone M, Goldsmith RE. Factors influencing intention to use personal health records. International Journal of Pharmaceutical and Healthcare Marketing. 2009 Apr; 3(1):8-25. doi: 10.1108/17506120910948485

[3] Gille F, Brall C. Limits of data anonymity: lack of public awareness risk trust in health system activities. Life Sci Soc Policy. 2021 Jul 26;17(1):7. doi: 10.1186/s40504-021-00115-9.

[4] Helou S, Abou-Khalil V, El Helou E, Kiyono K. Factors Related to Personal Health Data Sharing: Data Usefulness, Sensitivity and Anonymity. Public Health and Informatics. 2021 May;281:1051-1055. doi:10.3233/SHTI210345.

[5] Park HS, Kim K, Soh JY, Hyun YH, Jang SK, Lee S, Hwang GY, Kim HS. Factors Influencing Acceptance of Personal Health Record Apps for Workplace Health Promotion: Cross-Sectional Questionnaire Study. JMIR Mhealth Uhealth. 2020 Jun; 8(6): e16723. doi: 10.2196/16723.

[6] Abd-Alrazaq A, Bewick BM, Farragher T, Gardner P. Factors Affecting Patients' Use of Electronic Personal Health Records' in England: Cross-Sectional Study. J Med Internet Res 2019;21(7):e12373. doi:10.2196/12373.

[7] Weng C, Friedman C, Rommel CA, Hurdle JF. A two-site survey of medical center personnel's willingness to share clinical data for research: implications for reproducible health NLP research. BMC Med Inform and Deciss Mak. 2019 Apr;19(3):70. doi:10.1186/s12911-019-0778-z.

[8] Zargaran A, Ash J, Kerry G, Rasasingam D, Gokani S, Mittal A, Zargaran D. Ethics of Smartphone Usage for Medical Image Sharing. Indian J Surg. 2018 Jun; 80(3): 300-301. doi: 10.1007/s12262-017-1710-3.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210909

Assisting Data Retrieval with a Drug Knowledge Graph

Romain LELONG[a,b,1], Badisse DAHAMNA[a,b], Romain LEGUILLON[a], Julien GROSJEAN[a,b], Catherine LETORD[a,b], Stéfan J. DARMONI[a,b] and Lina F. SOUALMIA[b,c]
[a] *CHU Rouen, Department of Biomedical Informatics, F-76000 Rouen, France*
[b] *LIMICS U1142, Sorbonne Université, Paris, France*
[c] *Normandie Univ, UNIROUEN, TIBS-LITIS EA 4108, F-76000 Rouen, France*

Abstract. The Normandy health data warehouse EDSaN integrates the medication orders from the University Hospital of Rouen (France). This study aims at describing the design and the evaluation of an information retrieval system founded on a complex and semantically augmented knowledge graph dedicated to EDSaN drugs' prescriptions. The system is intended to help the selection of drugs in the search process by health professionals. The manual evaluation of the relevance of the returned drugs showed encouraging results as expected. A deeper analysis in order to improve the ranking method is needed and will be performed in a future work.

Keywords. Drug Information Retrieval; Knowledge Graph; Semantic Network.

1. Introduction

Data provided and/or exploited by drugs systems usually fall into two types: (a) factual drugs data and (b) knowledge drugs data. Factual drugs data mainly consist in drugs prescription and drugs administration data that are archived by hospitals usually as free text within the discharge letter or medication orders. Several methods have been proposed to perform information retrieval on factual drugs data: information extraction and free text search [1], machine learning [2]. However, the implementation of an effective information retrieval system requires the use of knowledge data in addition to factual data. Knowledge graph structure including the conceptual graph formalism [3] has been used for biomedical knowledge and data representation is particularly suited to drug knowledge data [4]. Existing medicinal drug databases such as Wikidata [5], Drug Bank[2], or GoodRx[3] contain valuable information but lack of comprehensiveness when taken separately and/or store some of this information as unstructured data [6]. In this study, the design of a system enabling the retrieval of prescription orders contained in the Normandy's Health Data Warehouse (EDSaN) [7] at the Rouen University Hospital (Normandy, France) is described. A conceptual graph of drug knowledge data was

[1] Corresponding Author, Romain LELONG, Department of Biomedical Informatics, Rouen University Hospital, 1 rue de Germont Rouen France; E-mail: romain.lelong@chu-rouen.fr.
[2] https://go.drugbank.com/releases/latest
[3] https://www.goodrx.com/

designed and used in the information retrieval process to retrieve the French Common Dispensing Unit (UCD) codes that are used to encode and bill administrated drugs in France. A first evaluation is conducted to assess the consistency of the concepts resulting from traversing the graph. The evaluation focused in this study on the UCD codes only.

2. Material and Method

2.1. The EDSaN Data Warehouse and HeTOP

The EDSaN data warehouse currently integrates 6,978,586 atomic prescriptions, *i.e.* of a single pharmaceutical specialty, distributed over 1,452,616 prescriptions' orders between 2011 and the end of August 2021. The data are loaded and maintained into EDSaN from the Rouen University Hospital Information System (HIS) and originate from either a dedicated database or in the care plan. Although various structured fields exist in the HIS to describe the prescriptions and administrations of drugs (dosage, time, number of administrations, etc.), the poor quality of these data, e.g. empty fields, wrong values, does not allow an accurate structured search. In EDSaN, every atomic prescription is associated to a UCD code that identifies the prescribed pharmaceutical product. The prescription orders include patient information (id, age, gender, and birthdate), stay information (id, entry and leaving dates, units) and prescription information (id, date), and a list of atomic prescriptions that corresponds to a single prescribed drug. The Health Terminology/Ontology Portal (HeTOP) (https://hetop.eu) is used as primary data source for drug knowledge data. It currently integrates terminological concepts from over than 70 health terminologies and/or ontologies as well as semantic relationships between those concepts. Since 2019, it includes a formal drug model suited for French specificities.

2.2. Structured Information Retrieval

The free and open-source search engine software library Apache Lucene was used to enable the querying of prescription orders. Lucene documents were generated for each order to provide Boolean search functionalities. A graphical module specifically dedicated to prescriptions were also added to the Java web application of EDSaN. It enables users to query the Lucene indexes through a form that provides a search field for each prescription order metadata. This form enables an accurate structured search since each metadata could be targeted. The search for prescription orders based on what drug were prescribed can notably be done by using the search field dedicated to UCD codes. This is essential to handle user queries for which the full text search does not return any answers. However, this task requires a pre-selection of the UCD codes to be targeted.

2.3. Modelling the Knowledge Graph

To assist in the selection of UCD codes, a knowledge graph was designed thanks to drug and related data (e.g., diseases) from HeTOP and its semantic network. Vertices are selected among HeTOP concepts and include MeSH (Medical Subject Heading) Descriptors, Anatomical Therapeutic Chemical (ATC) concepts, International Nonproprietary Names (INNs), virtual drugs from the MédicaBase

(https://www.medicabase.fr/index.htm) database, pharmacological roots, pharmacological specialties, drug components, drug components groups, and UCD codes. The edges between vertices were also taken from semantic relationships in the HeTOP. Some types of relationships were excluded due to their technical nature or low relevance in the context of this study (e.g., "not to be confused with"). The overall drug knowledge graph is composed of 131,277 vertices and 703,807 edges. When possible, the user's input is matched to a vertex of this graph. The paths starting from that node and leading to UCD codes are then aggregated and proposed to the user in the form of a selectable sub-graph. In order to provide the user with the most relevant UCD codes first, the paths resulting from the user's query are ranked according to a weight and then filtered. The weight of a path is calculated by summing the weights of the relationships traversed in that path. The weights of the relationships were assigned empirically in collaboration with a hospital pharmacist. Seven ascending ranges of path weight were defined in order to distribute the paths over seven corresponding relevance classes (C1 to C7). In the interface, a slider component allows the user to adjust the tolerance and then show or hide the paths assigned to these different classes in the resulted graph (see Figure 1).

Figure 1. Example of a path traversal of the knowledge graph at a Class 1 from the user query "neisseria meningitidis gpe C" which corresponds to INN 11565. It allows the proposal of paths leading to height relevant UCDs. Three types of paths were browsed in this example: INN > pharmacological specialty > UCD Code or INN > ATC Code > UCD Code or INN > ATC Code > pharmacological specialty > UCD Code.

3. Results

To evaluate the semantic graph modelling the drug knowledge, a set of n=88 terms was randomly drawn among the possible types of vertex. These terms led to more than 100,000 paths overall. For each term, only the UCD codes issuing from paths belonging to the first three non-empty relevance class proposed by the system were evaluated. This UCD codes assessment was done by a single expert, a hospital pharmacist and consisted in giving a score: 1 if the UCD code seemed to him unsatisfactory, 2 if it could be improved, and 3 if the result was consistent. At this evaluation stage, neither the path itself, nor the relevance class assigned to it by the system were considered. Table 1 summarizes the results of the evaluation.

Table 1. Average of scores given by the expert for the first non-empty relevance classes from C1 (min path weight) to C3 (deeper path). Nt: number of terms; Np: number of paths obtained. The shaded classes have not been taken into account in this study.

Type of term	Nt	Np	C1	C2	C3	C4	C5
drug components	7	1,514	**2.71**	1.08	1.04		
drug components groups	3	2,270	-	**3.00**	1.00	1.00	
MeSH Descriptors	11	20,665	-	**2.94**	1.49	1.05	
INN	7	2,232	-	-	**1.95**	1.01	1.00
virtual drugs	11	2,470	**3.00**	1.23	1.28		
Medical indication	6	3,341	-	-	**3.00**	2.82	1.52
ATC Code	14	79,094	2.25	-	**3.00**	1.00	
pharmacological specialties	7	3,198	**2.50**	1.59	1.89		
pharmacological roots	22	4,648	**3.00**	1.84	1.05		

4. Discussion and Conclusion

As the assignments of relevance classes is based on relationships weights that have been empirically chosen, the average score for each class needs to be assessed. From a general point of view, one can observe that for each type of term, the average scores of the paths tends to decrease with the level of relevance (i.e. the class) assigned to it. Although, exceptions can be found, especially with regards to the ATC terms (C1=2.25 whereas C3=3.00). This shows that the weights of the relationships have been consistently assigned. Nevertheless, significant variability and inconsistency can be observed between term types. The highest score of 3 is reached for several types of terms (e.g. Drug Component Groups (C2=3), Virtual Drugs (C1=3). One can observe that for Medical Indications and MeSH Descriptors, their score is still high, despite the fact that these terms are generic and consequently harder to semantically link to UCD codes than pharmacological specialties or ATC codes for instance. Moreover, the best possible relevance class that can be obtained varies among the types of terms (C3 for Medical Indications and INNs, C2 for drug components groups and MeSH descriptors, C1 for others) although the score of those class remain close to 3.00.

The main aim of the drug knowledge graph was to assist the user in selecting drugs of interest. From that perspective, this study showed encouraging results as the ranking of resulting UCDs codes were overall congruent with the expert judgment. However, some inconsistencies remain. Future work will therefore focus on the refinement of the weights assigned to the edges of the graph.

References

[1] Dietrich G et al. Replicating medication trend studies using ad hoc information extraction in a clinical data warehouse. BMC Med Inform Decis Mak. 2019 Jan 18;19(1):15.
[2] Corny J et al. A machine learning–based clinical decision support system to identify prescriptions with a high risk of medication error. J Am Med Inform Assoc. 2020 Nov 1;27(11):1688-1694.
[3] Sowa JF. Conceptual structures: information processing in mind and machine. Addison-Wesley Longman Publishing Co., Inc., 1984.
[4] Teng A, et al. Biomedical Graph Visualizer for Identifying Drug Candidates. bioRxiv, 2020.
[5] Vrandečić D & Krötzsch M. Wikidata: a free collaborative knowledgebase. Communications of the ACM, 57(10), 78-85.
[6] Mann M, et al. Open Drug Knowledge Graph. 2021.
[7] Lelong R, et al. Building a semantic health data warehouse in the context of clinical trials: development and usability study. JMIR Med Inform. 2019;7(4):e13917.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210910

Artificial Intelligence Models for Heart Chambers Segmentation from 2D Echocardiographic Images: A Scoping Review

Dawoud AL KINDI[a,b], Mowafa HOUSEH[a], Tanvir ALAM[a] and Zubair SHAH[a,1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*
[b] *Department of Cardiology, Heart Hospital, Hamad Medical Corporation, Doha, Qatar*

Abstract. Echocardiography (echo) is a non-invasive, safe, widely available imaging modality that is frequently used to assess the heart structure and function. Accurate heart chamber segmentation is an essential step to quantify certain parameters, including heart chamber volumes. In clinical practice, this task is manually done by echo experts, where it consumes considerable time and is subjective to both errors as well as intra-operator variability.. Artificial intelligence (AI) models have been used to automatically segment heart chambers. We conducted a scoping review to provide an overview of the AI models used for this task. Three bibliographic databases; PubMed, Embase, and Google Scholar were explored. Out of 640 initially retrieved studies, 36 studies were included. Multiple AI models used for echo images segmentation were identified, which can be broadly categorized into five methods: low-level image processing, deformable-based, statistical techniques, machine learning (ML), and deep learning-based (DL) techniques. The initial three categories were relatively simple and required less computational complexity compared to the ML and DL models. The convolutional neural network was the most widely used DL-based technique in most-recent publications. Generalizability of the models is a major concern that needs to be addressed in the future. Well-annotated larger 2D echo image datasets would be required to mitigate the challenges to some extent.

Keywords. Artificial intelligence; Deep learning; 2D echocardiography; segmentation

1. Introduction

Cardiovascular disease (CVD) is the leading cause of death worldwide [1]. Several imaging modalities are used for diagnosing this disorder, including echocardiography (echo), computed tomography (CT), cardiac magnetic resonance imaging (CMR), and cardiac nuclear imaging. Echo is a non-invasive, safe, and most widely-used imaging modality that plays a crucial role in the diagnosis and management of CVD. In 2D echo

[1] Corresponding Author, Zubair Shah, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar; Email: zshah@hbku.edu.qa.

studies, heart chambers images (left ventricle (LV), right ventricle (RV), left atrium (LA), right atrium (RA)) are acquired from multiple views, as well as during contraction (systole) and relaxation (diastole). The four basic views in echo are Apical two chambers (A2C), apical four chambers (A4C), parasternal long axis (PLAX) and parasternal short axis view (PSAX) [2]. The main drawbacks of echo, compared to other imaging modalities, are inconsistency and high inter- and intra-observer variation in image acquisition, analysis, and interpretation [3]. Heart chambers segmentation is done by manual boundary delineation, which is a time-consuming, subjective to errors task that depends on interpreter experience [4,5]. Therefore, it is critically importance to find methods to improve interpretation efficiency and reduce reporting time. Artificial intelligence (AI) models have been leveraged to execute tasks that are usually manually performed by echo experts (i.e., image segmentation and measurements of cardiac structural and functional indices) [6]. This review aims to explore different and the most recently used AI models for 2D echo images segmentation.

2. Methods

The review was conducted following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines [7]. A systematic search was conducted using three electronic databases, PubMed, Google Scholar, and EMBASE. The search terms were selected based on population, intervention, and outcome. We searched for terms related to 2D echo images and all AI segmentation models used for heart chambers segmentation. There were no age, gender, or health status restrictions. The search focused on the most recently-developed AI models; therefore, studies published from January 2010 to date were included. Because the Google Scholar search resulted in a large number of studies, only the first 100 studies were included. In addition, backward reference searching of the included studies and any relevant reviews was conducted. The titles and abstracts of all possible studies were reviewed, then full-text reading was conducted to identify the eligible studies. The data from the final list of included studies were extracted into an Excel sheet. Lastly, the data from the included studies were synthesized in a narrative approach.

3. Results

As shown in the PRISMA chart in Appendix 1, 36 studies satisfied the eligibility criteria and were included in this review from the initial 640 studies retrieved from the databases search. The included studies are listed in Appendix 2. The methods used for segmentation fell into five main categories: low level image processing (n=4), deformable-based (n=7), statistical techniques (n=6), machine learning (ML) (n=4) and deep learning (DL)-based (n=15) techniques [8]. The CNN U-net model was incorporated as the backbone architecture in 5 studies. Most of the studies (n=24) used AI models for LV segmentation, while only one publication discussed RV segmentation alone. Two studies provided a framework for comprehensive automated echo images analysis and interpretation and used algorithms to segment all heart chambers (LV, LA RV, and RA). One study included a dataset of children's echo images, while five studies used fetal images. The remaining studies included adult echo images (n=30). The characteristics of the included studies can be found in Appendix 3.

The AI models used in the low-level processing-based category include the watershed algorithm, thresholding, morphological appearance model, and the level-set algorithm. Active contouring was widely used for the deformable-based category; pSnakes, B-Spline, K-means clustering algorithms were used for this model. Other deformable-based models included phase-based level-set evolution and constrained level-set. The third category was based on statistical methods. An active appearance model was used for fetal heart segmentation. Two studies utilized an active shape model (ASM), while two others combined ASM with Random Forest. Classification algorithms were mainly used in ML approaches. The models used in this category included the shape regression machine model, Bayesian probability maps, and adaptive group sparse representation model. Another ML approach used a sparse matrix transform model combined with a level-set model. However, most echo image segmentation studies published in recent years focused on DL models. The majority of these models (n=14) are based on a convolutional neural network (CNN). Furthermore, many approaches were followed to design derivatives to CNNs for better segmentation performance such as encoder-decoder model, bilateral segmentation network and dynamic CNN. ML and DL models datasets included large data sets that ranged from 350 to 1500 echo images.

4. Discussion

4.1 Principal Findings

The objective of this scoping review is to summarize AI techniques used for 2D echo image segmentation. From the included 36 studies, we identified many AI models that had been experimented with. As LV parameters carry the most clinical importance, most studies focused on this chamber segmentation [3]. AI segmentation models fall into five main categories. One of the initial methods was low-level processing models. The performance of these techniques are acceptable and are not computationally exhaustive. But these techniques perform poorly in low-quality images. Deformable models require higher computational power and are sensitive to the initial contours. Segmentation accuracy was enhanced by statistical techniques compared to low-level processing and deformable models, but they were found to be sensitive to the presence of variations in shape or appearance. Conventional ML and DL methods have been extensively studied recently. These models require large, well-annotated datasets. U-Net has been incorporated as the backbone architecture in many studies for various medical imaging segmentation. In our review, we identified studies that either used U-Net architecture (n=5) or compared the performance of their proposed model to U-Net. The modifications in the U-Net achieved better performance than the vanilla U-Net.

Moreover, DL models can be designed to perform tasks related to image interpretation apart from segmentation alone. Zhang et al. [9] and Arafati et al. [10] proposed a framework that included all the four heart chambers segmentation while also measuring other important parameters. The model by Arafati et al. designed a fully convolutional neural network combined with adversarial training and post-processing optimization, which was compared to other DL models. Results showed superior performance of this model. Over the past decade, an overall trend to improve AI models' performance based on 2D echo image segmentation has been observed.

4.2 Practical Implications

The majority of AI models utilized for heart chamber segmentation based on 2D echo image are not only comparable to experts' manual performance, but also requiree lower execution time. Utilizing such AI-based frameworks in clinical setups would reduce the burden on clinicians and improve the diagnosis plan. We recommend that researchers focus on improving existing DL models. Despite the presence of few online public echo image datasets, there is still a need for larger and well-annotated datasets to achieve generalizable AI models.

5. Conclusions

In this study, we highlighted the AI-based techniques that are used for 2D echo image segmentation. DL-based techniques are the most recent and highly accurate methods to perform this task. Availability of large and well-annotated echo image datasets may help the community to improve existing DL models.

Appendix files are available at GitHub: https://github.com/DawdAlKindi/Appendix .

References

[1] Virani SS, Alonso A, Aparicio HJ, et al. American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association. Circulation. 2021 Feb 23;143(8):e254-e743.

[2] Mitchell C, Rahko PS, Blauwet LA, et al. Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: Recommendations from the American Society of Echocardiography. J Am Soc Echocardiogr 2019;32:1–64.

[3] Davis A, Billick K, Horton K, et al. Artificial Intelligence and Echocardiography: A Primer for Cardiac Sonographers. J Am Soc Echocardiogr. 2020 Sep;33(9):1061-66.

[4] Bhattacharyya S, Lloyd G. Improving appropriateness and quality in cardiovascular imaging. Circulation: Cardiovascular Imaging. 2015;8(12):e003988.

[5] Kusunose K, Shibayama K, Iwano H, et al. Reduced variability of visual left ventricular ejection fraction assessment with reference images: the Japanese Association of Young Echocardiography Fellows multicenter study. J Cardiol. 2018;72(1):74–80.

[6] Kusunose K. Steps to use artificial intelligence in echocardiography. J Echocardiogr. 2021 Mar;19(1):21-27.

[7] Tricco A, Lillie E, Zarin W, D et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Annals of Internal Medicine. 2018;169(7):467.

[8] Zamzmi G, Hsu LY, Li W, et al. Harnessing Machine Intelligence in Automatic Echocardiogram Analysis: Current Status, Limitations, and Future Directions. IEEE Rev Biomed Eng. 2021;14:181-203.

[9] Zhang J, Gajjala S, Agrawal P, et al. Fully Automated Echocardiogram Interpretation in Clinical Practice. Circulation. 2018 Oct 16;138(16):1623-1635.

[10] Arafati A, Morisawa D, Avendi MR, et al. Generalizable fully automated multi-label segmentation of four-chamber view echocardiograms based on deep convolutional adversarial networks. J R Soc Interface. 2020 Aug;17(169):20200267.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210911

Recent Developments in Artificial Intelligence-Based Techniques for Prostate Cancer Detection: A Scoping Review

Uzair SHAH[a], Md. Rafuil BISWAS[a], Mahmood Saleh ALZUBAIDI[a], Hazrat ALI[a], Tanvir ALAM[a], Mowafa HOUSEH[a] and Zubair SHAH[a, 1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*

Abstract. Artificial intelligence (AI) techniques can contribute to the early diagnosis of prostate cancer. Recently, there has been a sharp increase in the literature on AI techniques for prostate cancer diagnosis. This review article presents a summary of the AI methods that detect and diagnose prostate cancer using different medical imaging modalities. Following the PRISMA-ScR principle, this review covers 69 studies selected from 1441 searched papers published in the last three years. The application of AI methods reported in these articles can be divided into three broad categories: diagnosis, grading, and segmentation of tissues that have prostate cancer. Most of the AI methods leveraged convolutional neural networks (CNNs) due to their ability to extract complex features. Some studies also reported traditional machine learning methods, such as support vector machines (SVM), decision trees for classification, LASSO, and Ridge regression methods for features extraction. We believe that the implementation of AI-based tools will support clinicians to provide better diagnosis plans for prostate cancer.

Keywords. Prostate cancer, medical imaging, machine learning, deep learning.

1. Introduction

After lung cancer, prostate cancer is the second most common cancer in men [1]. Reports have projected that the number of prostate cancer cases may exceed the number of lung cancer cases in men in just over a decade [1], [2]. However, early diagnosis of prostate cancer can decrease the fatality and morbidity rates. In clinical practice, the diagnosis is typically performed by a transrectal ultrasound and blood tests for prostate-specific antigens [3]. Usually, the severity of prostate cancer is measured in terms of Gleason score (ranked from 6 to 10), with a higher score representing high-grade cancer that is more likely to spread in the tissue [4], [5]. Analyzing and grading prostate cancer scores require trained professionals, who usually rank the scores through manual screening and mutual consensus of many experts relying on their skills.

The AI research community has made progress in developing AI-based methods to support pathologists and radiologists, thus improving the overall efficiency of the process of diagnosing prostate cancer [6]. Typically, AI-based methods can enable quick

[1] Corresponding Author. Zubair Shah, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar. Email: zshah@hbku.edu.qa

processing and accelerate the diagnosis process while ensuring consistency [7]. In prostate cancer, AI assists in systematic pathological grading to evaluate prostate cancer stratification and care [8]. While many published methods exist proposing the use of AI to diagnose prostate cancer, there are few comprehensive review that may provide a quick insight to readers exploring the recent developments in AI's use for studying prostate cancer. This short review will serve as a quick reference for readers interested in studying and researching the role of AI methods in treating prostate cancer.

2. Methods

We performed a scoping review to highlight the advancements of AI-based tools detecting prostate cancer. We followed the guidelines from the *Preferred Reporting Items for Systematic Review and Meta-Analysis* (PRISMA-ScR) [9] to perform the review (Appendix 1). The search strategy, which included population (i.e., prostate cancer), target intervention (i.e., artificial intelligence), and target outcome (i.e., diagnosis), was applied to the most widely used databases (PubMed, Medline, Embase, and Google Scholar) in the medical domain. The search process was carried between February 10th and 14th, 2021 and extracted relevant articles published over the last three years covering the most recent publications in this domain.

3. Results and Discussion

A total of 1476 studies were retrieved by searching four databases (details in Appendix 4). Amongst these, 574 duplicate studies were excluded. Through title and abstract screening, 767 studies were excluded following the exclusion criterion of being non-English studies, non-peer-reviewed articles, or scoping review text. Through full-text screening of the remaining 135 studies, 66 studies were further removed following the exclusion criteria. A total of 69 studies were included in this review. The key characteristics of the articles surveyed are summarized in Table 1.

Table 1: Characteristics of AI-based techniques and data modalities reported for prostate cancer.

Characteristics	**Number of Studies**				
Model purpose	Diagnosis of prostate: 43		Grading prostate: 15		Segmentation:11
AI Branches	Deep Learning: 49		Machine Learning and Deep Learning: 2		Segmentation:11 Machine Learning:18
AI method	CNN:45	ANN:4	SVM:4	Random Forest: 3	Logistic Regression: 2
Feature Extraction Technique	CNN:45 LASSO:2 Radiomics :4	LB-FCN light:1 CSLBP:1 CFS:1	DCE:1 SFIT & EFDs:1 CDF & ADC:1	Genetic Algorithm:1 FCC:1 Mean Region of Interest:1	Histogram:1 Gray level co-occurrence matrix:1 NA:8
Imaging modality	MRI/PET/ CT: 34	Biopsy whole slide images: 13	Tissue Micro arrays:6	Histopathological data: 04	Genetic Data: 03

Index: ANN: Artificial Neural Network. CSLBP: completed and statistical local binary pattern. CFS: Correlation Feature Selection. DCE: Dynamic contrast-enhanced. SFIT: Scale-invariant feature transforms. EFDs: Elliptic Fourier Descriptors. CDF: Cumulative Distribution function. ADC: Apparent Diffusion Coefficient. FCC: Frequency Cepstral Coefficients. MRI: Magnetic Resonance Imaging. PET: Positron Emission Tomography. CT: Computed Tomography

Out of these 69 studies, 94% (n=65) were peer-reviewed journal articles, while the remaining were conference papers. Of these, 20% (n=12) were published in 2018, 39% (n=26) in 2019, and 36% (n=25) in 2020. The highest number of studies were published from the United States (n=16), followed by China (n=11) and Canada (n=8) (Appendix 2).

For the majority of the studies, the data was acquired from public repositories such as the ProstateX challenge dataset, PRMISE-12 dataset, and NCBI. The most dominant imaging modalities were magnetic resonance imaging (MRI), positron emission tomography (PET) scan, and computerized tomography (CT) scan. The datasets' descriptions are summarized in Table 2.

Table 2: Descriptions and Statistics of Public and Private Dataset for Prostate Cancer

Dataset	Host/Source	Type of data	Size (No. of samples)	Studies that reported the use (listed in Appendix 3)
Public Dataset	clinicaltrials.gov (NIH)	PET/MRI	122	42
	Registry of Catastrophic Illness Patient (subpart NIH)	Electronic Health Record (EHR)	20355	62
	Harvard Medical School and Brigham woman's hospital	MRI	682	52
	GLOBOCAN 2018	Ultrasound	1200	48
	ProstateX challenge data	mp-MRI	538	50, 20, 38
	National Centre of Biotechnology (NCBI)	Genetic dataset	179	25, 67
	MICCAI Prostate MR Image Segmentation 2012 (PROMISE12)	MRI	4050	32, 70
	ProstateGlandDB dataset	Biopsy whole slides images	35	13, 64
	Horosproject.org	PET Images	7336	71
	NCI-ISBI 2013 challenge	MRI	771	5
Private Datasets	Boramae Medical Center	EHR	3791	40
	The University of Alabama at Birmingham (UAB)	MRI Scan	1269	22
	Univ. of Texas Southwestern Medical Center	CT scan	85	19

Features extraction techniques were reported in 58 studies. Most of the included studies used CNN (n=45, listed as 3, 5, 6, 7, 9, 10 – 12, 14, 15, 18, 20 – 22, 24, 25, 28, 33 – 37, 39, 42, 43, 45 – 49, 50, 53, 55, 57, 58, 60 – 64, 66 – 70 in Appendix 3) followed by radiomics (n=4, listed as 30, 31, 38, 41 in Appendix 3) and LASSO (n=2, listed as 4, 44 in Appendix 3). Other reported methods for feature extractions were gray level co-occurrence matrix (54 in Appendix 3), mean region of interest (71 in Appendix 3), genetic algorithms (13 in Appendix 3), completed and statistical local binary pattern (52 in Appendix 3), apparent diffusion coefficient and cumulative distribution function (19 in Appendix 3), frequency cepstral coefficient (26 in Appendix 3), and correlation features selection algorithm (32 in Appendix 3).

It is challenging to compare the performance of the studies as each study utilized different feature extraction techniques, imaging modility and performance metrics. However, CNN were the prominent feature extraction techniques regardless of the imaging modility [10]. This may be in part because CNN is sensitive to the training data compared to radiomics and usually requires large data for better performance.

4. Conclusion and Future Direction

This review article identifies three main themes on prostate cancer detection, i.e., diagnosis, grading, and segmentation of histopathological images where AI-based methods were leveraged. We could not find significant application of the AI-based methods for the treatment and recommended medications for prostate cancer. The studies were categorized based on the usage of the AI methods, feature extraction techniques, and types of the dataset used. The use of these AI techniques are limited to academic and research purposes only and their real-life applications into clinical practice are currently limited – though a few cases are available where the AI-based tools have been used in clinical practices. Nevertheless, with the rapid progress of AI, technology readiness levels need to be improved for utilizing these methods in real-life diagnosis of prostate cancer.

The appendix is available at https://github.com/hazratali/appendix

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209-249. doi: 10.3322/caac.21660. Epub 2021 Feb 4. PMID: 33538338.

[2] Report by World Cancer Research Foundation, available at: https://www.wcrf.org/dietandcancer/prostate-cancer/ (Accessed: 25 August 2021)

[3] Aldoj N, Lukas S, Dewey M, Penzkofer T., "Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network," Eur. Radiol., vol. 30, no. 2, pp. 1243–1253, 2020, doi: 10.1007/s00330-019-06417-z

[4] H. Bonkhoff, "Gleason grading: diagnostic criteria and clinical implications.," Pathologe, vol. 26, no. 6, pp. 422–432, Nov. 2005, doi: 10.1007/s00292-005-0793-y

[5] Reda, Islam, Babajide O. Ayinde, Mohammed Elmogy, Ahmed Shalaby, Moumen El-Melegy, Mohamed Abou El-Ghar, Ahmed Abou El-fetouh, Mohammed Ghazal, and Ayman El-Baz. "A new CNN-based system for early diagnosis of prostate cancer." In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 207-210. IEEE, 2018. doi: 10.1109/ISBI.2018.8363556

[6] Pantanowitz, Liron, Gabriela M. Quiroga-Garza, Lilach Bien, Ronen Heled, Daphna Laifenfeld, Chaim Linhart, Judith Sandbank et al. "An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study." Lancet Digit. Heal., vol. 2, no. 8, pp. e407–e416, 2020, doi: 10.1016/S2589-7500(20)30159-X

[7] Egevad, Lars, Daniela Swanberg, Brett Delahunt, Peter Ström, Kimmo Kartasalo, Henrik Olsson, Dan M. Berney et al. "Identification of areas of grading difficulties in prostate cancer and comparison with artificial intelligence assisted grading." Virchows Arch., vol. 477, no. 6, pp. 777–786, 2020, doi: 10.1007/s00428-020-02858-w

[8] S. A. Harmon, S. Tuncer, T. Sanford, P. L. Choyke, and B. Türkbey, "Artificial intelligence at the intersection of pathology and radiology in prostate cancer.," Diagn. Interv. Radiol., vol. 25, no. 3, pp. 183–188, May 2019, doi: 10.5152/dir.2019.19125

[9] Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K.K., Colquhoun, H., Levac, D., Moher, D., Peters, M.D., Horsley, T., Weeks, L. and Hempel, S., "PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation," Annals of Internal Medicine, vol. 169, no. 7. pp. 467–473, Oct. 02, 2018, doi: 10.7326/M18-0850

[10] Daniel Truhn, Simone Schrading, Christoph Haarburger, Hannah Schneider, Dorit Merhof, and Christiane Kuhl "Radiomic versus Convolutional Neural Networks Analysis for Classification of Contrast enhancing Lesions at Multiparametric Breast MRI" Radiology 2019 290:2, 290-297, doi: 10.1148/radiol.2018181352

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210912

Findings from a Panel Discussion on Evaluation Methods in Medical Informatics

Felix HOLL[a,b1], Marina FOTTELER[a], Stefan MÜLLER-MIELITZ[c] and Walter SWOBODA[a]
[a] *DigiHealth Institute, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany*
[b] *IBE, Ludwig Maximilian University of Munich, Munich, Germany*
[c] *IEKF GmbH, Ibbenbüren, Germany*

Abstract. Healthcare systems are challenged by increasing costs. Digital technology can help to combat this trend. Evaluation of these technologies is uncommon or incomplete. Scholars have called for a standardized and holistic evaluation. We provide a synthesis of an online panel on medical informatics (MI) and stipulate a discussion on new guidelines for medical informatics project evaluations. The panel consisted of presentations and a discussion. The presentations gave the participants an overview of evaluation methods currently used in different medical informatics domains and their shortcomings. The presenters highlighted new evaluation methods such as a roadmap for economic analysis of eHealth projects and the German Digital Healthcare Act methods. Participants discussed the shortcomings of RCTs and methods that need to be included in eHealth evaluation and called for new evaluation methods. The discussion showed weaknesses of the currently used methods and underlined the need for a new, holistic evaluation standard for MI.

Keywords. Evaluation Research, eHealth, Mobile Health

1. Introduction

Improved life expectancy, driven by new drugs, treatments, and increased multimorbidity, have led to a dramatic rise in healthcare costs in the past decades and are challenging healthcare systems globally (1). The use of digital technologies (DT) in healthcare delivery can be a means to combat the trend of increasing cost (2). The potential of DT has fueled the development of new information technologies for healthcare and increased investment in new IT solutions, and overall substantial market growth in recent years (3). The growth includes traditional medical informatics domains such as clinical information technology (IT) systems like hospital information systems, but also emerging technologies like mobile health applications and smart assistive technologies (AT) for the elderly. While these new technologies, especially using mobile devices, may increase healthcare expenditure at first, they also can improve healthcare delivery and make it more efficient and potentially reduce costs in the long run (1).

There is substantial literature for the evaluation of MI projects from theoretical textbooks to reviews (4–7). The STARE-HI provides a recommendation for evaluation results reporting (8). Despite the existing methods and literature, new technologies and

[1] Corresponding Author, Felix Holl, DigiHealth Institute, Neu-Ulm University of Applied Sciences, 89231 Neu-Ulm, Germany; E-mail: felix.holl@hnu.de.

applications for health are often not evaluated systematically, and when evaluations are performed, they often not generalized and focus on individual aspects. Scholars have argued for the importance of a holistic, standardized, and comprehensive eHealth evaluation (9,10). The evaluation of MI projects has gained importance through recent developments such as the Digital Healthcare Act (DVG) in Germany and the European Medical Device Regulation (11,12).

This research aims to provide a synthesis of a panel about evaluation methods in medical informatics, highlights the shortcomings of the evaluation methods currently used and stipulate a discussion about new guidelines for evaluations of MI projects.

2. Methods

The panel had been accepted at the *Medical Informatics Europe 2020* conference, but panels were canceled due to the COVID-19 pandemic. The panel was held online as part of the *EFMI webinar series* in July 2020 with a duration of seventy minutes and consisted of four panel presentations with 10 minutes each and a moderated expert discussion with a duration of 30 minutes. The presentations aimed to give an overview of evaluation methods currently used in different areas of MI and related fields and to make participants aware of these methods' shortcomings. The second aim of the panel presentations was to inform participants about the research currently conducted to improve medical informatics evaluation methods. The expert discussion included the panelists and the audience and aimed to identify shortcomings of current evaluation methods, identify new approaches, and formulate recommendations for different evaluation methods. Participants were asked to share experiences with evaluation shortcomings and methodologies and frameworks they have used in their work. Participants were informed about the recording before the workshop. The panel recorded, transcribed and analyzed using grounded theory (13).

3. Results

3.1. Panel presentations

The panelists' diverse backgrounds, ranging from medicine and public health to medical informatics and economics, represented medical informatics interdisciplinarity. Panelist 1, with a background in MI and public health, is researching the socioeconomic impact of AT. Panelist 2 has a background in MI and global health with experience in the implantation and evaluation of mHealth technologies in Germany and low-and-middle-income countries. He is conducting research on evaluation methods for mobile health applications. Panelist 3 is an economist with a MI certificate and the CEO of a consultancy firm for eHealth & economics. Panelist 4 is a physician and computer scientist with 15 years of experience as a CIO of large hospitals. He is a research professor for MI at a university of applied sciences. All panelists were from Germany but previously have worked internationally in Denmark, the US, and southern African countries.

The first panelist shared the results on synthesizing the existing evidence for AT's effectiveness for the elderly, describing which types of technology are most feasible and beneficial and focus on the frailty of elderly people. Out of 11,400 records retrieved

during the search, only 19 trials met all inclusion criteria (the most important ones being: RCT, study population ≥65, no laboratory setting). Nine of those were pilot studies; five studies include caregivers, and only two considered AT's economic aspects. The second presenter gave an overview of the current state of the evaluation of mHealth applications. Key findings are that no published mHealth projects have been holistically evaluated, and most evaluations focus on individual aspects. User experience was assessed most often. No holistic evaluations were performed in published studies included in the analysis. The presenter defined holistic evaluations as including all relevant aspects in the assessment rather than looking at a single aspect. He also presented the DVG, which gives publicly insured patients to right to access digital healthcare services and the reimbursement of these services. The act includes a "fast-track" approval mechanism with a preliminary approval if an application fulfills a set of basic requirements and the obligation to prove a positive care effect (11). As an example of ongoing work on developing new evaluation methods in MI, the third panelist gave an overview of the work of the eHealth working group of the German Health Economic society (dggö) on developing a 12-point road map for the economic analysis of eHealth projects (14). The standard process allows better comparison of the results. These 12 points include all the aspects that have to be considered in an economic evaluation of eHealth applications in the opinion and consensus of the eHealth working group of the dggö. The final presenter highlighted the distinction between the evaluation of new treatments and drugs compared to the assessment of digital technologies in medicine, including a description of several additional aspects that need to be included in evaluating digital technologies. An example from the presenter's tenure as the CIO of a large university medical center concluded the presentations, where a holistic evaluation was performed before introducing a digital documentation system, in which improved job satisfaction was scored higher than increased costs and no additional medical benefit.

3.2. Expert discussion

The participants in the expert discussion came from the Netherlands, Malaysia, and Germany. There was consensus among the panelists and participants that the current evaluation methods used for digital solutions in medicine are inadequate and neglect several different aspects. Multiple participants stated that RCTs might not be the perfect solution. RCTs is a study design focused only on the outcome but not on additional dimensions of a digital solution. RCTs require extensive preparation and have a long duration. Moreover, RCTs are methodologically flawed for this field; blinding is challenging or impossible for eHealth interventions.

Technical aspects, such as reliability and usability, and other aspects like ethics and privacy, which are challenging to consider in an RCT, are neglected. A panelist concluded, "that RCTs should not be the standard anymore for the evaluation of telemedical and eHealth solutions." The participants agree that the evaluation should start as early as possible, ideally during the development process of new technologies and devices, and involve all relevant stakeholders. Aspects beyond the current patient-/user-focus need to be included in the evaluation. Ideally, a new study design should be created, which should be better suited to evaluating digital health solutions. Participants agreed that the development of new evaluation methods for eHealth has to be placed on the research agenda. One participant stated that "the most important aspect is that we as a scientific community have to try to establish a new, holistic evaluation standard that is broadly used and validated."

4. Discussion

With examples from different MI domains, the presentations highlighted shortcomings of currently used methods and showed examples of new techniques being developed. The review of AT for the elderly showed that existing studies are not holistic enough, in addition to the lack of evidence. Also, the methods used to evaluate mHealth applications are often insufficient, including recent developments such as the methods used in the DVG. The work on the 12-point road map for eHealth projects' economic analysis showed the complexity of holistic evaluation and highlighted their need. The closing presenter's remarks gave an overview of the aspects that need to be included and highlighted the need for holistic evaluation through a real-world example. The discussion underlined these themes and the agreement that RCT studies perfectly fit the evaluation of new drugs, but they are not ideally suitable for evaluating digital health devices. The task is now to develop a new method suitable for the holistic assessment of new digital solutions. Examples from different MI domains have shown that the evaluation methods currently used often only look at individual aspects and neglect important factors. As the gold standard for evaluating medical interventions, RCTs are too complicated for the fast-changing medical informatics environment while disregarding essential elements in the evaluation. A new, holistic evaluation standard for medical informatics that is broadly used and validated should be designed and established by the scientific community, especially for quickly evolving technologies such as mHealth.

References

[1] Rodriguez Santana I, Aragón MJ, Rice N, Mason AR. Trends in and drivers of healthcare expenditure in the English NHS: a retrospective analysis. Health Econ Rev. 2020;10(1):1–11.

[2] Kannry J, Smith J, Mohan V, Levy B, Finnell J, Lehmann CU. Policy Statement on Clinical Informatics Fellowships and the Future of Informatics-Driven Medicine. Appl Clin Inform. 2020 Oct 28;11(5):710–3.

[3] Koch S. Current Trends in Biomedical and Health Informatics. Methods Inf Med. 2019 Sep 12;58(02/03):061–2.

[4] Friedman CP, Wyatt J. Evaluation methods in biomedical informatics. Hannah KJ, Ball MJ, editors. New York, NY: Springer New York; 2006. 414 p. (Health Informatics).

[5] Brender J. Handbook of Evaluation Methods for Health Informatics. Handbook of Evaluation Methods for Health Informatics. Elsevier Inc.; 2006.

[6] Eslami Andargoli A, Scheepers H, Rajendran D, Sohal A. Health information systems evaluation frameworks: A systematic review. Int J Med Inform. 2017;97:195–209.

[7] Fanta GB, Pretorius L, Erasmus L. An evaluation of ehealth systems implementation frameworks for sustainability in resource constrained environments: A literature review. IAMOT 2015 - 24th Int Assoc Manag Technol Conf Technol Innov Manag Sustain Growth, Proc. 2015;1046–63.

[8] Talmon J, Ammenwerth E, Brender J, de Keizer N, Nykänen P, Rigby M. STARE-HI-Statement on reporting of evaluation studies in Health Informatics. Int J Med Inform. 2009;78(1):1–9.

[9] Catwell L, Sheikh A. Evaluating eHealth interventions: The need for continuous systemic evaluation. PLoS Med. 2009;6(8):1–6.

[10] Enam A, Torres-Bonilla J, Eriksson H. Evidence-based evaluation of ehealth interventions: Systematic literature review. J Med Internet Res. 2018;20(11).

[11] Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: lessons and opportunities for other countries. npj Digit Med. 2020;3(1):1–6.

[12] Becker K, Lipprandt M, Röhrig R, Neumuth T. Digital health – Software as a medical device in focus of the medical device regulation (MDR). it - Inf Technol. 2019 Oct 1;61(5–6):211–8.

[13] Birks M, Mills J. Grounded Theory A Pratical Guide. 2nd Editio. SAGE Publications Ltd; 2015.

[14] Köberlein-Neu J, Müller-Mielitz S. Roadmap zur Entwicklung eines Evaluationskonzeptes. In: Müller-Mielitz S, Lux T, editors. E-Health-Ökonomie. Springer Fachmedien Wiesbaden; 2017. p. 881–92.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210913

The Effect of Short Messaging System-Based Feedback on Physicians' Head CT Scan Ordering Behavior in Neurology and Neurosurgery Departments

Zahra MEIDANI[a,b], Afsaneh MOOSAVI[a], Hossein AKBARI[c], Felix HOLL[d,e,f] Sahar ZARE[a], Ehasn NABOVATI[a,b,1], Reza DANESHVAR[g], Ali MASOUD[g], Esmaeel FAKHARIAN[h] and Fakhrosadat MIRHOSSEINI[i]

[a] *Health Information Management Research Center, Kashan University of Medical Sciences, Kashan, Iran*
[b] *Department of Health Information Management &Technology, Kashan University of Medical Sciences, Kashan, Iran*
[c] *Department of Biostatistics and Epidemiology, Kashan University of Medical Sciences, Kashan, Iran*
[d] *IBE, Ludwig Maximilian University of Munich, Munich, Germany*
[e] *DigiHealth Institute, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany*
[f] *IGHS, University of California, San Francisco, San Francisco, CA, USA*
[g] *Department of Neurology, Kashan University of Medical Sciences, Kashan, Iran*
[h] *Trauma Research Center, Kashan University of Medical Sciences, Kashan, Iran*
[i] *Department of Anesthesia, Kashan University of Medical Sciences, Kashan, Iran*

Abstract. Short messaging system (SMS) works as one of the most popular strategies for physicians' behavior change via sending feedback and reminder messages. One of the areas in which SMS feedback can be effective on physicians' behavior is CT scan ordering. This study investigates the effect of mobile phone SMS feedback on residents' head CT scan ordering at a general teaching hospital in Iran. Through a three-month before-after experimental study, an intervention was conducted, and the CT scans ordered by an individual resident were evaluated every two weeks. Consequently, personal SMS-based feedback was provided to the residents, and the rate of CT per patient in the two phases of the study was analyzed. The mean CT scan ordered per patient decreased from 1.98 ± 1.09 to 1.74 ± 1.45, and this decrease was insignificant (P= 0.106). SMS-based feedback can reduce head CT scan ordering among residents; whereas this decline was not significant further studies are required to investigate its effectiveness.

Keywords. Cell phone, feedback, tomography, X-Ray computed, neurology, neurosurgery

[1] Corresponding author, Ehsan Nabovati, Health Information Management Research Center, Kashan University of Medical Sciences, Kashan, Iran; Email: Nabovati@kaums.ac.ir

1. Introduction

Mobile technologies as more affordable and accessible information technology in low-income countries present more opportunities to strengthen health services [1]. In recent years, SMS is one of the most popular strategies for physicians' behavior change via sending feedback and reminders messages [2]. One of the areas in which SMS-based feedback can change physicians' behavior is the CT scan examinations, following the steady CT utilization increase [3]. CT examinations increased from three million in 1980 to 80 million in 2014 [4]. The massive imaging volume imposes a tremendous cost to the health care system, and CT imaging contributes to 1.5–2% of all cancers [5-6]. Regardless of these adverse outcomes, it is argued that 20-50% of imaging procedures may be unnecessary [7]. The literature has reported that sending feedback can significantly improve physicians' behaviors regarding ordering and reduce hospital costs [8]. Therefore, this study was conducted to evaluate the effect of mobile phone SMS-based feedback on residents' head CT scan ordering behavior.

2. Method

This quasi-experimentally study was conducted during one-month pre-intervention and two months the intervention phases among the residents of neurology and neurosurgery departments in a general teaching hospital in Iran. To perform the intervention, the CT scans ordered by an individual resident were evaluated every two weeks. Personal SMS-based feedback containing information about CT scans ordered, including the number and cost, the patient exposure to radiation dose, and short educational tips, were provided to the residents. Consequently, the rate of CT per patient in the two phases of the study was analyzed using mean and standard deviation (SD) and chi-square, Fisher's exact test, and paired t-test. Sample content of the SMS feedback is presented in Figure 1.

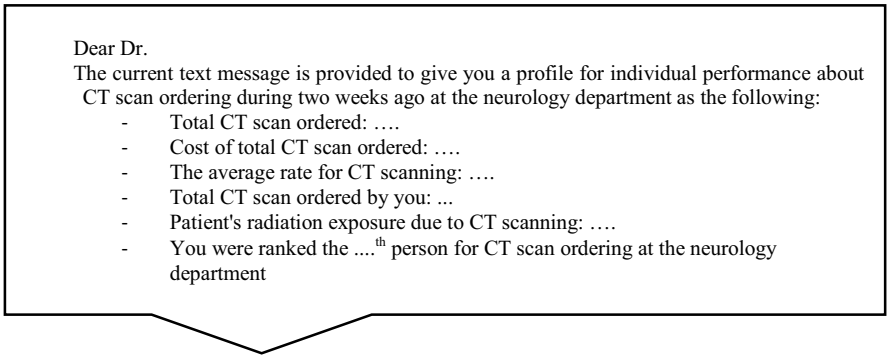


Figure 1. Sample content of the SMS feedback

The current study was supported by Vice-Chancellorship for Research & Technology at Kashan University of Medical Sciences [Grant number 97204] and National Agency for Strategic Research in Medical Education (NASR) [Grant Number 981867].

3. Results

The CT scan per patient was 1.98 in the pre-intervention phase and 1.74 post-intervention phase. Also, the CT scan ordered per resident in the two phases of the intervention was 11.67 and 11.64, respectively (Table 1).

Table 1. CT scan ordered per patient and per resident during phases of the study

Departments	Phases of study / Variables	Pre-intervention	Intervention	Total
Neurology	Number of CT	49	178	266
	Number of patients	47	132	179
	Number of residents	10	10	10
	CT scan per patient	1.68	1.41	1.48
	CT scan per resident	7.9	9.35	8.86
Neurosurgery	Number of CT	166	302	468
	Number of patients	77	149	226
	Number of residents	11	11	11
	CT scan per patient	2.15	2.02	2.07
	CT scan per resident	15.09	13.72	14.18
Total	Number of CT	245	489	734
	Number of patients	124	281	405
	Number of residents	21	21	21
	CT scan per patient	1.98	1.74	1.81
	CT scan per resident	11.67	11.64	11.65

Table 2 indicates the total ordered brain CT scans were 245 for 124 patients in the pre-intervention phase and 489 for 281patients in the intervention phase. The mean CT scan ordered per patient decreased from 1.98 ± 1.09 in the pre-intervention phase to 1.74 ± 1.45 during the intervention phase. However, the decrease of total CT scan ordered was not statistically significant during pre-intervention and intervention phases (P= 0.106).

Table 2. CT scan utilization per patient in two phases of pre-intervention and intervention

Departments	Phases of study	Number of patients	Number of CT scans	CT scan per patient (Mean ± SD)	p-Value
Neurology	Pre-intervention	47	79	1.68±0.84	0.074
	Post intervention	132	187	1.42±0.87	
Neurosurgery	Pre-intervention	77	166	2.16±1.19	0.564
	Post intervention	149	302	2.03±1.76	
Total	Pre-intervention	124	245	1.98±1.09	0.106
	Post intervention	281	489	1.74±1.45	

4. Discussion

The introduction of SMS-based feedback can reduce head CT scan ordering among residents, whereas this decline was not significant. However, the findings of the current study are in disagreement with Eccles et al. study, in that they reported the reminder messages feedback is ineffective and does not reduce requests for radiological examinations among general practitioners [3]. Sarafi Nejad's study confirms that the feedback via short text messages feedback has a positive impact on prescribing

parenteral steroids by general practitioners [9]. Since existing literature suggested a lack of consistency in reported results across the studies and the potential for bias, making any direct cause-and-effect relationship between the SMS-based feedback and behavior change by physicians may not be reliable. The limitations of this study include a small sample size and short time period of the study; the post intervention follow up to investigate the long-term effect of SMS intervention on physicians' behavior was cancelled due to the emergence of the COVID-19 pandemic in Iran. Given our poor understanding of the effectiveness of feedback, education [10-12] and SMS interventions on physicians" behavior, it is recommended that further research of higher quality be conducted to investigate the potential benefits of SMS-based feedback on physicians' behavior.

References

[1] Anstey Watkins JOT, Goudge J, Gómez-Olivé FX, Griffiths F. Mobile phone use among patients and health workers to enhance primary healthcare: A qualitative study in rural South Africa. Soc Sci Med. 2018 Feb;198:139-147. doi: 10.1016/j.socscimed.2018.01.011. Epub 2018 Jan 10. PMID: 29335160

[2] Kurumop SF, Bullen C, Whittaker R, Betuela I, Hetzel MW, Pulford J. Improving health worker adherence to malaria treatment guidelines in Papua New Guinea: feasibility and acceptability of a text message reminder service. PLoS One. 2013 Oct 7;8(10):e76578. doi: 10.1371/journal.pone.0076578. PMID: 24116122; PMCID: PMC3792049

[3] Eccles M, Steen N, Grimshaw J et al. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. Lancet 2001; 357: 1406–9

[4] Friedman DP, Smith NS. Impact of a Collaborative Radiology Utilization Management Program: Does the Specialty of the Referring Provider Matter? AJR Am J Roentgenol. 2016 Jul;207(1):121–5

[5] Fitzgerald R. Error in Radiology. Clin Radiol. 2001;56(12):938–46

[6] Sarma A, Heilbrun ME, Conner KE, Stevens SM, Woller SC, Elliott CG. Radiation and chest CT scan examinations: what do we know? Chest. 2012 Sep;142(3):750–60

[7] Sheng AY, Castro A, Lewiss RE. Awareness, Utilization, and Education of the ACR Appropriateness Criteria: A Review and Future Directions. J Am Coll Radiol. 2016 Feb;13(2):131–6

[8] Khalifa M, Khalid P. Reducing Unnecessary Laboratory Testing Using Health Informatics Applications: A Case Study on a Tertiary Care Hospital. Procedia Computer Science. 2014;37:253-60

[9] Nejad AS, Noori MR, Haghdoost AA, Bahaadinbeigy K, Abu-Hanna A, Eslami S. The effect of registry-based performance feedback via short text messages and traditional postal letters on prescribing parenteral steroids by general practitioners--A randomized controlled trial. Int J Med Inform. 2016 Mar;87:36-43. doi: 10.1016/j.ijmedinf.2015.12.008. Epub 2015 Dec 19. PMID: 26806710

[10] Meidani Z, Mousavi GA, Kheirkhah D, Benar N, Maleki MR, Sharifi M, Farrokhian A. Going beyond audit and feedback: towards behaviour-based interventions to change physician laboratory test ordering behaviour. J R Coll Physicians Edinb. 2017 Dec; 47(4):339-344. doi: 10.4997/JRCPE.2017.407. PMID: 29537404

[11] Meidani Z, Farzandipour M, Davoodabadi A, Farrokhian A, Kheirkhah D, Sharifi M, Khanghani ME. Effect of reinforced audit and feedback intervention on physician behaviour: a multifaceted strategy for targeting medical record documentation. J R Coll Physicians Edinb. 2017 Sep;47(3):237-242

[12] Davaridolatabadi N, Sadoughi F, Meidani Z, Shahi M. The Effect of Educational Intervention on Medical Diagnosis Recording among Residents. Acta Inform Med. 2013;21(3):173-5

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210914

Evaluative Frameworks and Models for Health Information Systems (HIS) and Health Information Technologies (HIT)

Amanda L. JOSEPH[a,1], Eleah STRINGER[a], Elizabeth M. BORYCKI[a,b] and Andre W. KUSHNIRUK[a]
[a] *School of Health Information Science, University of Victoria, Canada*
[b] *Michael Smith Foundation for Health Research, Canada*

Abstract. Evaluation criteria for health information systems (HIS) and health information technologies (HIT) is broad, diverse and lacks a gold standard approach that could be leveraged, to evaluate clinical systems at various stages of their system development life cycle (SDLC). Without generalizable tools such as frameworks or models, comparative analysis across HIS and HIT is not possible. This paper presents the findings from a scoping review, utilizing the Arksey and O'Malley methodology [1]. The objective of this review is two-fold: 1) to classify models and frameworks published between the years 2010-2020 according to their level of evaluative focus (e.g. micro, meso, macro, multi), 2) to identify the countries where these models and frameworks have been employed for the purpose of evaluation, using the International Medical Informatics Association (IMIA) Represented Regions [3]. The results demonstrated the heterogeneity of evaluation models and frameworks currently used in health informatics and reflected the necessity for more adaptive approaches to HIS and HIT evaluation.

Keywords. Health information system, health information technology, framework, model, evaluation

1. Introduction

Global digitization is advancing at a rapid pace and health information systems (HIS) and health information technologies (HIT) are gaining market prominence. The criticality of readily available, safe and usable technologies in healthcare is becoming increasingly vital. The diversity of the HIS and HIT available in the marketplace, is also reflected in the heterogeneity of evaluation models and frameworks currently used in health informatics. Standardization and cross-cultural instrument validation [2] in the creation of relevant models and frameworks is paramount. Such an approach would ensure that safe, reliable and efficient technological solutions are purchased and implemented appropriately in healthcare settings. The objective of this paper is two-fold: 1) to classify models and frameworks published between the years 2010-2020 according to their level of evaluative focus, 2) to identify the countries where these models and

[1] Corresponding author, Amanda L. Joseph, University of Victoria, British Columbia, Canada; E-mail: Amandalynnjoseph@uvic.ca.

frameworks have been employed for the purpose of evaluation of HIS and HIT, using the International Medical Informatics Association (IMIA) Represented Regions [3].

2. Methods

A scoping review following the Arksey and O'Malley methodology [1] was conducted in the EBSCOhost CINAHL®, Web of Science®, IEEE *Xplore*® and PubMed® databases. The keywords utilized were "evaluation AND (framework OR model OR theory)" AND "health information system." Prior to screening the articles, the researchers defined the terms model and framework. This was done to support consistent screening of the articles. Models identified key concepts and their relationships (in the context of systems) and were conceptualized as "the experiences, reflections and insights of scholars and practitioners" [4]. Frameworks were defined as "organizing structures that may be developed into models or theories over time" [4]. Following this, two researchers applied the inclusion and exclusion criteria (Table 1), screened the titles and abstracts of each article using Covidence®. A third researcher resolved the differences of opinions in article selection and then a full text review of all remaining articles was completed. Lastly, a market level analysis was conducted and the data extracted (Table 3) from the articles was categorized according to: country of use, IMIA represented region [3], level of evaluative focus (Table 2).

Table 1. Inclusion and exclusion criteria

Inclusion	Exclusion
English articles	Language other than English
Published between 2010-2020	Published outside of date parameters
Article abstract present	Article abstract absent
HIS or HIT as an intervention with an evaluative component that: a) Used a model or framework to evaluate a technology b) Developed a model or framework to evaluate a technology c) Tested a model or framework to evaluate a technology	Editorials and literature reviews that lacked an evaluative component that focused on: a) Clinical or organizational outcomes b) Patient risk factors and health conditions c) Surgery d) Medical devices e) Databases or data extracted from a database

Table 2. Definitions for micro, meso, macro and multi-levels of evaluative focus [5-8]

Level of evaluation	Micro	Meso	Macro	Multi
Definition	Evaluation of individual users interacting with technologies.	Users interacting as a team, group or organization using technologies within an organization.	Health system or inter-organizational level interactions using technologies.	Users interacting with and using technologies from micro, meso and macro perspectives.

3. Results

The initial search yielded 363 articles, 78 duplicates were removed, resulting in 285 articles screened for inclusion. From there 215 articles were excluded as they did not meet the inclusion criteria (Table 1), 70 articles were read in full for inclusion and 17

were excluded as they did not satisfy the inclusion criteria. The remaining 53 articles were then reassessed for inclusion and during the data extraction, phase 11 articles were excluded. These articles were excluded as they represented editorials or literature reviews that summarized the state of the literature but did not propose recommendations to address models or frameworks in HIS and HIT. The screening process of the scoping review was iterative and resulted in a final inclusion of 42 articles. As this was a review of existing publicly available literature, an ethics consult was not required. Some limitations of the study include that findings were guided by the search terms and were limited to articles in the English language only, therefore other relevant articles may have been omitted from the review based on this criteria.

Table 3. Categorical findings of market level analysis

Level	Classification	Country of Origin	IMIA Regions [3]
Micro-level	2 Frameworks [10,11] 0 Models	Indonesia [10], Netherlands [10]	Asia Pacific [9], European [10]
Meso-level	9 Frameworks [12-18,22,23] 4 Models [11,19-21]	France [11,15], Germany [12,17], Austria [12], Canada [13], Indonesia [14], Botswana [16], South Africa [18], Argentina [19], Australia [20], Finland [21], India [22], Cyprus [23]	European [11,12,15,17,21, 23], North America [13], Asia Pacific [14,20,22] African Region [16,18] Latin America and the Caribbean [19]
Macro-level	18 Frameworks [24-27,29-31,33-37,39-44] 4 Models [26,28,32,38]	Pakistan [24], Brazil [25], Ireland [26], Tanzania [27], United Kingdom [28,40], Sub-Saharan Africa [29], Somalia [30], Australia [31], India [31], Portugal [32], United States of America [33], France [33,45], Canada [33,44], Cyprus [34], Iran [35,36], Columbia [37], Libya [39], Sweden [40]	Asia Pacific [24], Latin America and the Caribbean [25,37], European [26,28,32-34,40], African Region [27, 29,30,39], Asia Pacific [31], North America [33,44], Middle East and North Africa [31,35,36]
Multi-level	6 Frameworks [45-50] 0 Models	France [45], Canada [50]	European [45], North America [50]

The results of the market level analysis (Table 3) indicated that: 50% of the studies applied a macro-level analysis, 31% utilized a meso-level perspective, 14% used a multi-level approach and 5% assessed HIS and HIT from a micro-level lens. The two articles [10,11] that contextualized HIS and HIT evaluation from a micro-level analysis came from the Asia Pacific and European IMIA represented regions [3], whereas meso and macro-level evaluations were dispersed across many diverse IMIA regions [3]. Of the multi-leveled articles, only two articles [45,50] utilized IMIA regions [3] in their analysis. The remaining four articles represented literature reviews that assessed HIS and HIT but did not specify a country or IMIA represented region [3].

With a collective total of 35 frameworks and eight models, the prominent theme and approach to evaluative design was the framework. However, as evidenced by O'Leary and colleagues [26], a holistic approach utilizing a model and a framework may be a more efficient and effective method. The findings revealed the need for comparability, when assessing HIS and HIT from various perspectives to ensure safety, usability and institutional applicability. However, this field is complex and evaluators must be cognizant of issues that exist in comparing differing nations and jurisdictions. Cultural,

social factors (e.g. language, time period, health literacy) and geographic customs [2] may alter the interpretation of the research questions and the overall success of the evaluation. Moreover, ignoring these diverse factors could impede the integrity and generalizability of frameworks or models, as each consideration may have direct influence on the outputs and approach to data collection.

4. Conclusion

In this scoping review, we have seen a range of models and frameworks that evaluated technology in healthcare settings. As evidenced by the heterogeneity of evaluative design and approaches currently used in health informatics, there is a need for more adaptive methods to HIS and HIT evaluation. To satisfy this critical gap in HIS and HIT assessment, future models and frameworks could be designed to incorporate patient, physician and caregiver journey mapping activities [51]. Additionally, a focus on clinical workflow, human factors (e.g. human information processing capabilities and limitations) and usability engineering could improve the safety and adoption of interactive clinical systems. An appropriate set of criteria (e.g. framework or model) could not only guide HIT implementations but could be leveraged to evaluate clinical systems at various stages of their system development life cycle (SDLC) [52]. As healthcare organizations are highly complex environments, integrating iterative usability testing into HIS and HIT assessment tools may be prudent. Furthermore, designing evaluative schemas from a socio-technical, cognitive [5] and organizational cultural approach may result in more effective HIS and HIT evaluation. Although, dynamic framework and model creation from a holistic and multifactorial lens could be challenging, such an approach may be the only feasible solution to adequately assess the fluid influx of technological innovation in healthcare. A dynamic evaluative approach could ensure that safe and usable technologies are procured and implemented into healthcare settings. Human centric, standardized and generalizable evaluative tools hold tremendous promise for improving healthcare service delivery, patient safety and the global health system.

References

[1] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol. 2005;8(1):19-32.
[2] Gjersing L, Caplehorn JRM, Clausen T. Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. BMC Med Res Methodol. 2010;10(1):13.
[3] IMIA Represented Regions [Internet]. IMIA. [cited 2021 Oct 1]. Available from: https://imia-medinfo.org/wp/regional-members/&gt.
[4] Borycki EM, Kushniruk AW, Bellwood P, Brender J. Technology-induced errors. The current use of frameworks and models from the biomedical and life sciences literatures. Methods Inf Med. 2012;51(2):95-103.
[5] Borycki EM, Kushniruk AW. Towards an integrative cognitive-socio-technical approach in health informatics: analyzing technology-induced error involving health information systems to improve patient safety. Open Med Inform J. 2010;4(1):181-7.
[6] Burns LR, Bradley EH, Weiner BJ. Shortell & Kaluzny's health care management: Organization, design, and behavior. Sixth ed. Clifton Park, NY: Delmar Cengage Learning; 2012.
[7] Carretero S, Stewart J, Centeno C. Information and communication technologies for informal carers and paid assistants: benefits from micro-, meso-, and macro-levels. Eur J Ageing. 2015;12(2):163-73.

[8] Macro-Level, Meso-level, and Micro-level Analysis- Course Hero [Internet]. Sociological Perspective. [cited 2021Aug28]. Available from: https://www.coursehero.com/sg/introduction-to-sociology/macro-level-meso-level-and-micro-level-analysis/.
[9] Erlirianto LM, Ali AHN, Herdiyanti A. The Implementation of the Human, Organization, and Technology–Fit (HOT–Fit) Framework to Evaluate the Electronic Medical Record (EMR) System in a Hospital. Procedia Comput Sci. 2015;72:580–7.
[10] Peute LWP, de Keizer NF, van der Zwan, E. P. A, Jaspers MWM. Reducing clinicians' cognitive workload by system redesign; a pre-post think aloud usability study. Stud Health Technol Inform. 2011;169:925-9.
[11] Hadji B, Martin G, Dupuis I, Campoy E, Degoulet P. 14 Years longitudinal evaluation of clinical information systems acceptance: The HEGP case. Int J Med Inf. 2016 Feb;86:20–9.
[12] Honekamp W, Ostermann H. Evaluation of a prototype health information system using the FITT framework. J Innov Health Inform. 2011 Mar 1;19(1):47.
[13] Heidebrecht CL, Kwong JC, Finkelstein M, Quan SD, Pereira JA, Quach S, Deeks SL. Electronic immunization data collection systems: application of an evaluation framework. BMC Med Inform Decis Mak. 2014;14(1):5.
[14] Faridah L, Rinawan FR, Fauziah N, Mayasari W, Dwiartama A, Watanabe K. Evaluation of Health Information System (HIS) in The Surveillance of Dengue in Indonesia: Lessons from Case in Bandung, West Java. Int J Environ Res Public Health. 2020 Mar 10;17(5):1795.
[15] Rejeb O, Pilet C, Hamana S, Xie X, Durand T, Aloui S, et al. Performance and cost evaluation of health information systems using micro-costing and discrete-event simulation. Health Care Manag Sci. 2018 Jun;21(2):204–23.
[16] Seitio-Kgokgwe O, Gauld RDC, Hill PC, Barnett P. Development of the National Health Information Systems in Botswana: Pitfalls, prospects and lessons. Online J Public Health Inform. 2015;7(2):e210.
[17] Sunyaev, A, Pflug, J. Research Toward the Practical Application of a Risk Evaluation Framework: Security Analysis of the Clinical Area within the German Electronic Health Information System (2011). Proceedings of International Bled eConference (Bled 2011), Bled, Slovenia, June 12-15, 2011, pp. 156-168.
[18] Hanmer LA, Isaacs S, Roode JD. Factors associated with health information system success: results of a survey of hospitals in South Africa. Stud Health Technol Inform. 2010;160(Pt 1):347-51.
[19] Borbolla D, Otero C, Lobach DF, Kawamoto K, Gomez Saldaño AM, Staccia G, Lopez G, Figar S, Luna D, Bernaldo de Quiros FG. Implementation of a clinical decision support system using a service model: results of a feasibility study. Stud Health Technol Inform. 2010;160(Pt 2):816-20.
[20] Yu P. A multi-method approach to evaluate health information systems. Stud Health Technol Inform. 2010;160(Pt 2):1231-5.
[21] Vahteristo A, Jylhä V. Effects of User Participation in the Development of Health Information Systems on Their Evaluation Within Occupational Health Services. Stud Health Technol Inform. 2020 Nov 23;275:207-211.
[22] Gudi N, Godinho MA, Saha D, Narasimhan P. Sociotechnical evaluation of a health management information system in South India. BMJ Innov. 2020 Oct;6(4):164–9.
[23] Stylianides A, Mantas J, Pouloukas S, Roupa Z, Yamasaki E. Evaluation of the Integrated Health Information System (IHIS) in Public Hospitals in Cyprus Utilizing the DIPSA Framework. Acta Inform Medica. 2019;27(4):240.
[24] Nawaz R, Khan SA, Khattak T, Nasir F, Abbas K. Interventional Study for Improving Health Information System in Khyber Pakhtunkhwa, Pakistan. Curēus. 2020;12(11):e11785.
[25] Mussi CC, do Valle Pereira CD, de Oliveira Lacerda RT, dos Santos EM. Pre-Implementation evaluation of a nationwide information system for university hospitals: lessons learned from a study in Brazil. Behav Inf Technol. 2018 Mar 4;37(3):217–31.
[26] O'leary P, Carroll N, Clarke P, Richardson I. Untangling the Complexity of Connected Health Evaluations. IEEE; 2015.
[27] Kiwanuka A, Kimaro H, Senyoni W, Thobias J. A framework for the acceptance process of District Health Information System for vertical health programmes. IIMC International Information Management Corporation Ltd; 2015.
[28] Cresswell KM, Worth A, Sheikh A. Actor-Network Theory and its role in understanding the implementation of information technology developments in healthcare. BMC Med Inform Decis Mak. 2010;10(1):67.
[29] Herman H, Grobbelaar SS, Pistorius C. The design and development of technology platforms in a developing country healthcare context from an ecosystem perspective. BMC Med Inform Decis Mak. 2020;20(1):55.
[30] Askar A, Ardakani M, Majdzade R. Bridging gaps in health information systems: a case study from Somaliland, Somalia. East Mediterr Health J. 2017;23(11):764-73.

[31] Tiwari VK, Kumar K, T.P. SR, Kulkarni PD. Standards, Frameworks and Practices in Health Management Information and Evaluation Systems (HMIES) in Australia and India: Lessons for Future Transition in India? J Health Manag. 2016 Mar;18(1):70–83.
[32] Grenha Teixeira J, Pinho NFd, Patrício L. Bringing service design to the development of health information systems: The case of the Portuguese national electronic health record. Int J Med Inform. 2019;132:103942.
[33] Kushniruk A, Beuscart-Zéphir M, Grzes A, Borycki E, Watbled L, Kannry J. Increasing the Safety of Healthcare Information Systems through Improved Procurement: Toward a Framework for Selection of Safe Healthcare Systems. Healthc Q. 2010;13(SP2):53-8.
[34] Stylianides A, Mantas J, Roupa Z, Yamasaki EN. Evaluation of an Integrated Health Information System (HIS) in a Public Hospital in Cyprus: A Pilot Study. Stud Health Technol Inform. 2017;238:44-7.
[35] Farzandipour M, Meidani Z, Nabovati E, Sadeqi Jabali M, Dehghan Banadaki R. Technical requirements framework of hospital information systems: design and evaluation. BMC Med Inform Decis Mak. 2020;20(1):61.
[36] Abdi Z, Majdzadeh R, Ahmadnezhad E. Developing a framework for the monitoring and evaluation of the Health Transformation Plan in the Islamic Republic of Iran: lessons learned. East Mediterr Health J. 2019;25(6):394-405.
[37] López DM, Blobel B, Gonzalez C. Quality evaluation of health information system's architectures developed using the HIS-DF methodology. Stud Health Technol Inform. 2010;156:21-31.
[38] Logan J. Electronic health information system implementation models - a review. Stud Health Technol Inform. 2012;178:117-23.
[39] El Oakley RM, Ghrew MH, Aboutwerat AA, Alageli NA, Neami KA, Kerwat RM, Elfituri AA, Ziglam HM, Saifenasser AM, Bahron AM, Aburawi EH, Sagar SA, Tajoury AE, Benamer HT; National Health Systems Conference. Consultation on the Libyan health systems: towards patient-centred services. Libyan J Med. 2013;8(1):20233-9.
[40] Prodinger B, Taylor P. Improving quality of care through patient-reported outcome measures (PROMs): expert interviews using the NHS PROMs Programme and the Swedish quality registers for knee and hip arthroplasty as examples. BMC Health Serv Res. 2018;18(1):87.
[41] Reynolds HW, Sutherland EG. A systematic approach to the planning, implementation, monitoring, and evaluation of integrated health services. BMC Health Serv Res. 2013;13(1):168.
[42] Gajanayake R, Sahama T, Lane B. The role of human factors when evaluating information accountability for eHealth systems. Stud Health Technol Inform. 2013;194:97-102.
[43] Kushniruk A, Borycki E, Kuo M, Kuwata S. Integrating technology-centric and user-centric system testing methods: ensuring healthcare system usability and safety. Stud Health Technol Inform. 2010;157:181-6.
[44] Bassi J, Lau F, Hagens S, Leaver C, Price M. Knowledge translation in eHealth: building a virtual community. Stud Health Technol Inform. 2013;183:257-62.
[45] Augusto V, Rejeb O, Xie X, Aloui S, Perrier L, Biron P, Durand, T. Performance evaluation of Health Information Systems using ARIS modeling and discrete-event simulation. IEEE; 2015.
[46] Delnord M, Tille F, Abboud LA, Ivankovic D, van Oyen H. How can we monitor the impact of national health information systems? Results from a scoping review. Eur J Public Health. 2020;30(4):648-59.
[47] Yusof MM. A Socio-Technical and Lean Approach Towards a Framework for Health Information Systems-Induced Error. Stud Health Technol Inform. 2019;257:508-512.
[48] van Mens HJT, Duijm RD, Nienhuis R, de Keizer NF, Cornet R. Determinants and outcomes of patient access to medical records: Systematic review of systematic reviews. Int J Med Inform. 2019;129:226-33.
[49] van Mens, Hugo J T, Duijm RD, Nienhuis R, de Keizer NF, Cornet R. Towards an Adoption Framework for Patient Access to Electronic Health Records: Systematic Literature Mapping Study. JMIR Med Inform. 2020;8(3):e15150.
[50] Lau F, Price M, Keshavjee K. From Benefits Evaluation to Clinical Adoption: Making Sense of Health Information System Success in Canada. Healthc Q.2011;14(1):39-45.
[51] Joseph AL, Kushniruk AW, Borycki EM. Patient journey mapping: Current practices, challenges and future opportunities in healthcare. Knowledge management & e-learning. 2020;12(4):387-404.
[52] Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Comput Biol Med. 2002;32(3):141-9.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210915

CARE REGIO – Digital Transformation and Technology in Nursing Care

Dominik FUCHS[a,1], Ann-Kathrin WAIBEL[b], Maximilian KARTHAN[b], Nadine SEIFERT[b], Michael ÖRTL[b], Sabrina BETHGE[c], Gina FISCHER[b], Luca WASCHBÜSCH[b], Steffen NETZBAND[d], Frank KRAMER[d], Andreas MAHLER[c], Alexandra TEYNOR[e], Dominik MERLI[e], Elisabeth Veronica MESS[e], Jennifer ÖTTINGER[b], Stefanie SCHMID[a], Amelie SCHEU[b], Walter SWOBODA[b] and Petra FRIEDRICH[a]

[a] *University of Applied Sciences Kempten,*
[b] *University of Applied Sciences Neu-Ulm,*
[c] *Augsburg University Hospital,*
[d] *University of Applied Sciences Augsburg,*
[e] *University of Augsburg*

Abstract. Digital technologies have the potential to improve the quality of nursing care. CARE REGIO is a Bavarian joint research project for digital transformation and technology in nursing care. The project goals are supporting the nursing staff, saving time, improving the quality of care as well as increasing the quality of life and safety of those in need of care. In Phase 1 of the project, literature and stakeholder analyses, and qualitative surveys were carried out. Subsequently, central fields of action were defined for Phase 2 of the project. CARE REGIO can make a significant contribution to evaluating existing digital solutions, developing new solutions, and accelerating their implementation into practice.

Keywords. Digitization, nursing care, care transition report, assistive systems, care data lake, nursing wiki

1. Introduction

Digital technologies have the potential to improve the quality of nursing care and make work processes more efficient. In its "State of the World's Nursing Report - 2020" [1], the World Health Organization (WHO) urges that the opportunities offered by digital health technology should be leveraged. The widespread use of technologies in nursing care still faces major challenges, such as financial hurdles, the lack of technically oriented training for caregivers, ensuring data privacy and data security, and the fear of losing one's job [2,3]. Currently, the use of isolated technical solutions prevails in German-speaking countries, which results in considerable additional administrative work for nursing staff [4]. When dealing with patient data, it is of the utmost importance to handle the collected patient-relevant data in a manner that complies with data protection regulations [5]. The CARE REGIO project aims to promote the digital

[1] Corresponding Author, Dominik Fuchs, University of Applied Sciences Kempten, Bahnhofstraße 6 87435 Kempten, Germany; E-mail:dominik.fuchs@hs-kempten.de.

transformation of care by implementing and evaluating broad and collaborative efforts. The project network consists of the three Bavarian Universities of Applied Sciences of Augsburg (UAA), Kempten (UAK) and Neu-Ulm (UAN), as well as the University of Augsburg (UA) and the Augsburg University Hospital (AUH). The project is funded by the Bavarian State Ministry of Health and Care (StMGP) and will take place in two phases: The goal of the one-year Phase 1, which was completed in September 2020, was to identify suitable fields of action through literature and market analysis, as well as through the project's own surveys, interviews, process and stakeholder analysis. The steps and results that led to selected central fields of action are presented below.

2. Methods

2.1. Qualitative Expert Interviews

Qualitative surveys were conducted at AUH and three cooperating nursing facilities. The focus was on the transfer of care data, the technical equipment for data transfer, and the communication servers used. Furthermore, the experts were asked about the education and training behavior of nurses. Specifically, if self-directed (online) learning could achieve better learning results than existing offline approaches. Corresponding guidelines for the interviews were created. In addition, a market analysis was conducted for digital learning offerings in the nursing context. The (technically focused) survey in the AUH focused on a query of the status quo concerning the nursing documentation systems and communication servers used. Basic requirements, wishes and concerns regarding partial digital automation of the care transition report (CTR) were also inquired. The survey of the cooperating care facilities focused on currently used data transfer systems, the respective service providers/vendors, currently used data formats, and existing interoperability standards. Likewise, future planned new acquisitions and changes were inquired.

Another independent series of interviews was conducted with two experts, which accompanied participants in a test series in which ten outpatient seniors tested a device-supported leg training for fall prevention over two weeks. For the survey, an interview guideline was developed in which facilitating factors and obstacles regarding the feasibility (implementation) and acceptance of leg training and an app-based assessment of fall risk were asked [6]. The expert interviews were evaluated using a qualitative content analysis based on Mayring [7].

2.2. Analysis of the Patient Transition Process

In a process analysis, the patient transfer process at the AUH was analyzed. The internal clinical processes and the processes in connection with the external facilities were examined. The goal was to determine the current status of the processes and to derive potentials for process improvement. Therefore, clinic staff and staff from external facilities were interviewed. The interviews were conducted and recorded using an interview guide. The results were then evaluated using qualitative methodology according to Mayring [7], and a process model was created from the results using Business Process Management Notation (BPMN 2.0). In a stakeholder analysis, the stakeholders significantly involved in the transition of the CTR were considered to record or define initial requirements and the resulting benefits.

3. Results

3.1. Qualitative Expert Interviews

The survey on the transfer of care data in three nursing facilities and the AUH determined that the transfer of nursing-relevant patient data from the analogous transfer forms to the electronic documentation is associated with a high time expenditure. As a guideline, "over 30 minutes" was mentioned. All of the facilities surveyed in the technically focused interviews used different systems for internal data transfer. External digital communication of nursing data did not take place as there is no data protection-compliant connection. According to the current status, data is transferred in advance by telephone and then only in printed form or as a fax. According to the respondents, a completely digital solution would mean significant added value for nursing care in the medium to long term, including reduced administrative effort and better preparation. Therefore, all facilities were in favor of an automated, electronic exchange of care data integrated into the respective documentation systems.

In the interviews about digital learning offerings in the nursing context, the potential success was deemed considerable and potentially larger with younger than older employees. Market analysis revealed that there is currently no project with a comprehensive digital compendium, a so-called ‘nursing wiki’. Corresponding German-language sources are currently hardly available, outdated, or no longer accessible.

In the qualitative analysis of the expert interviews regarding assistive systems for fall prevention, the following results could be identified: The technology-assisted movement training was well received. With regard to usability, the ease of use of the technical movement trainer was cited as a beneficial factor. Another advantage from the point of view of an expert was that the test persons could exercise in a seated position, so that the risk of injury can be minimized even without supervision. The adjustability of the pedaling resistance and the choice of speed allowed subjects to adapt the workout to their individual performance level and increase it gradually, which can also be considered a facilitating factor. Limited space in the participants' own homes was cited as a potential barrier. In addition, transporting the device was considered too difficult for older people without assistance due to its weight.

3.2. Analysis of the Patient Transition Process

One significant result of the process analysis was that there are many media discontinuities in the transition process, which means that work has to be repeated. Ambiguities regarding the responsibility of certain tasks can be more clearly regulated in a digitally recorded process. A process model was created with BPMN 2.0. The process modeling showed that there would be valuable opportunities for process optimization with regard to the way data and information are exchanged. In a subsequent analysis of the 'high-level activities', which focused on the transition of the CTR, it was possible to define possible requirements. These included automatic or manually confirmed data import and the ability to structure, filter, and search the data. In a stakeholder analysis, it could be shown that each stakeholder group could benefit from the digitization of the CTR.

4. Discussion and Future Directions

The surveys on the topic of patient discharge and transition management revealed a currently high time expenditure in the preparation of CTR and subsequent transition of patients. Furthermore, a lack of suitable software and interfaces for cross-sector digital communication was identified. The need to comply with data protection guidelines was also recognized. Accordingly, time savings are also a key funding factor following successful conversion to digital processes. In assistive systems and care transition, the high financial, time and personnel costs involved in implementation were mentioned as existing obstacles. The expert interviews and process analysis revealed issues, both for facilitating processes and the existing hurdles and obstacles. This applies in particular to the opportunities and hurdles identified in the area of digitization of discharge and transfer management and assistive technologies for fall prevention [3,8].

Following our analysis and surveys, the central fields of action for further work in Phase 2 were identified and narrowed down based on the results. In addition to the cross-cutting issue of scientific support (including data protection and ethics), the project has decided on the following key fields of action for Phase 2 of the project: Digitization of the CTR, assistive systems (in fall prevention), an online training and continuing education offering ('nursing wiki'), and a digital care research database (a nursing data lake) to increase the synchronization of data collection and transfer (interoperability and standardization) between care facilities. An important aspect that has hampered the sustainable use of similar projects is the networking of nursing processes at the data level. CARE REGIO, therefore, intends to make use of the German Telematics Infrastructure [9] and thus enable a secure and legally compliant form of documentation and forwarding of patient-relevant data in the field of nursing. Project developments will be made accessible and will be distributed through Open Access (e.g., 'nursing wiki') and/or Open Source in terms of software-related advancements.

Acknowledgment

This project is supported by funds from the Bavarian State Ministry of Health and Care (Bayerisches Staatsministerium für Gesundheit und Pflege – StMGP).

References

[1] Organization WH. State of the world's nursing 2020: investing in education, jobs and leadership. 2020.
[2] Kuhlmey A, Blüher S, Nordheim J, Zöllick J. Technik in der Pflege–Einstellungen von professionell Pflegenden zu Chancen und Risiken neuer Technologien und technischer Assistenzsysteme. Zentrum für Qualität in der Pflege, Berlin. 2019.
[3] Merda M, Schmidt K, Kähler B. Pflege 4.0–Einsatz moderner Technologien aus der Sicht professionell Pflegender. Forschungsbericht für die BGW, Hamburg. 2017.
[4] Rösler U, Schmidt K, Merda M, Melzer M. Digitalisierung in der Pflege. Wie intelligente Technologien die Arbeit professionell Pflegender verändern. Stand: März 2018, 1. Aufl. Berlin2018. 72 S. p.
[5] Ethikrat D. Big Data und gesundheit–Datensouveränität als informationelle Freiheitsgestaltung. 2017.
[6] Fuchs D, Tiebel J, Friedrich P. Device-supported training and assessment for fall prevention of community-dwelling elderly: a pre-post mixed methods study. Procedia Computer Science. 2020;176:2322-31.
[7] Mayring P. Qualitative Inhaltsanalyse. Grundlagen und Techniken, 12. Aufl. Beltz. Weinheim; 2015.
[8] Gesundheit Bf. Informations- und Kommunikationstechnologie für die Pflege. ePflege. Berlin2017. gematik. Telematikinfrastruktur - das digitale Gesundheitsnetz für Deutschland: gematik; 2021 [cited 2021 29.09.2021]. Available from: https://www.gematik.de/telematikinfrastruktur.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210916

Description of a Digital Health Platform for Emotional and Self-Management Support of Caregivers of Children Receiving Growth Hormone Treatment

Francisco José NÚÑEZ-BENJUMEA[a], Antonio De Arriba MUÑOZ[b], Marta Vara CALLAU[b] and Luis FERNANDEZ-LUQUE[a,1]
[a] *Adhera Health, Inc, Palo Alto, CA, USA*
[b] *Pediatric Endocrinology unit, Miguel Servet Children's University Hospital, Zaragoza, Spain*

Abstract. In this demo, we provide an overview of the digital platform ADHERA CARING which has been used for an intervention designed for emotional and self-management support of caregivers of children receiving growth hormone treatment (GHt). ADHERA CARING provides tailored emotional and self-management support to caregivers of children undergoing GHt to improve adherence to treatment through positive education, personalized motivational messages, and emotional support. This digital intervention has already been piloted in a clinical setting as part of an ongoing feasibility clinical study (NCT04812665).

Keywords. Digital Health, Growth Hormone, Behavioral Intervention, Caregiver

1. Introduction

Adherence to growth hormone treatment (GHt) among children is variable and remains a problem [1]. Prior research has been looking into digital health tools to support GHt treatment [2], that includes mobile applications [3], and more advanced digital health ecosystems that integrate mobile solutions with connected injection devices [4]. One of the key stakeholders to be supported are the caregivers, normally parents, of children that require long term pharmacological treatment since research shows the negative impact of stress on adherence and overall quality of life [5]. In this paper, we present a study exploring the feasibility of a mobile digital intervention for emotional and self-management support of caregivers of children receiving GHt. The intervention relies on a digital health platform powered by a Health Recommender System [6].

[1] Corresponding Author, Luis Fernandez-Luque Adhera Health, Inc, Palo Alto, CA, USA; E-mail: luis.tromso@gmail.com.

2. mHealth-based digital intervention

ADHERA CARING provides tailored emotional and self-management support to caregivers of children undergoing GHt to improve adherence to treatment through positive education, personalized motivational messages, and emotional support. Users follow an empowerment program with easy-to-digest educational, personalized, and actionable content included for them, and also perform mental well-being activities. Whenever a unit is tagged as read, its icon turns green to reflect the progress (Figure 1-A). Within each category, the last educational unit is a quiz (Figure 1-B).

Motivational messages are delivered into an inbox in the app (Figure 1-C). Every time the app receives a new message, a push notification is triggered so that the user can access the message directly. Messages can be rated by users with a score from 1 to 5 stars. Messages are delivered by a health recommender system [6]. This algorithm works in two steps. The first step is a knowledge-based algorithm to filter incompatible messages with the user profile. The remaining messages are then passed to a collaborative filtering algorithm, that uses the demographic profile and the user ratings to compute a similarity measurement between the user profile and each message.

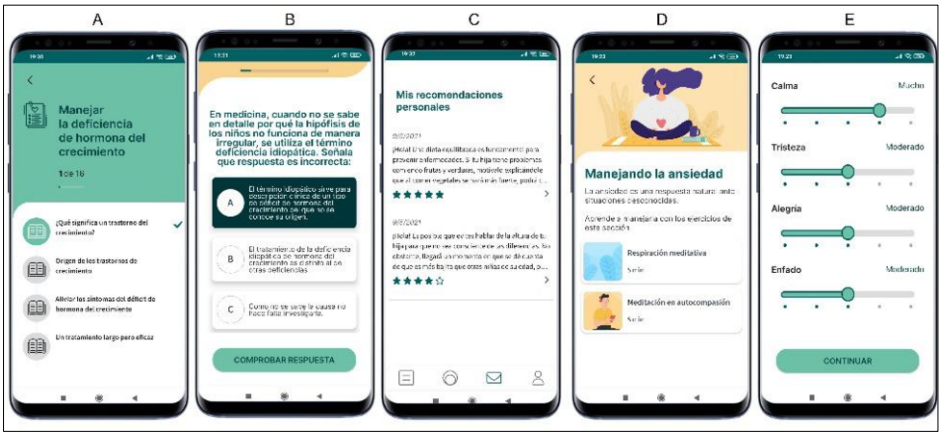


Figure 1. Screenshots of ADHERA CARING Platform

The mental well-being module (Figure 1-D) includes two exercises: 'Mindful breathing' and 'Self-compassion meditation'. These are audio-guided sessions that users can complete at their convenience. Before and after each session, the user is asked to complete a well-being assessment (Figure 1-E).

Acknowledgements

We would like to thank the rest of the team who contributed to this project and the caregivers for their participation.

Disclaimer: This investigator initiated study was financially supported by Merck Healthcare KGaA, Darmstadt Germany (CrossRef Funder ID: 10.13039/100009945). Merck reviewed the manuscript for medical accuracy only before submission. The authors are fully responsible for the content of this manuscript, and the views and opinions described in the publication reflect solely those of the authors.

References

[1] Haverkamp F, Gasteyger C. A review of biopsychosocial strategies to prevent and overcome early-recognized poor adherence in growth hormone therapy of children. J Med Econ. 2011;14(4):448–57.
[2] Dimitri P, Fernandez-Luque L, Banerjee I, Bergadá I, Calliari LE, Dahlgren J, de Arriba A, Lapatto R, Reinehr T, Senniappan S, Thomas-Teinturier C. An eHealth Framework for Managing Pediatric Growth Disorders and Growth Hormone Therapy. Journal of Medical Internet Research. 2021 May 20;23(5):e27446.
[3] Fernandez-Luque L, Labarta JI, Palmer E, Koledova E. Content Analysis of Apps for Growth Monitoring and Growth Hormone Treatment: Systematic Search in the Android App Store. JMIR Mhealth Uhealth. 2020 Feb 18;8(2):e16208.
[4] Koledova E, Le Masne Q, Spataru A, Bagha M, Dixon D. Digital Health in the Management of Pediatric Growth Hormone Therapy - 10 Years of Developments. Stud Health Technol Inform. 2021 May 27;281:926–30.
[5] Armstrong ML, Duncan CL, Stokes JO, Pereira D. Association of caregiver health beliefs and parenting stress with medication adherence in preschoolers with asthma. J Asthma. 2014 May;51(4):366–72.
[6] Hors-Fraile S, Malwade S, Luna-Perejon F, Amaya C, Civit A, Schneider F, Bamidis P, Syed-Abdul S, Li YC, De Vries H. Opening the black box: Explaining the process of basing a health recommender system on the i-change behavioral change model. IEEE Access. 2019 Dec 4;7:176525-40.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210917

Clinical Notes De-Identification: Scoping Recent Benchmarks for n2c2 Datasets

Taridzo CHOMUTARE[a,1]
[a]*Norwegian Centre for E-health Research, Tromsø, Norway*

Abstract. Publicly shared repositories play an important role in advancing performance benchmarks for some of the most important tasks in natural language processing (NLP) and healthcare in general. This study reviews most recent benchmarks based on the 2014 n2c2 de-identification dataset. Pre-processing challenges were uncovered, and attention brought to the discrepancies in reported number of Protected Health Information (PHI) entities among the studies. Improved reporting is required for greater transparency and reproducibility.

Keywords. Natural language processing, NLP, i2b2, de-identification

1. Introduction

Removing identifying information from data sources is defined as de-identification, where the goal is to make re-identification of individuals impossible. For healthcare, data privacy and security are the primary concerns. In this regard, a number of legislative guidelines exist in many territories around the world. For instance in USA, where Health Insurance Portability and Accountability Act (HIPAA) mandates that certain direct identifiers and quasi-identifiers, aptly named Protected Health Information (PHI), be removed from any health data before such data is shared. In Europe, the General Data Protection Regulation (GDPR) provides similar guidelines to protect the data of citizens.

For most datasets, a typical de-identification pipeline includes removing direct identifiers such as IDs or email. Even after these are removed, the data may still contain quasi-identifiers such as the date of birth. Therefore, an anonymization process is required to transform the data in order to reduce the risk of disclosure. For structured data, this process can be straightforward, for example by using statistical methods to reduce the risk of re-identification, e.g, by generalization or suppression.

In contrast, unstructured data, such as clinical text, requires more complex methods to reduce disclosure risk. There is an abundance of unstructured clinical text that could help shed light on some of the most key healthcare challenges today. In view of this important problem, a challenge to promote and disseminate natural language processing (NLP) methods for de-identifying clinical notes was launched in 2006, by Informatics for Integrating Biology and the Bedside (i2b2) [10], now called n2c2. Since then, the task has generated wide interest, and different research groups continue to work on improving performance. Even though much has been published on this topic, we still

[1] Corresponding Author, Taridzo Chomutare, Norwegian Centre for E-Health Research, 9019 Tromsø, Norway; E-mail: Taridzo.chomutare@ehealthresearch.no.

lack an overview of the progress. This study reviews some of the recent work on the topic.

2. Methods

A search was conducted for studies that had used the 2014 i2b2 de-identification datasets for benchmarking their new algorithms. The search string used variations of the core string "i2b2 AND de-identification". This excludes studies that participated in the challenge itself, and only includes studies conducted in the past five years (2017 - 2021). Search on Google Scholar, PubMed, IEEE and ACM were conducted. Studies that did not fully describe the properties of the data were excluded, for instance, if they did not provide the PHI entity numbers after pre-processing.

The following data items were collected: (i) methods used, (ii) data properties (number of PHI entities), and (iii) performance data. In addition to this data, it is also noted if each respective study publishes its code in a public repository.

3. Results

Based on the search hits, duplicates were screened out and 9 studies that met the selection criteria were included in this study. Results of the data properties are shown in Table 1, where we can observe only two studies that agree on the number of PHI entities for some of the HIPAA categories (bold print). The rest of the studies ended up with varying number of entities based on the same i2b2 dataset.

Table 1. Counts of entities as a total (*or just the test set) datasets as reported in the studies.

PHI	[1]	[2]*	[3]*	[4]	[5]	[6]	[7]	[8]	[9]
DATE	12482	4951	4980	-	-	12468	12381	12473	12532
NAME	7348	4131	2883	-	-	-	7258	7361	4839
AGE	**1997**	-	764	-	-	1991	2028	**1997**	790
CONTACT	**541**	171	218	-	-	-	610	**541**	419
ID	**1506**	576	625	-	-	1039	1549	**1506**	1126
LOCATION	4580	1177	1813	-	-	-	3986	4578	3001
PROFESSION	**413**	-	179	-	-	-	420	**413**	340
All entities	28867	10861	11462	28872	26787	28862	28205	28869	23047

Disregarding these discrepancies in number of entities, the reported performances are shown in Table 2. From the table, performance measured by F1, improved from a 2017 high of 0.983 [1], to a high of 0.985 [2] in 2021. However, because of the different evaluation methods, this improvement cannot be taken at face value. Studies used different evaluation methods, from token-based binary [1,6,7,8] classification to entity classification based on HIPAA PHI [5,9].

In terms of methods, all the studies were deep learning-based, and invariably used bidirectional long-short term memory (Bi-LSTM), Conditional Random Fields (CRF) and Gated recurrent units (GRU). With these base methods, studies developed multiple innovative ensemble methods through voting mechanisms [3], stacking [3,7] and novel attention mechanisms based on transformer models [7, 8], and use of rule-based methods and dictionaries [1, 2, 9].

Table 2. Performance as reported in the studies, and the respective methods used.

Study	Precision	Recall	F1	Evaluation	Code	Methods used
[1]2017	0.993	0.973	0.983	Binary token	no	Bi-LSTM, CRF, rule-based
[9] 2017	0.983	0.973	0.979	HIPAA PHI	Yes	Bi-LSTM, CRF, dictionaries
[6] 2018	0.989	0.972	0.981	Binary token	No	Bi-GRU
[8] 2019	0.990	0.983	0.987	Binary token	No	Bi-LSTM, CRF, transformers
[3] 2020	-	-	0.959	Strict entity	No	Bi-LSTM, CRF, voting, stacking
[7] 2020	0.980	0.984	0.982	Binary token	Yes	Bi-GRU, GRU-LSTM, stacking, self-attention
[2] 2021	0.979	0.992	0.985	-	No	DL, iterative fine-tuning, dictionaries
[4] 2021	0.947	0.918	0.933	Strict entity	No	Bi-LSTM, CRF, n-gram moving window
[5] 2021	0.839	0.818	0.828	HIPAA PHI	No	Bi-LSTM, CRF

4. Discussion

Perhaps the most unexpected finding was the large discrepancies in the reported number of PHI entities. It appears the problem stems from the need to re-format the original datasets, to satisfy the input format requirements for specific algorithms. This re-formatting or pre-processing during a typical de-identification pipeline, appear to yield significantly different results for each study. Therefore, it is difficult to measure the overall performance improvements if the data are not consistent. This is a key point since most reported improvements will only be small fractions of a percent. Even small discrepancies in test datasets will skew results.

This problem stems from the very nature of the dataset, which is generally considered sensitive and requires individuals to sign non-disclosure agreements. Therefore, the data or respective transformations into new data formats cannot be uploaded to the Internet. One solution could be to ask the data proprietor to update the repository with new updates on data formats. This could be useful since multiple studies reported minor errors in some annotations.

In terms of performance, different studies use different evaluation methods, and this has a large effect on the overall results, and makes it difficult to compare results across studies. While there is debate regarding the best evaluation method, it could be beneficial if multiple evaluation methods were used and reported. So far, however, the general reporting appears insufficient, since some studies are not specific about the evaluation methods used. Further, most of the studies provided neither the algorithm code nor the evaluation script. It is therefore nearly impossible to reproduce their work. This is especially important because there are many implementations of an algorithm, and small variations of an algorithm can have a significant impact. In addition, there are multiple combinations of hyper-parameters, and optimization processes were never reported. Advancement of scientific knowledge depends on full disclosure of such information, but only two studies provided a code repository [7, 9] of their work.

Turning to the methods used, a possible explanation for the common use of deep-learning is the scientific progress in the field, especially with the development of contextual embeddings and large language models like BERT [12]. These developments have changed the game for NLP, and much of emerging new innovation centers around their use. However, it is interesting to note the use of dictionaries, where some studies

use rule-based systems combined with deep learning algorithms as part of a whole system [1], or as part of a post-processing step.

5. Conclusions

While interest in this de-identification task appear to continue to increase, there are still challenges that distract the scientific community from fully realizing the ideals of shared datasets. Perhaps prioritizing better reporting and full code-sharing could be a starting point. This is an important step for reproduction of work and to make further scientific progress by building on current knowledge.

Acknowledgments

This work was partially supported by the Northern Norway Regional Health Authority, (Helse Nord); research grant HNF1395-18, NorKlinTekst project. The funding body does not have any role in the study.

References

[1] Liu Z, Tang B, Wang X, Chen Q. De-identification of clinical notes via recurrent neural network and conditional random field. Journal of biomedical informatics. 2017 Nov 1;75:S34-42.
[2] Murugadoss K, Rajasekharan A, Malin B, Agarwal V, Bade S, Anderson JR, Ross JL, Faubion Jr WA, Halamka JD, Soundararajan V, Ardhanari S. Building a best-in-class automated de-identification tool for electronic health records through ensemble learning. Patterns. 2021 Jun 11;2(6):100255.
[3] Kim Y, Heider PM, Meystre SM. Comparative Study of Various Approaches for Ensemble-based De-identification of Electronic Health Record Narratives. InAMIA Annual Symposium Proceedings 2020;2020:648.
[4] Lee K, Kayaalp M, Henry S, Uzuner Ö. A Context-Enhanced De-identification System. arXiv preprint arXiv:2102.08513. 2021 Feb 17.
[5] Ahmed A, Abbasi A, Eickhoff C. Benchmarking Modern Named Entity Recognition Techniques for Free-text Health Record De-identification. arXiv preprint arXiv:2103.13546. 2021 Mar 25.
[6] Zhao YS, Zhang KL, Ma HC. et al. Leveraging text skeleton for de-identification of electronic medical records. BMC Med Inform Decis Mak. 2018;18.
[7] Ahmed T, Al Aziz MM, Mohammed N. De-identification of electronic health record using neural network. Scientific reports. 2020 Oct 29;10(1):1-1.
[8] Tang B, Jiang D, Chen Q, Wang X, Yan J, Shen Y. De-identification of Clinical Text via Bi-LSTM-CRF with Neural Language Models. AMIA Annu Symp Proc. 2019.
[9] Dernoncourt F, Lee JY, Uzuner O, Szolovits P. De-identification of patient notes with recurrent neural networks. J Am Med Inform Assoc. 2017;24(3).
[10] Stubbs A, Uzuner Ö. Annotating longitudinal clinical narratives for de-identification: The 2014 i2b2/UTHealth corpus. J Biomed Inform. 2015;58.
[11] Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018 Oct 11.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210918

Predicting Hospital Admission for Emergency Department Patients: A Machine Learning Approach

Georgios FERETZAKIS[a,b,c,1], Aikaterini SAKAGIANNI[d], Evangelos LOUPELIS[c]
Georgios KARLIS[e], Dimitris KALLES[a], Lazaros TZELVES[f],
Rea CHATZIKYRIAKOU[g], Nikolaos TRAKAS[h], Stavroula PETROPOULOU[c],
Aikaterini TIKA[i], Zoi RAKOPOULOU[i], Ilias DALAINAS[i] and Vasileios KALDIS[j]
[a] *School of Science and Technology, Hellenic Open University, Patras, Greece*
[b] *Department of Quality Control, Research and Continuing Education, Sismanogleio General Hospital, Marousi, Greece*
[c] *IT department,Sismanogleio General Hospital, Marousi, Greece*
[d] *Intensive Care Unit, Sismanogleio General Hospital, Marousi, Greece*
[e] *Thoracic Diseases General Hospital Sotiria, Athens, Greece*
[f] *Second Department of Urology, National and Kapodistrian University of Athens, Sismanogleio General Hospital, Marousi, Greece*
[g] *Hematology Laboratory, Sismanogleio General Hospital, Marousi, Greece*
[h] *Biochemistry Department, Sismanogleio General Hospital, Marousi, Greece*
[i] *Administration, Sismanogleio General Hospital, Marousi, Greece*
[j] *Emergency Department, Sismanogleio General Hospital, Marousi, Greece*

Abstract. The objective of this study was to establish a machine learning model and to evaluate its predictive capability of admission to the hospital. This observational retrospective study included 3204 emergency department visits to a public tertiary care hospital in Greece from 14 March to 4 May 2019. We investigated biochemical markers and coagulation tests that are routinely checked in patients visiting the Emergency Department (ED) in relation to the ED outcome (admission or discharge). Among the most popular classification techniques of the scikit-learn library through a 10-fold cross-validation approach, a GaussianNB model outperformed other models with respect to the area under the receiver operating characteristic curve.

Keywords. emergency department, machine learning, artificial intelligence, critical care, patient admission, decision support, scikit-learn

1. Introduction and Background

A frequently encountered issue in any Emergency Department (ED) is critical overcrowding that leads to the exhaustion of emergency medical resources. This challenge, still present and exacerbated during the COVID-19 pandemic [1], stems from complex causes. Overcrowding in ED is associated with increased medical errors,

[1] Corresponding Author, Georgios FERETZAKIS, PhD, E-mail address: georgios.feretzakis@ac.eap.gr

increased waiting time, less favorable outcomes, and decreased patient satisfaction [2]. The number of ED visits in the USA increased from 128.97 million to 144.82 million between 2010 and 2016. Expressed as a population rate, ED visits per 1000 persons rose from 416.92 in 2010 to 448.19 in 2016 [3] and continues to increase faster than the population growth [3]. With the introduction of the electronic medical record (EMR) system, ED staff is confronted with a continuously increasing volume of data and information for each patient, including past ED visits and medical notes, previous laboratory results, and diagnostic imaging reports, and this results in overwhelming the capacity to manage and/or make use of it [4].

In the ED context, the medical personnel have to decide by proceeding rapidly in interpreting clinical data to classify patients and predict outcomes. This is fundamental to the emergency department (ED) clinical decisions and has a direct impact on cost, efficiency, and quality of medical care [2].

There do exist significant potential tools for improvement in ED decision-making through the application of Artificial Intelligence [5]. Machine learning (ML) can help predict admission to a hospital from the ED using variables collected as part of routine ED care. ML tools are cost-effective and may be used to help ED personnel make faster and more appropriate disposition decisions, decrease unnecessary testing, and alleviate ED crowding [4-6].

This study aims to present a low-cost ML approach, based on ED data and laboratory exams, that may contribute to the medical decision for patient hospital admission, further improving access to treatment and quality of care.

2. Materials and Methods

This observational retrospective research was conducted in the ED of a public tertiary care hospital in Greece and has been approved by the Institutional Review Board of Sismanogleio General Hospital (Ref. No 15177/2020, 5969/2021). The raw data was retrieved from a standard Laboratory Information System (LIS) and a hospital information system (HIS). After that, the data was processed by a DBMS software by running multiple SQL queries. For the machine learning part, we used scikit-learn [7], an ML library for Python. During the period (14 March – 4 May 2019), 3,204 ED visits were recorded. We investigated biochemical markers and coagulation tests that are routinely checked in patients visiting the ED in relation to the ED outcome (admission or discharge). The data set includes the following variables: serum levels of Urea (UREA), Creatinine (CREA), Lactate Dehydrogenase (LDH), Creatine Kinase (CPK), C-Reactive Protein (CRP), Complete Blood Count with differential, including white blood cells (WBC), neutrophil count (NEUT%), lymphocyte count (LYM%), hemoglobin (HGB), and platelets (PLT), Activated Partial Thromboplastin Time (aPTT), D-Dimer, International Normalized Ratio (INR), age, gender, ambulance use (Ambulance), triage disposition to ED unit and ED outcome (admission or discharge).

Our binary classification problem is considered with two classes of patients who visit the ED, i.e., those patients who are admitted and the rest who are discharged. We evaluated the most popular classification techniques (Logistic Regression, Linear Discriminant Analysis, k-nearest neighbors, Gaussian Naive Bayes, Decision Tree, Random Forest, Bagging classifier, C-Support Vector, XGBoost, LightGBM, Gradient Boosting, Multi-layer Perceptron) of the scikit-learn library through a 10-fold cross-

validation approach, which was applied to avoid overfitting. Missing values were filled in using the *SimpleImputer* method.

One-hot Encoding (OHE) was used for converting categorical features (gender, triage disposition to ED unit, ambulance use, and ED outcome) to numerical features.

Finally, to improve the performance of the final model, we applied a correlation-based feature selection method by evaluating the relationship between each input feature and the target variable (ED outcome). Based on the feature selection process, the numerical variables CPK, aPTT, INR, and PLT were excluded from our analysis since they had the smallest correlation with the ED outcome variable (after OHE).

The performance metrics of the ML techniques that were evaluated in this study were the following:

-Precision (*sklearn.metrics.precision_score*): The precision is the ratio TP / (TP + FP), where TP is the number of true positives and FP is the number of false positives. The precision refers to the classifier's ability not to categorize as positive a sample that is negative [8].

-Recall (*sklearn.metrics.recall_score*): The recall is the ratio TP / (TP + FN) where TP is the number of true positives and FN is the number of false negatives. The recall refers to the classifier's ability to find all the positive samples [8].

-f1 score (*sklearn.metrics.f1_score*): The F1 score can be interpreted as a weighted average of the precision and recall, with the best value being 1 and the worst being 0 [8].

-Accuracy classification score (*sklearn.metrics.accuracy_score*): Classification accuracy is a metric that measures a classification model's performance by dividing the number of correct predictions by the total number of predictions.

-Balanced accuracy score (*sklearn.metrics.balanced_accuracy_score*): The balanced accuracy in binary and multiclass classification problems to deal with imbalanced datasets. It is defined as the average of recall obtained in each class [9].

-ROC AUC (*sklearn.metrics.roc_auc_score*): Compute Area Under the Receiver Operating Characteristic Curve from prediction scores [10].

3. Results

Among the different classifiers that were evaluated through 10-fold cross-validation, a GaussianNB model with *var_smoothing=2e-6* outperformed other models with respect to ROC AUC. The GaussianNB model that implements the Gaussian Naive Bayes algorithm for classification [11] was built and optimized on the open-source Python toolkit scikit-learn. The likelihood of the features is assumed to be Gaussian:

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \, exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

The parameters σ_y and μ_y are estimated using maximum likelihood, and the parameter *var_smoothing* is a stability calculation that widens the curve to accommodate more samples that are further away from the distribution mean.

The performance metrics of the GaussianNB model are presented in the following table (Table 1).

Table 1. *Performance metrics of the GaussianNB model (10-fold cross-validation)*

	Precision	Recall	f1 score	Accuracy	Balanced accuracy	ROC AUC
GaussianNB	0.739	0.629	0.603	0.701	0.629	0.806

4. Discussion

Disposition decisions in the Emergency Department by triage physicians are often difficult, especially in cases with medium- or low-severity medical problems. The final physician's clinical judgment, sometimes biased, as well as social and safety issues, determines the decision to admit or discharge a patient referred to the ED. Machine learning techniques are increasingly applied in emergency medicine, especially for diagnosis and outcome prediction, supporting medical decision-making and potentially improving patient care [5, 6]. Still, there are some limitations in their use in routine clinical practice [12].

In the present study, we examined the performance of different ML models to predict emergency department disposition based on a multitude of readily available laboratory data. Among the most popular classification techniques of the scikit-learn library through a 10-fold cross-validation approach, a GaussianNB model outperformed other models with respect to area under the receiver operating characteristic curve and showed promising potential in assisting with predicting admission to the hospital.

This study's limitations were a relatively short time period examined, the fact that all patients' data came from a single study center, and the limited explainability of the proposed ML model. In future studies, additional data will be added by including patients from longer periods than the current study; however, the key take-away message is that physicians can be familiarized with ML techniques by presenting to them, retrospectively, the performance statistics of such models and allowing them to reflect on the prospect of harnessing existing data to compare admission rates over time and on the extent to which existing hospital protocols may be subject to human factors.

References

[1] Houshyar R, Tran-Harding K, Glavis-Bloom J, Nguyentat M, Mongan J, Chahine C, Loehfelm TW, Kohli MD, Zaragoza EJ, Murphy PM, Kampalath R. Effect of shelter-in-place on emergency department radiology volumes during the COVID-19 pandemic. Emerg Radiol. 2020 Dec;27(6):781-784. doi: 10.1007/s10140-020-01797-y.

[2] Sun BC, Hsia RY, Weiss RE, Zingmond D, Liang LJ, et al. Effect of emergency department crowding on outcomes of admitted patients. Ann Emerg Med. 2013 Jun;61(6):605-611.e6.

[3] Lane BH, Mallow PJ, Hooker MB, Hooker E. Trends in United States emergency department visits and associated charges from 2010 to 2016. Am J Emerg Med. 2020 Aug;38(8):1576-1581. doi: 10.1016/j.ajem.2019.158423.

[4] Moulik SK, Kotter N, Fishman EK. Applications of artificial intelligence in the emergency department. Emerg Radiol. 2020 Aug;27(4):355-358. doi: 10.1007/s10140-020-01794-1.

[5] Shafaf N, Malek H. Applications of Machine Learning Approaches in Emergency Medicine; a Review Article. Arch Acad Emerg Med. 2019 Jun 3;7(1):34.

[6] Stewart J, Sprivulis P, Dwivedi G. Artificial intelligence and machine learning in emergency medicine. Emerg Med Australas. 2018 Dec;30(6):870-874. doi: 10.1111/1742-6723.13145.

[7] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.

[8] Powers D. Ailab . Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. J Mach Learn Technol. 2011; 2: 2229-3981

[9] Brodersen KH, Ong CS, Stephan KE, Buhmann JM (2010) The balanced accuracy and its posterior distribution. 2010 20th International Conference on Pattern Recognition. doi: 10.1109/icpr.2010.764

[10] Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874.

[11] Gaussian Naive Bayes [cited 2021Aug20]. [Internet]. Available from: https://scikit-learn.org/stable/modules/naive_bayes.html#gaussian-naive-bayes

[12] Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med. 2019 Aug 16;2:77. doi: 10.1038/s41746-019-0155-4..

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210919

Automatic Classification of Diabetic Foot Ulcer Images – A Transfer-Learning Approach to Detect Wound Maceration

Jens HÜSERS[a,1], Guido HAFER[b], Jan HEGGEMANN[b], Stefan WIEMEYER[b], Mareike PRZYSUCHA[a], Joachim DISSEMOND [c], Maurice MOELLEKEN[c], Cornelia ERFURT-BERGE[d] and Ursula HÜBNER[a]
[a] *Health Informatics Research Group, Osnabrück University of AS, Germany*
[b] *Christian Hosptial Melle, Niels Stensen Hospitals, Germany*
[c] *Department of Dermatology, Venerology and Allergology, University Hospital of Essen, Germany*
[d] *Department of Dermatology, University Hospital Erlangen, Germany*

Abstract. Diabetic foot ulcer (DFU) is a chronic wound and a common diabetic complication as 2% - 6% of diabetic patients witness the onset thereof. The DFU can lead to severe health threats such as infection and lower leg amputations, Coordination of interdisciplinary wound care requires well-written but time-consuming wound documentation. Artificial intelligence (AI) systems lend themselves to be tested to extract information from wound images, e.g. maceration, to fill the wound documentation. A convolutional neural network was therefore trained on 326 augmented DFU images to distinguish macerated from unmacerated wounds. The system was validated on 108 unaugmented images. The classification system achieved a recall of 0.69 and a precision of 0.67. The overall accuracy was 0.69. The results show that AI systems can classify DFU images for macerations and that those systems could support clinicians with data entry. However, the validation statistics should be further improved for use in real clinical settings. In summary, this paper can contribute to the development of methods to automatic wound documentation.

Keywords: Clinical Decision Support System, Health Information Technology, Diabetic Foot Ulcer, Image Classification, Wound Care, Transfer Learning, Convolutional Neural Networks

1. Introduction

Diabetes mellitus has a high prevalence and is a global health threat. Among diabetic patients, 2% - 6% of them witness the onset of a diabetic foot ulcer (DFU). The IWGDF defines a DFU as "an infection, ulceration, or destruction of tissues of the foot" of diabetic patients [1]. DFU is a severe late-stage complication of diabetes as it can lead to pain, immobility, infection, and even foot and lower leg amputations. Short-term wound characteristics may indicate delayed healing, such as peri-wound skin

[1] Corresponding Author, Jens Hüsers, Osnabrück University of AS, Health Informatics Research Group, PO Box 1940, 49009 Osnabrück, Germany; E-Mail: j.huesers@hs-osnabrueck.de.

maceration [2]. Thus, information about the maceration status is essential for planning the wound care and the dressing strategy for health professionals.

In this context, maceration status is part of standardized interdisciplinary wound documentation [3]. The importance of wound documentation correlates with the efforts required to enter and update information. As wound images are easy to obtain and are usually taken by the health care provider anyway, they lend themselves to be used as a data source for automatic detection. For example, AI systems were employed to detect necrotic tissues or infection status of a wound in images [4].

Against this background, particularly the need for digitally supported documentation and the relevance of wound macerations for planning and conducting wound care, this study investigates the automatic classification of DFU images. The main objective is to train an AI system and evaluate its performance.

2. Methods

For this study, we collected 416 wound images that were part of the wound documentation at the Wound Care Center of Christliches Klinikum Melle Germany, a specialized in- and outpatient clinic for patients with DFU. The data preprocessing consisted of two steps. First, the wounds in the images were annotated using bounding boxes, which are frames around the DFU in an image indicating its location. Second, we cropped all DFUs in the images using the bounding box plus 75 pixels as an additional margin. This processing led to 434 images, each showing a single DFU; eleven images contained two, and one image showed three ulcers. Then, the 434 cropped images were classified regarding the maceration status by two health professionals, a physiotherapist and a wound specialist from Christliches Klinikum Melle.

The image classification system relied on the MobileNetV1 model, a convolutional neural network (CNN) for image classification. A key feature of MobileNetV1 among compared to other CNN arcitectures is the flexible adaption of its size to control the complexity of the system which we utilized in this study. The model training used the pre-trained weights based on the *imagenet* dataset, an open image database used for AI development and benchmark, thereby model training uses a transfer-learning approach. The input images were scaled to 224 by 224 pixels (plus three color channels). The top layer of the MobileNetV1 model was replaced with two fully connected layers and a final sigmoid output layer. The final model had 847,014 parameters.

Out of these 434 images, a random subset of 326 images (75%) served as the training set. The remaining 108 images formed the validation set (25%). We used a data augmentation pipeline for model training that randomly transformed the images before each training step to avoid overfitting. The pipeline rotated, sheared, and flipped the images horizontally and vertically. Additionally, the pipeline shifted the brightness randomly. To avoid overfitting, we also defined a dropout rate of 10% in all layers and an early stopping callback to stop training when there is no improvement on validation loss after 20 epochs. The model with the lowest loss was selected as the final model. The model was evaluated on the unaugmented validation set. A GPU (Tesla P100-PCIE-16GB) served as the computational backbone for the model training which was performed using the Python version of the open-source software library *TensorFlow 2.6*.

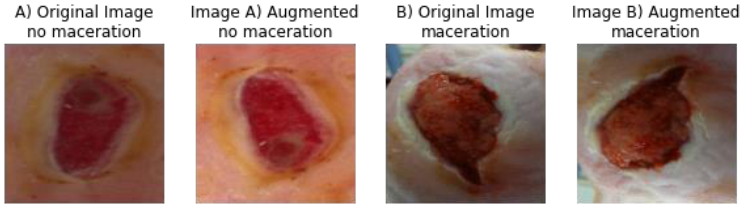


Figure 1. Subset of training images with corresponding labels and an augmented example

3. Results

The model training showed convergence, and the callback triggered early stopping after 93 training epochs. The monitored loss curves of the augmented training and validation losses showed the absence of overfitting. The final model yielded an F1-score of 0.71 on the 108 (unaugmented) validation images. This F1-score corresponds to a recall, also known as sensitivity, of 0.69 and a precision, also known as positive predictive value, of 0.67. The accuracy was 0.69, and the area under the receiver operating curve was 0.78. The images, the code of the training procedure, and the validation statistics are available online at [5].

4. Discussion

This study presents a system for classifying macerations in DFU images using a transfer learning approach. The validation showed that among all images for which the system identified a maceration, 67% were correct (precision). Among the images showing a maceration, 69% were correctly identified (recall). In light of these results, systems using artificial intelligence technologies such as CNNs promise to support the recording of DFU information. The findings are in line with similar initiatives that investigate methods to classify DFU images automatically. For example, the DFU Classification Challenge reached a F1-score of 0.73 for classifying necrotic tissues and wound infections of DFU images [4], which is comparable to our F1-score of 0.71 for macerations.

Although these validation statistics are promising for detecting macerations in DFU images from a scientific point of view, the overall accuracy of 69% is presumably not high enough for real clinical scenarios when used to automatically classify macerations. However, the current version can support semi-automatic recording by proposing the maceration status to a physician by pre-entering the information into the digital wound record, which the physician can accept or decline. Furthermore, the feedback from the physician could contribute to the continuous improvement of the CNN.

When applying classification systems like the one presented here, the context in which it was developed is essential and must be considered. For example, in this study, wound images used for model training showed DFUs without wound dressings and were not covered with cremes or gels. However, when physicians neglect this context, this might lead to unreliable classification. Thus, besides communicating the system's validity to physicians, they must be informed about the system's features and limits.

This study has limitations. Images from a single DFU center were used, and the performance of the final model was validated using the validation set rather than an additional external test set. Internal validity showed satisfactory results. We tried to improve external validity by using transfer learning, data augmentation, dropout, and a sparse model to force the system to learn the general pattern of macerations [6]. Thus, we expect this model to generalize well. Before this system is implemented in wound documentation for clinical use, it should be validated on an expanded image dataset from other clinical centers. Additionally, the number of training images should be increased to further improve the validity of the system.

In summary, the developed classification model showed satisfying validity for classifying wound images for macerations which must be further improved for clinical use in wound documentation to enable automatic wound documentation that promises to curtail the documentation time for clinicians.

Funding

This study was funded by the German Federal Ministry of Education and Research (BMBF) (grant: 16SV8616).

References

[1] van Netten JJ, Bus SA, Apelqvist J, Lipsky BA, Hinchliffe RJ, Game F, et al. Definitions and criteria for diabetic foot disease. Diabetes Metab Res Rev. 2020;36(S1):e3268.

[2] Haryanto H, Arisandi D, Suriadi S, Imran I, Ogai K, Sanada H, et al. Relationship between maceration and wound healing on diabetic foot ulcers in Indonesia: a prospective study. Int Wound J. 2017 Jun;14(3):516–22.

[3] Konsensusgruppe, Herberger K, Heyer K, Protz K, Mayer A, Dissemond J, et al. Nationaler Konsensus zur Wunddokumentation beim Ulcus cruris: Teil 2: Routineversorgung – Klassifikation der Variablenausprägungen. Hautarzt. 2017 Nov;68(11):896–911.

[4] Diabetic Foot Ulcer Challenge 2020 - Grand Challenge [Internet]. grand-challenge.org. [cited 2021 Aug 21]. Available from: https://dfu2020.grand-challenge.org/Evaluation/

[5] GitHub Gist for wound image classification for macerations [Internet]. [cited 2021 Aug 27]. Available from: https://gist.github.com/jnshsrs/1d4ce0b4477330b241d61156ab973f64

[6] Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. J Big Data. 2019 Jul 6;6(1):60.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210920

Effects of Computer-Aided Decision Support Systems on Appropriate Antibiotic Prescribing by Medical Interns: A Quasi-Experimental Study

Fatemeh RANGRAZ JEDDI[a,b], Mansooreh MOMEN-HERAVI[c,d], Ehsan NABOVATI[a,b], Felix HOLL[e,f,g], Hossein AKBARI[h,i] and Razieh FARRAHI[b,1]

[a] *Health Information Management Research Center, Kashan University of Medical Sciences, Kashan, Iran*

[b] *Department of Health Information Management & Technology, School of Allied Health Professions, Kashan University of Medical Sciences, Kashan, Iran*

[c] *Infectious Diseases Research Center, Kashan University of Medical Sciences, Iran*

[d] *Department of Infectious Diseases, School of Medicine, Kashan University of Medical Sciences, Kashan, Iran*

[e] *Digi Health Institute, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany*

[f] *IGHS, University of California, San Francisco, San Francisco, CA, USA*

[g] *IBE, Ludwig Maximilian University of Munich, Munich, Germany*

[h] *Social Determinants of Health (SDH) Research Center, Kashan University of Medical Sciences, Kashan, Iran*

[i] *Department of Biostatistics and Epidemiology, School of Public Health, Kashan University of Medical Sciences, Kashan, Iran*

Abstract. Literature suggests that the adoption of guidelines for antibiotic prescribing has a significant impact on improving prescription practices of physicians; thus, this study aimed to assess the effectiveness of computer-aided decision support systems (CA-DSS) on antibiotic prescribing among medical interns. A prospective before-and-after interventional study was conducted on 40 medical interns. The interns were asked to use the CA-DSS during a one-month internship course at the infectious disease department. The main outcome measure was the knowledge of medical interns regarding the type, name, volume, usual dosages, and administration route of antibiotics prescribed. Paired t-test was applied to assess the change of medical interns' knowledge before and after the study. There was a statistically significant difference between the mean score of interns' medical knowledge before 5.4±2 and after 9.1±2.8 using the CA-DSS (p= 0.000). CA-DSS as an IT-based training intervention was effective for the knowledge of medical interns to prescribe the right antibiotics for acute respiratory infections.

Keywords. Antibiotic, Computer, Prescription, Respiratory tract infection

[1] Corresponding author, Razieh Farrahi, Department of Health Information Management & Technology, School of Allied Health Professions, Kashan University of Medical Sciences, Kashan, Iran; Email: Farrahi-sr@kaums.ac.ir

1. Introduction

Acute respiratory infections are one of the major public health concerns [1] as well as one of the most important reasons for prescribing antibiotics in primary [2] and secondary care [3]. However, most of the prescribed antibiotics are reported as unnecessary and irrational worldwide [4]. Evidence indicates that unnecessary antibiotic prescription might be due to various reasons [5] such as lack of knowledge and failure to apply clinical guidelines reported among the major contributing factors in the unnecessary prescription of antibiotics in healthcare settings [6].

Therefore, developing targeted interventions in terms of training programs, clinical guidelines, and adopting state-of-the-art health information technology (HIT) is acknowledged as effective strategies to improve antibiotics prescription. There is increasing recognition that the education of medical students serves as one of the key approaches to combat the irrational prescription of antibiotics [7].

Given the rapid advancement of HIT and improvement of the digital knowledge and skills of students, instructors, and health professionals at healthcare settings, decision support systems (DSS) hold promises to promote knowledge and practice of rational use of antibiotics among physicians [8]. Adopting DSSs equip with antibiotics guidelines in a knowledge-base of an educational simulator can play a crucial role in improving medical students' knowledge about rational prescription of antibiotics [9]. Since the implementation of these interventions often is accompanied by resources utilization (cost, time, effort, human resources), it is important to investigate to what extent they obtained the desired objectives [10]. Furthermore, researchers have emphasized on background or context in which the intervention is implemented (e.g., hospitals, users, stakeholders, culture), because the impact of any HIT intervention is likely to be affected as much by the background, as the intervention features itself [11]. The purpose of the current study was to examine the effect of using computer-aided decision support systems (CA-DSS) on the appropriate prescription of antibiotics for acute respiratory infections by medical interns.

2. Method

We conducted a quasi-experimental study to evaluate the effect of CA-DSS on appropriate antibiotic prescribing by medical interns from March 2021 to July 2021 (5 months) at a general teaching hospital with 510 beds at Kashan University of Medical Sciences (KAUMS) in Iran. Every month, eight medical interns who were spending a one-month internship course at the infectious diseases department participated in the study; forty medical interns participated in the five-month study period. Clinical guidelines and attending physicians' clinical judgment were incorporated as the rules for the knowledge-base of the CA-DSS to perform reasoning and provide educational feedback to medical interns based on targeted scenarios. Sixty scenarios were written by the attending physician via the teacher interface. The correct answers for each scenario including the final diagnosis, signs & symptoms, diagnostic procedures, and the medication were determined based on guidelines and attending physicians' clinical judgment. The matching method to investigate the right answer was the decision tree via IF-THEN rules. The students were asked to use the CA-DSS while a one-month internship course at the department. After being logged in, a random scenario about acute respiratory infection was displayed to the participants; having read the scenario

description, the student goes through the four stages of determining the certain diagnosis and signs & symptoms in favor of diagnosis, diagnostic procedures, appropriate antibiotics, and medication prescription. After responding to each stage, the relevant content of the clinical guideline as well as attending physician recommendations were provided to the medical interns as educational feedback. Finally, the score of each medical intern and the average score of the other participants for the given scenario were also calculated. The main outcome measure was the knowledge of medical interns regarding the diagnosis and treatment of the infection's disease in addition to the appropriate antibiotic's prescription in terms of type, name, volume, usual dosages, and administration route of antibiotics before and after using the CA-DSS. The knowledge scores were measured using two paper-based exams on 14 scenarios, once before the intervention and once after one month applying the CA-DSS. Descriptive statistics, paired t-test and independent t-test were applied to assess the change of medical interns' knowledge score. This study was a part of Ph.D. thesis in the field of health information management that was approved by the Ethics Committee of KAUMS (IR.KAUMS.MEDNT.REC.1398.141) and funded by the deputy of research in Kashan University of Medical Sciences with the grant number (98225).

3. Results

These 40 medical interns included 21 men (52.5) and 19 women (47.5), of whom, 17(42.5%) were sixth-year medical students and 23 (57.5%) were seventh-year medical students. The CA-DSS was used 369 times by the medical interns to read the scenarios during the intervention period. The average studied scenarios for each medical intern were nine. Antibiotics were correctly selected by the medical interns for 248 scenarios. In 128 (34%) of the studied scenarios, students revised their answers in compliance with the CA-DSS feedbacks. The interns' knowledge score for antibiotic prescription increased from 5.4 ± 2 before the intervention to 9.1 ± 2.8 after the CA-DSS intervention ($P< 0.001$). The results of the independent t-test showed that although the percentage of changes in the score of interns who studied nine or more scenarios (95.7 ± 110) was higher in compared to interns who studied less than nine scenarios (85.3 ± 89.5), this difference was not statistically significant ($P=0.74$). The results of the independent t-test also showed that although the percentage of changes in the score of sixth-year students (116 ± 119) was higher in compared to seventh-year students (69 ± 74), these changes were not statistically significant ($P=0.14$).

4. Discussion

The CA-DSS as an IT-based educational intervention improved the medical interns' knowledge about the appropriate prescription of antibiotics for acute respiratory infections. The results of similar studies also suggest that DSSs can perform best in the most difficult clinical scenarios and improve physicians' knowledge and practice in antibiotic prescribing [12]. Conducted studies confirmed that training by DSS improves the knowledge of medical students significantly [13, 14]; and overcomes the clinical learning environment challenges in terms of limited access to the well-experienced attending physicians, and evidence-based guidelines to support appropriate antibiotic prescription [15]. Therefore, adopting CA-DSS and educational stimulators would

improve students learning through providing audit and feedback, facilitating access to clinical guidelines, and supporting evidence-based decision making. To our knowledge, this study represents the first attempt in Iran to develop a computer-aided decision support based on the clinical settings context and the preferences for assisting medical interns' antibiotics prescription. The web-based CDSS supports convenient access to the clinical guidelines at any time and any place. Our study has several limitations: the lack of performing regular pre-test and post-test due to the issues associated with the COVID-19 pandemic, in addition, the crowded services following the COVID-19 resulted in the medical interns' tiredness which might have led to the less usage of the system or usage without the required focus. Considering the positive impact of the intervention, it is suggested to design such educational decision aids for other diseases. Also using computer-aided stimulators is recommended as an alternative educational tool for training medical students in clinical environments.

References

[1] Dagne H, Andualem Z, Dagnew B, Taddese AA. Acute respiratory infection and its associated factors among children under-five years attending pediatric ward at University of Gondar Comprehensive Specialized Hospital, Northwest Ethiopia: institution-based cross-sectional study. BMC pediatrics. 2020;20(1):1-7.

[2] O'Connor R, O'Doherty J, O'Regan A, Dunne C. Antibiotic use for acute respiratory tract infections (ARTI) in primary care; what factors affect prescribing and why is it important? A narrative review. Irish Journal of Medical Science (1971-). 2018;187(4):969-86.

[3] Hussein RA. Patterns of Documenting Patients' Medical Records and Antibiotic Use for Cases with Acute Respiratory Infection in Basrah hospitals. University of Thi-Qar Journal of Medicine. 2018;15(1):1-11.

[4] Machowska A, Stålsby Lundborg C. Drivers of Irrational Use of Antibiotics in Europe. Int J Environ Res Public Health. 2018;16(1):27.

[5] Organization WH. WHO global strategy for containment of antimicrobial resistance. World Health Organization; 2001.

[6] Fleming A, Bradley C, Cullinan S, Byrne S. Antibiotic prescribing in long-term care facilities: a qualitative, multidisciplinary investigation. BMJ open. 2014;4(11): e006442.

[7] Laxminarayan R, Duse A, Wattal C, Zaidi AK, Wertheim HF, Sumpradit N, et al. Antibiotic resistance—the need for global solutions. The Lancet infectious diseases. 2013;13(12):1057-98.

[8] Wasserman S, Potgieter S, Shoul E, Constant D, Stewart A, Mendelson M, et al. South African medical students' perceptions and knowledge about antibiotic resistance and appropriate prescribing: Are we providing adequate training to future prescribers? South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde. 2017;107(5):405-10.

[9] Willis MH, Frigini LA, Lin J, Wynne DM, Sepulveda KA. Clinical decision support at the point-of-order entry: an education simulation pilot with medical students. Academic radiology. 2016;23(10):1309-18.

[10] Hawe P. Lessons from complex interventions to improve health. Annual review of public health. 2015;36:307-23.

[11] Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. Springer; 2018.

[12] Warner Jr H, Reimer L, Suvinier D, Li L, Nelson M, editors. Modeling empiric antibiotic therapy evaluation of QID. Proceedings of the AMIA Symposium; 1999: American Medical Informatics Association.

[13] Karakuş A, Duran L, Yavuz Y, Altintop L, Çalişkan F. Computer-based simulation training in emergency medicine designed in the light of malpractice cases. BMC Medical Education. 2014;14(1):155.

[14] Leung GM, Johnston JM, Tin KY, Wong IO, Ho LM, Lam WW, et al. Randomised controlled trial of clinical decision support tools to improve learning of evidence-based medicine in medical students. BMJ (Clinical research ed). 2003;327(7423):1090.

[15] Khumrin P, Ryan A, Juddy T, Verspoor K, editors. DrKnow: A diagnostic learning tool with feedback from automated clinical decision support. AMIA Annual Symposium Proceedings; 2018: American Medical Informatics Association.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210921

Remote AI Supported E-Multidisciplinary Oncology Conference in Breast Cancer as a Technology and Method to Optimize Outcomes in the Peripheries

Constantinos S. MAMMAS[a,b,1], Adamantia S. MAMMA[b]
George PAPAXOINIS[c], Irene GEORGIOU[c]
[a] *Elena El. Venizelou-Alexandra University Hospital of Athens, National and Kapodistrian University of Athens (NKUA)*
[b] *Program of Excellence 2014-16-Siemens Program for Greece*
[c] *Saint Savvas Anticancer Hospital of Athens*

Abstract. Aim: Feasibility-reliability control of Telemedicine Systems (TS) integrated with Multimedia Systems (MS) and Artificial intelligence (AI) for remote e-Multidisciplinary Oncology Conference in Breast Cancer. **Material and Methods:** Forty (n1=40) patients suffering from breast surgical oncology malignant (n2=32) and non-malignant (n3=8) diseases classified to seven categories: Nipple Discharge, Dominant Breast Mass, Occult Breast Lesion, Early Breast Carcinoma, Advanced Breast Carcinoma, Recurrent Breast Carcinoma) and treated clinically with the standard diagnostic (Mammography, US, MRI, Cytology, Pathology, BRCA1/2 Mutation Predisposition and Breast Cancer Risk Analysis) surgical, auxiliary therapeutic methods. Then clinical decisions compared to those proposed remotely by the virtual AI supported e-Oncology Conference for each patient. **Results:** In four (n4=4) out of forty patients (TS, MS and AI) supported decision making and surgical treatment proposal including postoperative Radiotherapy proposal was not as clear as expected. Non-output answer for non-malignant breast pathologies (n3=8) was accurately indicated by (MS and AI). Mean accuracy of (TS, MS and AI) for: **1.**Surgical Operative Planning including Rad=94.1%, **2.**Chem=96.8%, **3.**Horm=96.7% [In 95%, (Confidence interval: 85-99%)]. **Conclusion:** High feasibility-reliability of the virtual AI supported e-Multidisciplinary Oncology Conference for remote decision making and surgical planning and for optimum outcomes in Breast Cancer treatment makes it a clinical necessity especially for the periphery of Hellas.

Keywords. Breast Surgery**,** Tele-Radiology, Tele-Pathology, Tele-Cytology, AI, E-Multidisciplinary Oncology Conference

1. Introduction

The optimal management of patients with breast cancer (BC) requires the expertise of specialists from different disciplines. This has led to the evolution of multidisciplinary teams (MDTs), allowing all key professionals to jointly discuss individual patients and

[1] Corresponding Author, Constantinos St. Mammas, MD, MSc, PhD, 9, Kononos Ave., Pagrati, Athens, Greece; Email: csmammas@med.uoa.gr.

to contribute independently to clinical decisions. Also, it proved that MDTs for Multidisciplinary Oncology Conference in cooperation with Multidisciplinary Management of the breast cancer in high volume Breast Units optimize quality and have significantly better clinical outcomes for the benefit of the patients. However, MDTs for BC decision making in different regions in the periphery of Greece and in other countries are scarce [1,2]. The project searches feasibility-reliability of Telemedicine Systems (TS) integrated with Multimedia Systems (MS) and Artificial intelligence (AI) for remote e-Multidisciplinary Oncology Conference in Breast Cancer interrelated with Telepathology (TPE), Teleradiology (TRE) and Telecytology (TCE) remote evaluation [3,4,5].

2. Material and Methods

Experimentation included the development of an OTE-TS similar Experimental TS (Exp.-TS) for the simulation of the integrated TS, MS and AI based TRE and TPE and TCE virtual examination of each patient (pn 1009078, 34931, 34932, 34933) (Table 1.).

Table 1. Comparison of the Modules between OTE-TS and Exp.-TS

Modules	OTE-TS	EXP.-TS
a. Medical record process	+	+
b. Examinations results.	+	+
c. Capture scanning and imaging.	+	+
d. DICOM and PACS vision.	+	+
e.Real-time tele-conference	+	+
f. Chat and whiteboard facilities.	+	+
g. Application sharing.	+	+
h. Tele-secretary facilities.	+	+
j. Tele-Mentoring facilities	+	+
i. Telecommunication net	ISDN based	Internet based
k. Multimedia System	+	+
l.AI computation system	-	+

Simulation of the TRE upon N=40 Mammographic, MRI, CT, Breast and Upper Abdomen US digital images projected on the Exp.-TS in the internet (Cloud) for remote examination and decision making upon the virtual e-Multidisciplinary Breast Oncology Conference (Figure 1.). Simulation of the TPE including TCE based on the already worked out breast biopsies and their results for N=40 digital microscopic images.

Integrated AI supported Decision Making (including clinical interventions and genetic consultation) and Surgical Planning analysis using SPSS (version 17.0).

Forty (n1=40) patients suffering from breast surgical oncology malignant (n2=32) and non-malignant (n3=8) diseases classified to seven categories: Nipple Discharge, Dominant Breast Mass, Occult Breast Lesion, Early Breast Carcinoma, Advanced Breast Carcinoma, Recurrent Breast Carcinoma) and treated clinically with the standard diagnostic (Mammography, US, MRI, Cytology, Pathology, BRCA1/2 Mutation Predisposition and Breast Cancer Risk Analysis) surgical, auxiliary therapeutic methods. Then the diagnostic and therapeutic clinical decisions compared to those proposed by our (MS and AI) supported decision making and planning system and logic process for each patient.

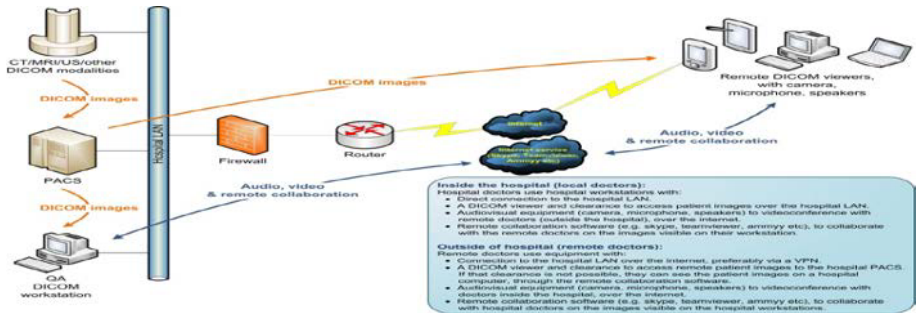


Figure 1. Simulation of the remote collaboration among specialists in the context of the remote virtual e-Multidisciplinary Breast Oncology Conference *via* internet (PACS and DICOM based remote examination).

3. Results

In four (n4=4) out of forty (n1=40) patients (TS, MS and AI) supported decision making and surgical treatment proposal including postoperative Radiotherapy proposal was not as clear as expected. Non-output answer for non-malignant breast pathologies (n3=8) was accurately indicated by (TS, MS and AI). Mean accuracy of (TS, MS and AI) for: **1.**Surgical Operative Planning including Rad=94.1%, **2.**Chem=96.8%, **3.**Horm=96.7% [In 95%, (Confidence interval: 85-99%)] (Tables 2, 3, 4).

Table 2. Simulated TRE (Mammography, MRI, CT, Breast and Upper Abdomen US)

	N	TP	FP	FN	TN	SENSITIVITY	SPECIFICITY	ACCURACY
Patients and Lesions	40	32	0	0	8	**100,0**	**100,0**	**100,0**

N=Number of digital images given for examination, TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative, Sensitivity(%), Specificity(%), Diagnostic accuracy (Efficiency)(%.)

Table 3. Simulated TPE including TCE

	N	TP	FP	FN	TN	SENSITIVITY	SPECIFICITY	ACCURACY
Patients and Lesions	40	32	0	0	8	**100,0**	**100,0**	**100,0**

N=Number of digital images given for examination, A=Number of Answers after Examination, TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative, Sensitivity(%), Specificity(%), Diagnostic accuracy (Efficiency)(%.)

Table 4. AI supported Decision Making and Surgical Planning

	N	TP	TN	ACCURACY+Rad	ACCURACY+Chem	ACCURACY+Horm
Surgical Planning	40	36	8	94.1%	96.8%	96.7%

N=Number of images given for examination, TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative, AI Supported Decision making and planning Accuracy (Efficiency)(%.).

4. Discussion

Given the lack of Multidisciplinary Oncologic Conferences in the periphery of Greece and the fact that the clinical outcomes of the patients suffering from breast cancer are better under multidisciplinary management in high volume breast units the abovementioned results are very promising to realize virtual e-Multidisciplinary

Oncology Conference to optimize quality of breast cancer management in the beginning. The results confirmed clinically in the periphery of Greece with a case of a female patient suffering from an exacerbation of right breast cancer (Extended loco-regional right breast angio-sarcoma) which treated initially with chemotherapy (May 2021). Then treated with an additional chemotherapy schedule based on the decision of the remote e-Multidisciplinary Oncology Conference in Breast Cancer using the above described technology and method (TS, MS and AI from the General Hospital of Kalymnos, June 2021). With regard to safety issues Morgan [6] proposed the Virtual private networks (VPN) as adequate solutions for authentication, access control and confidentiality. The secure hypertext transfer protocol (https) can be used to encrypt for web distribution. A public key infrastructure (PKI) solves all the issues mentioned above. Clinical efficacy research protocols for further study of the remote e-Multidisciplinary Oncology Conference in Breast Cancer is of high priority [7,8].

5. Conclusion

High feasibility-reliability of the virtual e-Multidisciplinary Oncology Conference in Breast Cancer integrated with AI supported decision making and treatment in Breast Cancer, makes it a clinical necessity for optimum clinical management and treatment of the breast cancer especially for the periphery of Hellas.

References

[1] Saini KS, Taylor C, Ramirez AJ, Palmieri C, Gunnarsson U, Schmoll HJ, et al. Role of the multidisciplinary team in breast cancer management: results from a large international survey involving 39 countries Annals of Oncol. 2012 Apr:23(4):853-9.

[2] Shao J, Rodrigues M, Corter AL, Baxter NN Multidisciplinary care of breast cancer patients: a scoping review of multidisciplinary styles, processes, and outcomes. Current Oncol. 2019 Jun;26 (3):385-397.

[3] Mammas CS, Mamma AS, Papaxoinis G, Georgiou I. AI supported decision making and planning in 40 breast surgery operations. In:Aristoteleio University of Thessaloniki Editor. Proceedings of the 32nd Panhellenic Congress of Surgery and International Forum; 2021 Jun 9-12;Thessaloniki:p.237.

[4] Mammas CS, Mandellos G, Economou GP, Lymberopoulos D. Structuring Expert-Leaded Medical Protocols for Telemedicine Systems. In: IEEE Editor. Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2001 Oct 25-28; Constantinople: p.p.3529-32.

[5] Karavatselou E, Economou GP, Chassomeris C, Daneli V, Lymperopoulos D. OTE-TS-a new value-added telematics service for telemedicine applications. IEEE Transactions on Information Technologies in Biomedicine. 2001 Sep;5(3):210-24.

[6] Morgan RH. Computer network security for the radiology enterprise Computer network. Radiology. 2001 Aug;220(2):303-09.

[7] Furness PN, Bamford WM. Tele-pathology. Curr Diag Pathol. 2001 Dec;7:281-291.

[8] Coiera E, editor. The Internet and Telemedicine. In: Coiera E, editor. Guide to Medical Informatics. 2nd ed. New York: CHM; c2007. 408 p.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210922

E-Multidisciplinary Oncology Conference for Liver Cancer as a Technology and Method to Optimize Personalization and Outcomes in the Peripheries

Constantinos S. MAMMAS[a,1] and Adamantia S. MAMMA[a]
[a]*Program of Excellence 2014-16-Siemens Program for Greece*

Abstract. Aim: To search feasibility and reliability of Telemedicine Systems (TS) in the remote multidisciplinary oncology conference for decision making and treatment of liver lesions. Material and Methods: By an experimental TS, twenty six (n=26) specialists based on a series of five (N=5) simulated remote image examinations, assessed sensitivity-specificity of the remote examination of the Liver (L) for neoplastic diseases and damages (Virtual Examination=VE). Results: Analysis showed: injuries (sensitivity=96%), injuries of the capsula (sensitivity=91.7%), hematomas (sensitivity=91,7%), non-neoplastic diseases (specificity=100%), neoplastic diseases (sensitivity=100%). Conclusion: The VE of the (L) in combination with high-tech visualization and multimedia and the remote participation of liver surgical oncology, oncology, radiology, pathology and cytology experts composes a feasible and reliable e-Multidisciplinary Oncologic Conference for a Personalized and Optimum Decision Making and Treatment in Liver Cancer.

Keywords. Tele-Medicine, Tele-pathology, Tele-radiology, Liver Cancer, E-Multidisciplinary Oncologic Conference

1. Introduction

The optimal management of patients with liver cancer (LC) requires the expertise of specialists from different disciplines. This has led to the evolution of multidisciplinary teams (MDTs), allowing all key professionals to jointly discuss individual patients and to contribute independently to clinical decisions. However, LC MDTs in different regions in the periphery and countries are scarce [1,2]. The project searches feasibility-reliability of Telemedicine Systems (TS) integrated with Multimedia Systems (MS) for remote e-Multidisciplinary Oncology Conference in LC interrelated with Teleradiology (TRE) and Telepathology (TPE) evaluation of the patients and the liver (L) for optimum decision making and surgical planning [3,4,5].

[1] Corresponding Author, Constantinos St. Mammas MD, MSc, PhD, 9, Kononos Ave., Pagrati, Athens, Greece; Email: csmammas@med.uoa.gr.

2. Material and Methods

Experimentation included.

The development of an OTE-TS similar Experimental TS (Exp.-TS) for the simulation of the integrated TS, MS and AI based TRE and TPE virtual examination of the patient and the liver viscera remotely (pn 1009078, 34931, 34932, 34933) (Table 1).

Table 1. Comparison of the Modules between OTE-TS and Exp.-TS

Modules	OTE-TS	EXP.-TS
a. Medical record process	+	+
b. Examinations results.	+	+
c. Capture scanning and imaging.	+	+
d. DICOM and PACS vision.	+	+
e.Real-time tele-conference	+	+
f. Chat and whiteboard facilities.	+	+
g. Application sharing.	+	+
h. Tele-secretary facilities.	+	+
j. Tele-Mentoring facilities	+	+
i. Telecommunication net	ISDN based	Internet based
k. Multimedia System	+	+
l.AI computation system	-	+

Simulation of the TRE of (Ls) on 15 abdominal MR digital images taken by the ACS-NT GYROSCAN MRI POWERTRACK 6000, 1.5T, (by PHILIPS) and projected on the Exp.-TS in the intranet of the Department of Radiology of the Aretaieion University Hospital for examination of damages and lesions by a radiologist (2012) (Figure 1).

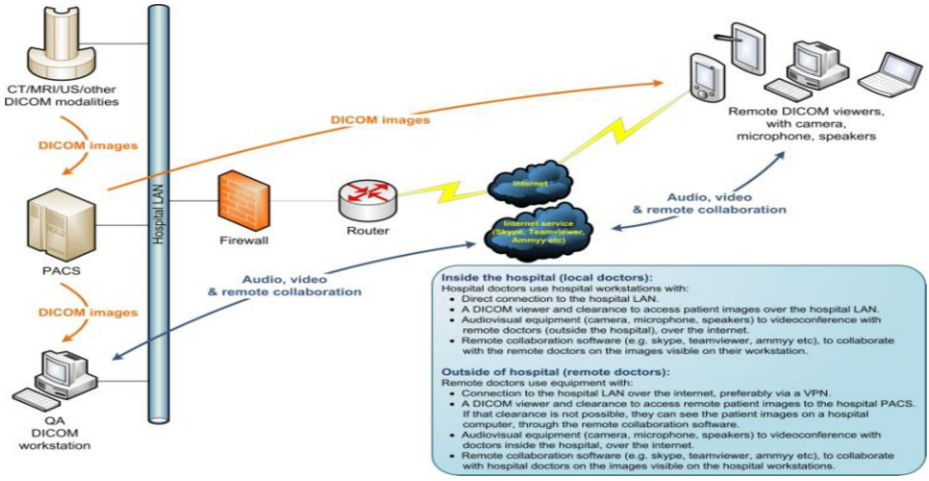


Figure 1. Collaboration between the hospital in the periphery and the Multidisciplinary Team for the remote examination of the patient's medical record and its medical data focused on Liver Cancer for the realization of the e-Multidisciplinary Oncology Conference for remote decision making and surgical planning *via* internet (DICOM and PACS based remote digital visualization of the necessary radiological examinations).

Experimental simulation of the remote virtual evaluation of the (L) upon n=5 digital histological images by 26 specialists in the intranet of the department of Pathology of the Medical School of Athens (2013).

Sensitivity-specificity analysis of the results by using the SPSS statistical software (version 17.0).

3. Results

All examiners defined (L) and diagnosed damages and lesions (Tables 2, 3, 4).

Table 2. Simulated TRE of the patients' (L)

	N	TP	FP	FN	TN	SENSITIVITY	SPECIFICITY	ACCURACY
Damages and Lesions	15	4	0	0	11	**100,0**	**100,0**	**100,0**

N=Number of images given for examination, TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative, Sensitivity(%), Specificity(%), Diagnostic accuracy (Efficiency)(%.)

Table 3. Simulated TPE of the patients' (L)

	N	A	TP	FP	FN	TN	SENSITIVITY	SPECIFICITY	ACCURACY
Damages and Lesions	26	26	24	0	1	0	**96,0**	**-**	**96,0**

N=Number of Photos given for examination, A=Number of Answers after Examination, TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative, Sensitivity (%), Specificity(%), Diagnostic accuracy (Efficiency)(%.)

Table 4. Integrated TRE and TPE of the patients' (L)

	N	TP	FP	FN	TN	SENSITIVITY	SPECIFICITY	ACCURACY
Damages and Lesions	41	28	0	1	11	**96,6**	**100,0**	**97,5**

N=Number of images given for examination, TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative, Sensitivity(%), Specificity(%), Diagnostic accuracy (Efficiency)(%.).

4. Discussion

The abovementioned very promising results confirmed clinically in the periphery of Greece with a clinical case of a female patient which had suffered from a right renal cancer that had been treated in the past but now suffered from an exacerbation characterized with a right lobe liver mass (July 2021). The patient treated with a surgical intervention based on the decision of the remote e-Multidisciplinary Oncology Conference in Liver Cancer using the above described technology and method focusing on tele-radiology. With regard to safety issues RH Morgan [6] proposed the Virtual private networks (VPN) as adequate solutions for authentication, access control and confidentiality. The secure hypertext transfer protocol (https) can be used to encrypt for web distribution. A public key infrastructure (PKI) solves all the issues mentioned above. Clinical efficacy research protocols for further study of the remote e-Multidisciplinary Oncology Conference in Liver Cancer is of high priority [7,8].

5. Conclusion

The VE of the (L) in combination with high-tech visualization and multimedia technology and the remote participation of liver surgical oncology experts for the realization of the e-Multidisciplinary Oncologic Conference seems feasible and reliable to optimize liver cancer personalized decision making and then multidisciplinary management in high volume liver units for the patients in the periphery of Hellas.

References

[1] Siddique O, Yoo ER, Perumpail RB et al.:The importance of a multidisciplinary approach to hepatocellular carcinoma. Journal of Multidisciplinary Healthcare 2017 Mar;10:95-100.
[2] Couri T, Pillai A. Goals and targets for personalized therapy for HCC. Hepatol Int. 2019 Mar;13(2):125-137.
[3] Mammas CS, Mamma AS, Papaxoinis G, Georgiou I. AI supported decision making and planning in 40 breast surgery operations. In:Aristoteleio University of Thessaloniki Editor. Proceedings of the 32nd Panhellenic Congress of Surgery and International Forum; 2021 Jun 9-12;Thessaloniki:p.237.
[4] Mammas CS, Mandellos G, Economou GP, Lymberopoulos D. Structuring Expert-Leaded Medical Protocols for Telemedicine Systems. In: IEEE Editor. Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2001 Oct 25-28; Constantinople: p.p.3529-32.
[5] Karavatselou E, Economou GP, Chassomeris C, Daneli V, Lymperopoulos D. OTE-TS-a new value-added telematics service for telemedicine applications. IEEE Transactions on Information Technologies in Biomedicine. 2001 Sep;5(3):210-24.
[6] Morgan RH. Computer network security for the radiology enterprise Computer network. Radiology. 2001 Aug;220(2):303-09.
[7] Furness PN, Bamford WM. Tele-pathology. Curr Diag Pathol. 2001 Dec;7:281-291.
[8] Coiera E, editor. The Internet and Telemedicine. In: Coiera E, editor. Guide to Medical Informatics. 2nd ed. New York: CHM; c2007. p.408.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210923

Use of Artificial Intelligence for Predicting COVID-19 Outcomes: A Scoping Review

Jinyan LYU[a,1], Wanting CUI[a] and Joseph FINKELSTEIN[a]
[a]*Icahn School of Medicine at Mount Sinai, New York NY, USA*

Abstract. During the COVID-19 pandemic, artificial intelligence has played an essential role in healthcare analytics. Scoping reviews have been shown to be instrumental for analyzing recent trends in specific research areas. This paper aimed at applying the scoping review methodology to analyze the papers that used artificial intelligence (AI) models to forecast COVID-19 outcomes. From the initial 1,057 articles on COVID-19, 19 articles satisfied inclusion/exclusion criteria. We found that the tree-based models were the most frequently used for extracting information from COVID-19 datasets. 25% of the papers used time series to transform and analyze their data. The largest number of articles were from the United States and China. The reviewed artificial intelligence methods were able to predict cases, death, mortality, and severity. AI tools can serve as powerful means for building predictive analytics during pandemics.

Keywords. Scoping Review, Artificial Intelligence, COVID-19

1. Introduction

In December 2019, the first COVID-19 case was reported in Wuhan, China. Millions of people have been infected with the virus, and billions of people are affected by the pandemic at work and in life [9]. The impact of this pandemic has lasted longer than many people expected. There are broad areas of research dedicated to combating the COVID-19 pandemic. Artificial intelligence (AI) and traditional mathematical prediction models are valuable tools to forecast the spread of disease and mortality rates.

In this paper, we collected all papers with artificial intelligence models that forecast outcomes. The goal of this paper was to perform a scoping review to explore papers that used AI to predict outcomes of the COVID-19 pandemic.

2. Method

We conducted the scoping review using the Arksey and O'Malley framework to identify papers that used artificial intelligence methods to forecast patients diagnosed with COVID-19. We performed a five-step scoping review: identify the research question, identify relevant studies, study selection, chart the data, and collect, summarize, and report the results [2,11]. We searched PubMed, the MEDLINE database, for published and preprint literature with full text between March 12, 2020, and May 24, 2021. There

[1] Corresponding Author, Jinyan Lyu, Icahn School of Medicine at Mount Sinai, 1770 Madison Ave, New York, NY, USA, 10035; E-mail: jinyan.lyu@mssm.edu.

were 1,057 papers in total. Moreover, we researched selected papers' reference lists. Five different search terms were used: forecasting, predicting, artificial intelligence, machine learning, and COVID-19.

Four inclusion criteria were proposed. The paper should use at least one artificial intelligence method to predict. The predictors are COVID-19 case count, mortality, or severity. The prediction must include at least one predictor, and it can be more than one predictor. The data should be identified data, de-identified data, or synonym data. Furthermore, three exclusion criteria were applied. If the data only comes from social media; if the data is not patient-related data; if a paper only proposed new approach with no modeling application.

Article selection made in two steps, first screening title and abstract, second, screening full text. However, when we searched related papers in MEDLINE, A relatively large result was received. And as the search engine would post all the results, including reference or paper suggestions, on that web page, all irrelevant results were deleted. Then, the two authors, Cui and Lyu, screened all abstracts and titles independently, which eliminated papers that did not use forecasting or prediction models. Secondly, they independently reviewed the remaining papers' full texts. Cohen Kappa tests had been conducted to examine the similarity between two reviewers' screening results, the abstract result was 0.86, and the full-text result was 1.

The data extraction form included an article's published date, dataset name and type, country, the start date of the dataset, the end date of the dataset, medical condition of patients, predictor, AI method, and other methods that were used. We focused on AI methods for prediction. AI is defined as a computational method that learned from current data and then given appropriate outputs.

3. Results

There were 1,057 research articles were found from MEDLINE. After removing duplicate results, irrelevant papers, pre-prints and reviews, there were 197 papers remained. By screening the title and abstract, 58 papers were selected, and then full-texts were reviewed. Based on our selection criteria (**Figure 1**), 17 papers were qualified for further analysis. The first AI forecasting paper was published in May 2020 [5], and most articles were published between Aug 2020 and Apr 2021. One paper could use more than one dataset [14,16,17], and also those datasets might have more than one country's data. Thus, there are 31 countries mention in the dataset. Also, datasets from China and the United States were the most popular. Datasets from Asia were the most analyzed among papers.

The purpose of using the artificial intelligence method was to predict cases, death, mortality, and severity. Since one paper could predict more than one variable, among the 17 papers, seven papers forecasted mortality [1,7,11,13-15,19], five papers forecasted daily report COVID-19 test-positive cases and death cases [4-6,12,16], four papers forecasted severity [3,9,17,18], and one paper predicted all of them [20]. Tree-based model was the most common choice, which was used seven times [9,11,12,13,15,16,19]. The neural network method was a new trend, and five papers applied this method [3,4,5,7,20]. Three papers used the regression method [1,6,17]. In addition, two papers used the deep learning method [14,18]. Furthermore, due to the time sensitivity of COVID-19, four papers used time series analysis first, and then applied the artificial

intelligence model [7,12,14,16]. ARIMA model was the most common choice for time series analysis [7,12].

To predict new confirmed or death cases, authors intended to use the r squared and RMSE to check the model fitness. Meanwhile, forecasting mortality and severity, authors would use the AUC-ROC curve and accuracy to test model performance. In the model fitness part, three articles used the r squared as a standard. The statistical neural network model and the long short-term memory networks model could predict COVID-19 confirmed cases and death cases, with the maximum r square score of 0.99. Moreover, the r squared of using artificial neural network model predict death cases was 0.92. To test the model performance, seven papers used the AUC score, and three papers used accuracy. The best AUC score of predicted mortality was 0.963 using Lasso model. The highest AUC score of severity prediction using random forest was 0.93. The highest accuracy rate was 98% on the three days mortality prediction using the time-aware long short-term memory neural network.

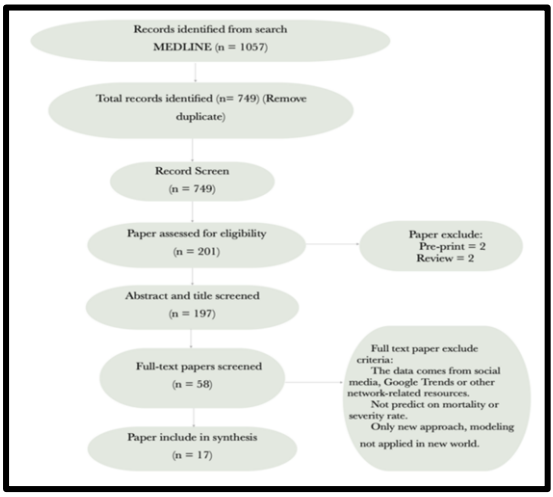


Figure 1. Paper selection process

4. Discussion

The scoping review is a good method to address undefined and broad questions. The majority of papers predicted mortality. In the study of mortality, neural network method was the most popular mortality prediction method. And the Lasso model generated the highest AUC score. Four papers predicted severity, three of which used random forest model, and this model had the best model performance. Five papers predicted daily confirmed cases and death cases. Neural network and random forest were the equally most common used models. Statistical neural network under this case had better fitness. Overall, the random forest model was the most popular model, and the Lasso model always tended to have a better performance. Due to the differences in datasets, it was difficult to compare models across datasets. Besides, we only chosen papers included AI models in forecasting cases, mortality, and severity. This topic was relatively narrow. In the future study, we will design a method to measure model performance and also extend our topics to other predictors that could be predicted by AI.

5. Conclusion

We conducted a scoping review of AI methods that were used to predict COVID-19 cases. We found that the tree-based model was the universal method. And most papers used official resources or EHR datasets. Asia was the most analyzed continent. Thus, scoping review is a valid method to explore the current trends of the use of AI in predicting COVID-19. Future study is warranted.

References

[1] An C, Lim H, et al. Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study, *Scientific reports* vol (2020), 10,1 18716. 30.
[2] Arksey H, O'Malley L, Scoping studies: towards a methodological framework, *Int J Soc Res Methodol* (2005), 8(1):19-32.
[3] Assaf D, Gutman Y, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19, *Intern Emerg Med* (2020),15(8):1435-1443.
[4] Braga MB, Fernandes RDS, et al, Artificial neural networks for short-term forecasting of cases, deaths, and hospital beds occupancy in the COVID-19 pandemic at the Brazilian Amazon, *PLoS One* (2021), 16(3):e0248161.
[5] Chatterjee A, Gerdes MW, Martinez SG. Statistical Explorations and Univariate Timeseries Analysis on COVID-19 Datasets to Understand the Trend of Disease Spreading and Death, *Sensors (Basel)* (2020), 20(11):3089.
[6] da Silva CC, de Lima CL, et al. Covid-19 Dynamic Monitoring and Real-Time Spatio-Temporal Forecasting. *Front Public Health* (2021), 9:641253.
[7] Dhamodharavadhani S, Rathipriya R, Chatterjee JM. COVID-19 Mortality Rate Prediction for India Using Statistical Neural Network Models, *Front Public Health* (2020), 8:441.
[8] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time, *Lancet Infect Dis* (2020), https://doi.org/10.1016/S1473-3099(20)30120-1.
[9] Heldt FS, Vizcaychipi MP, et al. Early risk assessment for COVID-19 patients from emergency department data using machine learning, *Sci Rep* (2021), 11(1):4200.
[10] Pham MT, Rajić A, et al. A scoping review of scoping reviews: advancing the approach and enhancing the consistency, *Res Synth Methods* (2014), 5(4):371-385.
[11] Rechtman E, Curtin P, et al. Vital signs assessed in initial clinical encounters predict COVID-19 mortality in an NYC hospital system, *Sci Rep* (2020), 10(1):21545.
[12] Saba T, Abunadi I, Shahzad MN, Khan AR. Machine learning techniques to detect and forecast the daily total COVID-19 infected and deaths cases under different lockdown types, *Microsc Res Tech* (2021), 84(7):1462-1474.
[13] Stachel A, Daniel K, et al. Development and validation of a machine learning model to predict mortality risk in patients with COVID-19, *BMJ Health Care Inform* (2021), 28(1):e100235.
[14] Sun C, Hong S, Song M, Li H, Wang Z. Predicting COVID-19 disease progression and patient outcomes based on temporal deep learning, *BMC Med Inform Decis Mak* (2021), 21(1):45.
[15] Vaid A, Somani S, et al. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation, *J Med Internet Res* (2020), 22(11):e24018.
[16] Watson GL, Xiong D, et al. Pandemic velocity: Forecasting COVID-19 in the US with a machine learning & Bayesian time series compartmental model, *PLoS Comput Biol* (2021), 17(3):e1008837.
[17] Wu G, Yang P, et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study, *Eur Respir J* (2020), 56(2):2001104.
[18] Yaşar Ş, Çolak C, Yoloğlu S. Artificial Intelligence-Based Prediction of Covid-19 Severity on the Results of Protein Profiling, *Comput Methods Programs Biomed* (2021), 202:105996.
[19] Yu L, Halalau A, et al. Machine learning methods to predict mechanical ventilation and mortality in patients with COVID-19, *PLoS One* (2021), 16(4):e0249285.
[20] Zawbaa HM, El-Gendy A, et al. A study of the possible factors affecting COVID-19 spread, severity and mortality and the effect of social distancing on these factors: Machine learning forecasting model, *Int J Clin Pract* (2021), 75(6):e14116.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210924

A Conceptual Framework to Predict Mental Health Patients' Zoning Classification

Sanjib Raj PANDEY[a], Alan SMITH[b], Edmund Nigel GALL[a], Ajay BHATNAGAR[c] and Thierry CHAUSSALET[a,1]
[a] *School of Computer Science & Engineering, University of Westminster, London, UK*
[b] *Business Intelligence Systems Team, Oxleas NHS Foundation Trust, London, UK*
[c] *Green Parks House (Inpatient Services), Oxleas NHS Foundation Trust, London, UK*

Abstract. Zoning classification is a rating mechanism, which uses a three-tier color coding to indicate perceived risk from the patients' conditions. It is a widely adopted manual system used across mental health settings, however it is time consuming and costly. We propose to automate classification, by adopting a hybrid approach, which combines Temporal Abstraction to capture the temporal relationship between symptoms and patients' behaviors, Natural Language Processing to quantify statistical information from patient notes, and Supervised Machine Learning Models to make a final prediction of zoning classification for mental health patients.

Keywords. mental health, zoning, temporal logic, natural language processing, machine learning

1. Introduction

Mental health (MH) is the biggest cause of disability in the UK: its estimated economic and social cost of £105.2 billion annually contributes up to 22.8% of the total expenditure [1]. Identifying, assessing, and mitigating mental health risk factors holds the potential to reduce the cost burden by correctly proportioning interventions according to the assessed risk. UK mental health trusts (MHTs) use a mostly evidenced-based manual process for zoning classification, a clinical assessment of risk factors to assign patients to zones using established zoning criteria [2]. This provides essential and pragmatic support to the Health Care Professionals (HCPs) who are required to regularly assess, implement, and evaluate the treatment plans of the MH patients in their care. In this study we introduce an automated framework to facilitate the design and development of an evidence-based decision support tool to predict and inform MH zoning classification.

Patients are diagnosed as suffering from one or more MH conditions as described by the International Classification of Diseases Version 10 (ICD-10) [3]. Within each ICD-10 category a patient may experience a series of symptoms and behaviors, each with defined start and end time points (episodes), of varying intensities that describe the severity of the impact the episode has on the patient's wellbeing. By combining episode profile of ICD-10 descriptors, intensity, and environmental triggers, a time-based overall descriptor of patient acuity can be derived. The zoning process aims to classify the patient

[1] Corresponding Author, Thierry Chaussalet, School of Computer Science & Engineering, University of Westminster, 115 New Cavendish St, London, W1W 6UW, UK; E-mail: chausst@westminster.ac.uk.

into one of three risk levels (Red: High; Amber: Medium; and Green: Low), based on a set of guidelines that map changing patient acuity to changing risks, during regular meetings, where discussions of patient condition relative to risk ensure that appropriate care management plans are in place. However, issues were identified regarding the current zoning approach. First, there is no real time reflection of caseload and patients' risk profiles. Second, manual assessments are resource intensive. Third, subjective assessment can result in inconsistent ratings. Fourth, manual practices impede on optimal sharing of this key information with all staff.

There are various approaches to extract clinical data for automated predictions from Electronic Health Records (EHRs). A hybrid approach comprising supervised Machine Learning (ML), rule-based annotation, and unsupervised natural language processing (NLP) for identifying disorder mentions achieved a 0.776 F-score [4]. Zhang et al. [5] proposed a system incorporating temporal modelling, and ML approaches for an early anxiety/depression detection framework. Neural networks and NLP were applied to predict the presence/absence of MH conditions using clinical notes [6]. NLP was used to capture severe mental illness symptoms from clinical text within discharge summaries [7]. Clinical data extraction divides into two categories: (a) ML and rule-based approaches, and (b) information extraction applications are both discussed in [8].

This research will investigate techniques required to develop a tool and assess the feasibility of its use within an NHS MHT, as a replacement/augmentation of the current system. The tool should have flexibility to allow Trusts to tailor it to their requirements.

2. Methodology

We propose a hybrid approach to build an automated decision support tool that will classify patient zoning stages, comprising a Temporal Model, an NLP Model and Supervised ML Model, as shown in Figure 1. Each model's specific responsibilities and functionalities are discussed below.

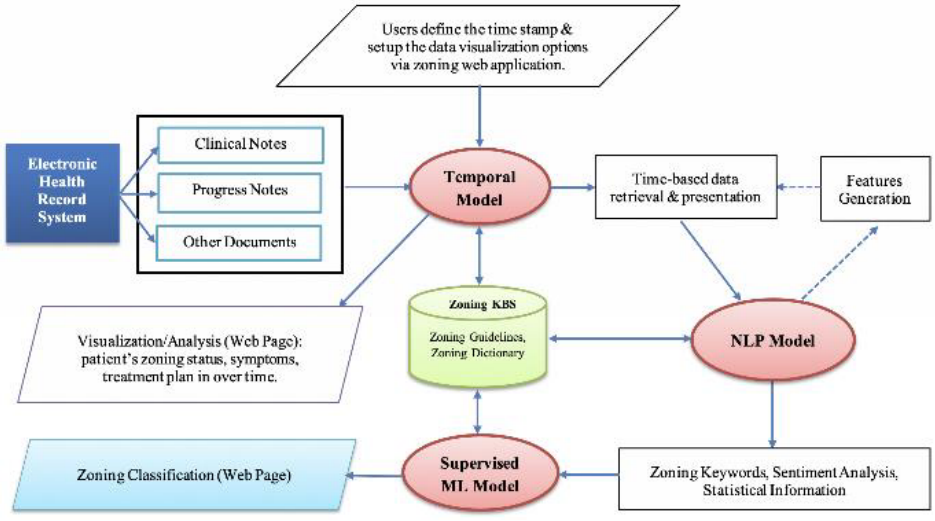


Figure 1. Hybrid approach for automated decision support tool to predict and inform zoning classification.

Both absolute time (the date and time of an event) and relative time (the intervals between events), respectively known as Point-based and Interval-based temporal approaches, are key elements of clinical data management. In this framework we propose using point-based temporal primitives, which are good for computation and more appropriate for the clinical reasoning process, to derive the Temporal Model, a temporal abstraction and representation of zoning classification. This shall feed the NLP Model, as shown in Figure 2. The NLP Model performs the following tasks: keywords extraction based on the Zoning Knowledge-based System (KBS); sentiment analysis (e.g., with the contextual window); and calculates statistical information about the zoning keywords.

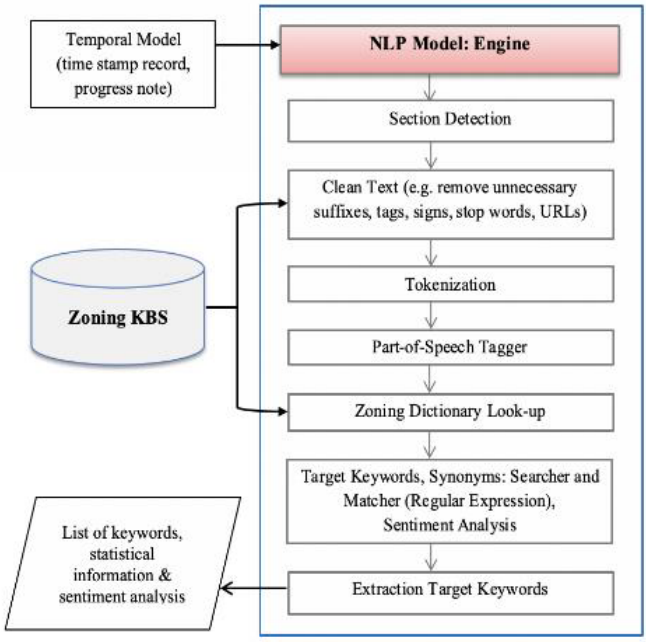


Figure 2. The NLP Model Process.

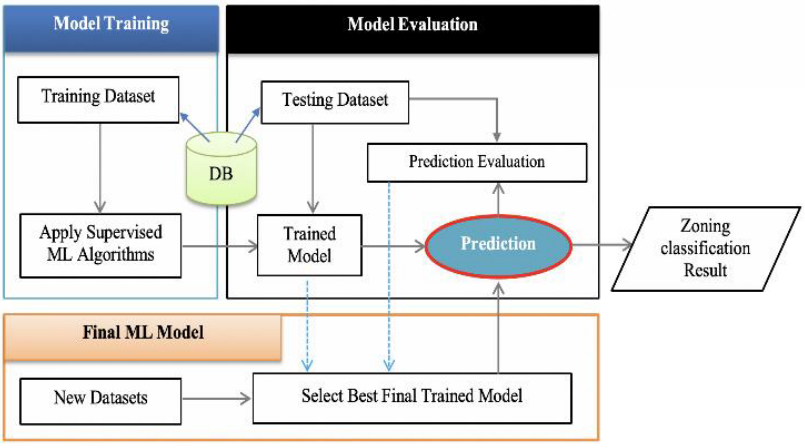


Figure 3. The Supervised ML Model Process.

The Supervised ML Model, shown in Figure 3, reads the target keywords, sentiment analysis and statistical information from the NLP Model, generates required ML and prediction datasets, produces the final tuned ML model, and predicts zoning classifications. We will use several supervised ML algorithms, as have already been applied in medical information applications, e.g. logistic regression, decision tree, support vector machine, neural network, and random forest [9]. The output of each algorithm will be evaluated; one algorithm will be selected for the final development based on the evaluated results.

3. Conclusion and Future Work

In this paper we introduced a new conceptual framework using a hybrid approach, combining Temporal, Natural Language Processing, and Supervised Machine Learning Models for patient zoning classification in mental health. Ultimately, we will develop a real-time artificial intelligence-based zoning classification tool for treatment of mental health patients, in collaboration with domain users (e.g., doctors, nurses, other HCPs and ex-patients) to ensure the project meets their needs whilst applying our proposed hybrid approach.

References

[1] HMG/DH. No health without mental health: a cross-government mental health outcomes strategy for people of all ages [Internet]. London: Department of Health; 2011 Feb [cited 2021 Feb 11]. 103 p. Available from: https://www.gov.uk/government/publications/no-health-without-mental-health-a-cross-government-outcomes-strategy.

[2] Croucher S, Williamson GR. Risk assessment in mental health: Introducing a traffic light system in a community mental health team. The Open Nursing Journal. 2013;7:82–8. doi:10.2174/1874434620130529004.

[3] World Health Organization, editor. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines [Internet]. Geneva: World Health Organization; 1992 [cited 2021 Jan 12]. 362 p. Available from: https://apps.who.int/iris/handle/10665/37958

[4] Wang C, Akella R. A Hybrid Approach to Extracting Disorder Mentions from Clinical Notes. AMIA Jt Summits Transl Sci Proc. 2015 Mar 25 [cited 2021 Mar 4];2015:183–7. Cited in PubMed; PMID 26306265.

[5] Zhang J, Xiong H, Huang Y, Wu H, Leach K, Barnes LE. M-SEQ: Early detection of anxiety and depression via temporal orders of diagnoses in electronic health data. In: 2015 IEEE International Conference on Big Data (Big Data). 2015;2569–77. doi:10.1109/BigData.2015.7364054.

[6] Tran T, Kavuluru R. Predicting mental conditions based on "history of present illness" in psychiatric notes with deep neural networks. Journal of Biomedical Informatics. 2017 Nov 1;75:S138–48. doi:10.1016/j.jbi.2017.06.010.

[7] Jackson RG, Patel R, Jayatilleke N, Kolliakou A, Ball M, Gorrell G, Roberts A, Dobson RJ, Stewart R. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. BMJ open. 2017 Jan 1;7(1):e012012.

[8] Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, Liu S, Zeng Y, Mehrabi S, Sohn S, Liu H. Clinical information extraction applications: a literature review. Journal of biomedical informatics. 2018 Jan 1;77:34-49.

[9] Pandey SR, Ma J, Lai CH, Regmi PR. A supervised machine learning approach to generate the auto rule for clinical decision support system. Trends in Medicine. 2020 May 22;20(3):1–9. doi:10.15761/TiM.1000232.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210925

Stroke Units Necessity for Patients, Web-Based "SUN4P" Registry: Descriptive Characteristics of the Population

Georgios MAVRAGANIS[a,1,], Eleni KOROMPOKI[a,b], Evangelos TSAMPALAS[c], Dafni GAREFOU[c], Helen ALEXOPOULOU[c], Maria LYPIRIDOU[c], Ioannis KALLIONTZAKIS[d], Anastasia FRAGKOULAKI[d], Aspasia KOURIDAKI[d], Argyro TOUNTOPOULOU[e], Ioanna KOUZI[e], Sofia VASSILOPOULOU[e], Efstathia KARAGKIOZI[f], Anna-Maria LOUKA[f], Efstathios MANIOS[a], Anastasia VEMMOU[a], Christos SAVOPOULOS[g], Gregorios DIMAS[g], Athina MYROU[g], Haralampos MILIONIS[h], Georgios SIOPIS[h], Hara EVAGGELOU[h], Athanasios PROTOGEROU[i], Stamatina SAMARA[i], Asteria KARAPIPERI[i], Nikolaos KAKALETSIS[j], Iliana KARAGKOUNI[j], Olympia KONSTANTAKOPOULOU[k], Petros GALANIS[k], Daphne KAITELIDOU[k], Stefanos PAPASTEFANATOS[k], Kostas VEMMOS[b,*], George NTAIOS[b,f,*] and Olga SISKOU[k,*]

**equal last authors*

[a] *Department of Clinical Therapeutics, Alexandra Hospital, National and Kapodistrian University of Athens, Athens, Greece*
[b] *Hellenic Stroke Organization, Greece*
[c] *Department of Neurology, Panarkadikon General Hospital, Tripoli, Greece*
[d] *Department of Neurology, General Hospital of Chania, Creta, Greece*
[e] *1st Department of Neurology, Eginition Hospital, National and Kapodistrian University of Athens, Athens, Greece*
[f] *Department of Internal Medicine, Faculty of Medicine, School of Health Sciences, University of Thessaly, Larissa, Greece*
[g] *1st Medical Propedeutic Department of Internal Medicine, Aristotle University of Thessaloniki, AHEPA Hospital, Thessaloniki, Greece*
[h] *Department of Internal Medicine, School of Medicine, University of Ioannina, Ioannina, Greece*
[i] *Cardiovascular Prevention & Research Unit in Laiko General Hospital of Athens at the Medical School of the National & Kapodistrian University of Athens, Greece*
[j] *Second Propedeutic Department of Internal Medicine, Aristotle University of Thessaloniki, Hippokrateion General Hospital of Thessaloniki, Greece*
[k] *Center for Health Services Management and Evaluation, Department of Nursing National and Kapodistrian University of Athens, Athens, Greece*

Abstract. The aim of this study was to present the descriptive characteristics of the Stroke Units Necessity for Patients (SUN4P) registry. **Methods:** The study population derived from the web-based SUN4P registry included 823 patients with first-ever acute stroke. Descriptive statistics were used to present patients' characteristics. **Results:** The vast majority of patients (80.4%) had an ischemic stroke, whereas 15.4% had a hemorrhagic stroke. Hypertension was the leading risk

[1] Corresponding author, Georgios Mavraganis; E-mail: giomavraganis@gmail.com.

factor in both patients. The patients with ischemic stroke had higher prevalence of traditional cardiovascular risk factors such as diabetes mellitus, dyslipidemia and smoking and most commonly cryptogenic stroke (39%). National Institutes of Health Stroke Scale (NIHSS) was higher among patients with hemorrhagic in comparison to those with ischemic stroke (10.5 vs 6 respectively). Moreover, all patients had similar rate of disability prior to stroke, as shown by Modified Rankin Scale (mRS=0). **Conclusions:** These data are in accordance with current evidence and should be thoroughly assessed in order to ensure optimal therapeutic management of stroke patients.

Keywords. Stroke, Organization of acute stroke care, Stroke care, Disease management

1. Introduction

Stroke is one of the most prominent causes of death and the leading cause of disability in adults worldwide [1]. Aging of the population and projected trends in risk factors are expected to heighten the prevalence of stroke as the latter increases with advancing age in both genders [2]. Given the high rate of mortality and disability in patients suffering from stroke, there is emerging need to optimize current medical treatment, ameliorate the quality of care and identify preventable risk factors that could significantly mitigate the burden of stroke on healthcare systems. The aim of Stroke Units Necessity for Patients (SUN4P) study is to provide essential data on acute stroke management that will contribute to achievement of the aforementioned issues. Herein, we present the descriptive characteristics of the population of SUN4P registry.

2. Methods

The study population derived from the web-based "Improving Stroke Care in Greece in Terms of Management, Costs and Health Outcomes- (SUN4P)" registry (ClinicalTrials.gov, NCT04109612). SUN4P is an ongoing observational non-invasive prospective cohort multicenter study of patients with first-ever acute stroke, hemorrhagic and ischemic, admitted to seven public hospitals [3]. Descriptive statistics were used in order to present patients' characteristics. The SUN4P design was based on the European General Data Protection Regulation (GDPR) and aligned with the Declaration of Helsinki.

3. Results

The study population consisted of 823 patients admitted from July 2019 to July 2021. The vast majority of patients (80.4%) had an ischemic stroke, whereas 15.4% had a hemorrhagic stroke, as demonstrated in Table 1. Hypertension was the leading risk factor in both patients with ischemic (67%) and hemorrhagic stroke (66%). Of note, the patients with ischemic stroke had a higher prevalence of other traditional cardiovascular risk factors such as diabetes mellitus (27% vs 17%), dyslipidemia (37% vs 28%) and smoking (26% vs 20%) in comparison with patients hospitalized for hemorrhagic stroke. Furthermore, in these patients, cryptogenic stroke (39%) followed by cardioembolic stroke except for patent foramen ovale (29%) were the most frequent. Notably,

prevalence of atrial fibrillation was comparable between patients with ischemic and hemorrhagic stroke, being present at almost one fourth of subjects (25% vs 27% respectively). Moreover, all patients had similar rate of presence of symptoms at their daily activities prior to stroke, as shown by Modified Rankin Scale (mRS) (median mRS= 0 in both patients with ischemic and hemorrhagic stroke). Nonetheless, patients with hemorrhagic stroke were substantially more impaired than those with ischemic stroke in the acute phase given the higher NIHSS (median NIHSS=10.5 vs 6 in patients with hemorrhagic and ischemic stroke respectively). Baseline demographic and clinical characteristics of the study's population are shown in detail in Table 1.

Table 1. Descriptive characteristics of the SUN4P population (n=823)

	Ischemic stroke (n=662)	**Hemorrhagic stroke (n=127)**	**Non-classified (n=34)**
Gender (male)	328 (50%)	73 (57%)	20 (59%)
Age (mean± SD)	75.2 (13.7)	75.6 (13.2)	80.3 (10.2)
Self-employed	571 (86%)	113 (89%)	19 (56%)
mRS prior to admission	0 (0-1)	0 (0-1)	0 (0-1)
NIHSS at admission	6 (3-10)	10.5 (4-22)	4 (2.5-10.5)
Body mass index (kg/m^2)	27.64 (5.0)	27.86 (5.4)	27.0 (5.4)
History of:			
Hypertension	443 (67%)	84 (66%)	17 (50%)
Diabetes mellitus	178 (27%)	21 (17%)	5 (15%)
Dyslipidemia	248 (37%)	35 (28%)	13 (38%)
Smoking	174 (26%)	25 (20%)	2 (6%)
Coronary artery disease	81 (12%)	16 (13%)	5 (15%)
Atrial fibrillation	164 (25%)	34 (27%)	15 (44%)
Valve replacement	21 (3%)	3 (2%)	11 (32%)
Heart failure	47 (7%)	6 (5%)	2 (6%)
Symptomatic PAD	16 (2%)	1 (1%)	2 (6%)
Active malignancy	35 (5%)	4 (3%)	1 (3%)
Alcohol addiction	55 (8%)	13 (10%)	1 (3%)
Carotid artery disease	16 (2%)	3 (2%)	4 (12%)
Transient ischemic attack	57 (9%)	15 (12%)	2 (6%)
Modified Ischemic TOAST classification			
Large-artery Atherosclerosis (≥50%)	19%		
Cardioembolic (except for patent foramen ovale)	29%		
Lacunar stroke	11%		
Other rare conditions	3%		
Cryptogenic stroke	*39%*		
ESUS	9%		
Multiple causes	6%		
Incomplete investigation	24%		

Notes: Continuous variables are presented as mean (SD) and nominal as count (absolute percentages). For NIHSS and mRS, median values with interquartile ranges are provided.

**Percentages do not add up to 100% as in case of some patients multiple reasons existed. Abbreviations: mRS, Modified Rankin Scale; PAD, peripheral artery disease; SD, standard deviation; NIHSS, National Institutes of Health Stroke Scale; ESUS, embolic stroke of undetermined source.*

4. Discussion

Preliminary results of SUN4P cohort, the first multicenter stroke registry in Greece, demonstrate that patients presented with acute ischemic and hemorrhagic stroke in a similar rate to existing data from other registries [2]. Moreover, stroke prevalence was almost equal between male and female patients, in accordance with current evidence

corroborating that the disparities in prevalence of stroke among younger patients tend to diminish at oldest age groups, especially at patients over 75 years old [4]. Among patients with ischemic stroke, cryptogenic and cardioembolic stroke were the most prevalent, congruent with current literature [5]. Notably, in our study the NIHSS at admission was markedly higher among patients with hemorrhagic stroke; this comes in accordance with previous data reporting higher relative frequency of hemorrhagic stroke with increasing stroke severity [6].

5. Conclusions

In conclusion, the present study demonstrates that our data are in accordance with existing evidence, with a significant proportion of SUN4P patients presenting with ischemic stroke. These patients had higher rate of traditional cardiovascular risk factors, whereas patients with hemorrhagic stroke had more severe clinical presentation on admission. Given the prevalence of traditional cardiovascular risk factors at our registry, it can be concluded that a significant proportion of strokes could have been prevented, highlighting the current gaps in primary prevention of these patients. These data should be thoroughly assessed in order to ensure high value care and optimal therapeutic management during not only the acute phase of stroke patients but also for effective secondary prevention.

Acknowledgments

This project has received funding from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under grant agreement Nb538.

References

[1] Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, et al. Executive summary: heart disease and stroke statistics--2014 update: a report from the American Heart Association. Circulation. 2014;129(3):399-410. Epub 2014/01/22. doi: 10.1161/01.cir.0000442015.53336.12. PubMed PMID: 24446411.

[2] Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, et al. Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. Circulation. 2020;141(9):e139-e596. Epub 2020/01/30. doi: 10.1161/cir.0000000000000757. PubMed PMID: 31992061.

[3] Improving Stroke Care in Greece in Terms of Management, Costs and Health Outcomes- SUN4P. . Available at: http://wwwsun4patientsgr/portal/en/homeaspx (18/05/2020)

[4] Reeves MJ, Bushnell CD, Howard G, Gargano JW, Duncan PW, Lynch G, et al. Sex differences in stroke: epidemiology, clinical presentation, medical care, and outcomes. Lancet Neurol. 2008;7(10):915-26. Epub 2008/08/30. doi: 10.1016/s1474-4422(08)70193-5. PubMed PMID: 18722812; PubMed Central PMCID: PMCPMC2665267.

[5] Kolominsky-Rabas PL, Weber M, Gefeller O, Neundoerfer B, Heuschmann PU. Epidemiology of ischemic stroke subtypes according to TOAST criteria: incidence, recurrence, and long-term survival in ischemic stroke subtypes: a population-based study. Stroke. 2001;32(12):2735-40. Epub 2001/12/12. doi: 10.1161/hs1201.100209. PubMed PMID: 11739965.

[6] Andersen KK, Olsen TS, Dehlendorff C, Kammersgaard LP. Hemorrhagic and ischemic strokes compared: stroke severity, mortality, and risk factors. Stroke. 2009;40(6):2068-72. Epub 2009/04/11. doi: 10.1161/strokeaha.108.540112. PubMed PMID: 19359645.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210926

Machine Learning, Clinical Notes and Knowledge Graphs for Early Prediction of Acute Kidney Injury in the Intensive Care

Iacopo VAGLIANO[a,1], Wei-Hsiang HSU[a] and Martijn C. SCHUT[a]
[a] *Dept. of Medical Informatics, Amsterdam UMC, Location AMC, The Netherlands*

Abstract. Acute kidney injury (AKI) is an abrupt decrease of kidney function which is common in the intensive care. Many AKI prediction models have been proposed, but an analysis of what is the added value of clinical notes and medical terminologies has not yet been conducted. We developed and internally validated a model to predict AKI that includes not only clinical variables, but also clinical notes and medical terminologies. Our results were overall good (AUROC > 0.80). The best model used only clinical variables (AUROC 0.899).

Keywords. Acute kidney injury, clinical models, ICU, natural language processing

1. Introduction

Acute kidney injury (AKI) is an abrupt decrease of kidney function, with a prevalence of up to 50% in the intensive care unit (ICU). Early recognition of AKI is crucial as the efficacy of intervention greatly depend on it. Several AKI prediction models have been built using clinical variables, e.g., vitals and laboratory measurements [2]. Few studies used the rich information contained in clinical notes, possibly because it is not well known how to represent such information in AKI prediction models.

Knowledge graphs may enrich the representation of clinical notes in prediction models. A knowledge graph is a graph-based abstraction of knowledge to represent data from diverse sources. Medical terminologies, which are inherently complex, are often expressed as knowledge graphs, for example, the Unified Medical Language System (UMLS) [5] and SNOMED CT [6]. To the best of our knowledge, one study has used notes and knowledge graphs to predict AKI, but without using clinical variables [7]. So, a comparison of which type of information among clinical variables, clinical notes, and medical terminologies is most effective in AKI prediction modeling is not yet available.

We aim to investigate whether adding information extracted from clinical notes and knowledge graphs into machine-learning prediction models to predict AKI can improve predictive performance. We developed and internally validated a model to predict AKI within the first 48 hours of admission, which includes not only clinical variables, but also clinical notes to provide a comprehensive view of the patients' pathophysiologic condition. Furthermore, we aim to study what is the effect of enriching notes with external knowledge from UMLS or SNOMED CT on the models' performance.

[1] Corresponding Author, Iacopo Vagliano, Dept. of Medical Informatics, Amsterdam UMC, Location AMC, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands; E-mail: i.vagliano@amsterdamumc.nl.

2. Method

Data and population: We used data from the publicly available critical care database, Medical Information Mart for Intensive Care III (MIMIC-III). This database integrates comprehensive clinical data of patients admitted to an ICU at the Beth Israel Deaconess Medical Center in Boston, during 2001 to 2012 [8]. AKI was defined according to the KDIGO guidelines [9]. We included patients who aged 18 years or older at the time of ICU admission, have at least one measurement of serum creatinine or urine output, and whose length of stay in the ICU was at least 48 hours. To make sure patients have both clinical variables and notes available for our model, only ICU stays containing at least one note were retained. Data preprocessing is illustrated in the supplementary material.[2]
Model development: Our AKI prediction model is based on Long Short Term Memory (LSTM) networks [10]. First, two independent LSTM models were built separately, one uses clinical notes as input and the other uses clinical variables. Then, the output of these two models were concatenated and given as input to a final layer to deliver an overall prediction which combines information from both clinical variables and clinical notes (the latter, optionally enriched with external information from knowledge graphs).

Clinical notes were represented with Word2vec. To enrich notes with external knowledge from UMLS and SNOMED CT, we extracted three subgraphs, i.e., UMLS synonyms, SNOMED CT synonyms, and SNOMED CT parent-child relationships. We used retrofitting to generate a refined representation of clinical notes, using relational information from each subgraph to encourage linked words to have similar representations. More details are available in Section 1.1 of the supplementary material.[2]

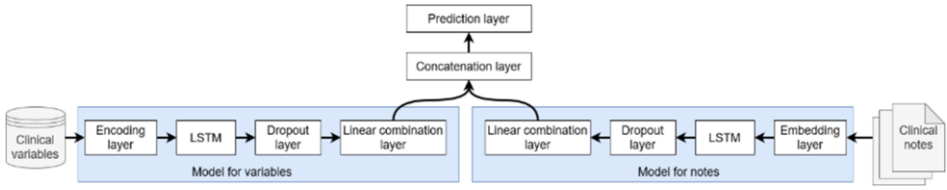


Figure 1. Architecture of the LSTM models used.

The architecture of the overall LSTM model is shown in Figure 1. We used the same architecture in the two LSTM models for variables and notes, except that the LSTM for clinical notes has an embedding layer, whilst the LSTM for clinical features has an encoding layer. The former layer acts as a lookup table to return the word embeddings learnt as parameters by the model during training, while the latter is a linear layer to compress the high-dimensional and sparse input variables into a lower-dimensional continuous representation, easier to manipulate by the model. The LSTM layers are followed by a dropout layer, to help prevent models from overfitting, and another linear layer, which creates a second learning filter for our model before concatenating the output of the two models, when relying on both clinical variables and clinical notes. The overall model ends with a projection layer that returns the probability of a patient having or developing AKI. The model parameters are shown in the supplementary material.[2]

Internal validation and performance measures: The dataset was randomly split into 80% training, 10% validation, and 10% test sets. We measured discrimination with the

[2] https://osf.io/thp53/, last access October 1, 2021.

area under the receiver operating curve (AUROC) and the area under the precision-recall curve (AUPRC); calibration with calibration curves.

3. Results

The final dataset consisted of 46,985 ICU stays of 33,795 unique patients. Descriptive statistics of the population are in Section 2.1 of the supplementary material.[2]

Table 1 outlines the discrimination (AUROC and AUPRC) of the models. Using only clinical variables achieved the best results. Exploiting clinical variables and notes performed better than only notes, and models with external knowledge achieved similar results to models without external knowledge. The calibration curves of the models are available in the supplementary material. Using clinical variables and notes retrofitted with parent-child relations yielded the best calibration, with a slight improvement on the models with only notes and with notes retrofitted with SNOMED CT synonyms.

Table 1. Models' discrimination with various inputs. We used six different input sets. Syn. stands for synonyms.

Input sets	Clinical variables	Clinical notes	Variables + notes	Variables + notes + UMLS syn.	Variables + notes + SNOMED syn.	Variables + notes + SNOMED parent-child
AUROC	**0.899**	0.801	0.821	0.821	0.819	0.816
AUPRC	**0.957**	0.898	0.910	0.910	0.909	0.908

4. Discussion

Our results shows overall good performances (AUROC > 0.80). The best-performing model (AUROC 0.90) used only clinical variables. Clinical notes retrofitted with parent-child relations used together with clinical variable yielded the best calibration.

When exploiting clinical notes, we used only 3,000 words to represent each patient, while the patients' notes included over 13,000 words on average. This might be the main reason why clinical notes did not improve the discrimination. Clinical notes may still be a valuable source of information on AKI prediction. We examined important words by ranking the most frequent words in the clinical notes (see the supplementary material[2]). In most cases, these top ranking words appear to be clinically meaningful. For example, *heparin* is a medication used to prevent blood clotting during kidney dialysis. *Edema*, which is a sign of AKI, happens when failing kidneys do not remove extra fluid which builds up in the patient's body causing swelling in the legs, ankles, feet, and/or hands. *Lasix* is one diuretic that can treat fluid retention and swelling, which might be caused by kidney dysfunction. With more advanced approaches to preprocess raw text from the notes, it is possible that structured clinical features and unstructured clinical notes exist as complementary sources of information for machine learning models to predict AKI.

Similar issues could explain why knowledge graphs did not affect the models' discrimination. For example in UMLS, only 7,093 out of 359,080 medical concepts (roughly 0.02%) are single-word. Clinical notes often refer to multi-words concepts, such as *end-stage chronic renal failure* or *end-stage kidney disease*. Since our word embeddings represent single words, multi-word concepts cannot be captured. The representation of multi-word concepts may improve models' performance.

Our study has some limitations. First, the MIMIC III database includes US patients; thus our results may not generalize to other populations. Second, we performed a simple

train/validation/test split, which does not take into account the variability of train/validation/test sets. Third, we did not perform full hyperparameter tuning but relied on a set of parameters for our models pre-selected in preliminary experiments.

As a strength, our study relies on a public dataset to encourage reproducibility and our code is available at bitbucket.org/aumc-kik/ml-cn-kg-4-aki-prediction.

Future work includes exploring more-advanced preprocessing of notes, ranging from using more words per patient to represents multi-word concepts, as well as study different models, such as convolutional neural networks. A prospective validation is also needed to assess clinical utility and effect on patient outcomes.

5. Conclusions

Early recognition of AKI is essential for effective treatment of this disease in the ICU. In contrast to previous work, we used various types of information, i.e., clinical variables, clinical notes and knowledge graphs, and we studied what is the added value of clinical notes and knowledge graphs to predictive performance. We provided effective models to predict AKI in the ICU. All the models achieved good results. The best discrimination was achieved by using only clinical variables, the best calibration with retrofitted clinical notes and clinical variables. Our work contributed to combining clinical variables, notes and knowledge graphs, which may also be useful in other settings and populations.

Acknowledgments

We thank Miguel Ríos Gaona for the help to extract clinical notes.

References

[1] Hoste EA, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive care medicine. 2015 Aug;41(8):1411-23.
[2] Park S, Lee H. Acute kidney injury prediction models: current concepts and future strategies. Current Opinion in Nephrology and Hypertension. 2019;28(6):552-559.
[3] Sun M, et.al. Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes and Structured Multivariate Physiological Measurements. MedInfo. 2019 Aug 21;264:368-72.
[4] Le S, et.al. Convolutional Neural Network Model for Intensive Care Unit Acute Kidney Injury Prediction. Kidney international reports. 2021 May 1;6(5):1289-98.
[5] Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research, 2004;32(suppl_1):D267-D270.
[6] Rogers J, Bodenreider O. SNOMED CT: Browsing the Browsers. In KR-MED, 2008:30-36.
[7] Li Y, et al. Early prediction of acute kidney injury in critical care setting using clinical notes. In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2018 Dec 3 (pp. 683-686). IEEE.
[8] Johnson AE, et al. MIMIC-III, a freely accessible critical care database. Scientific data. 2016 May 24;3(1):1-9.
[9] Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. Nephron Clin. Pract. 120. 2012;c179–c184.
[10] Hochreiter S, et al.. Long short-term memory. Neural Computation, 9(8):1735–1780.
[11] Mikolov T, et al. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems 2013 (pp. 3111-3119).
[12] Faruqui M, et al. Retrofitting word vectors to semantic lexicons. In North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 2015. (pp. 1606-1615). ACL

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210927

Contribution of Artificial Intelligence in Pregnancy: A Scoping Review

Israa ABUELEZZ[a], Asma HASSAN[a], Bassam Ali JABER[a], Mohamad SHARIQUE[a], Alaa ABD-ALRAZAQ[a], Mowafa HOUSEH[a], Tanvir ALAM[a] and Zubair SHAH[a,1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*

Abstract. For the past ten years, the healthcare sector and industry has witnessed a surge in Artificial Intelligence (AI) technologies being used in many different medical specialties. Recently, AI-driven technologies have been utilized in medical care for pregnancy. In this work, we present a scoping review that explores the features of AI-driven technologies used in caring for pregnant patients. This review was conducted using the Preferred Reporting Items for Systematic review and Meta-Analyses extension for Scoping Reviews. Our analysis revealed that AI techniques were used in predicting pregnancy disorders such as preeclampsia and gestational diabetes, along with managing and treating ectopic pregnancies. We also found that AI technologies were used to assess risk factors and safety surveillance of pregnant women. We believe that AI-driven technologies have the potential to improve the healthcare provided to pregnant women.

Keywords. Artificial intelligence, machine learning, pregnancy

1. Introduction

Mothers pass through much anxiety during their pre-and post-pregnancy periods [1]. During this challenging time, the mother's health should be managed and monitored for her and the fetus's safety. Pregnant women should seek well-designed prenatal care that considers a comprehensive view of the clinical data, regular laboratory tests, ultrasound images, and any other related data that can help practitioners in making the correct clinical decisions for a healthy delivery [1]. The data provided by the clinical readings during a women's pregnancy, such as ultrasound images, laboratory tests, and Electronic Health Records, present complex, contemporary challenges in analyzing these divergent data sources [2]. AI-based technologies have been touted as a possible aid in analyzing heterogeneous data sources. It can help in medical diagnostics and assist physicians in making informeds decision on treatment options for pregnant patients. Little is known about the state and use of AI technologies during pregnancy. To the best of our knowledge, there is no review in the current literature on AI techniques and their applications in pregnancy. This scoping review aims to explore features of AI-driven technologies used for pregnant women.

[1] Corresponding Author: Dr. Zubair Shah, Hamad Bin Khalifa University, College of Science and Engineering, Doha, Qatar; E-mail: ZShah@hbku.edu.qa.

2. Methods

We conducted this scoping review following the PRISMA-ScR guidelines [3]. All primary studies on the use of AI in pregnancy were reviewed. Articles that addressed the use of AI in childbirth and postpartum were excluded. We excluded all studies published in a language other than English. No restrictions were applied regarding the year or country of publication, and settings. Google Scholar and PubMed were used to retrieve the studies. The first 60 articles in Google Scholar were reviewed. Backward-reference list checking was also performed to retrieve additional relevant studies. The terms used to search the databases were related to the target intervention (e.g., artificial intelligence) and target health condition (e.g., pregnancy). The study selection process was based on three main phases: identification phase (i.e., removing duplicates), screening phase (i.e., screening titles and abstracts), and eligibility phase (i.e., reading full texts). The extracted data included the characteristics of the study (e.g., author, year, country, and publication type), features of the AI technique (e.g., branch, model, and model validation type), and features of the dataset used in the model (e.g., data source, data type, and dataset size). The study selection and data extraction were independently performed by authors AH and IA. Any disagreements between the two reviewers were resolved by discussion. The extracted data was then synthesized using a narrative approach.

3. Results

A total of 1,753 articles were retrieved from the two databases. Of these, we identified and removed 38 duplicates. In the screening phase, 1,595 articles were excluded after scanning their titles and abstracts. Of the remaining 120 articles, 96 were excluded after reviewing their full texts, as they reported on irrelevant interventions (e.g., using statistical tests rather than AI). A total number of 24 studies were included. The flowchart of the study selection process is shown in Appendix A.

More than one-third of the included studies were published in the USA (n=9, 37.5%). 22 studies were journal articles and 2 were conference proceedings. The majority of the studies (n=22) were published between 2016 and 2020. AI techniques were used for predicting pregnancy disorders/complications in about 75% (n=18) of the included studies [4-21]. Specifically, the techniques discussed were utilized for predicting preeclampsia [5, 7, 9, 15], preterm birth [6, 13, 19], gestational diabetes [8, 14, 21], gestational age [4, 18], patient's metabolomics profile [12, 20], suicidal behavior [11], uterine contractions [16], labor due date [17], and hypertensive disorder [10]. Additionally, of the 24 studies, four (n=4, 17%) employed AI techniques for treatment and management of ectopic pregnancies [22], gestational diabetes [23], late-onset preeclampsia [5], and hypertensive disorder [10]. Further, five studies (n= 5, 21%) used AI to assist with patients' safety outcome [13, 19, 24-26]. Specifically, AI was used to assess risk factors [13, 19, 26] and safety surveillance for pregnant women in online social media networks [24, 25].

Further, machine Learning (ML) techniques were used in most of the included studies (n=18, 75%). Linear Regression model, which outperformed than other ML models, was employed in six ML studies (n = 6, 33 %) [27 14, 15, 17, 18]]. 12 out of 18 studies (67%) used the ML models for prediction purposes [4, 5, 7-10, 14, 15, 18, 21, 23, 25]. In the second AI technique, deep learning (DL) -based techniques were implemented in eight studies (n=8, 33%) [6, 8, 9, 16, 19, 22, 24, 27]. Similar to the ML studies, DL

was used mainly for prediction purposes (n=7, 87%) [6, 8, 13, 16, 19, 22, 27]. Natural language processing techniques were implemented in one study [25] to understand how women seek information from a social network community. Clinical databases (i.e., database of health centers and hospitals) were the most commonly incorporated data source for model implementation and validation (n=13) [4-7, 10-12, 16, 17, 19, 22, 23, 26]. About 87.5% (n = 21) of the included studies reported the used model validation techniques. K-fold cross validation was the most commonly used technique (n=17) [4-8, 10, 12, 14-17, 19, 21, 22, 24, 26, 28].

4. Discussion

This review summarizes the contribution of AI in pregnancy. AI techniques were used in about 75% of the included studies for predicting pregnancy disorders/complications. We discovered a few issues that have not been addressed in the reviewed studies, which indicates a gap in the literature. First, did not find studies discussing the effectiveness of models used in treating disorders such as preeclampsia or gestational diabetes. Second, little information was available regarding public data sources for the AI models' training. As such, it is essential to prioritize monitoring hypertension and recommending new AI applications in the diagnosis/management of preeclampsia in future investigations. Further assessments are also needed to outline the validation approach and algorithms utilized by AI technology during pregnancy.

This study has a few limitations. We did not assess the effectiveness of AI models in predicting and diagnosing pregnancies. Therefore, we encourage future researchers to conduct systematic reviews that assess the effectiveness of AI models for the same. Moreover, we only focused on studies concerning the pregnant woman and pregnancy; we did not consider the childbirth or postpartum period. Therefore, we may have overlooked some other sub-domains within the field of pregnancy and childbirth. Only two databases were used to retrieve the literature whereas more databases may have been more comprehensive. Lastly, the search strategy was limited to studies written in English. Therefore, we may have missed AI and pregnancy-related studies published in a non-English language.

5. Conclusions

We conclude that ML and LR were identified as the most commonly utilized AI technique and models in the studies included in this review. AI-driven technologies may improve healthcare services provided to pregnant patients. It is recommended for future studies to assess the effectiveness of models used for the well-being of the pregnant woman. Appendix Files are available at GitHub: (https://github.com/AHassan2/AI-in-Pregnancy-Scoping-Review).

References

[1] Mischi M. et al, *Pregnancy monitoring*. 2014, Hindawi.
[2] Emin EI, et al, Artificial intelligence in obstetrics and gynaecology: is this the way forward? in vivo, 2019. 33(5): p. 1547-1551.

[3] Tricco, A.C., et al., *PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation.* Annals of Internal Medicine, 2018. **169**(7): p. 467-473.
[4] Liang, L., et al., *Metabolic dynamics and prediction of gestational age and time to delivery in pregnant women.* Cell, 2020. **181**(7): p. 1680-1692. e15.
[5] Jhee, J.H., et al., *Prediction model development of late-onset preeclampsia using machine learning-based methods.* PLoS One, 2019. **14**(8): p. e0221202.
[6] Gao, C., et al., *Deep learning predicts extreme preterm birth from electronic health records.* Journal of biomedical informatics, 2019. **100**: p. 103334.
[7] Marić, I., et al., *Early prediction of preeclampsia via machine learning.* American Journal of Obstetrics & Gynecology MFM, 2020. **2**(2): p. 100100.
[8] Shen, J., et al., *An Innovative Artificial Intelligence–Based App for the Diagnosis of Gestational Diabetes Mellitus (GDM-AI): Development Study.* Journal of Medical Internet Research, 2020. **22**(9): p. e21573.
[9] Nair, T.M., *Statistical and artificial neural network-based analysis to understand complexity and heterogeneity in preeclampsia.* Computational biology and chemistry, 2018. **75**: p. 222-230.
[10] 1Yang, L., et al., *Predictive models of hypertensive disorders in pregnancy based on support vector machine algorithm.* Technology and Health Care, 2020. **28**(S1): p. 181-186.
[11] Zhong, Q.-Y., et al., *Use of natural language processing in electronic medical records to identify pregnant women with suicidal behavior: towards a solution to the complex classification problem.* European journal of epidemiology, 2019. **34**(2): p. 153-162.
[12] Bahado‐Singh, R.O., et al., *Artificial intelligence and amniotic fluid multiomics: prediction of perinatal outcome in asymptomatic women with short cervix.* Ultrasound in Obstetrics & Gynecology, 2019. **54**(1): p. 110-118.
[13] Idowu, I.O., et al. *Artificial intelligence for detecting preterm uterine activity in gynecology and obstetric care*. in *2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing*. 2015. IEEE.
[14] White, S.L., et al., *Early antenatal prediction of gestational diabetes in obese women: development of prediction tools for targeted intervention.* PloS one, 2016. **11**(12): p. e0167846.
[15] Sufriyana, H., Y.-W. Wu, and E.C.-Y. Su, *Prediction of preeclampsia and intrauterine growth restriction: development of machine learning models on a prospective cohort.* JMIR medical informatics, 2020. **8**(5): p. e15411.
[16] Peng, J., et al., *Preliminary study on the efficient electrohysterogram segments for recognizing uterine contractions with convolutional neural networks.* BioMed research international, 2019. **2019**.
[17] Blobel, B. *Prediction of a Due Date Based on the Pregnancy History Data Using Machine Learning*. in *pHealth 2020: Proceedings of the 17th International Conference on Wearable Micro and Nano Technologies for Personalized Health*. 2020. IOS Press.
[18] Papageorghiou, A.T., et al., *Ultrasound‐based gestational‐age estimation in late pregnancy.* Ultrasound in Obstetrics & Gynecology, 2016. **48**(6): p. 719-726.
[19] Lee, K.-S. and K.H. Ahn, *Artificial neural network analysis of spontaneous preterm labor and birth and its major determinants.* Journal of Korean medical science, 2019. **34**(16).
[20] Delplancke, T.D., et al., *Analysis of sequential hair segments reflects changes in the metabolome across the trimesters of pregnancy.* Scientific reports, 2018. **8**(1): p. 1-12.
[21] Pustozerov, E., et al., *The role of glycemic index and glycemic load in the development of real-time postprandial glycemic response prediction models for patients with gestational diabetes.* Nutrients, 2020. **12**(2): p. 302.
[22] Fernández, A.D.R., D.R. Fernández, and M.T.P. Sánchez, *A decision support system for predicting the treatment of ectopic pregnancies.* International journal of medical informatics, 2019. **129**: p. 198-204.
[23] Werner, E.F., et al., *Screening for gestational diabetes mellitus: are the criteria proposed by the International Association of the Diabetes and Pregnancy Study Groups cost-effective?* Diabetes care, 2012. **35**(3): p. 529-535.
[24] Sarker, A., et al., *Discovering cohorts of pregnant women from social media for safety surveillance and analysis.* Journal of medical Internet research, 2017. **19**(10): p. e8164.
[25] Wexler, A., et al., *Pregnancy and health in the age of the Internet: A content analysis of online "birth club" forums.* PloS one, 2020. **15**(4): p. e0230947.
[26] Chu, R., et al., *Predicting the risk of adverse events in pregnant women with congenital heart disease.* Journal of the American Heart Association, 2020. **9**(14): p. e016371.
[27] Rigla, M., et al., Gestational diabetes management using smart mobile telemedicine. Journal of diabetes science and technology, 2018. 12(2): p. 260-264.
[28] Fiset, S., et al., *Prediction of spontaneous preterm birth among twin gestations using machine learning and texture analysis of cervical ultrasound images.* University of Toronto Medical Journal, 2019. **96**(1): p. 6-9.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210928

Introducing Gamification in eHealth Platforms for Promoting Wellbeing

Antonios PARDOS[a,b], Andreas MENYCHTAS[a,b] and Ilias MAGLOGIANNIS[b,1]
[a] *Bioassist S.A.*
[b] *Department of Digital Systems, University of Piraeus*

Abstract. Gamification techniques are adopted by IT systems and applications in order to facilitate their adoption and motivate users to take advantage of specific application features. The current work presents a modern approach for the effective implementation of gamification features in a prototype eHealth application which encourages the daily use of the application, endorses the users to continuously monitor their health and promotes a healthier lifestyle. The implementation of this approach is modular and flexible in order to be easily applied in any similar system and tailor the provided features for user activity monitoring, analysis, feedback, and interactivity, to the specific requirements of the different usage scenarios.

Keywords. Gamification, eHealth, mHealth, IT systems

1. Introduction

eHealth solutions are considered as a type of intervention that can reach large numbers of people however, the low usage of the system from the participants is a recognized barrier to their sustainability and long-term success [1]. Romeo et. al. [2] mentioned that apps used to monitor physical activity are most effective in the short term, and the engagement declines over time which in turn limits the intervention effectiveness. The "use of game mechanics in non-gaming contexts" [3] is the technique known as gamification and is adopted for a variety of purposes in different contexts. The advantages of adopting those strategies might vary depending on the users' characteristics [4] or the scope of the application [5], and are related to both the user experience and how the users tend to engage with the app.

The most common and, historically, the traditional form of a gamification approach [6] is the Points - Badges - Leaderboards (PBL) triptych. More specifically, the user is awarded with numerical rewards (points) each time he completes a specific action within the application. Different users are ranked in descending order based on the points they have accumulated creating "Leaderboards", a concept borrowed from gaming design. A widely used approach is the *"badges"*, which are given to users when they complete specific achievements. These achievements also demonstrate the level of familiarity of the user with the application and the change of his behavior to meet its purposes [6]. In this work we present an implementation of the badges approach, in the context of an eHealth application, as a fully flexible and modular IT solution which can be applied and

[1] Corresponding Author, Ilias Maglogiannis, Department of Digital Systems, University of Piraeus, 80, M. Karaoli & A. Dimitriou St., 18534 Piraeus, Greece; E-mail: imaglo@unipi.gr.

integrated in any similar system. The combination of eHealth systems with gamifications features has the advantage of motivating the users to change their daily behaviors in a healthier way. Physical activity boost is a major outcome that the users of eHealth applications can achieve through gamification, and in the proposed proof-of-concept scenario it is also combined with the continuous measurement of other biosignals and the daily use of the application. Technically, the approach focusses on two aspects, an advanced model for efficient data management and analysis, and also the multimodal mechanisms required to calculate the badges levels. Both are considered of major importance for the efficient and effortless integration into any eHealth system.

2. Implementation

The implementation of the approach has two main elements: a) the *modeling* of the gamification aspects in order to be interoperable and easily manipulated from the various components, and b) the use of the proper calculation techniques so that it can be easily integrated with existing eHealth solutions, which requires the implementation of specific *functions* as software modules. Core part of the badges model is the composite objects containing the number of levels for the various badges and their definition.

Each badge has a category depending on the activity that the user has to achieve. The proposed model of the badges' model includes the following four activities:

- Daily usage: The users have to use the app (or specific app features) daily in order to upgrade his badge levels.
- Measurements: The measurements are processed in different timeframes and aggregation functions (e.g. sum, average, max, min) are applied to calculate the level.
- Sleep: The users have to achieve a certain sleep score every night (e.g. duration, quality) to be promoted to the next level.
- Steps: The users have to make a number of steps every day and stay active.

The system is able to calculate different aggregation metrics which are related with the specific activities of each user. This approach gives the system operators the option to define any number of badges, in any category, referencing in the model the function, or the combination of functions which can be used to calculate the metrics of the user activity. Aggregation functions such as *min*, *max, sum, count* on raw data can be applied in different time periods for each activity in order to easily perform the calculation of the levels for complex badge types. An example on this approach is to set badges for periods P in {Daily, Weekly, Monthly}, with aggregation function *func=sum*, and *data=STEPS* or *data=SLEEP*. The data are then summarized for the requested periods (e.g. one of Daily, Weekly, Monthly), the calculation functions will produce the intermediate scores and finally the badge levels. The users' Personal Health Records (PHRs), or other system data like the time the user accessed the app, are the raw input for the calculations and are produced by querying the data storage of the eHealth system. The raw data are transformed into objects that only hold the amount of information required for calculations. It should be noted that the proposed model, also supports caching of previous calculations outcomes allowing for incremental implementation of calculation operations. This means that for long periods, the data are continuously calculated, as they arrive, and if there are no changes on the level, they are cached in order to avoid querying

again the data from the database and aggregating the different periods from the beginning. The general approach for the calculation of the users' progress in such badges is illustrated in **Figure 1**.

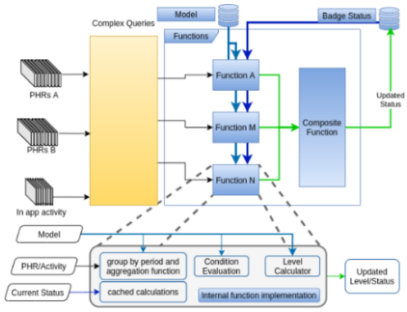


Figure 1. Model-driven composition of functions for the calculation of level status

3. System in practice

The approach has been validated in the context of eHealth applications of BioAssist SA [9]. The model and the calculation functions have been implemented as cloud functions according to the design concept of microservices that is followed by the particular platform. The operation of the solution demonstrates a fully flexible and modular design which can be applied and integrated in any similar system. The badge models for this prototype system are focused on the interaction with users and are tailored to the need of remote patient monitoring as presented in the **Figure 2**.

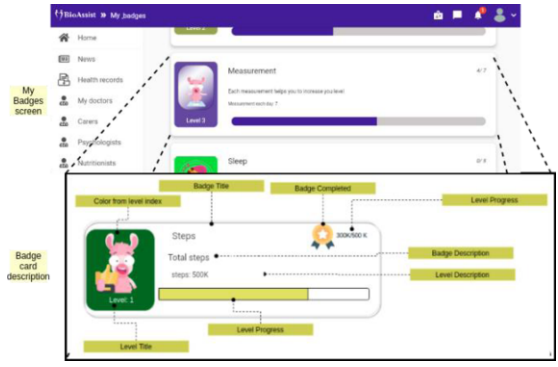


Figure 2. The "My Badges" screen and a sample badge card

The user interface was designed in order to address the main objective to provide feedback on the user's progress and educate them how to accomplish the various health and wellbeing related goals. Interviews with users and professionals, and analysis of their feedback was fundamental for the improvement of the overall design and of the experience. All badges have a number of levels to ensure the long-term gamification feature within the application. Every badge has an icon and the levels in the same badge are distinguished by the background color of the icon. The model also supports different icons and messages for each level. On technical and operation level, each model is stored in a key-value database as a json format and contains all the badges and the levels that each badge has. Twice a day and also after specific user interactions with the app, the

system executes the function that is defined in the model which calculates the scores and updates the status of the badges. The results are stored in the database and the users are notified in-app and via email for their progress.

4. Conclusions and Future Work

The role of gamification as a feature in eHealth applications seems to have positive effects on the motivation of users in order to achieve their targets, increase their physical activity, and adhere to a healthcare plan. In this work we presented an implementation of a gamification approach based on "badges", in the context of an eHealth application, as a fully flexible and modular IT solution which can be applied and integrated in any similar system. The extensive evaluation of the approach is ongoing in the context of pilot activities of the application which are currently operational. The proposed gamification strategy is expected to promote the use of the application, motive the users to exploit specific features of the eHealth system and most importantly help them improve their health and wellbeing. Extensions are already designed and will be presented in a future work. The extended approach will be focused on the implementation of leaderboards and user competitions which are expected to further enhance the user experience and provide more incentives for system use.

Acknowledgements

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: MediLudus Personalised home care based on game and gamify elements T 2EKΔK-03049).

References

[1] Hutchesson MJ, et al. eHealth interventions targeting nutrition, physical activity, sedentary behavior, or obesity in adults: A scoping review of systematic reviews. Obesity Reviews. 2021 Jun 23.
[2] Romeo A, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. Journal of medical Internet research. 2019;21(3):e12053.
[3] Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining" gamification". InProceedings of the 15th international academic MindTrek conference: Envisioning future media environments 2011 Sep 28; p. 9-15.
[4] Koivisto J, Hamari J. Demographic differences in perceived benefits from gamification. Computers in Human Behavior. 2014 Jun 1;35:179-88.
[5] Rajanen M, Rajanen D. Usability benefits in gamification. In GamiFIN. 2017;87-95.
[6] Hamari J, Koivisto J, Sarsa H. Does gamification work?--a literature review of empirical studies on gamification. In 2014 47th Hawaii international conference on system sciences, IEEE. 2014, January;3025-3034).
[7] Costa R, et al. SmartLife smart clothing gamification to promote energy-related behaviours among adolescents. In2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC), IEEE.2017 Jun 27;1489-1495).
[8] Tabak M, De Vette F, Van DIjk H, Vollenbroek-Hutten M. A game-based, physical activity coaching application for older adults: Design approach and user experience in daily life. Games for health journal. 2020 Jun 1;9(3):215-26.
[9] BioAssist | Innovative eHealth solutions. (2021). Retrieved 9 September 2021. Available from: https://bioassist.eu/.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210929

Use of Scratch as ICT Educational Tool in Health

Theodora ARGYRI[a,1], Emmanouil ZOULIAS[a], Joseph LIASKOS[a] and John MANTAS[a]
[a] *Health Informatics Laboratory, Faculty of Nursing, National and Kapodistrian University of Athens, Athens, Greece*

Abstract. Education adds knowledge, experience, and skills to the human personality, from childhood to old age. This paper focuses on adult education, presenting the difficulties, techniques and results that have been demonstrated to produce educational training program, including material using Scratch. Referring to the planned education from the school, university or any other institution, the educators have a duty to share their knowledge with the students, to show them the way of knowledge, learning, and exploration. They are constantly looking for new training techniques and methods that will attract anyone interested. An important role in this continuous effort is called to play the technology, which nowadays evolves. This paper addresses some of the advantages and disadvantages of technology in health training, how technology can be used and how it can be combined with training. In addition, important issues for health training are mentioned. The training program, the target group to which it is addressed, the means to be used and the objectives are presented in detail. Finally, technological training programs were created, which will enrich the specially designed training program, for which the methodology used, and their purpose will be analyzed.

Keywords. Health training, Health Informatics, Artificial Intelligence, Games

1. Introduction

As a pedagogical scientific approach of adult education and training nowadays is «the systematic and organized process of teaching and learning» [1], based on a specially designed program, specific methods, within a certain time frame, educational objectives and with tools that are defined beforehand. [2]. Researchers of the Assessment and Teaching of 21st Century Skills (ATC21S) framework, concluded that 21st century skills can be grouped into four broad categories: (i) ways of thinking, (ii) ways of working, (iii) tools for working and (iv) skills for living in the world [3]. It has been proven that technology promotes the adult educational and training process [4].

Several studies on health training exploiting Information and Communication Technology (ICT) tools rely either on training patients to comply with prevention and treatment [5,6,7], or on training health professionals to improve quality of care [6,8]. Mobile tablet-based educational tools [7] or online interactive educational tools [5,6,8] have been used for the training of health providers. These studies have concluded that

[1] Corresponding Author, Theodora Argyri, Health Informatics Laboratory, Faculty of Nursing, National and Kapodistrian University of Athens, Athens, Greece; E-mail: dora.argyri@gmail.com.

the use of ICT tools enriches learners' knowledge and promotes the correct response to health-related events [6,8].

This work presents the exploitation of Scratch as an ICT tool to produce interactive educational/training material for the training needs of potential health professionals (e.g., student nurses, student physicians), as well as non-professionals, like patients, in the introduction of the cardiovascular system anatomy. Scratch is a graphical dynamic programming tool for trainers and teachers to produce interactive educational/training material [9]. According to the literature search in major databases, no previous application appears in health training.

2. Methods

During the design phase of this study the researchers predefined the scenarios, texts and images that would be embodied in the produced training material. The first scenario designed, assuming that a potential user selects a region of interest of the heart's graphical representation, and the corresponding information is displayed. The second one presents an introduction of the parts of the heart, and it is initiated when the user presses the corresponding button onto the heart's graphical representation. The third scenario illustrates the automatic translation of the training material presented made by Scratch in several languages. During the design phase, the online version of Scratch was used (https://scratch.mit.edu/). The development of the training materials performed by a registered nurse, the main researcher of this work, having no programming experience but high management ability of ICT.

The tests performed were limited, initially within the research team members, and then by a small number of 15 volunteers, consisted of health experts' trainers, health expert trainees, public. These participants who tested, also evaluated the educational/training tool and material, and they were interviewed through unstructured interviews. The interviews included open questions regarding the functionality, content, error detection, and aesthetics of the tool. However, answers were not recorded systematically.

3. Results

For the teaching needs of cardiac anatomy, three different educational games were created, as interactive or animation teaching/ learning tools, using Scratch. The first one (Figure 1) is an interactive tool that presents the graphical representation of heart, where the user may select a specific point of the cardiac graphical representation, which corresponds to a specific part of human heart that interests him, and the corresponding information is automatically displayed. This Scratch game created, is available online on the link https://scratch.mit.edu/projects/437563118.

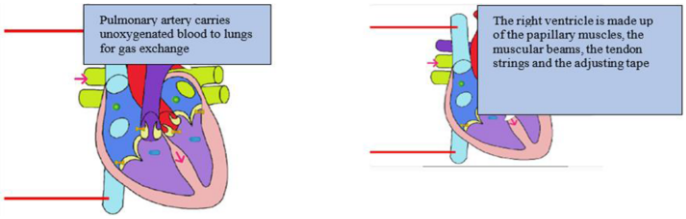


Figure 1. Heart anatomy interactive learning tool.

The second one, is an animation tool that further enhances the basic heart anatomy training by naming and listing the parts of human heart in its graphical representation (Figure 2). It is available on the link https://scratch.mit.edu/projects/437599456.

Figure 2. Heart anatomy animation tool, and its Scratch code.

The third one automatically translates the training material to several languages available, adapting the course to various languages speakers' trainees (Figure 3). It is available on the link https://scratch.mit.edu/projects/522423819.

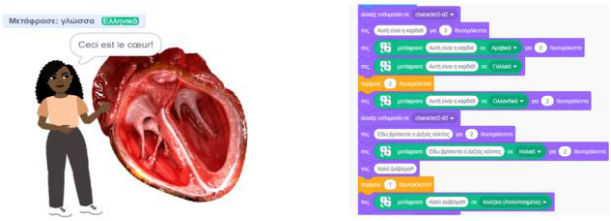


Figure 3. Animation tool translating the training material in several languages, and its Scratch code

The results of the evaluation of the material showed that in general it is an easy to use and understandable tool. Few people encountered difficulties in using it, which were overcome by explaining how the tool works.

4. Discussion

Our study is expected to assist interactive training of health professionals and students to provide better quality of care [6,8], and train patients to better understand aspects of their health or still comply to their treatment [5,6,7]. There are studies that prepare presentation of health issues targeting students [9]. In this work the training material created additionally offers interactive, animated, and multilingual material. Several difficulties have been encountered for the design and creation of the training tool, which teaches the «Anatomy of the Heart». As far as the first game/ tool created is concerned, one difficulty was the integration of interactivity. Another one was to provide suitable concise, simple, and easily understandable training material, which provides all the

necessary information of heart anatomy. On the other hand, the other two animated games/ tools were created easier and faster than the first one.

The main limitation of our study is the small number of trainees evaluated the training tools and material, and the fact that their responses were not systematically recorded.

5. Conclusions

In this study, educational/ training material was created, related to basic anatomy of the heart, for the training needs of both health and non-health professionals. Three educational material, as interactive or animated educational tools were created using Scratch as ICT tool. The introduction of ICT gaming tools like Scratch in health training may provide an alternative interactive educational approach, enhance the learning experience, and motivate the trainees. The use of Scratch for creating educational material requires no programming skills, and its learning curve has been proved low. The type of educational material presented in this study, is suitable for distance learning, for people with special needs, and for trainees who work, by considering their available time and pace; moreover, it facilitates home training and learning in movement restrictions, like those imposed during COVID-19 pandemic [10].

In our next steps, we plan to evaluate the training tools created with a more structured and systematical way; moreover, to create training tools and material in health, suitable for the learning needs of disabled people. Such example would be the design of a training tool, which exploits a computer-based built-in camera for detecting the facial movement of the trainee, in order to automatically display the learning information.

References

[1] Xochelis P, Introduction in Pedagogical: Funtamental problems in Pedagogical Science, Thessaloniki: Kiriakidis (In Greek). 2010.
[2] Pedagogical psychological encyclopedia - Dictionary. Athens: Greek Letters; 1989:1–601.
[3] Binkley M, Erstad O, Herman J, Raizen S, Ripley M, Miller-Ricci M, and Rumble M. Defining Twenty-First Century Skills Book Assessment and Teaching of 21st Century, 2012 Editors: Patrick Griffin Barry McGaw Esther Care, Springer 10.1007/978-94-007-2324-5.
[4] Roger M. The purposes of education. Athens: Metaichmio; 2003:1–368.
[5] Silva Junior GB, Oliveira JGR, Araújo EMR, Albuquerque ACRMM, Mota FAX, Freitas Filho RA, Vasconcelos Filho JE, Rodrigues MAF. Renal Health: Providing Information and Technological Tools to Empower Patients to Live Better with Kidney Disease. Stud Health Technol In-form. 2021 May 27;281:674-678.
[6] Hendricks M, Varathan O, Cassim F, Kidd M, Moodley K. Impact of implementing an online interactive educational tool for future HIV cure research in an HIV clinic waiting room in Cape Town, South Africa. AIDS Care. 2020 Aug;32(8):965-969.
[7] Bowman C, Lunyera J, Alkon A, Boulware LE, St Clair Russell J, Riley J, Fink JC, Diamantidis CA Patient Safety Educational Tool for Patients With Chronic Kidney Disease: Development and Usability Study. JMIR Form Res. 2020 May 28;4(5):e16137.
[8] Choueiry J, Reszel J, Hamid JS, Wilding J, Martelli B, Harrison D. Development and Pilot Evaluation of an Educational Tool for the FLACC Pain Scale. Pain Manag Nurs. 2020 Dec;21(6):523-529.
[9] Ramos VP. ICT in the curriculum using scratch with students of the alternative curriculum. 2015 International Symposium on Computers in Education (SIIE); 2015;9-72, doi: 10.1109/SIIE.2015.7451650.
[10] Iatronet. Home education: The first learning impulse in the child [Internet]. 2014 [cited 9 January 2021]. Available at: https://www.iatronet.gr/ygeia/paidiatriki/article/25755/ekpaidefsi-apo-to-spiti-i-prwti-mathisiaki-wthisi-sto-paidi.html.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210930

The Journey of Clinical Registries Through Various Phases of the Digital Age: A Technical Perspective

Mansoor Ali BAIG[a,b,1], Mohammad Shouki BAZARBASHI[a], Haifa AlDakhil[a] and Saniyah M. BAIG[c]
[a] *King Faisal Specialist Hospital & Research Center, Riyadh, Saudi Arabia*
[b] *Department of Biostatistics Epidemiology & Scientific Computing, KFSHRC, KSA*
[c] *Ilia State University, Tbilisi, Georgia*

Abstract. The concept of registry-based medical research goes back to the mid of 18th century where data was collected in actual physical registers and analyzed using manual counts in a very primitive way until computing technologies took over to digitize information, to change the process all the way from data collection to data analysis. This digital age of technology can be hypothetically classified in 3 eras; the Digitization Era, the Integration Era, and finally the Futuristic, Smart Intelligence Era. This study would highlight the changes in the fundamental aspects of a medical registry under each of these digital eras.

Keywords. Clinical registries, Disease registries, healthcare data, EDC

1. Introduction

The National Leprosy Registry, in Norway represented the world's first national patient register for any disease, established in 1856 [1], and eventually it became a practice to establish various disease registry programs to carry out evidence-based clinical research. These data repositories played a significant role in the control of the disease by evaluating the trends in disease incidence and other characteristics and thus helping the providers to plan accordingly. Medical registries have been identified as a source of information and evidence for patient monitoring, promoting evidence-based clinical decision-making [2]. Coupled with advancements in tools and technology, there has been a significant improvement and development in the way the registries are managed. It is worth studying and benchmarking the achievements of each phase of these advancements in relation to the other.

2. Methods

This study involves observation of various aspects of a medical registry under different progressive time frames (digital eras). These aspects not only include the collection,

[1] Corresponding Author, Mansoor Ali BAIG, BESC, King Faisal Specialist Hospital & Research Center, Riyadh, Saudi Arabia; E-mail: mansoorbaig@hotmail.com, Mbaig97@kfshrc.edu.sa.

entry, storage, extraction, processing, audit, reporting, analysis and integration of data, but also include predictive and prescriptive modeling of data. These aspects may be common to all eras or may be specific to a particular era.

3. Results

The digital revolution which started during the end of the 20th century, fueled the exponential growth of data. This revolution also created a great demand for advancements in tools and techniques to simplify data collection, storage and extraction. Ease in accessibility and availability of research quality data along with simplified, seamless and unified access (locally as well as on global platforms) was desired as the outcome of these advancements. This exponential growth in the data digitization efforts can be categorized into 3 major digitization eras, namely the Early Digitization Era, the Centralization or Integration Era and the Futuristic Smart Intelligence Era. This study would highlight the changes in the fundamental aspects of a medical registry under each of these digital eras.

3.1. Early Digitization Era

This was the most critical and cumbersome phase and the first step towards transformation to a modern age, where one had to give up the traditional way of doing things as the modern system would demand a 360-degree change as to how things would be done.

In the early phases, the registries had to run a parallel process along with the traditional methods to carry out the data collection, storage, extraction and analysis. Primitive systems were developed for the electronic data capture, which required manpower with computing skills, which were scarce during the early phase. The paper based case report forms were still used as a primary data capture instrument, and later as an instrument to audit or validate the electronic data. Certain issues like system portability, distributed data sourcing and data availability was more of a challenge. In this phase, the data was collected on paper based instruments (case report forms), verified by the research registrar and manually entered in the electronic data capture system. A random audit of the electronic data was carried out, which was also based on the primary data capture instrument. While on the other hand, it was very difficult or in some cases almost impossible to verify or audit the primary data, which was collected on paper. It took a lot of time and effort to streamline and adapt towards an electronic data capture system to manage and maintain the registry data. Some of the ancient registry programs had an additional burden of transforming the retrospective data. A few followed the traditional data entry approach and a few tried to transform or rather extract their paper based data using the OCR (Optical character recognition) technology [3].

The electronic data capture systems were an interface to a data repository, which was either maintained as a file based system or a database management system (DBMS) or a relational database management system (RDBMS), whereas, some registries mostly with limited or linear dataset (flat format) preferred to capture their data in a spreadsheet to avoid the complicated data extraction step and ease the use of carrying out the data analysis and reporting. Lastly, the data collected had to be reported as well as analyzed to provide the necessary evidence for the research hypothesis. In the early age, some primitive 3rd party data analysis tools were available, but mostly the EDC would provide

predesigned as well as custom reporting tools which had the capability of carrying out the basic analytics on the fly.

During this era, the data was digitized and stored in a silo with a closed system which did not have the ability to communicate or exchange data with other systems or entities.

3.2 Data Integration Era:

With a rapid growth in the internet penetration, it was now possible to move, transfer, interlink, interchange, interact with and within the data repositories, following an integrated approach. During this era, there were various governmental mandates on healthcare providers to quickly adapt to a Healthcare Information System (HIS) in order to maintain an electronic health record for the patients. During this period, data was still being captured redundantly in the Electronic registry as well as the HIS. During the primary phase of HIS implementations, the quality of data was not trusted upon and besides this, there were various technical and ethical issues with access to the HIS data, but during the mid-phase most organizations facilitated access to the HIS, which was used by the registries as a secondary data source, mostly to carry out data audits. But in the later phases, the registries collaborated with the medical records, medical informatics and the IT teams to come up with various data integration methods and strategies to avoid or minimize manual data capture. New methods/tools were introduced to capture the data directly from the data warehouse or. importing the most common minimum dataset from existing registries [2], or setting up an online-offline model to download the required data to be manually imported in the registry database.

Dynamic reporting features were made available, as a part of the registry systems. Also a number of sophisticated data analysis and bio-statistical tools like IBM SPSS, SAS, STATA etc. were available for carrying out any detailed and advanced data analysis.

3.3 The Smart Intelligence Era:

We have currently approached the beginning of this era, where various integrated systems communicate seamlessly through minimal human intervention, in order to manage and maintain the required data through a single interface deployed on an integrated environment. Number of cohorts can be run on the fly to identify or validate a research hypothesis, based on which the data can be sourced. These systems would also take a data based approach to generate various research hypothesis, each of which could be taken up as a distinct research project.

The concept of a silo based registry need not be followed and instead HIS itself can be a primary source once the necessary tools, software or platforms are implemented. These tools would act as an honest broker [4]; taking care of the good clinical practice. Further, based on the IRB approval, the required data can be sourced.

With a vision of establishing and intelligent enterprise, some healthcare entities have embraced the concept of a Data Science Hub [9] to assist researchers with their data problems, covering the descriptive, predictive and prescriptive approaches.

Systems like I2B2 [5] can be very instrumental in running cohorts on the HIS data to speed up the research process. I2B2 SHRINE [6] is another platform which helps healthcare organizations around the globe to share de-identified data or cohorts, enabling the healthcare providers to carry out disease specific benchmarking studies or validate their research hypothesis. Another futuristic approach by CLINERION [8] is to globalize

or rather centralize the de-identified clinical data (in form of counts), to be made available primarily to the pharma companies to globally identify patient samples through discovery cohorts in order to carry out phase 1/2/3 clinical trials. This same approach can be extended towards registries. Such initiatives accelerate development of innovative medical treatments by improving efficiency in clinical research.

The later phases of this era would be futuristic and based on Machine Learning, AI and collaborative research helping the global healthcare community, carry out benchmarking studies and personalized care. A registry system in the near future would just be a virtual interface to a data view generated through a well-organized data mart arranged through integrated systems, further integrated with advanced AI enabled analytical engines [7] where not only descriptive, but even predictive analysis could be done on the fly. These advancements in technology would act as a vital catalyst in speeding up research based activities in any healthcare setup, drastically reducing the research turnaround time, ensuring the highest quality coupled with an innovative approach.

4. Conclusion

The study provides a progressive view of the various aspects of registries in the light of the different phases of the digital age and also anticipates the possible advancements and implications of the concept of registries in the future. This study also highlights some unique platforms which works towards globalization of healthcare data along with strict adherence to the ethics and good clinical practice. These advanced tools and platforms would eventually virtualize the concept of managing and maintaining registry data or a registry based system. It would also provide a tool to the providers to define and design a registry in a matter of days if not hours.

5. References

[1] Irgens LM. The origin of registry-based medical research and care. Acta Neurol Scand Suppl. 2012;(195):4-6. doi: 10.1111/ane.12021. PMID: 23278649.
[2] Baig MA, Alzahrani SJ, Almuhaizea MA, Al-Shagathrh F. Fundamentals of Implementing a National Clinical Registries Program. Stud Health Technol Inform. 2019 Jul 4;262:384-387. doi: 10.3233/SHTI190099. PMID: 31349248.
[3] Paulsen A, Harboe K, Dalen I. Data entry quality of double data entry vs automated form processing technologies. Health Sci Rep. 2020 Nov 29;3(4):e210. doi: 10.1002/hsr2.210. PMID: 33283058
[4] Choi, H. J. et al. (2015). Establishing the role of honest broker: bridging the gap between protecting personal health data and clinical research efficiency. PeerJ, 3, e1506. https://doi.org/10.7717/peerj.1506
[5] Klann, J. G. et. al. (2016). Data interchange using i2b2. Journal of the American Medical Informatics Association: JAMIA, 23(5), 909–915. https://doi.org/10.1093/jamia/ocv188
[6] McMurry et. al. (2013). SHRINE: enabling nationally scalable multi-site disease studies. PloS one, 8(3), e55811. https://doi.org/10.1371/journal.pone.0055811
[7] PAT Research, Top 18 Artificial Intelligence Platforms, 2020.
[8] https://www.predictiveanalyticstoday.com/artificial-intelligence-platforms
[9] Clinerion,[n.d] Ideas and insights from our clinical research and real-world data experts
[10] https://www.clinerion.com/index/Experts-Corner
[11] Baig MA, Alzahrani SJ. Developing a Framework for a Healthcare Data Science Hub; Challenges and Lessons Learned. Stud Health Technol Inform. 2019. doi: 10.3233/SHTI190008. PMID: 31349257.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210931

Patient's Unique Identifier for Efficient and Secure Monitoring of Pregnant Women in Burkina Faso

Seydou Golo BARRO[a,b,1], Adrien UGON[c,d], Nadège R. NANA[a] and Pascal STACCINI[c]
[a] *Nazi BONI University, Bobo-Dioulasso, Burkina Faso*
[b] *RETINES Lab - Cote d'Azur University, Nice, France*
[c] *ESIEE, Paris, France*
[d] *Sorbonne Université, CNRS, LIP6, F-75005 Paris, France*

Abstract: The objective of this work is to set up a device allowing to identify the pregnant woman in a univocal and reliable way during her pregnancy follow-up. This study is a continuation of a project to improve the electronic monitoring of pregnancy in pregnant women in Burkina Faso. The methodology is based on the scientific work of the GMSIH of France (1). The work has lead to the design and implementation of a model that allows to assign a "Unique Identifier" to each pregnant woman from her first prenatal visit. The Patient ID is developed from the person's identification trait profile. It consists of a sequence of 20 characters and a security "key" of 2 characters. After the design, a reliability test of the model was performed to take into account identity anomalies (duplicates, collisions).

Keywords: Unique Identifier, Patient Identification System, pregnancy monitoring

1. Introduction

Pregnancy monitoring is a major problem in Burkina Faso but also in most developing countries. With the development of digital technology and particularly of "communicating computing", it is now possible to envisage a dematerialization of the exchange of information on the patient. Under these conditions, the follow-up of the patient or the pregnant woman requires fluid access to information wherever it may be and especially without getting the wrong person. However, the diversity of identification systems within healthcare systems entails increased risks of errors with enormous and sometimes fatal consequences. In the US, a collective of 27 organizations led by the American Health Information Management Association (AHIMA) published an open letter to representatives on June 11, 2019, reporting that misidentifications cost US $ 17.4 million (15.5M €) per year on average at a healthcare center, due to loss of income and refused reimbursement requests (2). This is what justifies the need for a reliable identification system in any healthcare system, which is an essential condition for secure information exchanges. Identifying the patient in the information system consists of assigning him a new identifier or finding an existing identifier, from his external characteristics called identification "traits", supposed to represent a person in order to identify him uniquely in this. system. There are three types of traits: strict traits, extended traits, complementary traits. These three types of traits are distinguished according to their more or less easy to obtain, more or less stable, more or less confidential and more

[1] Corresponding author, Seydou Golo Barro, Nazi BONI University, Bobo Dioulasso; E-mail: seydou.golo@gmail.com.

or less discriminating character. The unique identifier is a sequence of numeric or alphanumeric characters. It is the "key" to access information concerning the person admitted to the care system. Even in developed countries, even today, the electronic exchange of health data raises many debates concerning patient privacy, the choice of patient identifier and associated technical devices (3).

Problematic of the unique patient identifier: The issue of the unique patient identifier is raised in most countries around the world and each country has its own identification process. Patient identification is an important part of providing care and is the first step in a process that lasts throughout the care journey. It is one of the important parts in the health information system and contributes to the evolution of the health care system. A study report by the Group for the Modernization of the Hospital Information System (GMSIH), "Principles and process of patient identification": International experiences of health systems and identification policy", presents the experience of some countries in patient identification. The report classifies countries into three categories (4).

- The countries where the national unique identification of the patient is organized or planned: countries of the North of Europe (Denmark, the Netherlands, Belgium, Great Britain, Ireland) and New Zealand, Germany?
- Countries where the patient identification is more of the order of the "region": Southern European countries (Spain and Italy), federal countries (Canada and Australia).
- The recent undecided countries: the United States of America, which is moving towards the federation of identifiers.

In Africa, for many health facilities, the precise identification of patients remains a challenge. This is due to the insufficient quality of information systems and the absence of national systems for the identification of persons (5) (6). In Burkina Faso, the unique patient identification system is not yet used. This raises the issue of the security of information exchange and the continuity of care when the patient returns to the system for further care. Initiatives are underway for the establishment of a unique identification system for Burkinabé citizens and a unique national identifier. However, none of these initiatives have yet come to fruition. It is in the light of all these issues that we decided to design a unique patient identification model in a Burkinabé context. As part of this work, the unique identifier model is being tested in the monitoring of pregnancy and childbirth.

2. Methods

The method used for the design and implementation of the patient ID system is based on the scientific work of the GMSIH of France.

2.1. Creation of the Burkinabè model of unique identifier (patient ID)

According to the work of the GMSIH, a unique identifier of quality must be built with reliable features, always available. These traits must allow a bijective relationship between the physical person and his identity. There are three types of identifying traits: Strict traits, extended traits, complementary traits. These three types of traits are distinguished according to their more or less easy to obtain, more or less stable, more or less confidential and more or less discriminating character. Some traits are more stable than others. These are retained to constitute the "strict features": generally the names,

first names, date of birth and sex. Traits that are more unstable over a person's life, such as marital status or address, form the "extended traits". They serve as discriminating elements rather of an administrative nature, when the strict features are insufficient to remove any doubt about an identity. Finally, "complementary traits" are health information, which allows healthcare professionals to discriminate against people even more finely. In our case, we have retained the following traits: "Strict traits": family name, first name (s), date of birth, sex. "Extended features": place of birth. "Complementary features": Address of residence.

2.2. Criteria for creating the patient ID

To achieve a valid identification it is necessary to respect a certain number of rules: We have therefore developed an identification protocol to create a unique identifier in Burkina Faso taking into account the local context. It breaks down into four stages:

1) Use of identification traits. It consists in recovering the strict and secondary traits of the patient necessary for the creation of the unique identifier.
2) Standardization of characters. This step allows you to format the identification traits. It is:
   - Standardize the name and first name fields in an appropriate format (upper case, no accents, no spaces for compound names);
   - Perform a first hash (using the sha512 hash algorithm) of the first and last name;
   - Code the sex field in the format 1 for male and female;
   - Standardize the date of birth in this format: ddmmyyyy (8 digits);
   - Retrieve the code of the place of birth for the "place of birth" field. If the persons were born abroad the place of birth code is 00099;
   - Retrieve the city of residence code for the residence address field. If the persons reside abroad, the residence code is 00099;
   - Carry out a concatenation on the result of the preceding operations according to the following format: **(**sex_code.ddmmyyyy.hash (namefirstname) .birth_ocalitecode.codelocalite_residence.sector_num);
   - Keep only the digits of this concatenation and convert the letters into their corresponding in-tier according to the ASCII encoding.
3) Hash: perform a second hash of the result obtained with the "sha512" algorithm, the goal being to strengthen the security of the identification system. After this concatenation, take the first 20 characters from the index 50.
4) Calculation of the control key: The control key ensures the validity of the identifier. It is determined from the constituent characters of the identifying code. By performing a Euclidean division by 97 of the numbers resulting from the hash, we obtain a remainder which will be subtracted from 97. The result of this subtraction is the control key.

3. Results

Implementation of the patient ID system: For the implementation of the system, two types of libraries "hashlib" and "unicode" which respectively allow hash function and character normalization. The intermediate functions are "Normalize (string)" which uses the unicode library with the unicodeda-ta.normalize and unicodedata.category functions. this function returns the deleted string of uppercase, underscores, spaces, numbers and

special characters; the Con-vers_ASCII function (string, L0 = ['0', '1', '2', '3', '4', '5', '6', '7', '8', '9']) which takes as input a character string and a L0 list of digits from 0 to 10. It decomposes the string into a list of simple elements using .strip; it goes through this list, and converts the letters to integers using the ord () function; it returns a list concatenated with the letters of the string converted to ASCII integers. The main functions are: 1) "trait_d_information ()". It allows you to enter the identification traits of the patient. it returns a list of traits concatenating the entered identification traits: traits = [surname_of_family, first_name, date_of_birth, code_sex, code_localite_birth, code_localite_residence, sector_num]. 2) "unique_identifier (traits)" which takes as input the list of previous traits, then normalizes the first and last names fields, last name and first name hash, first con-catenation of standardized fields, conversion of the letters of the hash into their integer ASCII, new hash, selection of the first 20 characters of the hash from index 50, reconversion of the letters into their ASCII in-tier and selection of the first 20 digits, to give a first identifier, calculation of the control key, concatenation the identifier and the control key. It is also this function that revolves the Unique Patient Identifier.

Integration of the unique patient identifier method in an HIS: Patient ID works within a hospital information system. We test our model by integrating it into the electronic pregnancy monitoring system of the University Hospital of Bobo-Dioulasso (e-Lafia) (7). E-Lafia is a Pregnancy Tracking Information System. Patient information is collected in e-Lafia through a registration form. This information is stored in a MySQL database. E-Lafia HIS has been selected since it uses healthcare interoperability standards such as HL7v3, FHIR, DICOM, thus facilitating interaction with our unique identification system.

4. Conclusions

This pilot experience of designing and implementing a unique patient identification system within a pregnancy monitoring system is an innovation and a basis for computerization and data exchange projects on the patient. Although functional, our patient ID system could be improved to integrate identity reconciliation procedures with traceability of all operations; the creation of an identity-vigilance unit which will ensure the conditions for correct data collection, the reporting of anomalies and duplicates as well as the inclusion of corrections.

References

[1] GMSIH. Identification du patient : Guide pour l'élaboration des politiques d'identification et de rapprochement : Annexe 1 : Charte d'identification. 2002.
[2] USA : vers un identifiant patient unique au niveau national [Internet]. Innovation Mutuelle. 2019 [cité 24 sept 2020]. Disponible sur: https://www.innovation-mutuelle.fr/actualite/usa-vers-un-identifiant-patient-unique-au-niveau-national/
[3] Dumont V. Controverses autour de l'échange électronique de données de santé. La question de l'identifiant du patient. Termin Technol Inf Cult Société. 1 juin 2011;(108-109):125-38.
[4] GMSIH. Principes de l'identification du patient – Tome 2. Paris: GMSIH; 2002 avr p. 80. (Modele de rapprochement).
[5] Bacher SC. L'identification du patient. :106.
[6] Bediang G, Amougou EE, Essi M-J. Identification des Personnes Physiques et des Patients: Une Étude de Cas au Cameroun. 2019.
[7] Kaboré AA, Kaboré BE, Barro SG. e-Lafia: Suivi électronique de la grossesse. Université Nazi BONI; 2020.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210932

The Implementation of a Mobile APP for Cancer Care Management at the Brazilian National Cancer Institute

Emmanuel Tenorio Cavalcante PIRES[a], Cezar CHENG[a], Sandro Luís Freire de CASTRO SILVA[a,1], Sandra Regina GOIS[b], Francisco Santos SABBADINI[c] and Antônio Augusto GONÇALVES[a,b]
[a] *Instituto Nacional de Câncer - COAE Tecnologia da Informação, Rua do Resende 195, Rio de Janeiro, 20230-092, Brazil*
[b] *Universidade Estácio de Sá - MADE, Avenida Presidente Vargas 642, Rio de Janeiro, 200071-001, Brazil*
[c] *Universidade do Estado do Rio de Janeiro - UERJ Rod Pres Dutra km 298 , Resende, 27537-000, Brazil*

Abstract. There is increasing recognition of the potential to use mobile health (mHealth) technologies such as smartphone apps to support clinical care. Mobile apps are progressively being implemented to manage chronic diseases like cancer to improve patient care. The app deployed at the Brazilian National Cancer Institute enables patients to access the appointments/exam booking, medication prescription, and cancer-prevent educational resources, improving self-management, autonomy, and cancer treatment outcomes in resource-limited environments. Despite acknowledging the importance of mHealth, research is still scarce regarding patients' views on using these innovations in cancer care management.

Keywords. Mobile App, mHealth, Cancer Care.

1. Introduction

Population aging, improvements in information and communication technologies (ICT), and advances in public health care have introduced additional telecare services requirements. In this regard, ICT innovations revolutionize chronic patient care, providing real-time monitoring for health professionals [1].

There is increasing recognition of the potential of mobile health (mHealth) technologies such as smartphone apps in optimizing clinical care. Mobile phones currently represent a technology present in patients' daily lives. They enable access to different information and healthcare services independent of the time and place. This connectivity offers excellent opportunities to use mobile technology for improving healthcare services [2,3].

Mobile apps are progressively being implemented to manage chronic diseases, improving patient care. Although the potential benefits of mHealth seem particularly

[1] Corresponding Author, Sandro Luís Freire de Castro Silva, Instituto Nacional de Câncer - Tecnologia da Informação, Rua do Resende 195, Rio de Janeiro, 20230-092, Brazil; Email: sandrofreire@gmail.com.

suitable for monitoring chronic diseases like cancer, the treatment success depends on patient compliance. Cancer treatment is accomplished through intensive and prolonged ambulatory procedures that frequently occur outside the hospital units, representing a significant challenge [4]. Mobile health (mHealth) is a broadly accessible technology that provides health services through portable devices [5]. The implementation of mHealth initiatives has been positive, particularly in cancer care management and follow-up [6,7]. Despite acknowledging the importance of mobile health apps, research still scarce regarding patients' views on using these innovations in cancer care management.

2. Method

The project was conducted at the Brazilian National Cancer Institute between January 2020 to November 2020. The technical team has developed the INCA App as shown in figure 01, through the following steps: 1) Assess INCA patients' cancer care requirements, 2) develop an app tailored for cancer care management, 3) validate the app's usability, and 4) App installation on INCA patients' smartphone.

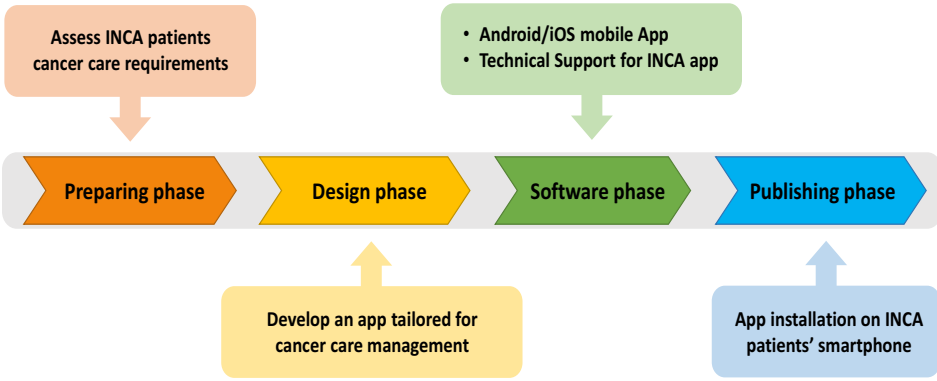


Figure 1. INCA App Implementation Phases.

For this implementation, a survey was carried out at INCA hospital units through more than 50 semi-structured interviews with patients and family members. The APP requirements implemented at the first stages of the project remarked that the application should be available on the most used platforms.

The cancer treatment management application was developed to enhance the patient-clinician relationship and guarantees patients' data consistency and accuracy. The app can effectively track and manage INCA patients during cancer treatment in an interactive and engaging framework.

This technology can reduce administrative burdens and improve both efficiency and quality of care delivery in the clinic. Among the patients identified as mHealth users, approximately half of the respondents used INCA mobile technologies for automation and decision-making support to schedule an appointment, access personal information, and read test results.

3. Results

INCA patients can instantly receive notifications of appointments/exam booking, medication prescription, and cancer prevention education messages through the application. Patients could access supportive cancer care with effective clinician interaction, regardless of place and time. Among the main functionalities created to facilitate the INCA patient's cancer care, as shown in figure 02, the following can be highlighted:

Figure 2. INCA App Screenshot.

- Digital patient's registration card. Its purpose is to facilitate the identification of the patient when he goes to INCA or any other health units;
- Appointments and exams scheduling to inform and alert the user on which dates, times, and units should attend for consultations and exams;
- Medical information briefly presents the type of neoplasia, diagnosis date, staging, and patient's situation. It is a facilitator so that the patient, whenever necessary, has the information readily available to present to health professionals who provide care in other health units;
- List of medications in use during his treatment at INCA. Likewise, it is a facilitator so that the patient, whenever necessary, has the information readily available to present to health professionals.
- Ombudsman Channel facilitates communication between patients and the Ombudsman sector to make complaints, suggestions, and requests.

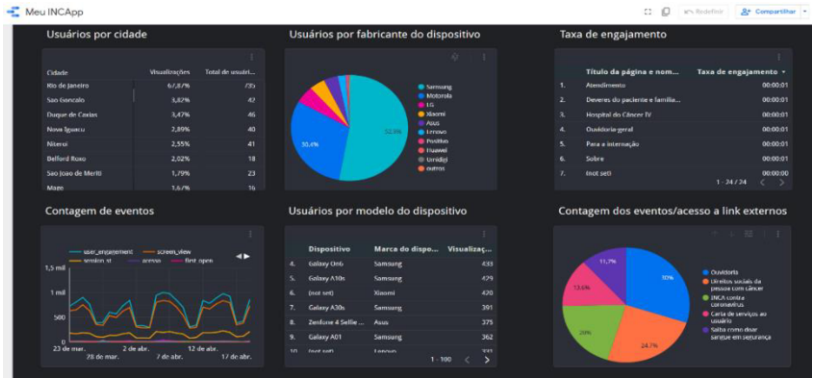


Figure 3. INCA App Panel Screenshot.

The resources used to monitor the INCA App provide information and indicators for decision-making of future improvements, made available through a panel with graphics when the application is in use. The Panel was created on the development platform Firebase, Google, whose purpose is to collect information recorded on each cell phone of INCA patients to provide statistical data, as shown in figure 03.

4. Discussion

Modern technologies like mobile apps are increasingly improving healthcare services with innovative solutions. Advances in wireless technology can lead to the development of mHealth, which offers an unprecedented opportunity for health care organizations to deliver high-quality care. These innovations empower cancer patients to be strategic actors in managing their treatment through mobile devices.

Implementing mobile applications in healthcare organizations can support cancer care management, increase satisfaction, and strengthen trust and communication between patients, physicians, and clinicians. Mobile app facilitates treatment follow-up and improves patients data consistency.

Advances in information technology and the development of mobile apps may enable the deployment of powerful tools to deliver personalized cancer care management. The INCA app tracks patients' treatment data, provides cancer-prevent education and identifies evidence-based recommendations for cancer treatment interventions personalized to the INCA environment.

References

[1] Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? J Am Med Assoc Dec 2013;310(22):2395-2396.

[2] Tarricone R, Cucciniello M, Armeni P, Petracca F, Desouza K, Hall L, Keefe D. Mobile Health Divide Between Clinicians and Patients in Cancer Care: Results from a Cross-Sectional International Survey. JMIR mHealth uHealth. 2019;7(9).

[3] Noah B, Keller MS, Mosadeghi S, Stein L, John S, Delshad S. Impact of remote patient monitoring on clinical outcomes: an updated meta-analysis of randomized controlled trials. NPJ Digit Med. Jan 2018;1(1).

[4] Collado-Borrell R, Escudero-Vilaplana V, Calles A, Garcia-Martin E, Marzal-Alfaro B, Gonzalez-Haba E. Oncology patient interest in the use of new technologies to manage their disease: a cross-sectional survey. Jouneal Med Internet Res. 2018;20(10).

[5] Jongerius C, Russo S, Mazzocco, K. Research-tested mobile apps for breast cancer care systematic review. JMIR Mhealth Uhealth. 2019;7;e10930.

[6] El Shafie RA, Weber D, Bougatf N, Sprave T, Oetzel D, Huber PE, Debus J, Nicolay NH. Supportive care in radiotherapy based on a mobile app: a prospective multicenter survey. JMIR MHealth and UHealth. 2018;6:1–14.

[7] de Castro Silva SL, Gonçalves AA, Cheng C, Martins CH. Mobile Technology Applications in Cancer Palliative Care. InpHealth 2018 Jun 5 (pp. 199-202).

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210933

A Self-Management App for Patients with Schizophrenia: A Pilot Study

Buket CINEMRE[a], Mahmut GULERCE[b,1] and Kemal Hakan GULKESEN[b]
[a] *Department of Psychiatry, Faculty of Medicine, Akdeniz University, Antalya, Turkey*
[b] *Department of Biostatistics and Medical Informatics, Faculty of Medicine, Akdeniz University, Antalya, Turkey*

Abstract. We aimed to observe the effects of a self-management mobile app for patients with schizophrenia. A mobile app was designed to record and remind users to take medication and some daily activities. The patients were asked to use the app for one month. Before starting to use the app, and after completion of one month, the patients were assessed by some psychiatric scales. Twelve patients completed the study. The mean number of automatic reminders per patient was 918/month or 29.6/day. The mean entry number per patient was 158.4/month or 5.1/day. The numbers of daily recorded activities showed a decreasing tendency over time. The pre-study PANSS score was 16.2±5.8 (mean±SD) and the post-study score was 14.9±6.9 (p=0.040). The pre-study FROGS daily life skills score was 20.3±3.8 and the post-study score was 19.7±4.1 (p=0.012). The results suggest that a mobile app may be helpful for patients with schizophrenia.

Keywords. Schizophrenia, smartphone, mobile applications, reminder systems

1. Introduction

Treatment of schizophrenia is mainly pharmacological, however, patient non-adherence to medication is a serious problem. Three-fourths of patients with schizophrenia become non-adherent within two years of hospital discharge [1]. In addition to pharmacological treatment, psychosocial treatments have been suggested for the treatment of schizophrenia. Some of the evidence-based practices include illness self-management training, and social skills training [2]. Mobile phones may support adherence to taking medication and psychological interventions. In a review on apps to support coordinated specialty care for prodromal and early course schizophrenia disorders, the authors found 21 eligible studies on 16 apps. Only three of these studies were intervention studies [3].

The primary objective of the present study is to observe if a self-management mobile app for patients with schizophrenia affects their Positive and Negative Syndrome Scale (PANSS) [4] score after one month of use. The secondary objective is to observe their mobile app usage patterns.

[1] Corresponding Author, Mahmut GULERCE, Department of Biostatistics and Medical Informatics, Faculty of Medicine, Akdeniz University, 07070 Antalya, Turkey; E-mail: mahmutta@gmail.com.

2. Methods

The patient activities were divided into four groups; 1) Self-care activities (eating, shaving, taking shower, make-up, etc.), 2) Household duties (cooking, cleaning, shopping, etc.), 3) Social activities (meeting with friends, calling friends, forming a new acquaintance, etc.), and 4) Drug adherence. In the following meetings, detailed lists of activities under each group were prepared. The system was designed with three components 1) Database and core software, 2) Mobile app, and 3) Web interface for physicians. The mobile app has the function of sending reminders for each activity and recording performed activities based on input from the patients. The physicians could adjust the activities for each individual patient.

The operating system of the web server was CentOS v6.5. Apache webserver 2.2, MySql database server service 5.7.21, and some 1st part libraries for PHP programming language were installed on the core component. The system was hosted on a cloud-based VPS server. The app was developed on the Apache Cordova development platform, which provides a single development environment for Android, IOS and Windows operated mobile phones. We did not use a third-party service for notification messages (reminders) because of concerns about security. We developed an isolated notification engine and integrated it into the system. The patient data was not stored in the patients' smartphones. All the communications between the app, the physician interface and the server were secured by an SSL certificate. The physician user interface was restricted by IP control.

Before the study, usability tests for the app were performed. Heuristic evaluation was completed by five experts. After re-designing the app according to the results of the heuristic evaluation, we performed a second usability test with five patients with schizophrenia using the think-aloud method. The app was re-designed after the second step of the usability study.

There was a default activity list for the patients. They were expected to have breakfast every day, go shopping every three days, clean the home every week, etc. The physician could tailor the default activities and add new activities for each patient. For example, if the patient had a daily shaving habit, it was also added to the patient's activity list. The medications were also entered by the physician. The patients could see their activity list in the mobile app and they were expected to record their activities (Figure 1). If they did not enter an activity in the expected period, the mobile app reminded them.

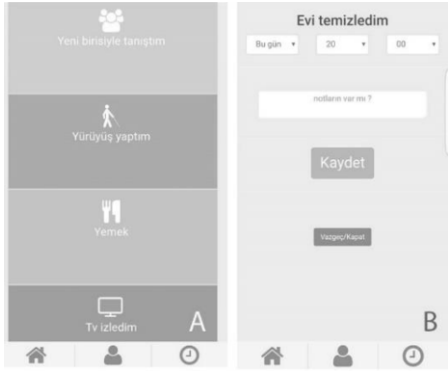


Figure 1. Activity menu (A), and activity record screen (B) of the app.

The study included patients with a diagnosis of schizophrenia according to the DSM-V, who satisfied the following criteria: age 18–65 years; active smartphone users; clinically stable patients who regularly visited our outpatient center over the previous three months. All patient treatments, including medication types and doses, were maintained for the duration of the study period. The ethical permission was obtained from the Akdeniz University Clinical Research Ethical Committee (2015/130). The number and type of reminders were tailored for each patient with the help of the web interface. The patients were asked to use the app for one month after user training. They were given a telephone number to contact in case of technical problems, but no additional support or stimulation was provided except smartphone reminders.

Before they began using the app, and after completion of one month, the psychopathological status of the patients was assessed by the PANSS (used for measuring symptom severity of patients with schizophrenia), the Clinical Global Impressions Scale (CGI, used for measuring symptom severity, treatment response and the efficacy of treatments in treatment studies of patients with mental disorders) [5], the Instrumental Activities of Daily Living (IADL) Scale (evaluates the activities that allow an individual to live independently in a community) [6], and The Functional Remission of General Schizophrenia (FROGS) Scale (it is specific to schizophrenia, which assesses daily and social functioning comprehensively, regardless of the psychotic symptoms) [7]. We recruited 14 patients into the study, but we lost two of them in the follow-up. So, we completed the study with 12 patients, resulting in a 0.71 post-study power for detection of the difference in 16 (Cohen's d: 0.8) of the PANNS score. Data were compared using paired t-test or the Wilcoxon test. Spearman's rho test was used for correlation analysis.

3. Results

One patient used the app less than 30 days and one patient did not participate in post-study tests, so we completed the study with 12 patients. On average, the patients were 34.6±10.8. Eight (66.7%) of them were female and four (33.3%) of them were male. In one month, 12 patients recorded a total of 1901 activities. The mean of automatic reminders per patient was 918/month or 29.6/day. The mean entry number per patient was 158.4 (median 116.5) and for each patient, the mean entry number per day was 5.1. The number of recorded activities of the patients in the study showed great variation (minimum 22, maximum 388). The most frequent entries were; Ate 171 (9.0%), washed my face 163 (8.6%), brushed my teeth 140 (7.4%), had breakfast 114 (6.0%), and made my bed 104 (5.5%). We observed an ongoing resistance to recording their medication. The average monthly entry for each patient is only five times.

The numbers of daily recorded activities showed a decreasing tendency over time (Figure 2). Pre-study and post-study test results of the patients are shown in Table 1.

4. Discussion and Conclusion

The most striking result of our study is the observation of the patients' decrease in app use over time. Despite being a relatively young group of participants, the patients used the app an average of 5.1 times a day. This number is similar to 5.2 in FOCUS [8] and 4.2 in A4i studies [9]. Better adherence in the first week of the study in patients with schizophrenia was also observed in other previous studies [10, 11]. The app produced a

mean of 918 reminders for each patient and they responded to 17.3% of them. This behavior reminds us of "alert fatigue", a phenomenon that is discussed for clinical decision support systems [12].

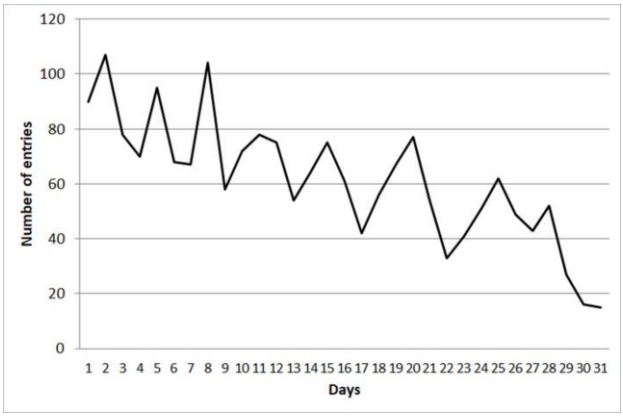


Figure 2. The number of daily activity records during usage of the patients for 31 days.

Table 1. Pre-study and post-study test results of the patients *Data with parametric distribution is shown as mean±standard deviation, and data with non-parametric distribution are shown as median (Q1-Q3).

Test	Pre-study*	Post-study*	p
PANSS positive signs	16.2±5.8	14.9±6.9	0.040
PANSS negative signs	16.0±5.5	15.1±5.6	0.374
PANSS general psychopathology scale	29.3±7.2	27.8±11.6	0.409
PANSS Total	61.4±16.4	57.8±20.0	0.103
FROGS social functioning	17.8±4.5	17.7±4.5	0.809
FROGS health and treatment	13.3±2.8	13.3±2.9	0.586
FROGS daily life skills	20.3±3.8	19.7±4.1	0.012
FROGS occupational functioning	5.8±1.8	5.6±2.0	0.339
FROGS Total	56.9±11.6	56.2±12.3	0.069
CGI Severity of illness	3 (2-4)	3 (2-4)	0.157
IADL total	7.5 (7.0-8.0)	7.5 (7.0-8.0)	0.317

Previous studies showed a positive effect of the mobile intervention on patients with schizophrenia [8-9, 13-14]. In the present study, when pre-study and post-study scores of PANNS, FROGS, CGI, and IADL were compared, statistical significance was observed only for PANNS positive signs (mean pre-test 16.2 and post-test 14.9) and FROGS daily life skills (mean pre-test 20.3 and post-test 19.7). Interestingly, the PANNS positive signs score shows improvement while the FROGS daily life skills score shows deterioration during the study period. The general effect seems to be positive because the size effect of PANNS positive signs is larger than FROGS daily life skills score, but due to the small sample size and absence of a control group, these findings must be carefully interpreted. These changes may be random or a result of other factors such as medication.

A major limitation of our study is our small number of patients. The study was originally planned for 14 patients, but we finished the study with 12 patients. This number would be moderate in case of good patient adherence, but we observed an apparent decrease in patient usage during the one-month study period. Because of the decreasing usage problem, we are suspicious if pre and post-test comparisons reflect the

real effect of such an app. Another limitation is the absence of a control group. The observed difference in test scores may be due to the natural course of the disease or result of the patients' medication. The absence of a post-study questionnaire also limits the value of the present study. A feedback questionnaire would add some detailed information on the user perspective and experience.

The results of the present study show a decrease in time of usage of an app for self-management by patients with schizophrenia. A reason for this problem may be the high number of reminders produced by the app. With the help of a more carefully selected smaller set of reminders, better adherence to the app may be obtained. Another approach may be the addition of some data which are automatically produced by smartphones. Modern smartphones carry several sensors including localization and acceleration sensors which can produce valuable data about the patient activity. We also observed quite variable usage patterns among patients. In future studies, larger groups would be helpful to stratify patients to observe different usage patterns and understand the type of patients who will benefit from the intervention. A control group would also be very helpful to isolate the effect of intervention from other possible factors.

References

[1] Cañas F, Alptekin K, Azorin JM, et al. Improving treatment adherence in your patients with schizophrenia: the STAY initiative. Clin Drug Investig. 2013 Feb;33(2):97-107.
[2] Mueser KT, Deavers F, Penn DL, Cassisi JE. Psychosocial treatments for schizophrenia. Annu Rev Clin Psychol. 2013;9:465-97.
[3] Camacho E, Levin L, Torous J. Smartphone Apps to Support Coordinated Specialty Care for Prodromal and Early Course Schizophrenia Disorders: Systematic Review. J Med Internet Res. 2019 Nov 12;21(11):e16393.
[4] Kay SR, Opler LA, Lindenmayer JP. The Positive and Negative Syndrome Scale (PANSS): rationale and standardisation. Br J Psychiatry Suppl. 1989 Nov;(7):59-67.
[5] Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont). 2007 Jul;4(7):28-37.
[6] Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. 1969 Autumn;9(3):179-86.
[7] Llorca PM, Lançon C, Lancrenon S, Bayle FJ, Caci H, Rouillon F, Gorwood P. The "Functional Remission of General Schizophrenia" (FROGS) scale: development and validation of a new questionnaire. Schizophr Res. 2009 Sep;113(2-3):218-25.
[8] Ben-Zeev D, Kaiser SM, Brenner CJ, et al. Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia. Psychiatr Rehabil J. 2013 Dec;36(4):289-296
[9] Kidd SA, Feldcamp L, Adler A, et al. Feasibility and outcomes of a multi-function mobile health approach for the schizophrenia spectrum: App4Independence (A4i). PLoS One. 2019;14(7):e0219491.
[10] Palmier-Claus J, Rogers A, Ainsworth J, et al. Integrating mobile-phone based assessment for psychosis into people's everyday lives and clinical care: a qualitative study. BMC Psychiatry. 2013 Jan 23;13:34.
[11] Ainsworth J, Palmier-Claus J, Machin M, et al. A comparison of two delivery modalities of a mobile phone-based assessment for serious mental illness: native smartphone application vs text-messaging only implementations. J Med Internet Res. 2013 Apr 05;15(4):e60.
[12] Ancker JS, Edwards A, Nosal S, et al. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Med Inform Decis Mak. 2017 Apr 10;17(1):36.
[13] Ben-Zeev D, Brenner CJ, Begale M, et al. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull. 2014 Nov;40(6):1244–1253.
[14] Schlosser DA, Campellone TR, Truong B, Etter K, Vergani S, Komaiko K, Vinogradov S. Efficacy of PRIME, a Mobile App Intervention Designed to Improve Motivation in Young People With Schizophrenia. Schizophr Bull. 2018 Aug 20;44(5):1010-1020.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210934

The Use of Robotics in Dementia Care: An Ethical Perspective

Evangeline WAGNER[a,1] and Elizabeth M. BORYCKI[a,b]
[a] *School of Health Information Science, University of Victoria, Canada*
[b] *Michael Smith Foundation for Health Research, Canada*

Abstract. Dementia and other related diseases are becoming increasingly diagnosed and are placing a serious strain on the healthcare system. Robotic technology research is currently underway to provide the much needed support to patients, caregivers, and health providers. This includes examining the ethical implications of robots use in healthcare. This scoping review explores the current state of the literature regarding robotics and dementia with a special lens on ethics. More specifically, this paper strives to gain an understanding of the current ethical considerations, and propose an intervention for evaluating ethical considerations prior to implementation. This research was conducted using PRISMA guidelines, extracting data from articles. Our findings revealed that further attention to policies and guidelines that are currently in place for general use of the technology should be utilized, and applied specifically to the context where the technology will be used.

Keywords. Robotics, dementia, ethics

1. Introduction

Dementia and other related diseases are becoming increasingly diagnosed in Canada, placing a serious burden on patients, families, caregivers and health practitioners [1]. 402,000 Canadian seniors aged 65 and over are living with dementia, and the number of individuals diagnosed with the disease increases by 76,000 people each year. The incidence rate is 14.3 cases per 1,000 senior individuals and over two thirds of those diagnosed with the disease are women [2]. There is a growing need for support and assistance for patients and families, who have been affected by dementia and other related diseases, due to the severe economic challenges that the disease poses to families and the healthcare system. Currently, the annual healthcare costs for dementia patients in Canada are $8.3 billion, and by 2031 these costs are expected to double to over $16.5 billion [2]. The use of technological solutions in healthcare in general can significantly reduce federal and provincial spending. The use of technology in any way should be done in an ethical and thoughtful manner. Therefore, there is a need to consider the technology in the context of the patient's best interests [3]. The objective of this research is to examine the current state of the literature regarding the use of robotic technology in dementia care with a special emphasis on ethical considerations. Specifically, the aim is to explore ethical awareness surround the technology among designers, developers and implementers.

[1] Corresponding Author, Evangeline Wagner; E-mail: evangelinemwagner@uvic.ca.

2. Background

Robotic technology has the potential to assist many individuals who are experiencing limitation and their caregivers. Limitations can include physical, cognitive, and/or emotional impairments that affect the patient as a result of their disease or condition [4]. By definition, a robotic device is a "physically embodied system capable of enacting physical change in the world". It should be added that robots can assist patients with cognitive tasks such as problem-solving and financial management, as well as basic activities of daily living (ADL) tasks such as grooming, feeding and moving [4]. Robotic devices can help with ADLs such as hygiene and instrumental activities of daily living (IADLs) such as going grocery shopping. Robotic devices can support an individual with dementia throughout the progression of their disease by helping them to adapt to the cognitive challenges associated with navigating differing contexts or settings [5].

The use of robotics and other medical technologies can provide life changing support to patients, families, caregivers and healthcare providers. However, ethical discussions must be understood in order to ensure that the rights of patients and involved parties are respected. The IMIA Code of Ethics includes the Fundamental Ethical Principles and the General Principles of Informatics Ethics, which provide for the planning and consideration of ethics in the delivery of modern healthcare solutions [3]. It should be noted that the IMIA Code of Ethics encompasses several, if not all medical (or health) technologies. It is argued that the guidelines can be interpreted differently for varying types of diseases and technologies that are used to treat or support those diseases.

3. Methods

A scoping review was undertaken to assess the current state of the literature. The review employed Arksey and O'Malley's [6] and Levac's [7] guidelines, and adhered to the Preferred Reporting Items for Systematic Reviews (PRISMA) guidelines.

3.1. Literature Search

A comprehensive search of four electronic databases: MEDLINE®, PubMed®, IEEEXplore® and Web of Science® was conducted. The search terms used were "dementia", "ethics" and "robotics", and were entered into each of the databases. Articles were extracted between the years of 2005 and 2021 in order to obtain relevant results. Following the database searches, all articles were extracted using Zotero® software. After extraction, the cumulative search results were imported into Covidence® for title screening and abstract screening, full text review and data extraction. Prior to importing the articles, the search results were visually checked for accuracy and unrestricted ability to view and download documents.

3.2. Inclusion Criteria and Exclusion Criteria

For an article to be included in the scoping review, it must be an empirical study focused on the ethical implications of robotics use in home care, and specifically in dementia care. Articles were included between the years 2005 and 2021. Additionally, if the article mentioned mild cognitive impairment and home care it was were considered for

inclusion and review. Articles were excluded if: (1) they did not mention ethics in either the objectives or discussion, (2) did not include robotics in the study (3) studied the use of a robot outside of a home care context (e.g. a hospital) and (4) not in English or French. Posters, abstracts, pamphlets and infographics were excluded. Studies that did not mention dementia or mild cognitive impairments were excluded, as well as studies which were primarily concerned with the function of the device rather than the ethical implications of its use.

3.3. Procedure

The article screening process was conducted using PRISMA guidelines for scoping reviews [8]. One reviewer screened titles and abstracts to fit the inclusion and exclusion criteria using Covidence®. The reviewer downloaded the final set of articles that were accepted into the study that were fully reviewed. All included articles fit the inclusion criteria and relevant data were extracted. Extracted data from the included articles following the full text review included the year of publication, country of origin, study design and number of participants. Additionally, specific details that were related to the inclusion criteria were extracted such as ethical issues and mitigation strategies that were mentioned. Any gaps in the research as well as the findings and impact were extracted.

4. Results

Following an initial search, 49 studies were imported into Covidence® for screening. After 11 duplicates were removed, 38 titles and abstracts were screened. Following title and abstract screening, a total of 31 studies were screened as full text. A further 17 studies were excluded based on the criteria, as outlined above (see Figure 1). Following the full text review, 12 articles were included in the study.

4.1. Article Characteristics

The articles included in the study were published between the years 2010 and 2020. Studies were included from several different journals (n=10). Many studies were published in *Ethics and Information Technology* (25%[3/12]). Half of the studies (50%[6/12]) were published in ethics journals. The studies employed survey, interview and focus group designs. Some researchers published literature reviews focused on this topic. The first authors of these research works represented nine countries, with the most common being England (25%[3/12]) and Switzerland (17%[2/12]). The remaining studies were conducted in Australia, Finland, Sweden, Ireland, United States, Norway and Belgium.

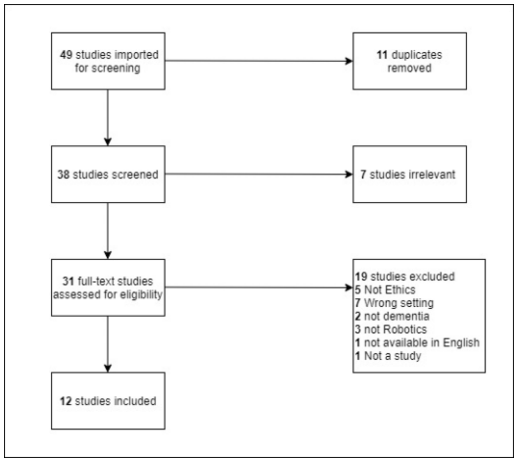


Figure 1. PRISMA Diagram

4.2. Themes

Three themes were identified: ethical considerations, robotic functions and healthcare system impact. The majority of the studies discussed ethical considerations that should be taken into account prior to robot implementation. The most common ethical considerations which were discussed were the potential for reduced human contact, privacy, affordability, infantilization, safety, independence, justice and informed consent [9-13]. The robots that were studied performed a number of functions. Robot functions included providing companionship [9], assistance (with basic movement, preventative measures) [10, 15] or were involved with video surveillance (i.e. surveillance) robots [14]. The remainder of studies did not state the specific functions of the robots. The principle impacts were safety, standardization and providing ethical solutions to health problems. Additionally, there were observed calls for policy change medical professionals in the field of robotics involved with caring for patients with cognitive impairments and dementia.

5. Discussion

The results of the review displayed the current state of the literature involving empirical studies of dementia and robotics used in healthcare with ethical considerations. Early research studies on the topic focus on the specific functions and desired effects of the use of robotic devices in healthcare for different diseases, and have now evolved into trials and studies that examine more than functionality. The role of robotics in the healthcare delivery process has raised questions and concerns about ethics and patient safety [10]. A gap in the research is the lack of standardization between regional codes of ethics and robotic implementations. For example, in Canada, the IMIA Code of Ethics is used; however, there are no guidelines or policies that specifically address the current research findings at the intersection of robotics and dementia care that fully reflect the relevant IMIA ethical principles. The potential implications of ethical breaches and mishaps are discussed in the literature. However, there is limited research regarding ethics in the conception, design and development of robotic devices *prior* to implementation.

6. Conclusion

Robots are considered a treatment and symptom management option for patients with dementia and have been seen to have numerous benefits to patients and caregivers. Further investigation is required to assess the ethical implications of robotic devices, and the potential ramifications that they may have on patients, their families and society. Additionally, a guide should be developed in order to evaluate the suitability of a device before implementation to reduce the chances adverse events from occurring.

Acknowledgements

Evangeline Wagner received funding from the Jamie Cassel's Undergraduate Research Award to conduct this research. Elizabeth Borycki receives funding from the Michael Smith Foundation for Health Research, Vancouver Canada to support her research work.

References

[1] Duong S, Patel T, Chang F. Dementia: What pharmacists need to know. Canadian Pharmacists Journal/Revue des Pharmaciens du Canada. 2017 Mar;150(2):118-29.
[2] Government of Canada [Internet]. Dementia in Canada, including Alzheimer's Disease: Highlights from the Canadian Chronic Disease Surveillance System - Canada.ca. / Gouvernement du Canada; 2017 [cited 2021 Aug 18]. Available from: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/dementia-highlights-canadian-chronic-disease-surveillance.html
[3] Kluge E-H. The IMIA code of ethics for health information professionals [Internet]. The IMIA Code of Ethics for Health Information Professionals. 2016. Available from: https://imia-medinfo.org/wp/wp-content/uploads/2015/07/IMIA-Code-of-Ethics-2016.pdf
[4] Riek LD. Healthcare robotics. Communications of the ACM. 2017 Oct 24;60(11):68-78.
[5] Clotet E, Martínez D, Moreno J, Tresanchez M, Palacín J. Assistant personal robot (APR): Conception and application of a tele-operated assisted living robot. Sensors. 2016 May;16(5):610.
[6] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International journal of social research methodology. 2005 Feb 1;8(1):19-32.
[7] Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implementation science. 2010 Dec;5(1):1-9.
[8] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MD, Horsley T, Weeks L, Hempel S. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Annals of internal medicine. 2018 Oct 2;169(7):467-73.
[9] Bradwell HL, Winnington R, Thill S, Jones RB. Ethical perceptions towards real-world use of companion robots with older people and people with dementia: survey opinions among younger adults. BMC geriatrics. 2020 Dec;20(1):1-0.
[10] Burmeister OK. The development of assistive dementia technology that accounts for the values of those affected by its use. Ethics and Information Technology. 2016 Sep;18(3):185-98.
[11] Ienca M, Wangmo T, Jotterand F, Kressig RW, Elger B. Ethical design of intelligent assistive technologies for dementia: a descriptive review. Science and engineering ethics. 2018 Aug;24(4):1035-55.
[12] Jaakola J. Ethics by Other Means? Care Robot Trials as Ethics-in-Practice. TECNOSCIENZA: Italian Journal of Science & Technology Studies. 2021 Mar 9;11(2):53-72.
[13] Johansson-Pajala RM, Gustafsson C. Significant challenges when introducing care robots in Swedish elder care. Disability and Rehabilitation: Assistive Technology. 2020 Jun 10:1-1.
[14] Portacolone E, Halpern J, Luxenberg J, Harrison KL, Covinsky KE. Ethical issues raised by the introduction of artificial companions to older adults with cognitive impairment: A call for interdisciplinary collaborations. Journal of Alzheimer's Disease. 2020 Jan 1;76(2):445-55.
[15] Sætra HS. The foundations of a policy for the use of social robots in care. Technology in society. 2020 Nov 1;63:101383.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210935

Development of Algorithms for Automated Timed Up-and-Go Test Subtask and Step Frequency Analysis

Ane DIZ FELIPE[a,b,c], Andreas ZIEGL[a,b,1], Dieter HAYN[a,d] and Günter SCHREIER[a,b]
[a] *AIT Austrian Institute of Technology GmbH, Graz, Austria*
[b] *Graz University of Technology, Austria*
[c] *Universidad Politécnica de Madrid, Spain*
[d] *Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*

Abstract. Frailty is one of the major problems associated with an aging society. Therefore, frailty assessment tools which support early detection and autonomous monitoring of the frailty status are heavily needed. One of the most used tests for functional assessment of the elderly is the "Timed Up-and-Go" test. In previous projects, we have developed an ultrasound-based device that enables performing the test autonomously. This paper described the development and validation of algorithms for detection of subtasks (stand up, walk, turn around, walk, sit down) and for step frequency estimation from the Timed Up-and-Go signals. The algorithms have been tested with an annotated test set recorded in 8 healthy subjects. The mean error for the developed subtask transition detection algorithms was in between 0.22 and 0.35 s. The mean step frequency error was 0.15 Hz. Future steps will include prospective evaluation of the algorithms with elderly people.

Keywords. Frailty, timed up-and-go test, tele-health, monitoring

1. Introduction

The growth in life expectancy in the past decades has led to increasing frailty, which is a risk factor for many diseases [1]. To prevent the associated negative health outcomes, an early identification of the frailty signs and symptoms is needed. There are various approaches for automatically assessing frailty with different devices, which are very well summarized in [2] and [3]. One of the most promising frailty-assessment tools especially for home-based, autonomous scenarios, is the Timed Up-and-Go (TUG) test [4]. While the total TUG time has already proven to be a good frailty measure, it has been shown that considering also sub-tasks of the TUG test [5] and gait speed [6] can further improve frailty assessments. In a previous work, we have developed an ultrasound-based TUG measurement device to be used autonomously by patients at risk of frailty [7]. Initial experiments for subtask detection showed promising results [8], but still left some room for improvement. This paper describes the development of advanced algorithms for subtasks detection and for step frequency estimation from TUG signals.

[1] Corresponding author: Andreas Ziegl, Center for Health & Bioresources, AIT Austrian Institute of Technology, Reininghausstr. 13, 8020 Graz, Austria; E-mail: andreas.ziegl@ait.ac.at.

2. Methods

2.1. Recording of supervised TUG tests

We recorded healthy volunteers performing supervised TUG tests. TUG device data were captured together with audio and video data using a camera in an Android mobile phone. Data automatically derived from the TUG recordings were compared to a ground truth, which was manually extracted from the audio and video recordings.

2.2. Subtask classification

We designed and compared algorithms based on three different approaches for subtask detection. For filtering and smoothing, a Gauss distribution was fit to the original TUG signal. Four transitions in between the five subtasks of the TUG test were annotated: stand up/walk forward (T1), walk forward/turn around (T2) turn around/walk back (T3) walk back/sit down (T4). Figure 1 shows an example of a test signal with the moments of the transitions marked by the vertical lines.

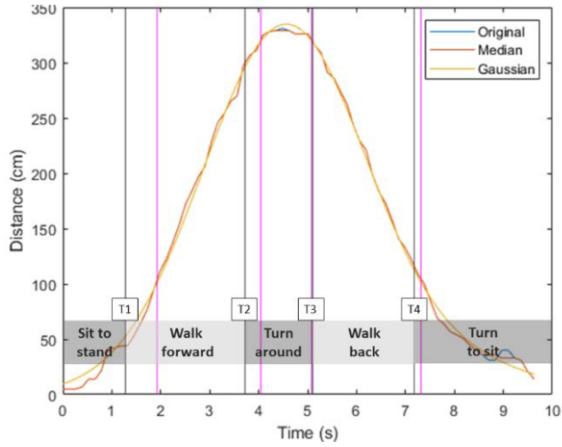


Figure 1. Example of a test signal. The curve represents the distance from the subject to the device placed in the backrest of the chair. TUG subtasks and the transitions between them according to reference are represented in gray. Transition times estimated with one of the algorithms are marked by pink vertical lines.

Approach 1 was based on our previous approach published in [8], which separated the subtasks depending on specific distances of the subject to the device. However, no reference subtask annotations were available when developing the approach in [8]. Therefore, for approach 2, we optimized the threshold distances based on the recorded videos. For approach 3, T1 and T4 were estimated before the maximum and at the minimum of the polynomial fitting of the derivatives of the signals. Table 1 presents an explanation of the different criteria established for the detection of the subtask transitions.

Table 1. Explanation of three different approaches to separate subtasks: stand up/walk forward (T1), walk forward/turn around (T2) turn around/walk back (T3) walk back/sit down (T4).

Approach	T1	T2	T3	T4
1	1 m (way out)	3.2 m (way out)	3.2 m (way back)	1 m (way back)
2	0.5 m (way out)	3 m (way out)	3.3 m (way back)	1.2 m (way back)
3	Before the maximum of the derivative's polynomial fitting	-	-	At the minimum of the polynomial fitting

2.3. Step frequency

We considered two different approaches for estimating the step frequency. In the frequency domain, after applying some processing steps to the signals, we transformed them to the frequency domain via a fast Fourier transform. We considered the highest peak in the spectrum as the estimated frequency. In the time domain, we derived the number of steps made during the test from the derivative curve of the signal. Specifically, the number of peaks higher than a certain value in the derivative signal was assessed with each of the peaks corresponding to one step. The peaks were counted for the walking period. After estimating the number of steps for each test, we calculated the step frequency dividing the number of steps in a test by the duration of the test.

3. Results

3.1. Dataset

Eight healthy volunteers (20 to 29 years, 5 female) each performed 10 tests, resulting in 80 TUG, video and audio signals. The average test duration was of 10.01 s. The average difference between total TUG time measured by the device and respective reference values from the videos was 0.19 ± 0.16 seconds (mean relative error: 1.50%).

3.2. Analysis of the subtask classification strategies

Table 2 summarizes the results for all the approaches. The mean time difference refers to the comparison between the real subtask transition times (extracted from the videos) and the times obtained by establishing the corresponding approach.

Table 2. Comparison of subtask times as calculated based on criteria with the ground truth derived from video recordings including correlation coefficients R.

Approach 1	**T1**	**T2**	**T3**	**T4**
R (p-value)	0.65 ($p < 0.001$)	0.93 ($p < 0.001$)	0.87 ($p < 0.001$)	0.97 ($p < 0.001$)
Mean ± std	0.71 ± 0.34 s	0.28 ± 0.24 s	0.41 ± 0.48 s	0.34 ± 0.34 s
Approach 2	**T1**	**T2**	**T3**	**T4**
R (p-value)	0.48 ($p < 0.001$)	0.91 ($p < 0.001$)	0.86 ($p < 0.001$)	0.95 ($p < 0.001$)
Mean ± std	0.25 ± 0.27 s	0.25 ± 0.23 s	0.35 ± 0.48 s	0.28 ± 0.37 s
Approach 3	**T1**	**T2**	**T3**	**T4**
R (p-value)	0.62 ($p < 0.001$)	-	-	0.97 ($p < 0.001$)
Mean ± std	0.22 ± 0.17 s	-	-	0.27 ± 0.22 s

3.3. Analysis of the step frequency calculation strategies

Estimated step frequencies were compared with the reference frequencies extracted from the videos. In the frequency domain, the difference between the measurements was considerably high and with respect to the correlation, the results were not significant. For the analysis in the time domain of the step frequency of the participants, the estimated number of steps was translated to frequency by dividing it by the total duration of the test. The mean difference and standard deviation between the real and the estimated frequencies for all subjects was 0.15 ± 0.12 Hz ($R=0.45$, $p < 0.05$).

4. Discussion

Regarding the validation of the TUG device, the level of accuracy ensures that the small error would not influence the identification of the subject's frailty. Overall, it can be stated that the TUG device is also useful to detect where each of the subtasks starts and ends autonomously, avoiding the use of complex camera systems [9] or sensors [10]. For the gait speed estimation, in the frequency domain, the results cannot be considered worthwhile since a p-value of $p > 0.05$ leads to discard this method. In the time domain, the results were significantly improved as the mean error of the estimation as well as the p-value were considerably decreased.

The newly developed algorithms have been optimized and tested based on the same test set. A prospective study is currently planned, to validate the results and to exclude potential overfitting. There may be imprecision in the results coming from the inaccuracy at the determination of the "real" parameters. If different people had determined the values for those parameters, they may not have given the exact same results.

All the investigation proceeded in this work should be extended to elderly people. It should be studied if the designed algorithms can also be applied to signals obtained from elderly people which may also have mobility limitations. All the processing steps developed could be applied to previously recorded TUG device signals for other studies. In [7] fall risk of an elderly population was assessed by discrimination between fallers and non-fallers. For this purpose, different features of the TUG test were considered. This investigation could be extended by evaluating the differences in gait speed or in the duration of specific subtasks between the two groups.

References

[1] Sukkriang N, Punsawad C. Comparison of geriatric assessment tools for frailty among community elderly. Heliyon. 2020 Sep 14;6(9):e04797.

[2] Cobo A, Villalba-Mora E, Pérez-Rodríguez R, Ferre X, Rodríguez-Mañas L. Unobtrusive Sensors for the Assessment of Older Adult's Frailty: A Scoping Review. Sensors (Basel). 2021 Apr 23;21(9):2983.

[3] Vavasour G, Giggins OM, Doyle J, Kelly D. How wearable sensors have been utilised to evaluate frailty in older adults: a systematic review. J Neuroeng Rehabil. 2021 Jul 8;18(1):112.

[4] Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991 Feb;39(2):142-8.

[5] Ansai JH, Farche ACS, Rossi PG, de Andrade LP, Nakagawa TH, Takahashi ACM. Performance of Different Timed Up and Go Subtasks in Frailty Syndrome. J Geriatr Phys Ther. 2019 Oct/Dec;42(4):287-293.

[6] Jung HW, Jang IY, Lee CK, Yu SS, Hwang JK, Jeon C, Lee YS, Lee E. Usual gait speed is associated with frailty status, institutionalization, and mortality in community-dwelling rural older adults: a longitudinal analysis of the Aging Study of Pyeongchang Rural Area. Clin Interv Aging. 2018 Jun 6;13:1079-1089.

[7] Ziegl A, Hayn D, Kastner P, Löffler K, Weidinger L, Brix B, Goswami N, Schreier G. Quantitative falls risk assessment in elderly people: results from a clinical study with distance based timed up-and-go test recordings. Physiol Meas. 2020 Dec 17;41(11):115006.

[8] Ziegl A, Kastner P, Modre-Osprian R, Schreier G. Automated Measurement and Subtask Analysis of the Timed Up-and-Go Test in the Field of Geriatrics. Annu Int Conf IEEE Eng Med Biol Soc. 2018 Jul;2018:1526-1529.

[9] Ansai JH, Farche ACS, Rossi PG, de Andrade LP, Nakagawa TH, Takahashi ACM. Performance of Different Timed Up and Go Subtasks in Frailty Syndrome. J Geriatr Phys Ther. 2019 Oct/Dec;42(4):287-293. doi: 10.1519/JPT.0000000000000162. PMID: 29210935.

[10] Coulthard JT, Treen TT, Oates AR, Lanovaz JL. Evaluation of an inertial sensor system for analysis of timed-up-and-go under dual-task demands. Gait Posture. 2015 May;41(4):882-7. doi: 10.1016/j.gaitpost.2015.03.009. Epub 2015 Mar 21. PMID: 25827680.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210936

Digital Health Platform for Emotional and Self-Management Support of Caregivers of Children Receiving Growth Hormone Treatment

Gabriel SIGNORELLI[a], Francisco José NÚÑEZ-BENJUMEA[a], Antonio De Arriba MUÑOZ[b], Marta Vara CALLAU[b], Maitena PIERANTONELLI[c], Alba JIMÉNEZ-DÍAZ[c] and Luis FERNANDEZ-LUQUE[a,1]
[a] *Adhera Health, Inc, Palo Alto, CA, USA*
[b] *Pediatric Endocrinology unit, Miguel Servet Children's University Hospital, Zaragoza, Spain*
[c] *Labpsitec, Universitat de València, Valencia, Spain*

Abstract. In recent years there has been growing research on the combination of evidence-based behavioral change techniques with mobile-based recommender systems. In this paper, we have focused on understanding the psychological burdens experienced by caregivers of children undergoing growth hormone treatment (GHt) and the perceived barriers to and drivers of the adoption of a digital health solution. This is a mixed-methods formative research study looking into technical acceptance aspects of using digital health for the emotional support of parents of children undergoing GHt. After one month using the ADHERA CARING platform (Adhera Health, Inc., Palo Alto, CA), individual semi-structured interviews were conducted. ADHERA CARING provides tailored emotional and self-management support to caregivers of children undergoing GHt to improve adherence to treatment through positive education, personalized motivational messages, and emotional support. A preliminary thematic analysis and categorization were carried out, based on the Behavioral Intervention Technology (BIT) model. The majority of participants were female. All caregivers positively valued having the tool, especially at the beginning of treatment. Information provided in the educational module was useful and improved self-efficacy. Motivational messages contributed to commitment and reinforced the educational content, thus promoting continuity of treatment and potentially improving treatment efficacy. Most participants (n=10, 80%) accessed all educational units and completed all the 27 quiz questions. Regarding the motivational messages, the overall average rating was 4.55 out of 5.00. ADHERA CARING has the potential to help caregivers to understand the treatment journey. Nevertheless, users have identified that some types of educational content are more valuable at specific stages of the treatment journey, which suggests that personalization of educational content is required to adapt to different stages of the patient journey.

Keywords. Digital Health, Growth Hormone, Behavioral Intervention, Caregiver

[1] Corresponding Author, Fernandez-Luque, Adhera Health, Inc, Palo Alto, CA, USA; E-mail: luis@adherahealth.com

1. Introduction

Adherence to growth hormone treatment (GHt) among children is variable and remains a problem [1]. Prior research has been looking into digital health tools to support GHt treatment [2], that includes mobile applications [3], and more advanced digital health ecosystems that integrate mobile solutions with connected injection devices [4].

One of the key stakeholders to be supported are the caregivers, normally parents, of children that require long term pharmacological treatment since research shows the negative impact of stress on adherence and overall quality of life [5]. In this paper, we present a study exploring the feasibility of a mobile digital intervention for emotional and self-management support of caregivers of children receiving GHt. The intervention relies on a digital health platform powered by a Health Recommender System [6] that incorporates self-management educational content, motivational messages, and interactive mental well-being content addressing aspects such as stress management, self-esteem [7] and compassion [8].

2. Methods

This is a mixed methods formative research study looking into technical acceptance aspects of using digital health for the emotional support to parents of children undergoing GHt. The study was conducted in the Pediatric Endocrinology unit of the Hospital Miguel Servet (Zaragoza, Spain). It received the clearance from the hospital ethical board and the protocol is publicly available at clinicaltrials.gov (NCT04812665).

After a one-month period of using the ADHERA CARING platform (Adhera Health, Inc., Palo Alto, CA), individual semi-structured interviews were conducted. These interviews addressed the psychological burdens experienced by caregivers of children receiving GHt and the barriers and facilitators for the adoption of the mobile solution were discussed in relation to cargivers' individualized needs and expectations related to children's health outcomes.

The interviews were audio recorded and transcribed by one research assistant, and a preliminary thematic analysis and categorization was carried out, based on the Behavioral Intervention Technology Model (BIT) [8], which provides an integrated conceptual and technological framework for eHealth interventions. A preliminary thematic analysis and categorization was carried out, based on the BIT model [8], which provides an integrated conceptual and technological framework for eHealth and mHealth interventions. Demographic and engagement data was also analyzed using descriptive statistics.

2.1. mHealth-based digital intervention

ADHERA CARING provides tailored emotional and self-management support to caregivers of children undergoing GHt to improve adherence to treatment through positive education, personalized motivational messages, and emotional support. Once onboarded, users follow an empowerment program with easy-to-digest educational, personalized, and actionable content included for them, and also perform mental well-being activities.

Education is provided via different educational pieces of text classified by categories and organized in units. These contents were fully accessible from day one of user

enrollment. Whenever a unit is tagged as read, its icon turns green to reflect the progress (Figure 1-A). A progress bar is also displayed and advances as the units are tagged as read by the users. Within each category, the last educational unit is a quiz. The quiz is made up of several questions with multiple choices (Figure 1-B). An explanation of the answer is provided right after answering the question. The quiz is only available once all the other units within the category have been marked as read.

Motivational messages are delivered into an inbox in the app (Figure 1-C). Every time the app receives a new message, a push notification is triggered so that the user can access the message directly. Messages can be rated by users with a score from 1 to 5 stars. Messages are delivered by a health recommender system [5]. This algorithm works in two steps. The first step is a knowledge-based algorithm to filter incompatible messages with the user profile. The remaining messages are then passed to a collaborative filtering algorithm, that uses the demographic profile and the user ratings to compute a similarity measurement between the user profile and each message.

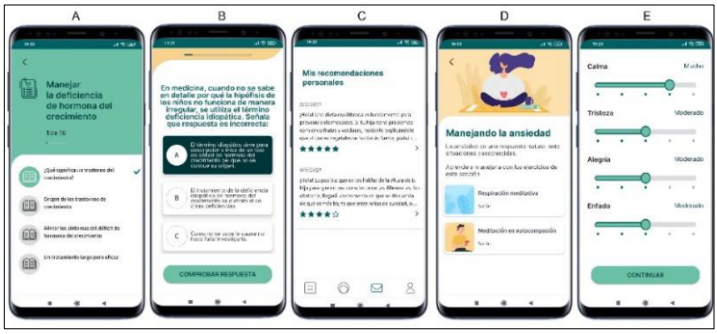


Figure 1. Screenshots of the ADHERA CARING Platform.

The mental well-being module (Figure 1-D) includes two exercises: 'Mindful breathing' and 'Self-compassion meditation'. These are audio-guided sessions that users can complete at their convenience. Before and after each session, the user is asked to complete a well-being assessment (Figure 1-E).

3. Results

The majority of participants were female (80%) with an average age of 45 years. The interviewees positively valued having a tool that provides them with the necessary information and support for emotional well-being to face stressful situations and approach the treatment in a more natural and calm way, avoiding the distress transfer from the caregiver to the child. Notably, many of the caregivers highlighted the importance of having the support of the solution especially at the early stages of treatment.

Regarding the behavior change strategies used to achieve the objectives, the caregivers liked the structure of small topics to facilitate the understanding of each theme. Therefore, the educational content seems to be important to reinforce and remind caregivers of some aspects of patient treatment. The emotional management tips and educational content were also raised as a good initiative. The contents provided by the ADHERA CARING health recommender system (AI based) were perceived useful by the users. Caregivers liked the messages and their frequency. They also found positive

the strategy to reinforce the educational learnings. In general, the interviewees agreed that ADHERA CARING very intuitive and easy to use, with an attractive design.

Engagement data of the ADHERA CARING platform (Table 1) indicates that 80% of the total of 10 participants have accessed all 39 available educational units and completed all the 27 available quiz questions. The overall average rating was 4.55 out of 5.00.

Table 1. Engagement Data

Participant	Education units accessed	Education units completed	Quiz Questions completed	Messages rated	Average Rating
1	39 (100%)	39 (100%)	27 (100%)	-	-
2	39 (100%)	39 (100%)	27 (100%)	17	5.00
3	39 (100%)	39 (100%)	27 (100%)	41	5.00
4	39 (100%)	39 (100%)	27 (100%)	46	5.00
5	39 (100%)	39 (100%)	27 (100%)	32	4.03
6	39 (100%)	39 (100%)	27 (100%)	6	5.00
7	39 (100%)	39 (100%)	27 (100%)	36	4.11
8	39 (100%)	4 (10%)	0 (0%)	-	-
9	39 (100%)	15 (38%)	9 (33%)	6	3.33
10	39 (100%)	39 (100%)	27 (100%)	-	-

4. Conclusions

The ADHERA CARING platform has potential to help caregivers to understand the treatment journey, specially at the beginning of the GHt, with a friendly interface and easy to digest educational content organized in small units and facilitating the reinforcement of information through personalized messages and quizzes. Nevertheless, users have identified that some types of educational content are more valuable at specific stages of the treatment journey than others, which suggest that deeper personalization adapting to the stage of patient journey is required.

Acknowledgements

We would like to thank the rest of the team who contributed to this project and the caregivers for their participation. Disclaimer: This investigator-initiated study was financially supported by Merck Healthcare KGaA, Darmstadt Germany (CrossRef Funder ID: 10.13039/100009945). Merck reviewed the manuscript for medical accuracy only before the conference submission. The authors are fully responsible for the content of this manuscript, and the views and opinions described in the publication reflect solely those of the authors.

References

[1] Haverkamp F, Gasteyger C. A review of biopsychosocial strategies to prevent and overcome early-recognized poor adherence in growth hormone therapy of children. Journal of medical economics. 2011 Jan 1;14(4):448-57.

[2] Dimitri P, Fernandez-Luque L, Banerjee I, Bergadá I, Calliari LE, Dahlgren J, de Arriba A, Lapatto R, Reinehr T, Senniappan S, Thomas-Teinturier C. An eHealth Framework for Managing Pediatric Growth

Disorders and Growth Hormone Therapy. Journal of Medical Internet Research. 2021 May 20;23(5):e27446.

[3] Fernandez-Luque L, Labarta JI, Palmer E, Koledova E. Content analysis of apps for growth monitoring and growth hormone treatment: systematic search in the android app store. JMIR mHealth and uHealth. 2020 Feb 18;8(2):e16208.

[4] Koledova E, Le Masne Q, Spataru A, Bagha M, Dixon D. Digital Health in the Management of Pediatric Growth Hormone Therapy–10 Years of Developments. Studies in Health Technology and Informatics. 2021 May 1;281:926-30.

[5] Armstrong ML, Duncan CL, Stokes JO, Pereira D. Association of caregiver health beliefs and parenting stress with medication adherence in preschoolers with asthma. Journal of Asthma. 2014 May 1;51(4):366-72.

[6] Hors-Fraile S, Malwade S, Luna-Perejon F, Amaya C, Civit A, Schneider F, Bamidis P, Syed-Abdul S, Li YC, De Vries H. Opening the black box: Explaining the process of basing a health recommender system on the i-change behavioral change model. IEEE Access. 2019 Dec 4;7:176525-40.

[7] Germer CK, Neff KD. Self‐compassion in clinical practice. Journal of clinical psychology. 2013;69(8):856-867.

[8] Bedrov A, Bulaj G. Improving self-esteem with motivational quotes: opportunities for digital health technologies for people with chronic disorders. Frontiers in psychology 9. 2018:2126.

[9] Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The Behavioral Intervention Technology Model: An Integrated Conceptual and Technological Framework for eHealth and mHealth Interventions. J Med Internet Res 2014;16(6):e146. doi: 10.2196/jmir.3077.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210937

Evaluation of Meditation Apps Available on Google Play and Apple Store: An App Review

Raghad MUHIYADDIN[a], Alaa ABD-ALRAZAQ[a], Zubair SHAH[a], Tanvir ALAM[a] and Mowafa HOUSEH[a,1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*

Abstract. Many meditation apps have been used to improve the mental wellbeing of individuals. However, little information is available regarding the quality of the applications. This study aims to evaluate meditation apps using the Mobile Applications Rating Scale (MARS). A systematic search for meditation apps was performed on both Android Google Play and Apple iOS Store. We used two keywords to search both app stores: meditation and mindfulness. Out of 623 apps identified, 334 apps were excluded due to language, containing only reminders to meditate, or for not being accessible. A total number of 289 apps remained, of which 280 apps were excluded for being information-only focused, containing religious practices, eating habits, exercises, or for not being free. Therefore, nine apps were included in this review for evaluation. The MARS ratings used in this app review were based on scores from a prior study conducted. The mobile app Headspace had the highest average (4), which is rated as 'good' based on MARS. The remaining apps were rated as acceptable with averages that ranged from 3.2-3.7.

Keywords. Meditation apps, mindfulness apps, Mobile Applications Ratings Scale (MARS), app review

1. Introduction

There is significant research discussing the impact of meditation on an individual's well-being. Mindfulness has been found to be beneficial to mental health and is being adopted in medicine both for therapeutic applications as well as in creating protocols for treating conditions such as chronic pain and stress [1]. Meditation is a skill that requires sustained practice and efforts to be effective; this is a challenge for training that implemented in traditional face-to-face or app-based settings. Apps allow for ease of access to practice meditation; additionally, apps designed to interact with users and are aesthetically pleasing are more likely to regularly engage users in performing meditation-based practice activities.Several meditation-based apps have the potential to improve mental wellbeing. Although these apps are easily accessible on various app stores, limited information is available regarding the quality of the applications. The objective of this app review is to evaluate meditation apps available on Android Google Play and Apple iOS Store using the Mobile Applications Rating Scale (MARS) [2].

[1] Corresponding Author: Dr. Mowafa Househ, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar; Email: mhouseh@hbku.edu.qa

2. Methods

2.1. App Search and Selection

A systematic search for meditation apps was conducted in February 2021. For Android, the Google Play Store was used to conduct this search, while the App Store was used for iOS devices. The terms used in searching the app stores were meditation and mindfulness. As for the eligibility criteria, apps were included if they were accessible, included meditation exercises, being written in English, and available for free with in-app-purchases. They were excluded if they only contained reminders for users to meditate, or if they incorporated religious practices, eating habits, exercises, or focused on pregnancy. They were also excluded if they were only information-based and were not for free. In the first phase, after the apps were found, apps were included or excluded based on language and accessibility. In the second phase, the apps were included or excluded based on the app's cost and content.

2.2. App Evaluation

The quality of the apps was assessed by using the MARS rating scale. The MARS scale consists of 23 items classified into 3 categories; classification, satisfaction, and quality of the app. The MARS adopts a 5-point scale, with the ratings noted to include 1-inadequate, 2-Poor, 3-Acceptable, 4-Good, 5-Excellent. Included apps were rated using the satisfaction category, and four subscales of the app quality category. These include functionality, engagement, information quality, and aesthetics. The MARS score is obtained through computing the mean of the app quality subscales and overall mean score; subjective items are rated separately. MARS ratings used in this review were based on scores from a previously conducted study [3].

3. Results

3.1. Search Results

By searching Google Play Store and App Store, we identified 246 and 377 apps, respectively. A total of 623 apps were found (Appendix 1). In the preliminary screening phase, we excluded 334 apps for the following reasons: they were not easily accessible (n=10), they were not available in English (n=38), they only reminded users of time to meditate, guided meditation tracks, games, etc. (n=286). In the second phase, we excluded 280 apps because they incorporated religious practices, eating, exercise, or pregnancy (n=183), were information-focused (n=84), or not free (n=13). A total of nine apps were included for evaluation.

3.2. App Quality Based on MARS

Appendix 2 shows the rating of the MARS in each category: engagement, functionality, aesthetics, information, satisfaction, and then the overall score of all these categories. Among the apps, Headspace had the highest overall score (4), which indicates a 'good' quality based on MARS. On the other hand, the apps Mindfully Me and Mindfulness

Coach had the lowest overall score (3.2), which indicates an 'acceptable' quality based on MARS. The overall score ranged between 3.3 and 3.7 for the remaining, indicating an 'acceptable' quality. Engagement and satisfaction were the lowest-rated aspects of the apps with an average of 3.2 whereas functionality was the highest-rated aspect of the apps with an average of 3.2.

3.3. Features of Meditation Apps

Our results show that apps with timers, tracking facilities, the ability to set reminders and link to social media are considered to be high-quality meditation apps [3]. The features we looked for in apps were in-app purchases, timers, reminders, tracking, and social media features (Appendix 3). Only three apps had all the above-mentioned features: Headspace, Mindfulness Daily, and Rhythm Free. On the other hand, one app (i.e., Complete Mindfulness) did not have any of these features. Tracking facilities were the most available feature in the apps (n=8). In-app purchases was the least available feature in the apps (n=3).

4. Discussion

4.1. Principal Findings

Of the reviewed apps, different methods were used in explaining the concept of meditation and mindfulness. However, there was variance to the extent of explanation, with apps like Headspace and Smiling Mind noted to adopt unique visual methods in explaining the benefits of meditation and mindfulness.

Meditation is a skill that requires sustained practice and efforts for it to be effective, and this is a challenge for both training that is implemented in a traditional face-to-face setting or app-based training. Apps allow for ease of access to meditation practice; apps which have been designed to be interactive and are aesthetically pleasing are most likely to engage users in regularly performing meditation-based practice activities. With regards to engagement, 3.0 was rated to be the minimum acceptable score on the MARS. All of the included apps have met this score. These apps were noted to have high-quality graphics, with interfaces that are simple to adopt and voices that are soothing when guided meditation tasks are being performed. For example, Headspace is noted to use short videos to aid the meditation tracks or serve as a guide during meditation.

Additionally, an app that features a user-base community may motivate users to engage in healthy activities, as these communities could aid users in discussing and sharing their experiences and challenges of meditation practice. This could serve as a support or replacement for the provided support usually present in face-to-face meditation practices. Although more than 60% of the apps reviewed provided some sort of link to other social media platforms, Headspace was the only app that included community support as part of its features.

4.2. Practical and Research Implications

Our review did not evaluate the effectiveness of mediation apps on mental health. We recommend that systematic reviews should be conducted to examine the effectiveness of

meditation apps on mental health. In addition, we only focused on free or in-app purchase apps, so further app reviews are needed to evaluate the quality of non-free apps. Additionally, we did not search other app stores such as HUAWEI AppGallery. Researchers should consider exploring other app stores in addition to the App Store and Google Play store. Healthcare professionals may be interested in recommending the meditation apps that have the highest overall score and the largest number of features to patients. We also recommend that app developers consider user engagement and satisfaction when creating future apps. Finally, companies may benefit by providing their employees with meditation apps that are free to use (or havein-app purchases), rather than asking workers to use paid apps.

4.3. Strengths and Limitations

The study adopted the use of MARS in the measurement of the quality of apps. This was dependent on four objective subscales including information quality, visual aesthetics, functionality and engagement, and one subjective scale of satisfaction. Only the objective scales are incorporated in the total app score as per the quality scales. This indicates the reliability of the measure as to the rating of the quality of apps. It should be noted that MARS scores adopted in this study were based on prior MARS results for the apps identified. We may have excluded several meditations apps because we studied free, English-written apps available at the App Store and Google Play Store.

5. Conclusions

Most mediation apps that are available to the public and in the English language have an acceptable quality. However, there is still room to improve mediations apps in order to achieve a higher quality. Healthcare professionals may want to consider recommending mediation apps that have the highest overall score and the largest number of features to their patients. App developers should consider features that would best appeal to user engagement and satisfaction in addition to the remaining aspects of MARS. Systematic reviews should be conducted to evaluate the effectiveness of mediation apps. Appendix files are available at GitHub: https://github.com/rmuhiyaddin/Appendix.git.

References

[1] Choudhary N, Gill C. Universal Panacea or Organizational Fad? The Role of Context on the Effectiveness of Mindfulness. Academy of Management Proceedings. 2020;2020(1):15355. Crossref, doi:10.5465/ambpp.2020.15355abstract.

[2] 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 and uHealth. 2015 Mar 11;3(1):e3422.

[3] Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR mHealth and uHealth. 2015 Aug 19;3(3):e4328.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210938

Features of Meditation Apps: A Scoping Review

Raghad MUHIYADDIN[a], Alaa ABD-ALRAZAQ[a], Mohannad ALAJLANI[b], Zubair SHAH[a], Tanvir ALAM[a] and Mowafa HOUSEH[a,1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*
[b] *Institute of Digital Healthcare, WMG, University of Warwick, Warwick, United Kingdom*

Abstract. This review aims to provide an overview of the features of meditation apps as described in empirical literature. Nine databases were searched for this review. Search terms were related to all types of meditation. Study selection and data extraction of the included studies were conducted by two reviewers. We included 93 studies in this review. Headspace was the most common app among studies and the most common type of meditation was mindfulness. Stress was the most targeted health condition by the studies. Future research needs to focus on different mental conditions other than stress to understand the effect of meditation apps on mental health.

Keywords. Meditation, mindfulness, mobile applications, scoping review

1. Introduction

Meditation is a well-known practice that helps in reducing stress, depression, anxiety, and other mental disorders [1]. It is the process of an individual focusing their awareness to relax the mind and body [2]. In the past decade, meditation has become accessible ditially. People can access meditation applications through their phones, tablets, or computers. The main advantage of online-based meditation programs is convenience because of their ubiquitousness and low cost. There are many meditation apps on the market, including well-known apps such as Headspace, Calm, and Insight Timer. However, most studies conducted investigating these meditation apps focus on mental disorders. No scoping review has explore and summarize the features of meditation apps, despite the importance of this information. To fill this gap, this review aims to provide an overview of the features of meditation apps as described in the empirical literature.

2. Methods

The guidelines of the Joanna Briggs Institute (JBI) methodology were followed to conduct this scoping review [3]. Nine bibliographic databases were searched for this review: MEDLINE, EMBASE, PsycINFO, CINAHL, Scopus, IEEE Xplore, Cochrane

[1] Corresponding Author: Dr. Mowafa Househ, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar; Email: mhouseh@hbku.edu.qa.

Library, Google Scholar, and ACM Digital Library. Because Google Scholar recommended thousands of results, only the first 100 studies were included. Search terms that were chosen are based on the intervention (meditation) and the platform (mobile applications). The search strings used to search all of the above databases can be found in Appendix 1. No restrictions were applied on population (i.e., users of the app), measured outcome, setting, country of publication, and year of publication. The study selection process consisted of screening titles and abstracts of all retrieved studies, then reading full texts of the remaining studies. Data were extracted from the included studies using a data extraction form created in Excel (Appendix 2). The study selection and data extraction processes were carried out by two reviewers (RM & MA) independently. A third reviewer (AA) was consulted in cases of disagreement. Extracted data were summarized using two tables and narratively described.

3. Results

3.1. Search Results

As shown in Appendix 3, 1,201 studies were retrieved from searching the nine databases, and 342 duplicates were removed before starting the study selection process. After scanning the titles and abstracts of the 859 remaining studies, 748 studies were excluded studies due to irrelevant intervention, irrelevant platform, type of publication, and non-English studies. Reading the full text of the remained 111 studies led to excluding 18 studies due to the irrelevant intervention, irrelevant publication type, unavailable full text, and non-English studies. Overall, 93 studies were included in the data synthesis. Appendix 4 shows the list of all included studies.

3.2. Characteristics of Studies

As seen in Appendix 5, most studies included were journal articles (n=77, 82.8%). The included studies were published in 22 different countries; however, about two-thirds of the studies were published in the United States (n=61, 66%). The majority of studies were conducted between 2017 and 2019 (n=80, 86%). About 63% of studies had a sample size of less than 100 (n=59), while the remaining studies had a sample size of 100 or more (n=33, 35.5%). The mean age of participants was only reported in 70 studies, in which the average was 35.2 yrs. Only 85 studies reported the sex of the participants, where 25.8% of participants were males. Almost half of the participants were recruited from the community (n=45, 48.4%). Appendix 6 shows the characteristics of each included study.

3.3. Features of Meditation Apps

As shown in Appendix 7, 50 meditation apps were mentioned in the 93 studies included in this review. The most common app used in the included studies was Headspace (n=17) followed by Calm (n=9), AEON (n=4), and Pacifica (n=4). The studies' featured meditation apps were used for 16 different health conditions. The most common health condition targeted by the meditation apps was stress (n=39), followed by mental health (n=8), depression (n=7), and cancer (n=7). There were other health conditions targeted

such as nicotine dependence (n=3), essential hypertension, and dementia (n=1) (Appendix 7). We identified six different types of meditation delivered by the apps. Mindfulness meditation was the most common type (n=55) followed by guided meditation (n=19), and Mindfulness-Based Stress Reduction (MBSR) (n=8). The majority of apps in the included studies could be operated on both Android as well as iOS (n=72) devices. The majority of the reported apps were in English only (n=29) or multi-language (n=30). Some apps were developed in non-English languages (n=11). More than half of the apps were still available in the app store at the time of this study (n=60). Some were unavailable (n=24), while the availability status for some were unknown because the studies did not mention the app used (n=9). As for the cost, 35 apps were free with in-app purchases, 23 apps were completely free, while the costs of 32 apps are unknown. Appendix 8 highlights the characteristics of all included meditation apps.

4. Discussion

4.1. Principal Findings

This scoping review aimed to report the features of meditation apps as reported in the empirical literature. In the 93 included studies, around 50 different apps were mentioned. The top three apps mentioned were Headspace, Calm, and AEON. Headspace is currently the leading app in the market due to many factors such as ease of use, variety of exercises offered, reminders, aesthetics, and being scientifically supported [4]. Although meditation apps were used for various health conditions, they were used mainly for alleviating stress. It is known that meditation has a substantial impact on reducing stress, other health conditions have yet to be targeted. This allows room for improvement for future app developers or conducting future research on different populations with health conditions. Additionally, only one app targeted people with dementia, but more studies discussing meditation's impact on dementia should be conducted, as it has been proven that meditation can help with improving neural markers and improving working memory [5]. The most common type of meditation reported in studies was mindfulness meditation, followed by guided meditation.

4.2. Strengths & Limitations

This review encaptures a large number of studies that explore app features and the effect they have on different health conditions. This review is considered to be inclusive for a few reasons. Firstly, it included 93 studies, which can be considered comprehensive. Secondly, nine major databases were searched. Thirdly, no restrictions were applied on type of meditation, target health condition, population, measured outcome, year of publication, setting, country of publication. In addition, the study selection and data extraction processes were done by two reviewers, which reduces the selection bias.

Although this review was inclusive in many ways, it was at times exclusive. There were many types of meditation-related activities that were excluded such as yogic practices, religious practices, martial-arts training, and exercise programs. Additionally, only studies written in English were included, thereby, we might have missed some studies. We also only included empirical studies with a certain type of publications such as peer-reviewed articles, thesis, conference proceedings. Lastly, we did not assess the

effectiveness and safety of using meditation apps, and, to the best of our knowledge, no previous study investigated this. Therefore, systematic reviews are needed to evaluate the effectiveness and safety of using meditation apps.

4.3. Research & Practical Implications

The majority of the included studies were conducted in developed countries. Future research should be conducted in developing countries, where people suffer more from mental health issues and have insufficient resources to combat them (e.g., psychiatrists) [6-8]. Additionally, further research should be conducted on health conditions other than stress. Because mindfulness meditation was mentioned in more than half of the studies, we encourage researchers to explore types of meditation apps for future research. Specific guidelines should be followed when conducting future research as there was inconsistency in many of the population (e.g., the mean age, and sex) and meditation apps (e.g., language of the app and the cost) characteristics reported. Lastly, app developers should build multi-language meditation apps to grow a user base, rather than limiting their app to one language or region in which users can access the app.

5. Conclusions

This study identified and explored 50 different meditation apps. Some of the most popular apps were Headspace, Calm, Aeon, Pacifica, Craving to Quit, and Situated Interactive Mindfulness App. Although there were 16 health conditions identified, stress was the most targeted health condition. Future research should focus on health conditions other than stress to be able to understand the effect of meditation on many different mental health conditions. Appendix files are available at GitHub: https://github.com/rmuhiyaddin/Appendix.git.

References

[1] Schreiner I, Malcolm JP. The benefits of mindfulness meditation: Changes in emotional states of depression, anxiety, and stress. Behaviour Change. 2008 Sep;25(3):156-68.
[2] Edenfield TM, Saeed SA. An update on mindfulness meditation as a self-help treatment for anxiety and depression. Psychology research and behavior management. 2012;5:131.
[3] Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil HA, Munn Z. Chapter 11: Scoping reviews. JBI Manual for Evidence Synthesis [Internet]. Adelaide: JBI, 2020 [cited 2019 Aug 14]. doi.org/10.46658/JBIMES-20-12.
[4] PsyberGuide. Professional Review of the App Headspace. One Mind PsyberGuide. 10 Dec. 2020. onemindpsyberguide.org/expert-review/headspace-professional-review.
[5] Abbasi J. Meditation App Improves Attention in Young Adults. JAMA. 2019;322(6):495. Crossref, doi:10.1001/jama.2019.11649.
[6] Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The lancet. 2012 Dec 15;380(9859):2197-223.
[7] Oladeji BD, Gureje O. Brain drain: a challenge to global mental health. BJPsych international. 2016 Aug;13(3):61-3.
[8] Anthes E. Mental Health: There's an App for That. Nature. 2016;532(7597):20-23. Crossref, doi:10.1038/532020a.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210939

A Multilingual Browser Platform for Medical Subject Headings

Raphael SCHEIBLE[a,b,1], Philip STRECKER[b], Suhail YAZIJY[b], Fabian THOMCZYK[b], Ruslan TALPA[c], Andrea PUHL[d] and Martin BOEKER[a,b]
[a] *Institute of Artificial Intelligence and Informatics in Medicine, Chair of Medical Informatics, University Hospital rechts der Isar, Technical University of Munich, Munich, Germany*
[b] *Institute of Medical Biometry and Statistics, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany*
[c] *subZero LLC, Oradea, Romania*
[d] *Cochrane Germany, Cochrane Germany Foundation, Freiburg, Germany*

Abstract. The National Library of Medicine (NLM) controls and publishes the thesaurus Medical Subject Headings which is used for indexing PubMed. Besides an XML export, the NLM offers a web based MeSH browser. The platform contains English terms. The German Institute for Medical Documentation and Information (DIMDI) partially translated and published these terms. Recently, the German National Library of Medicine (ZB-MED) overtook the translation of MeSH. However, there is no dedicated platform which focuses on MeSH and covers multiple languages. Here, we address this gap, by offering a modern multilingual searchable MeSH browser. A modular platform using open source technology is presented. The frontend enables the user to search and browse terms and switch between different languages. The current version of the presented MeSH browser contains English and German MeSH terms and can be accessed at https://mesh-browser.de.

Keywords. MeSH-Browser, Web Service, Databases, Medical Subject Headings

1. Introduction

Medical Subject Headings (MeSH) is a hierarchical thesaurus for medical terminology which is controlled and published by the National Library of Medicine (NLM). It is used for indexing PubMed and annually released in XML format. For public access, the NLM offers a web-based MeSH Browser for English terms. The German Institute for Medical Documentation and Information (DIMDI) has been translating MeSH for many years, revising it annually and adding new terms. However, after the National Library of Medicine (NLM) stopped providing the translation system for MeSH in 2019, DIMDI halted its translation activities after releasing their translated 2019 version of MeSH. Mid 2020, the German National Library of Medicine (ZB-MED) decided to take on the MeSH translation task with a semi-automated approach [1]. Nevertheless, new datasets were not published until now, making the aforementioned DIMDI 2019 version the most

[1] Corresponding author, AIIM, Klinikum rechts der Isar der Technischen Universität München, Ismaninger Str. 22, 81675 München, Germany; E-mail: raphael.scheible@tum.de.

recent and reliable German MeSH data to work with. The objective of this work is to present a modern multilingual MeSH browser, fully implemented with open-source technology, based on the DIMDI 2019 dataset in combination with the 2021 English version of MeSH.

2. Methods

2.1. System Design

PostgreSQL has been selected as database software because it is designed to handle a wide range of workloads, from single machines to large-scale web services, while natively supporting JSON. To store the data of MeSH exports, we modeled a multilingual database layout. With PostgREST [2] we added a RESTful API on top of the database. Dedicated views and procedures, which access multiple internal data tables, were exposed as API endpoints for which an OpenAPI documentation was generated. ElasticSearch (ES) was used to provide a fuzzy full-text search. For this purpose, a specially defined search container with searchable content from various data tables was connected to ES by ZomboDB [3], a third party PostgreSQL extension. Unfortunately, PostgREST does not support ZomboDB's operator and query DSL. Thus, to adjust requests to the search endpoint, we leveraged OpenResty. The entire implementation of the backend was accelerated by the PostgREST Starter Kit [4] with its associated CLI tool [5]. The frontend was developed as a single page application (SPA) using the React Admin framework [6]. In order to communicate to an API, the React Admin app requires a compatible data provider which generates HTTP requests depending on the API architecture. In this case, a custom data provider for PostgREST was developed and used [7]. The overall system design is illustrated in Figure 1.

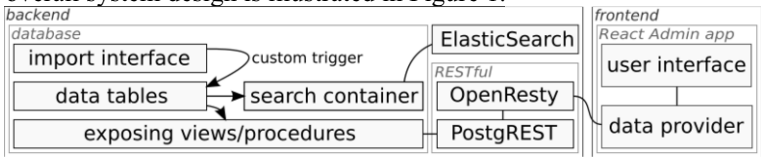


Figure 1. The architecture of the application.

2.2. Data Import

The NLM designed an XML format in which both the English and German MeSH exports are published as individual files. Consequently, a generic import procedure was developed. Specifically, each XML file was flattened into CSV files, partially containing JSON content in order to represent nested structures of the XML format. Subsequently, CSV data was directly written into the database using custom views of the underlying PostgreSQL database. JSON content was processed by trigger functions distributing nested data into the designated data tables. As the translations of MeSH qualifiers were missing in the German XML file, a CSV database export published by the DIMDI was processed to integrate the missing data. The import workflow was implemented using snakemake [8].

3. Results

3.1. REST Interface

The REST interface exposes 10 public endpoints: 7 of them to query specific structures from underlying data tables, 1 to provide answers to search queries and 2 to serve database procedures for querying the tree structure of MeSH. The internal structures consist of 16 tables, 5 views, 5 triggers and 8 procedures. One of the 5 views is for importing data using a customized trigger. Another one is the search container.

3.2. User Interface

The current version of the interface is built with minimal visual complexity and contains four pages to show MeSH information: (1) a landing page with a search field, (2) a list of search results, (3) detailed view for MeSH descriptors (cf. Figure 2) and (4) a MeSH tree browser. The interface of these pages is designed to minimize the number of steps required to reach a desired MeSH entity. Both language selection and link to the MeSH tree browser are accessible on each page. On the landing page, the query input field lists suggestions by utilizing ElasticSearch. When querying, the results are displayed in a list that allows further customization of the search term and sorting by score, MeSH term and entity type. On (3) and (4), detailed information about a specific MeSH descriptor, including related concepts and associated MeSH qualifiers, is displayed. Generally, the user interface labels are internationalized and designed for adding more translations in the future. Additionally, the web application offers consistent URLs and views of MeSH terms, providing shareable links. As a bonus, an integrated Swagger UI displays the OpenAPI documentation and enables users to consume the API endpoints directly through the browser.

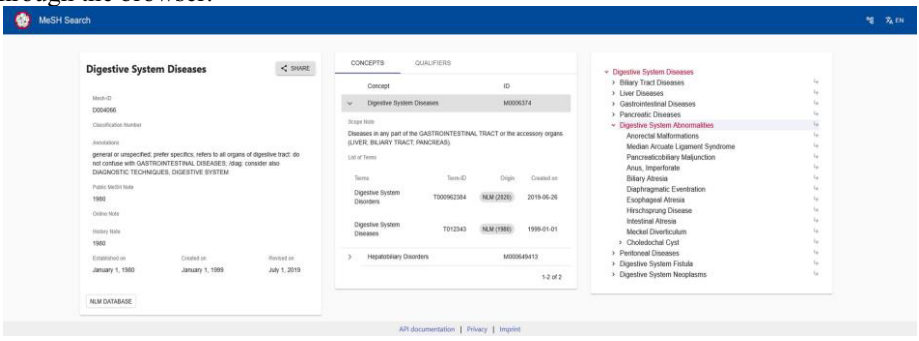


Figure 2. Screenshot of the detailed view of a MeSH descriptor showing the tree and detailed information.

4. Discussion

The proposed system design provides a feasible solution to implement a multilingual MeSH browser. Based on fuzzy search functionality of ES, the platform provides a simple, yet powerful access to the MeSH thesaurus and offers streamlined adaptation of additional languages. This initial use case was restricted to help the team at "Wissen Was Wirkt" [9] - a German-language blog run by Cochrane Germany - to efficiently find German tags and categories for their blog posts using the comprehensive MeSH

vocabulary. Another application could be the preparation of English search queries for MEDLINE, similar as in the abandoned work of the French MeSH Browser [10].

5. Conclusion

The presented MeSH browser was designed for multilingual functionality with modern open-source technology. In addition to a modern user interface, a RESTful API with OpenApi documentation was realized. The current published version contains English and German descriptors and is publicly available at https://mesh-browser.de.

Acknowledgements

This work was supported by the German Ministry for Education and Research (BMBF FKZ 01ZZ1801B, 01ZZ1804A). Ruslan Talpa is the CEO of subZero LLC, the company that develops the PostgREST Starter Kit and its associated CLI tool.

References

[1] Langnickel L, Baum R, Wollnik-Korn G, Fischer-Wagener B, Madan S, Fluck J. The future of German MeSH: a new semi-automatic translation process and new services for search and annotation. GMS Med Inform Biom Epidemiol. 2020 Aug 25;16(2):Doc07.
[2] PostgREST/postgrest: REST API for any Postgres database [Internet]. [cited 2021 Sep 1]. Available from: https://github.com/PostgREST/postgrest
[3] ZomboDB. zombodb/zombodb [Internet]. 2021 [cited 2021 Aug 24]. Available from: https://github.com/zombodb/zombodb
[4] PostgREST Starter Kit [Internet]. subZero; 2021 [cited 2021 Aug 24]. Available from: https://github.com/subzerocloud/postgrest-starter-kit
[5] subZero CLI [Internet]. subZero; 2021 [cited 2021 Aug 24]. Available from: https://github.com/subzerocloud/subzero-cli
[6] marmelab/react-admin: A frontend Framework for building B2B applications running in the browser on top of REST/GraphQL APIs, using ES6, React and Material Design [Internet]. [cited 2021 Aug 24]. Available from: https://github.com/marmelab/react-admin
[7] PostgREST Data Provider For React-Admin [Internet]. raphiniert.com; 2021 [cited 2021 Aug 24]. Available from: https://github.com/raphiniert-com/ra-data-postgrest
[8] Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, et al. Sustainable data analysis with Snakemake. F1000Research. 2021 Jan 18;10:33.
[9] Blog Wissen Was Wirkt [Internet]. Wissen Was Wirkt. [cited 2021 Aug 23]. Available from: https://wissenwaswirkt.org/
[10] Thirion B, Pereira S, Névéol A, Dahamna B, Darmoni S. French MeSH Browser: a cross-language tool to access MEDLINE/PubMed. AMIA Annu Symp Proc AMIA Symp. 2007 Oct 11;1132.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210940

Data Privacy, Regulations and Legal Issues on COVID-19 Tracking Apps: A Scoping Review

Utoomporn WONGSIN[a,1], Arnat WANNASRI[b,c], Usman IQBAL[d] and Tuo-Yu CHEN [d]

[a] *PhD student in Global Health and Health Security, Taipei Medical University, Taiwan*
[b] *Nopparat Rajathanee Hospital, Department of Medical Services, Ministry of Public Health, Thailand*
[c] *National Yang Ming Chiao Tung University, Taiwan*
[d] *Masters and PhD Program in Global Health Department, College of Public Health, Taipei Medical University, Taiwan*

Abstract. It cannot be deniable that smartphone apps have grown exponentially and are playing a crucial role in the response to the COVID-19 pandemic in many countries. This paper aims to investigate data privacy, regulations and legal issues on COVID-19 tracking apps. A literature search will be followed the PRISMA guidelines extension for a scoping review. The search will be conducted on PubMed and Google Scholar. A total of 38 articles from 7,626 articles were reviewed. Mostly articles report on data privacy. Not many articles report on regulations and legal issues. However, there are many challenges on COVID-19 applications such as security risks, privacy issues, political, ethical, and legal risks, and standardization issues.

Keywords. Covid19 Apps, data privacy, legal issues COVID-19, Mobile Application, Global Health, Medical Informatics.

1. Introduction

As of 31 January, 2019 nCoV declared global health emergency by World Health Organization (WHO), and WHO named the disease Coronavirus Disease 2019 (COVID-19) on 11 February 2020[1-2]. The COVID-19 outbreak has spread worldwide [3], and all attacked countries have been making efforts to control this threat [4]. Since the COVID-19 attacked our world, no-contact tracing using mobile applications has grown exponentially and played a crucial role in the response to the COVID-19 pandemic in several countries.

Mobile apps for COVID-19 are to track and monitor people who are suspected to have COVID-19 infection and then report directly to government agencies. However, there is no evidence to provide a big picture of data privacy, regulation, and legal issues on Mobile application for COVID-19. Therefore, this paper aimed to investigate these aforementioned issues.

[1] Corresponding Author, Utoomporn Wongsin, PhD student in Global Health and Health Security, College of Public Health, Taipei Medical University, Taiwan; Email: d537108007@tmu.edu.tw

2. Methods

2.1. Search strategy

The literature search was conducted on PubMed and Google Scholar from February 11, 2020 to December 31, 2020. A literature review was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines extension for a scoping review [5].

2.2. Eligibility criteria

The inclusion criteria were: 1) the paper should describe the mobile app clearly, 2) had a clear demonstration concerning data privacy, regulation and legal issues on COVID-19 applications, and 3) language is restricted to English language.

2.3. Study selection

The relevant keywords were used for a search as following: “COVID-19” “coronavirus”, “nCOV19”, “contact tracing”, “COVID-19 apps”, “symptom tracking”, “mobile health application”, “mobile apps,” “mobile applications,” “smartphone apps”, “mobile phone apps”, “Applications”, “App data privacy”, “Security”, “Regulations”, “legal issues on apps”, “data privacy apps”, “legal issues on COVID-19 tracking apps” “data privacy on COVID-19 tracking apps”

3. Results

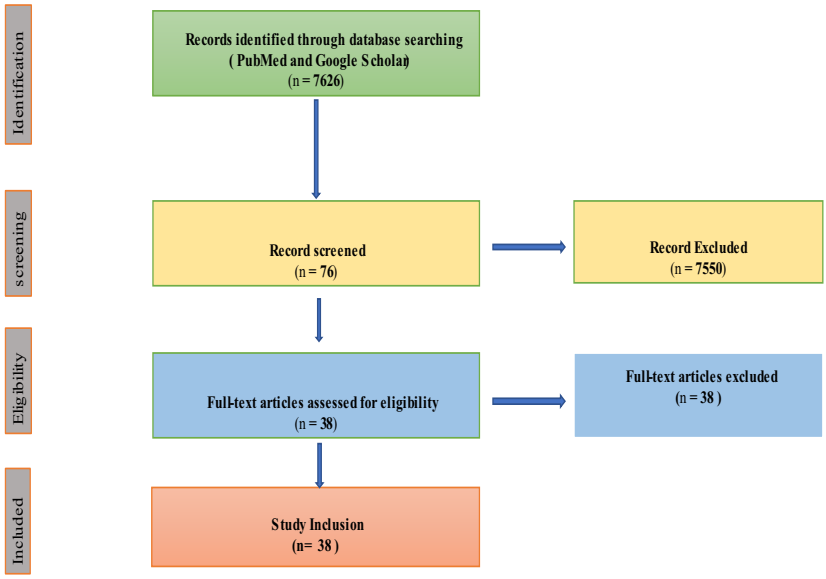


Figure 1. Research Framework according to PRISMA guidelines.

3.1. Literature search results

According to the PRISMA guidelines authors screened the abstracts of all discovered papers and assessed the available full text papers based on the eligibility criteria to finalize the selected papers to be included in the review. A total of 38 articles from 7,626 searched articles were included in this scoping review study.

3.1. Data privacy

Currently, it is undeniable that numerous digital technologies for contract tracing are adopted by the government around the globe to combat COVID-19. However, there are several concerns and challenges that needed to be addressed and data privacy is considered as the top issue. Results indicate that data of individuals' privacy were mainly reported (74% of 38 articles) for disease control policy purpose. Data privacy is considered a backbone of COVID-19 tracking application. However, the data privacy in deploying the applications for COVID-19 tracking are varied from countries to countries. Furthermore, although the protection of individual information privacy and identity is vague, yet the privacy of users should be protected to minimize the privacy concern. This is in line with the standard approach to protect patient's privacy that is to limit access to patients' information [6].

3.2. Regulation and Legal Issues

Although the novel contact-tracing technologies have been receiving attention in many countries because of its significant power for combating COVID-19, the applications are supposed to subject to laws and regulations to prevent abuse or exploitation. Findings show that besides data privacy issue, regulations and legal issues were also reported (26% of 38 articles). Law enforcement should be implemented for data privacy. Although data are necessary for epidemic control, its use should be regulated under data privacy regulations [7]. Individuals' data should be protected by law enforcement agencies.

4. Discussion

As the COVID-19 pandemic has drastically worsened, contract-tracing strategy has established as one of the COVID-19-related responses. Mobile applications are the novel powerful tools that were applied in many countries around the world to fight against COVID-19. The government can collect data from the mobile application users to identify those who are potentially exposed to the coronavirus. The contract-tracing using mobile apps have also caused public concerns and doubts on data its intrusion on data privacy. The government should make use of the collected data transparently.

5. Conclusions

It can be concluded that, on the one hand, governments use data for controlling the COVID-19 situation, but on the other hand the privacy pertaining to user information

should be protected by law. To the best of our knowledge, this is the first scoping review on data privacy, regulations and legal issues of COVID-19 tracking apps. Although this study used only two main databases, yet the databases covered an overwhelming number of published papers related to our review issues.

References

[1] WHO | Pneumonia of unknown cause – China [Internet]. WHO. World Health Organization; [cited 2020 Sep 22]. Available from: http://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/

[2] Sun P, Lu X, Xu C, Sun W, Pan B. Understanding of COVID-19 based on current evidence. J Med Virol. 2020;92(6):548-551. doi:10.1002/jmv.25722

[3] WHO. Coronavirus disease 2019 (COVID-19): situation report – 78. World Health Organization. 2020. [cited 2020 August 21]. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200407-sitrep-78-covid-19.pdf

[4] Mahmood S, Hasan K, Colder Carras M, Labrique A. Global preparedness against COVID-19: we must leverage the power of digital health. *JMIR Public Health Surveill.* 2020;6(2):e18980. doi: 10.2196/18980. https://publichealth.jmir.org/2020/2/e18980/

[5] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018 Sep 4;169(7):467–73.

[6] Kasperbauer TJ. Protecting health privacy even when privacy is lost. J Med Ethics. 2020 Nov;46(11):768–72.

[7] Blockchain for COVID-19: Review, Opportunities, and a Trusted Tracking System - PubMed [Internet]. [cited 2021 Jan 10]. Available from: https://pubmed.ncbi.nlm.nih.gov/33072472/

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210941

Assessing Staff's and Stroke Patients' Experiences in 8 Hospitals in Greece: Results from a Prospective Multi-Center Study ("SUN4Patients")

Petros GALANIS[a], Olympia KONSTANTAKOPOULOU[a,1], Iliana KARAGKOUNI[a], Paris GALLOS[a], Evangelos TSAMPALAS[b], Maria LYPIRIDOU[b], Anastasia GAMVROULA[b], Georgios MAVRAGANIS[c], Efstathios MANIOS[c], George NTAIOS[d,e], Efstathia KARAGKIOZI[e], Haralampos MILIONIS[f], Chara EVANGELOU[f], Argyro TOUNTOPOULOU[g], Ioanna KOUZI[g], Sofia VASSILOPOULOU[g], Athanasios PROTOGEROU[h], Stamatina SAMARA[h], Asteria KARAPIPERI[h]. Christos SAVOPOULOS[i], Apostolos I. HATZITOLIOS[i], Athina MYROU[i], Ioannis KALLIONTZAKIS[j], Aspasia KOURIDAKI[j], Stefanos PAPASTEFANATOS[a], George PAPASTEFANATOS[k], Kostas VEMMOS[d], Panayiota SOURTZI[a], Thalia BELLALI[l*], Eleni KOROMPOKI[c,d*], Daphne KAITELIDOU[a*] and Olga SISKOU[a*]

**equal last authors*

[a] *Center for Health Services Management and Evaluation, Nursing Department, NKUA, Athens, Greece*

[b] *Department of Neurology, Panarkadikon General Hospital, Tripoli, Greece*

[c] *Department of Clinical Therapeutics, Alexandra Hospital, NKUA, Athens, Greece*

[d] *Hellenic Stroke Organization, Athens, Greece*

[e] *Department of Internal Medicine, Faculty of Medicine, University of Thessaly, Larissa, Greece*

[f] *Department of Internal Medicine, School of Medicine, University of Ioannina, Greece*

[g] *Department of Neurology, Eginition Hospital, NKUA, Athens, Greece*

[h] *Cardiovascular Prevention & Research Unit, Laiko General Hospital, Medical School NKUA, Athens, Greece*

[i] *1st Propaedeutic Internal Medicine Department, Aristotle University of Thessaloniki, AHEPA Hospital, Thessaloniki, Greece*

[j] *Department of Neurology, General Hospital of Chania, Creta, Greece*

[k] *ATHENA Research Center, Athens, Greece*

[l] *Department of Nursing, International Hellenic University, Thessaloniki, Greece*

Abstract. To assess stroke patient-reported experiences and hospital staff experiences, during hospital stay. Methods: Stroke patient-reported experiences (n=387) were recorded using the translated and culturally adapted NHS-Stroke Questionnaire into Greek and staff experiences (n=236) were investigated using the Compassion Satisfaction and Burnout subscales of the ProQOL questionnaire. Results: Staff's mean compassion satisfaction score was 39.2 (SD=6.3) and mean burnout score was 24.3 (SD=5.6). Only 38.5% of the staff stated that there is smooth cooperation with healthcare professionals of other specialties/disciplines.

[1] Corresponding Author: Olympia Konstantakopoulou, 123 Papadiamantopoulou str., 115 27 Athens, Greece; E-mail: olympiak1982@hotmail.com.

Personnel working in an NHS Hospital was more satisfied and less burned-out when compared to personnel working at a University Hospital ($p=0.02$ and $p<0.001$, respectively). Mean total patient-reported experiences score was 81.9 (SD=9.5). Bivariate analysis revealed statistically significant differences for total patient-reported experiences among the eight study hospitals ($p>0.001$). Conclusions: Health policy planners and decision-makers must take into consideration the results of such self-reported measures to establish innovative techniques to accomplish goals such as staff-specialization, continuous training and applying formal frameworks for efficient cooperation amongst different disciplines.

Keywords. Stroke Patient Experiences, Staff Experiences, Stroke Care

1. Introduction

The evaluation of stroke patients' and health professionals' working in stroke clinics/centers experiences has been an emerging research topic recently. However, most of the studies conducted so far are qualitative using open interviews and semi-structured questionnaires. Regarding stroke patients, the main focus has been on assessing the quality of medical and nursing care they receive during their hospitalization and rehabilitation [1]. In regards to the respective staff, work satisfaction, burnout, engagement and personal perceptions are more often studied [2, 3, 4, 5]. In order to enhance patients' and staff experiences, the use of universally accepted validated quantitative tools may prove useful to stir health policy and decision making towards improving quality of care. The aim of this study was to assess stroke patient-reported and hospital staff's experiences in 8 Hospitals in Greece, in the context of the SUN4Patients prospective multi-center study.

2. Methods

A cross-sectional study was conducted including 387 stroke patients, who were admitted in eight NHS and University Hospitals in Greece and were interviewed via telephone one month following hospital discharge. Physicians (n=139) and nursing personnel (n=97), working in the SUN4Patients study centres, were recruited to participate in the staff experiences survey. The translated and adapted into the Greek language NHS-Stroke Questionnaire [6] composed of items on diagnosis, hospital admission, medical and nursing staff, the provided care during hospitalization and hospital discharge was used. The Compassion Satisfaction and Burnout subscales of the ProQOL questionnaire were used to measure staff's experiences while caring for stroke patients. Descriptives were calculated using absolute and relative frequencies and mean (SD) and/or median (IR). Bivariate analyses were conducted and multivariate linear regression analyses were performed to control for confounding effects (for staff experiences: demographics, type of hospital, working conditions and resources, level of cooperation, etc. and for patient experiences: demographics, hospital type, health status, days of hospitalization, etc.). The research protocol was approved by the Bioethics Committee of the NKUA Nursing Department, and the Scientific Committees of the study centres/hospitals. Patients and hospital staff gave their informed consent, adhering to the EU 2016/679 GDPR provisions and the Declaration of Helsinki.

3. Results

Hospital staff's mean age was 39.2 years old (SD=9.7). Staff was taking care of stroke patients for an average of 7.73 years (SD=8.1), the median number of stroke in-patients who were treating on a monthly basis was 10 (IR=11), while 25.4% were employed in a NHS and 74.6% at a University Hospital/Clinic. Mean compassion satisfaction score was 39.2 (SD=6.3) (max=50) and mean burnout score 24.3 (SD=5.6) (max=50). More than half of the participants (57%) stated that there is almost always the possibility to refer the patient directly to rehabilitation following discharge. More than 1 out of 5 (22.9%) and almost 1/3 participants stated that nursing and medical staff are numerically adequate, respectively, while 32.6% and 57.3% stated that nursing and medical staff are knowledge-and-skills-wise equipped to care for stroke patients, respectively. Finally, only 38.5% stated that there is smooth cooperation with other specialties/disciplines towards effective management of stroke patients.

Patients' mean age was 73.3 years old (SD=14.4) and 29.5% were admitted in a NHS Hospital. Mean total patient-reported experiences score was 81.9 (SD=9.5) (max=100). Almost 2/3 of the participants (64.9%) showed absence of symptoms before admission, while only 26.1% showed absence of symptoms one-month later. Also, 66.9% and 65.9% of the participants stated that the level of communication with nursing and medical staff, respectively, was full and comprehensive and less than half of the participants stated that nursing staff was numerically adequate at all times/almost at all times. Almost 2/3 stated that nursing staff (63.3%) and medical staff (68%) had sufficient knowledge and skills to take care of stroke patients and almost 4/5 received the necessary instructions on life-style changes, medication and adverse-events.

Table 1: Comparison between study scores and hospitals

Hospitals	Health Personnel		Stroke Patients
	*Mean compassion satisfaction score (max=50) (SD)**	*Mean burnout score (max=50) (SD) **	*Mean experience score(max=100) (SD) **
Hospital 1-UH/C	37.7 (5.4)	24.0 (5.0)	84.3 (5.9)
Hospital 2-UH/C	39.8 (6.2)	23.7 (6.1)	79.7 (11.7)
Hospital 3-UH/C	38.8 (7.6)	26.3 (5.1)	84.0 (9.1)
Hospital 4-UH/C	39.1 (5.0)	25.2 (3.7)	76.9 (14.1)
Hospital 5-NHS	41.4 (4.7)	22.2 (4.8)	81.7 (5.8)
Hospital 6-NHS	44.5 (4.3)	17.1 (4.4)	82.0 (8.6)
Hospital 7-UH/C	36.9 (8.0)	27.4 (6.4)	78.3 (10.5)
Hospital 8-UH/C	34.8 (5.4)	28.0 (3.0)	82.3 (8.2)

UH/C=University Hospital/Clinic, NHS=NHS Hospital, *ANOVA test, p-value<0.001

According to the bivariate analysis (Table 1) and multivariate model results (Table 2), personnel working in NHS Hospitals was more satisfied and less burned-out compared to those working at a University Hospitals/Clinic (p=0.02 and p<0.001, respectively).

Table 2: Multivariate Analysis Results for Staff Experiences (only significant relations are shown, a=0.05)

Characteristic /Independent Variable	Dependent variable	
	Compassion satisfaction score *(R^2=0.180)*	Burnout score *(R^2=0.250)*
	b / 95% CI / P value	*b / 95% CI / P value*
NHS in relation to University Hospital	2.2/0.3-4.1/0.02	3.6/2.03-5.2/<0.001
Years of experience in stroke patients' care	0.1/0.001-0.2/0.049	-0.2/-0.3-0.04/0.013
Possibility to refer the patient directly to a rehabilitation following discharge	1.4/0.37-2.41/0.007	-0.93/-1.76-0.09/0.029
Quality of cooperation	1.3/0.16-2.52/0.026	-1.8/-2.8--0.85/<0.001

More years of experience in managing stroke patients was related to increased satisfaction and decreased burnout ($p=0.049$ and $p=0.013$, respectively), as well as for the possibility to refer the patient directly to a rehabilitation centre following discharge ($p=0.007$ and $p=0.029$, respectively) and the quality of cooperation with health professionals of other specialties/disciplines ($p=0.026$ and $p<0.001$, respectively). Bivariate analyses revealed that increased age, days of hospitalization in an ICU and mRS (Modified Rankin Score) scores (prior to admission and at one-month follow-up) were correlated to decreased total patient-reported experiences score. Statistically significant differences were found among the eight study hospitals ($p>0.001$) (Table 1), but no difference between NHS Hospitals and University Hospitals/Clinics ($p=0.948$).

4. Discussion

According to the results of the study, a significant deviation was observed between patients' and staff's views on numerical and scientific adequacy of the nursing staff in stroke care. Patients were more satisfied in comparison to health professionals. This finding could be interpreted either by the customary trusting relationship between patients and their "carers", or by the professionals' acknowledgement for more specialized training. Higher compassion satisfaction was noted for NHS Hospitals' staff, a trait that could be explained by the better cooperation with health professionals of other specialties/disciplines. Putting emphasis on staff professional development in stroke care, as well as improving cooperation amongst them high-quality stroke care will be succeeded [7, 8].

5. Conclusions

Health policy planners and decision makers must take into consideration the results of such self-reported measures, so as to establish innovative techniques to accomplish goals such as staff-specialization, continuous training and applying formal frameworks for efficient cooperation amongst different disciplines, in favor of improving patients' and health professionals' experiences and, ultimately, the quality of the provided care to stroke patients.

Acknowledgements: The project was funded from the Hellenic Foundation for Research and Innovation and the General Secretariat for Research and Technology, grant No 538.

References

[1] Lutz, B. et al., The crisis of stroke: experiences of patients and their family caregivers. Topics in stroke rehabilitation 18, (2011), 786-797.
[2] Fiabane, E. et al., Occupational stress and job satisfaction of healthcare staff in rehabilitation units. Med Lav 103, (2012), 482-492.
[3] Catangui, E.J. And Roberts, C. J. The lived experiences of nurses in one hyper-acute stroke unit. British Journal of Nursing, 23, (2014). 143-148.

[4] Busis, N.A., et al., Burnout, career satisfaction, and well-being among US neurologists in 2016. Neurology 88, (2017). 797-808.
[5] Theofanidis, D. And Gibbon, B. Exploring the experiences of nurses and doctors involved in stroke care: a qualitative study. Journal of clinical nursing. 25, (2016), 1999-2007.
[6] Konstantakopoulou O., et al., Translation and transcultural adaptation of the NHS-Stroke Questionnaire on capturing and assessing stroke patient experiences, Nursing Care and Research 59, (2021), 41-51.
[7] Struwe, J. H., et al., How is nursing care for stroke patients organised? Nurses' views on best practices. Journal of Nursing Management 21, (2013), 141–151.
[8] Clancy, L., Povey, R., & Rodham, K. "Living in a foreign country": Experiences of staff-patient communication in inpatient stroke settings for people with post-stroke aphasia and those supporting them. Disability and Rehabilitation 42 (2018), 324–334.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210942

Quantifying Heterogeneity in Tumors: Proposing a New Method Utilizing Convolutional Neuronal Networks

Georg PROKOP[a,1], Michael ÖRTL[b], Marina FOTTELER[b], Peter SCHÜFFLER[c], Johannes SCHOBEL[b], Walter SWOBODA[b], Jürgen SCHLEGEL[a] and Friederike LIESCHE-STARNECKER[a]
[a] *Department of Neuropathology, Institute of Pathology, Technical University of Munich, GERMANY*
[b] *DigiHealth Institute, Neu-Ulm University of Applied Sciences, GERMANY*
[c] *Department of Computational Pathology, Institute of Pathology, Technical University of Munich, GERMANY*

Abstract. Heterogeneity is a hallmark of glioblastoma (GBM), the most common malignant brain tumor, and a key reason for the poor survival rate of patients. However, establishing a clinically applicable, cost-efficient tool to measure and quantify heterogeneity is challenging. We present a novel method in an ongoing study utilizing two convolutional neuronal networks (CNN). After digitizing tumor samples, the first CNN delimitates GBM from normal tissue, the second quantifies heterogeneity within the tumor. Since neuronal networks can detect and interpret underlying and hidden information within images and have the ability to incorporate different information sets (i.e. clinical data and mutational status), this approach might venture towards a next level of integrated diagnosis. It may be applicable to other tumors as well and lead to a more precision-based medicine.

Keywords. Tumor heterogeneity, Glioblastoma, Convolutional Neuronal Network, Neuropathology, Digital Pathology

1. Introduction

Cancer heterogeneity is ubiquitous: For many years, human cancers are known to be heterogenous in their gene expression and morphological appearance [1, 2]. Since tumor heterogeneity is believed to play a key role in acquiring therapy resistance and thus leading to a poor prognosis in patients, there is great interest in it [3, 4]. Most studies use large-scale mutational characterization to qualify and quantify cancer heterogeneity. While these methods are useful to describe and discover intra- and intertumoral differences, they lack extensive accessibility. Thus, there is a need for easily available and clinically applicable methods to quantify tumor heterogeneity, ideally aided by computational technologies [5].

[1] Corresponding Author, Georg Prokop, Department of Neuropathology, School of Medicine, Institute of Pathology, Technical University Munich, Trogerstraße 18, 81675 München, Germany; E-mail: georg.prokop@tum.de.

Artificial intelligence is a promising and powerful tool: Among these computational technologies, artificial intelligence (AI) is one of the most promising emerging approaches. Certain types of AI, especially convolutional neuronal networks (CNN), have successfully been applied in different areas of medical image analysis. Among others, they perform at least equal to a dermatologist in detecting skin-cancer in images [6, 7] and predict patient outcome from colorectal cancer tissue [8]. It is also possible to train neuronal networks to integrate different types of data, such as images and genomic biomarkers to generate more accurate predictions of survival [9]. With this powerful tool at hand, one can tackle other complex problems and questions like the automatic quantification of heterogeneity [10].

Heterogeneity is a hallmark of glioblastoma: As its original name "glioblastoma multiforme" suggests, glioblastoma (GBM) is known to be inter- and intratumorally heterogenous [11]. It is also infamous for its poor median survival rate of 15 months [12]. Currently, GBM is diagnosed according to the WHO classification based on histological and molecular criteria creating a so called *integrated diagnosis* [13]. Since heterogeneity is considered a hallmark of GBM and several studies describing it have been published, it is an ideal specimen to test our novel approach for automated quantification of heterogeneity [11, 14].

In the ongoing study described in this manuscript, we combine histological images with mutational features that account for tumor heterogeneity, thus resulting in a next level of integrated diagnostics. By using easily available hematoxylin and eosin (HE) slides as input (standard diagnostic slides), we ensure our trained networks to be applicable in various settings, helping on the road towards precision-based medicine. This manuscript focuses on the main approach and methodology in our current research.

2. Method

2.1. Overall Layout

We quantify heterogeneity by using CNNs to detect underlying and hidden information in digital HE slides. We train two separate networks using supervised machine learning. The first network differentiates between tumor tissue and surrounding brain tissue in HE slides. The second network is set up to measure heterogeneity in the tumor regions in the HE slides defined by the first CNN. As network architectures, we test both pre-trained networks like the Google Inception V3 and an untrained network architecture. To avoid overfitting, we augment our data by rotating and flipping the input images and use dropout regularization. The ground truth for training are manually labelled HE slides and heterogeneity data derived from immunohistochemical (IHC) staining of seven proteins. However, the final pipeline will only use HE slides to both detect tumor lesions and quantify their heterogeneity (see Figure 1). This study was approved by the local ethics committee (vote no. 164/19 S-SR, chair Prof. Dr. G. Schmidt).

2.2. Tumor detection

Acquisition and pre-processing of image data: As a training data set, we use 58 GBM tissue samples stained with HE. 54 tumor-free reference slides from brain autopsies

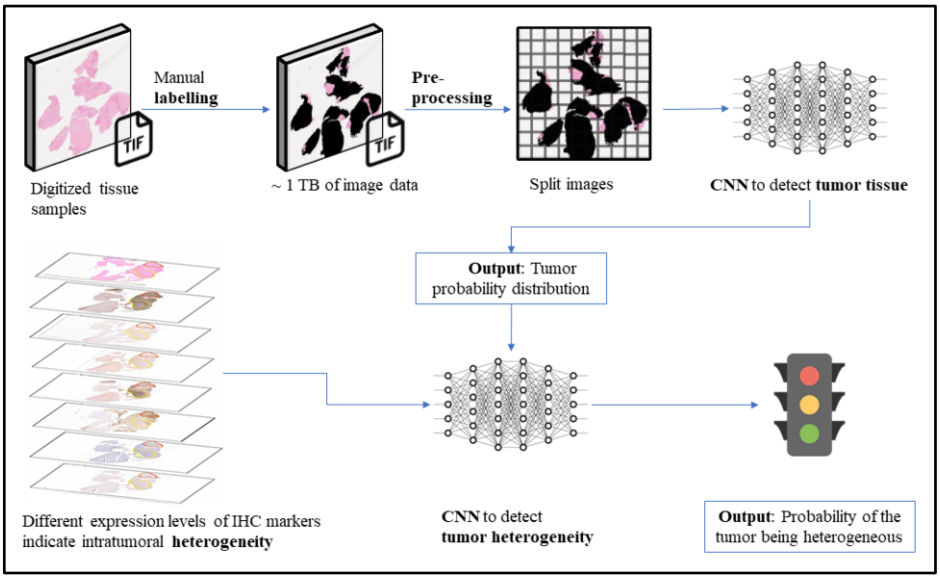


Figure 1. *Training the Tumor Heterogeneity Measurement Pipeline:* Together with control samples from brain autopsies, the tumor images are used to train a CNN to delimitate healthy and tumor tissue and secondly to measure tumor heterogeneity based on a immunohistochemically derived classification system.

served as controls. After digitizing the slides, the core regions of the tumorous lesions were manually labelled and separated from the surrounding brain tissue. Altogether this resulted in over 1 TB of image data (available as TIF files). As pre-processing steps, the images were split into smaller images patches creating approx. 278.900 training samples per slide.

Processing and analyzing the images: The aforementioned smaller patches are analyzed by a CNN in a sliding window approach. Starting with a 300 by 300 pixel frame, the network determines whether the central core region of that patch contains tumor tissue by using all the information in the whole frame. Thus, all the information in the frame is condensed onto its core. Next, the frame is shifted to the side by the central core region. Thereby, the whole image is analyzed creating a probability distribution of the core regions containing background, tumor or healthy tissue.

2.3. Heterogeneity measurement

We determined the overall heterogeneity of the tumor sample using a previously described method based on the different expression levels of seven IHC markers [14]. Training the second CNN incorporates the heterogeneity data, provided as a classification of the tumors in ‘homogenous’, ‘intermediate’ or ‘heterogenous’, and the previously derived tumor distribution to determine the probability of the HE slide being derived from a heterogenous or homogenous tumor sample. Thus, this probability serves as a surrogate parameter of heterogeneity itself.

2.4. Validation

To validate our pipeline, we chose various approaches: Firstly, we randomly split our training data into two patient groups. The first group, about 80% of the total patients, is

the actual training data set, the second, about 20%, serves as a validation data set. Here, the network can prove its capabilities. Secondly, the detected tumor tissue is shown in the original image using a color code, indicating different levels of tumor probability. In the future this may help pathologists to identify tumor areas more easily. Additionally, we will validate our networks in publicly available image sets like the TCGA data set or the Ivy Atlas. Finally, we will correlate our findings to patient survival. The source code will be made publicly available.

3. Conclusion

This article describes the method of an ongoing project to quantify heterogeneity in tumors and GBM in particular. As heterogeneity, especially of GBM, is of utmost importance for therapy decisions and thus patient-survival, this project will contribute to a better treatment in neurooncology. By incorporating patient data into a multimodal learning algorithm, we want to advance precision-based medicine and provide a next level of integrated diagnosis.

References

[1] Heppner GH. Tumor Heterogeneity. Cancer Research. 1984;44(6):2259-65.
[2] Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. Nat Med. 2016;22(1):105-13.
[3] Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. 2018;15(2):81-94.
[4] Lee J-K, Wang J, Sa JK, Ladewig E, Lee H-O, Lee I-H, et al. Spatiotemporal genomic architecture informs precision oncology in glioblastoma. Nature Genetics. 2017;49(4):594-9.
[5] McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. 2017;168(4):613-28.
[6] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115-8.
[7] Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. Eur J Cancer. 2019;113:47-54.
[8] Bychkov D, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. Sci Rep. 2018;8(1):3395.
[9] Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. Proceedings of the National Academy of Sciences. 2018;115(13):E2970-E9.
[10] Zhong Q, Ruschoff JH, Guo T, Gabrani M, Schuffler PJ, Rechsteiner M, et al. Image-based computational quantification and visualization of genetic alterations and tumour heterogeneity. Sci Rep. 2016;6:24146.
[11] Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010;17(1):98-110.
[12] Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med. 2005;352(10):987-96.
[13] Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol. 2021;23(8):1231-51.
[14] Liesche-Starnecker F, Mayer K, Kofler F, Baur S, Schmidt-Graf F, Kempter J, et al. Immunohistochemically Characterized Intratumoral Heterogeneity Is a Prognostic Marker in Human Glioblastoma. Cancers (Basel). 2020;12(10).

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210943

The Need for Data Contextualization in Urban-Water Systems in Terms of Environmental and Behavioural Health

Panagiotis KATRAKAZAS[a,1], Theodora KALLIPOLITOU[a] and Ilias SPAIS[a]
[a] *Zelus P.C., Athens, Greece*

Abstract. The current paper addresses the need for making scientific knowledge easily accessible, comprehensible, and tailored for citizens, especially in urban-water habitats, enabling their behavioural change and consequent climate change resilience. It proposes a schema that integrates data from different sources and highlights their relevance to citizens (aiming to raise their awareness), the impact on the citizens' Quality of Life as well as the way they (will have to) perform various activities. Targeted bibliographical research through online digital libraries was conducted to capture the scientific coverage and validation of this need. As an outcome, the complexity and interdependencies of environmental and behavioural health issues growth has been confirmed, and public health programs have begun to identify the need for the integration of data from diverse sources. Therefore, the proposed schema could be used for enabling better design of public health policy making.

Keywords. data contextualization, environmental health, urban-water system, locality impact, behavioural engagement

1. Introduction

As highlighted by the COVID-19 pandemic [1] there is a pressing need to get a holistic approach towards climate mitigation and adaptation pathways, to understand the effects of anthropogenic activity in local environments where the marine and urban ecosystems are strongly interrelated and drive human activity. This holds especially true for urban water systems, e.g., a port-city, where, despite boosting the economy, there is a negative impact on the environment, hence the health of residents of the surrounding areas [2,3]. Living near the port areas is associated with low socioeconomic status, while the environmental pollution caused by port activities increases the risk of developing a number of diseases, including heart and lung diseases, as well as increased stress levels, resulting in negatively affecting the residential Quality-of-Life levels [4].

2. Methods

Three digital online libraries (i.e., IEEEXplore, PubMed and ScienceDirect) were used to identify studies of relevant interest to the topic exploration. Targeted bibliographical

[1] Corresponding Author, e-mail: p.katrakazas@zelus.gr

research was performed to highlight the different perspectives taken on behavioural, environmental, and public health linkage with respect to data contextualisation needs in these areas. Our results are hereinafter provided.

2.1 Bridging the gap between environmental, behavioural public health space

A key factor that contributes to climate related anxiety (the so-called eco-anxiety) is knowing danger is coming but not having any appropriate scripts, skills, or direct agency in place to mitigate it [5]. Two of the key messages for policymakers is a) "to enhance the personal responsibility for people to behave more climate-friendly by providing information about the impact and the consequences of personal behaviour" [6,7] and b) to promote and engage on support for neighbourhood and community-level greening [8]. Complementary to that and given the current Covid-19 crisis, increased need for environmental health services [9] and attention should be given to duty and heedfulness of social justice and environmental health issues, in order to eliminate discrimination against vulnerable populations [10].

As environmental exposures pose an important health determinant [11], public health interventions should take into consideration the links between behavioural aspects and implementation of health in all policies aiming to improve public health. However, given the recent failure of behavioural change approaches given the Covid-19 pandemic [12] and the environmental impact yielded by it, approaches like the one presented by B.M. Stieb et al [13] should be brought into the spotlight, especially when urban-water systems are concerned, as highlighted also by the lack of relevant articles [14].

2.2 A suggestion towards a multi-level approach schema

Considering the different perspectives and the different levels of complexity among port-cities, we need to take into consideration the micro-, meso- and macro- system perspective, including, but not limited to, each port-city's different governance mechanisms, competitiveness, spatial and strategic interdependencies, and economical role, complementarily to the Sustainable Development Goals (SDGs). In more detail:

The **micro-level** refers to the extraction of relationships affecting the inner systemic elements of the port-city. To do so, bringing an advanced behavioural services framework for the transition of the port-city actors' activities using geospatial data insights and end users' input (e.g., a way of analyzing the characteristics of locality in urban-water systems can be outlined. The Locality Pattern System [15] or simple perception mapping [16] can be applied and adapted in the planning and design processes, resulting in future-orientated localities with biodiversity in mind. The establishment of these mechanisms will enable and facilitate the inclusion and participation of port-city end users in tackling the struggles of their own area (locality characterization) and the research-based design of actions that can be implemented at different levels (i.e., consumption and lifestyle) for the public health policy makers.

The **meso-level** tackles the conflicting nature of port-city by examining the port and city entities. Promotion of a constructive and evidence-based dialogue among these two systemic elements can be achieved by deploying related services for monitoring air emissions, urban built-up and public health footprint of port-city activities (e.g. like the ones used in [17] and [18]. These services can provide open, comprehensible, and wide access information and recommendations for action to the key users and stakeholders for an interactive and meaningful decision-support.

Finally, the **macro-level** examines the interrelation of the port-city as an entity located in a global map, therefore it is about promoting knowledge and information collection, sharing and management from natural, social, engineering, and maritime sciences via a high-level knowledge system, so as to increase the socioeconomic, health and climate resiliency of port-city environments in a multidisciplinary and systemic approach. Adaptation of existing multi-level frameworks [19–21] will allow a better monitoring of the environmental health and behaviour engagement, in accordance with the Sustainability Development Goals (SDGs) set for 2030, which point towards implementing a sustainable development of the society.

3. Discussion: The need for data contextualization

Collecting data and observing biodiversity behaviour in urban-water areas is more than crucial for identifying correlations and quantifying the impact of observed changes on people's health. The schema will facilitate the monitoring of the relative active level of urban-water system development and the anthropogenic activity levels, including industrial and spatial integration to identify and characterize the urban-water habitation and correlation factors. A better monitoring of these influencing parameters and action mechanisms from a dynamic and regional comparison perspective among the heterogeneous flows of actors, assets and structures involved, will test what is already known about mitigation and adaptation behaviours being more engaging for citizens with high environmental awareness or concern, versus those with lower awareness or concern.

This will provide the basis for "a breadth of rationales for actions bringing joint mitigation and adaptation benefits – and the provision of practical information to help citizens modify existing practices" [8]. Contextualizing information about air, water, noise and other types of environmental disturbance will reinforce and build on emergent insights on the relation between each citizen's practices (whether they are individuals or organizations living and operating in it) and the wider social and cultural context. It will also question the extent to which extant social and cultural practices within an urban-water area may be able to cope with the pace and extent of climate change and its consequences on people's wellbeing. In practical terms, this will lead to a holistic vision of the biogeography of the urban-water systems helping to generate prospective proposals for localized management and long-term transformation.

However, data contextualization in this context will not be easy as opposing social, political and public health processes have to be faced. As technological progress and attractions of urban-water habitats will sustain migration into them, their infrastructure will not or hardly be able to keep up with population increase. Crowdedness, unemployment, pollution, health and security risks will hinder the pathway towards a sustainable development. Therefore, having a holistic view of the urban-water habitat elements interactions, strengthening any stabilising interrelations and gaining data insights is more than necessary in the operation of the global society system, especially for providing the missing feedbacks, meaning the creation of a schema, like the one proposed in the current paper, which assists towards the successful coverage of common human wellbeing and their micro-, meso- and exo-systemic interests.

References

[1] Sharifi A, Khavarian-Garmsir AR. The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management. Sci Total Environ. 2020 Dec 20;749:142391.

[2] Manisalidis I, Stavropoulou E, Stavropoulos A, Bezirtzoglou E. Environmental and Health Impacts of Air Pollution: A Review. Front Public Health [Internet]. 2020 Feb 20 [cited 2020 Nov 17];8. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7044178/.

[3] European Court of Auditors. Air pollution: our health still insufficiently protected [Internet]. LU: Publications Office; 2018 [cited 2021 Jan 10]. (Special report No (European Court of Auditors. Online). Available at: https://data.europa.eu/doi/10.2865/80097.

[4] European Commission. Directorate General for the Environment., University of the West of England (UWE). Science Communication Unit. Links between noise and air pollution and socioeconomic status. [Internet]. LU: Publications Office; 2016 [cited 2020 Nov 17]. Available at: https://data.europa.eu/doi/10.2779/200217.

[5] Ingle HE, Mikulewicz M. Mental health and climate change: tackling invisible injustice. Lancet Planet Health. 2020 Apr 1;4(4):e128–30.

[6] Jakučionytė-Skodienė M, Liobikienė G. Climate change concern, personal responsibility and actions related to climate change mitigation in EU countries: Cross-cultural analysis. J Clean Prod. 2021 Jan 25;281:125189.

[7] Odou P, Schill M. How anticipated emotions shape behavioral intentions to fight climate change. J Bus Res. 2020 Dec 1;121:243–53.

[8] Kondo K, Mabon L, Bi Y, Chen Y, Hayabuchi Y. Balancing conflicting mitigation and adaptation behaviours of urban residents under climate change and the urban heat island effect. Sustain Cities Soc. 2021 Feb 1;65:102585.

[9] Rodrigues MA, Silva MV, Errett NA, Davis G, Lynch Z, Dhesi S, Hannelly T, Mitchell G, Dyjack D, Ross KE. How can Environmental Health Practitioners contribute to ensure population safety and health during the COVID-19 pandemic?. Safety Science. 2021 Apr 1;136:105136.

[10] Aschner M, Paoliello MMB, Tsatsakis A, Bowman AB, Dorea JG, Hartung T, et al. Social injustice in environmental health: A call for fortitude. Environ Res. 2021 Mar 1;194:110675.

[11] Rojas-Rueda D, Morales-Zamora E, Alsufyani WA, Herbst CH, AlBalawi SM, Alsukait R, et al. Environmental Risk Factors and Health: An Umbrella Review of Meta-Analyses. Int J Environ Res Public Health. 2021 Jan;18(2):704.

[12] West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. Nat Hum Behav. 2020 May;4(5):451–9.

[13] Stieb DM, Huang A, Hocking R, Crouse DL, Osornio-Vargas AR, Villeneuve PJ. Using maps to communicate environmental exposures and health risks: Review and best-practice recommendations. Environ Res. 2019 Sep 1;176:108518.

[14] Zheng Y, Zhao J, Shao G. Port City Sustainability: A Review of Its Research Trends. Sustainability. 2020 Jan;12(20):8355.

[15] Prominski M, Busche K. Locality Pattern Systems as Design Tools for Water-Related Open Spaces. In: Wang F, Prominski M, editors. Water-Related Urbanization and Locality: Protecting, Planning and Designing Urban Water Environments in a Sustainable Way [Internet]. Singapore: Springer; 2020 [cited 2021 Sep 9]:3–26. Available at: https://doi.org/10.1007/978-981-15-3507-9_1.

[16] Delphine, Busquet M, Santpoort R, Witte P, Spit T. Living on the edge: Identifying challenges of port expansion for local communities in developing countries, the case of Jakarta, Indonesia. Ocean Coast Manag. 2019 Apr 1;171:119–30.

[17] Andrade MJ, Costa JP, Blasco López J. 3DPortCityMeasure: Methodology for the Comparative Study of Good Practices in Port–City Integration. Sustainability. 2020 Jan;12(3):880.

[18] Liu J, Kong Y, Li S, Wu J. Sustainability assessment of port cities with a hybrid model-empirical evidence from China. Sustain Cities Soc. 2021 Dec 1;75:103301.

[19] Cao L, Tang Y, Cai B, Wu P, Zhang Y, Zhang F, et al. Was it better or worse? Simulating the environmental and health impacts of emissions trading scheme in Hubei province, China. Energy. 2021 Feb 15;217:119427.

[20] Alvarez CH, Evans CR. Intersectional environmental justice and population health inequalities: A novel approach. Soc Sci Med. 2021 Jan 1;269:113559.

[21] Milojevic-Dupont N, Creutzig F. Machine learning for geographically differentiated climate change mitigation in urban areas. Sustain Cities Soc. 2021 Jan 1;64:102526.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210944

Usability Inspection of a Mobile Cancer Telerehabilitation System

Joseph FINKELSTEIN[a], Xingyue HUO[a], Irena PARVANOVA[a], Matthew GALSKY[a]
[a] Icahn School of Medicine at Mount Sinai, New York, NY, USA

Abstract. Cancer rehabilitation (CR) has been shown to address physical impairments and progressive disablement experienced by patients with cancer. Access to CR is limited by multiple barriers. Previous studies demonstrated that telemedicine approaches may facilitate access to rehabilitation services however usability and acceptance of cancer telerehabilitation has not been systematically evaluated. This goal of this study was usability inspection of a mobile cancer telerehabilitation system based on cognitive walkthroughs and heuristics evaluations, supplemented by surveys to capture health literacy, patient engagement, and acceptance. The System Usability Scale provided a standardized reference for usability and satisfaction, and the mean result of 83.2±16.9 placed this mobile interface in the top 10th percentile. A semi-structured qualitative interview provided actionable feedback, which informed the next iteration of this project. Overall, this mobile telerehabilitation system was considered by cancer patients easy-to-use, satisfying, and engaging with 91% of participants planning to use it in the future.

Keywords. Cancer rehabilitation, telemedicine, usability inspection, mobile health

1. Introduction

Cancer rehabilitation has been shown to result in significant improvements in physical, social, psychological and vocational functioning of cancer survivors [1]. Number of cancer survivors is expected to increase due to a growing aging population and an expanding spectrum of effective cancer therapies. It is estimated that the population of cancer survivors will reach 19 million by 2024 [2]. Access to cancer rehabilitation is limited due to multiple existing barriers [3]. Telemedicine approaches can improve access to rehabilitation programs [4] however their acceptance by cancer patients has not been systematically assessed. High acceptance of telerehabilitation by older adults [5] and patients with chronic conditions have been demonstrated in our previous work [6-7]. We designed a telerehabilitation system based on the principles of iterative patient-centered design [8-9]. This project is aimed at assessing usability and acceptance of this system in cancer patients.

2. Methods

The study was designed to evaluate the usability and acceptance of the cancer telerehabilitation system by eliciting quantitative and qualitative feedback from cancer patients. Participants were patients with metastatic urogenital cancer receiving outpatient oncology care at the Mount Sinai Health System in New York. Upon sitting down with

a tablet computer, participants were given a packet of surveys, followed by instructions to complete 3 tasks, then an additional package of evaluations and surveys to record their feedback. The tablet computer included a touchscreen, a stylus, and a keyboard with trackpad, and it was preconfigured to display an icon to log in to a telerehabilitation system on screen (Figure 1).

Figure 1. User Interface

Surveys began with a socio-demographic form and the Rapid Estimate of Adult Literacy in Medicine (REALM). Participants then performed three tasks as part of the cognitive walkthrough. Afterwards, participants provided feedback on post-task questionnaires, a heuristics evaluation form, and the System Usability Scale (SUS). A semi-structured qualitative exit interview completed the session.

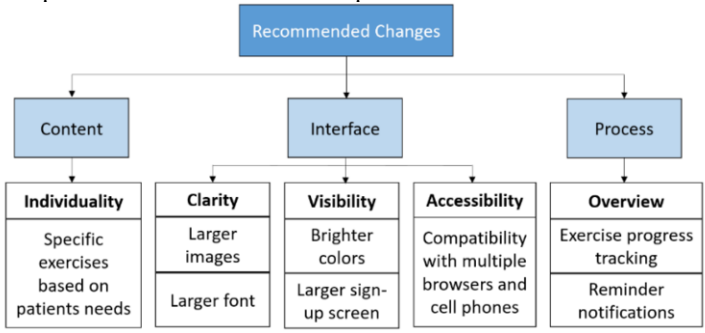


Figure 2. Usability Concept Map of User's Suggestions

The three tasks for the cognitive walkthrough could be completed with or without assistance from the research associate; if a participant needed help, this request was recorded. Task 1 instructed the user how to log in to the system. Task 2 required users to complete a symptom survey. Task 3 asked users to review exercise plan, and find and perform a particular exercise (sit-to-stand).

After completing the tasks 1 and 2, the participant was asked to grade each task on a scale of 1 (very difficult) to 5 (very easy) using a survey that included the following questions: 1) How difficult or easy was it to review the content and finish the task? 2) How difficult or easy was it to answer the questions/complete the form? 3) How satisfied are you with using this system to complete this task? 4) How would you rate the amount of time it took to complete this task? 5) Is the system visually appealing? 6) Is the system easy to navigate? After completion of the task 3, only questions 3 to 6 were answered. There were also two open-ended questions that ask the participant to share any other feedback and to describe problems they might have encountered. An exit semi-structured qualitative interview followed an interview guide comprising open-ended inquiries on content, interface and process related to the system.

3. Results

Overall, 11 consecutive patients attending genitourinary medical oncology clinic between April and August of 2021 agreed to participate in this study and provide feedback on usability and acceptance of the cancer telerehabilitation system. Average age of the participants was 68.1±11.2 years old ranging from 42 to 85 years; 7 participants identified as White (70%), 2 identified as Black (20%), and one did not provide a response for this question; 2 patients identified themselves as Hispanic (18%); 8 patients reported daily computer use at home (73%) and 3 patients claimed they never use computer at home (27%); English proficiency was self-reported as good (18%) and excellent (82%); Participants reported a range of familiarity with cancer rehabilitation with 6 patients (54.5%) stating that they never heard about cancer rehabilitation. The average REALM score of 65.5±0.8 indicated a 'High School' level of health literacy.

Task self-assessment results are presented in Table 1 as average scores, with a score of 5 indicating highest satisfaction. Overall post-task scores demonstrated high acceptance of the user interface with total scores of 4.7±0.6; 4.6±0.6; 4.7±0.5 for tasks 1, 2, 3 respectively. The mean satisfaction score for these three tasks was 4.8±0.4; 4.6±0.7; 4.9±0.3, and the mean "easy to navigate" score was 4.5±0.7; 4.7±0.5; 4.7±0.5, respectively. All study subjects (100%) were able to successfully complete the three test tasks with mean completion time in seconds of 87.5±101.5; 92.5±93.1; 183.1±502, respectively. Minor assistance was provided for 3, 1, 3 subjects to complete tasks 1, 2, 3, respectively. The main reason for assistance was difficulties in entering login name and password. Ten out of eleven study subjects (90.9%) were certain that they would like to use the cancer telerehabilitation system in the future, one subject was not sure.

The heuristic evaluation demonstrated that the interface functionality is self-explanatory and can be easily recognized and carried out by the cancer patients. The overall score for heuristics evaluation was 4.5±0.7 with the highest score assigned for control and recognition (4.7±0.5) and the lowest – for aesthetics (4.2±1.2).

System Usability Scores (SUS) were normalized according to the standard scoring algorithm. Mean SUS was 83.2±16.9, which corresponds with an 'above average' usability rating, and places this system usability at greater than the 90th percentile.

The semi-structured qualitative interview captured instructive feedback that described the cancer telerehabilitation platform as easy-to-use, engaging, and supportive for regular exercise. Tailored interactive delivery of multimedia exercise materials was perceived as a major facilitator of daily exercise. The participants stated "I liked the fact that you can either read or watch the video and the video was good. And the video is almost as a partner. You're not alone." and "I like the pictures and videos, even if you don't speak the language very well, you can still do exercises and trust that they can help." Regarding the interface, the patients commented "I am not excessively technology minded, but it was clear to me" and "I didn't really feel that there is necessity for any extra instructions on the exercises. It was very well presented. Once I had my signing in information, I could do it by myself, the whole program is intuitive." The patients believed that the telerehabilitation system "will get you to actually do the exercise." One patient commented: "I think that this will be a great alternative for people like me who are not that much into exercising on a daily and weekly basis." Most of the participants' suggestions were related to improved aesthetics and help options such as: "make the images a little bigger and expand it," "the colors are little washed out," "some guidance may be helpful as I'm not used to touchscreen interface for someone my age,

"touchscreen with the re-sizing and scrolling requires some learning." A summary of patient suggestions are depicted in Figure 2.

Table 1. Task Self-Assessment

Task 1: Log in to System	Mean (SD)
Login Difficulty	4.7 (0.5)
Navigation Difficulty	4.7 (0.5)
Satisfaction	4.8 (0.4)
Amount of Time	4.8 (0.6)
Visually Appealing	4.5 (1.0)
Easy to Navigate	4.5 (0.7)
Task 2: Complete Symptom Survey	
Content Difficulty	4.6 (0.5)
Questions Difficulty	4.7 (0.5)
Satisfaction	4.6 (0.7)
Amount of Time	4.5 (0.8)
Visually Appealing	4.5 (0.8)
Easy to Navigate	4.7 (0.5)
Task 3: Start and Finish Exercise	
Satisfaction	4.9 (0.3)
Amount of Time	4.5 (0.7)
Visually Appealing	4.7 (0.5)
Easy to Navigate	4.7 (0.5)

4. Discussion

The mobile cancer telerehabilitation system has been enthusiastically accepted by cancer patients who reported high usability and satisfaction rankings and expressed interest using such a system in the future. The results of usability inspection are largely consistent with our previous work on this platform in older adults [4] and people with chronic neurodegenerative disorders [6].

This is the first study assessing cancer telerehabilitation usability and acceptance in patients with metastatic urogenital cancer. Limitations of this pilot study include a limited sample size, however within this convenience sample patients from broad range of age, computer and health literacy were represented. This limitation will be mitigated in our future studies by broadening the patient base and geographical locations.

5. Conclusions

The mobile cancer telerehabilitation platform has demonstrated a high usability rating with excellent levels of user-reported satisfaction. Next steps should include addressing patient suggestions on improving the system functionality and evaluating clinical impact of cancer telerehabilitation in a randomized controlled trial.

References

[1] Cheville AL, Mustian K, Winters-Stone K, et al. Cancer Rehabilitation: An Overview of Current Need, Delivery Models, and Levels of Care. Phys Med Rehabil Clin N Am. 2017 Feb;28(1):1-17.

[2] American Cancer Society: Cancer treatment & survivorship: facts & figures 2014-2015, Atlanta, 2015.

[3] Alfano CM, Cheville AL, Mustian K. Developing High-Quality Cancer Rehabilitation Programs: A Timely Need. Am Soc Clin Oncol Educ Book. 2016;35:241-9.
[4] Jeong IC, Liu J, Finkelstein J. Association between System Usage Pattern and Impact of Web-Based Telerehabilitation in Patients with Multiple Sclerosis. Stud Health Technol Inform. 2020;272:346-349.
[5] Finkelstein J, Liu J. Usability of Telerehabilitation System Supporting Multipronged Exercise in Patients with Multiple Sclerosis. Stud Health Technol Inform. 2018;251:281-284.
[6] Jeong IC, Liu J, Finkelstein J. Factors Affecting Adherence with Telerehabilitation in Patients with Multiple Sclerosis. Stud Health Technol Inform. 2019;257:189-193.
[7] Finkelstein J, Liu J. Designing Telerehabilitation System for Multipronged Exercise in Patients with Multiple Sclerosis. Stud Health Technol Inform. 2018;254:16-23.
[8] Bedra M, Finkelstein J. Feasibility of post-acute hip fracture telerehabilitation in older adults. Stud Health Technol Inform. 2015;210:469-73.
[9] Finkelstein J, Jeong IC, Doerstling M, et al. Usability of Remote Assessment of Exercise Capacity for Pulmonary Telerehabilitation Program. Stud Health Technol Inform. 2020;275:72-76.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210945

Using Automated Text Processing to Assess the Patient Experience of an On-Demand Tele-Urgent Care

Saif KHAIRAT[a,1], Xinyi ZHANG[a], Makena BOYD[b], Barbara EDSON[b] and Robert GIANFORCARO[b]
[a] *University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*
[b] *UNC Health - Chapel Hill, NC, USA*

Abstract. Novel methods are needed to evaluate the perceptions of patients using telehealth. Automated text processing methods presents a golden opportunity to classify and analyze unstructured survey responses from patients. This study analyzed 585 unstructured entries from telehealth patients. Satisfied patients who returned for a second visit applauded the efficiency and physician interactions. While unsatisfied patients who did not return for a second visit complained of misdiagnosis and inefficiencies in e-prescription. Patient experience was significantly different between weekdays and weekends ($p<0.05$). Overall, tele-urgent are convenient for patients however, there are current facilitators related to patient-provider interaction and health information exchange that need further optimization.

Keywords. Telemedicine, Patient, Experience

1. Introduction

The use of patient surveys for care is one of the most effective ways that providers can measure patient safety and satisfaction. While the typical use of Likert scale questions provides valuable feedback for providers and health care administrators, the free text parts of these patient surveys are often underutilized (1). The information within these free text comments on the quality of care provide context in which the patient scored different aspects of care on the more structured questions of the survey. Given high the variation in patient comments on surveys, the use of word clouds and natural language processing (NLP) can be used to automatically identify themes of care. Automatic analysis of free text can be used to identify themes in patient tweets after their visit with to the hospital (2).

As more people are exposed to virtual care, the more opportunity there is to learn how to improve current operations to better suit patients. It is paramount to understand the nuances and experiences of patients using a virtual care system to improve the overall quality of care. The goal of this study was to evaluate the patient experience within a tele-urgent care center using automated text analysis of patient survey comments.

[1] Corresponding Author, Saif Khairat, PhD, MPH, University of North Carolina at Chapel Hill, 428 Carrington Hall, Chapel Hill, NC 27514 USA; E-mail: saif@unc.edu.

2. Methods

A Southeastern Medical Center in the U.S. utilized telehealth to provide on-demand care to individuals using board-certified physicians. The 24-hour virtual urgent care service was developed to aid individuals with medical needs who may need urgent care and need help when walk-in clinics are closed or because they have limited independent mobility. The online, on-demand service is available to the public regardless of their demographics.

After the tele-urgent encounter, patients received a satisfaction survey about their telemedicine experience including a question about their overall satisfaction (measured by 5-point likert scale) and open text comments about the patient's experience. We measured satisfaction as According to the outcomes above, we created 2 new variables. One is a binary variable with two levels: Satisfied/Dissatisfied and it is based on the overall experience of the encounter. If the overall experience is Excellent/Very good/ Good/ Fair, the value of this variable would be "Satisfied". If the overall experience is Poor, the value of it would be "Dissatisfied". The other is also a binary variable with two levels: Revisit and NotRevisit. The level of revisit means after the first encounter, the patients re-used the service at least once. The level of NotRevisit means after the first encounter, the patients never revisited. Institutional Review Board approval was obtained. Data was collected between 2018-2019 through the web-portal and imported into a secure, HIPAA compliant business intelligence server. The data was reviewed and cleaned by two graduate students. We applied the ordinal logistic regression function named polr from R (version 3.5.1) package polr to determine if there is a significant relationship between overall experience and the day of the week (weekday/weekend).

First, we analyzed the comments made by all patients during the first encounter to detect the reason why patients revisited or never revisited after their first visit. First, we split these comments into 4 different groups: (1) comments from patients who revisited and were satisfied with the first encounter, and (2) comments from patients who were unsatisfied with their first encounter and never revisited. In order to find the reasons why patients revisited or never revisited, we analyzed the patient comments and summarized the keywords of comments across different patient groups based on word frequencies using package nltk and wordcloud in Python (version 3.6.5). The outcome of this study was the perceptions of patients who revisited or never revisited after their first encounter.

Finally, we used word clouds to visualize the keywords that the comments are about in each group. In the plots, terms that appeared more frequently in the cleaned text would stand out. In addition, we provided a frequency table of common words in each group using function CountVectorizer. By comparing keywords across different groups, we can conclude what people liked or disliked about the virtual service in general.

3. Results

A total of 585 survey responses were analyzed from 546 unique patients, of which 400 (73%) were female, and 207 (38%) patients were in 35-49 age group. Of the 602 encounters, 431(73.7%) encounters were completed on weekday, and 154 (26.3%) were on weekend. The percentage of poor overall experience on weekday encounters was significantly higher than weekend encounters, figure 1. Excellent patient satisfaction levels were significantly higher on weekends (71%) compared to weekdays (63%) (p-value<0.05). Poor ratings of the service on the weekdays (8%) were double the ratings on weekends (4%). An association was found between the volume of visits and patient

satisfaction levels. Weekdays had substantially higher visit volume compared to the weekends and patient satisfaction was lower on weekdays compared to the weekends.

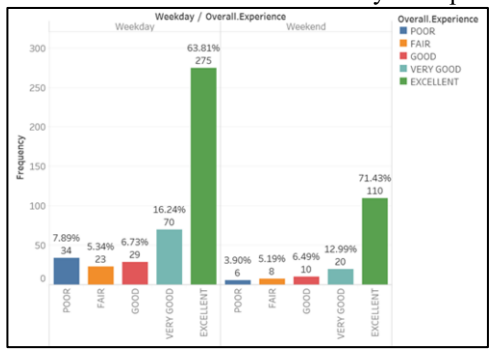


Figure 1. Patient satisfaction on the weekdays versus weekends

There were 37 patients who were satisfied with their first visit and revisited later, we noticed that 41% patients adopted the keyword "excellent". The cohort mainly cared about physician and time in terms of encounter. The keywords like "convenient", "time" and "quick home" are all related to the online convenience property of virtual service, Figure 1. Examples of patients' responses who were satisfied after their first encounter and revisited for a second visit included "how quickly the md returns your call", "I didn't have to take time off of work", and "it was quick and easy".

Figure 2. NLP-based word cloud of the perceptions of satisfied patients who revisited for a second visit.

There were 31 unsatisfied patients who did not return, Figure 3. Feedback from these patients included "I have waited almost 24 hours for my daughter's prescription to be called in", "repeatedly waiting for multiple calls and consultations", "Physician offered no advice", and "did not agree with diagnosis".

Figure 3. NLP-based word cloud of the perceptions of unsatisfied patients who did not revisit.

4. Discussion

We found that weekdays and video visits had lower patient experience ratings compared to weekends and phone visits. Patients who returned for a second visit had high satisfaction in their first visit. Factors that drove high satisfaction included convenience, not missing work, and excellent recommendations from the treating physician.

Patients who had high satisfaction in their first visit but did not return for a second visit reported that short wait times, pleasant physician interaction, and ease of receiving prescription as drivers of high satisfaction. We observed that patients who did not return for a second visit were mainly unsatisfied with their physician interaction, misdiagnoses, long wait times, and the duration of the visit or feeling rushed during the visit. The findings support our previous analysis of patient satisfaction after using telehealth (3-5).

Patient experience in the telehealth space may be influenced by various factors such as gender, the telehealth modality used, and the specialty of the visit (6). Similarly, provider telehealth competencies play an important role in how patients evaluate their telehealth experience such that telehealth visits that are short or the patient cannot clearly see the provider have lower satisfaction scores.

This study had limitations. We assumed that the reason for some patients not returning for a second visit was due to a satisfaction-related issue, which may not be true in many cases. Although it may be true, patient's health condition may be a confounding factor such that some patients may not have been sick again and thus, did not need a second visit. Also, we did not account for providers' perception of using tele-urgent care to compliment patient views.

In conclusion, tele-urgent care clinics may be beneficial and convenient to patients however, this study reveals that there are certain patient concerns around the quality of patient-provider interaction and privacy concerns that require further enhancements. Patients reported significantly higher satisfaction on weekends compared to weekdays. Satisfied patients who returned for a second visit applauded the efficiency and physician interactions. While unsatisfied patients who did not return for a second visit complained of misdiagnosis and inefficiencies in e-prescription.

References

[1] Doing-Harris K, Mowery DL, Daniels C, Chapman WW, Conway M. Understanding patient satisfaction with received healthcare services: A natural language processing approach. AMIA Annu Symp Proc. 2016;2016:524-33. Epub 2017/03/09. PMID: 28269848.

[2] Greaves F, Laverty A, Ramirez-Cano D, Pulman S, Moilanen K, Darzi A, et al. Analysis Of Patients' Comments About Hospitals In The English Nhs Via Twitter, And Comparison With Patient Surveys. Bmj Quality Safety. 2014;23(4):348. doi: 10.1136/bmjqs-2014-002893.5.

[3] Khairat S, Tirtanadi K, Ottmar P, Gudhe R, Austin CA. Would Geriatric Patients Accept Using a Telemedicine Platform for Post ICU-Discharge Follow-Up Visits? Stud Health Technol Inform. 2019;264:1233-7. Epub 2019/08/24. doi: 10.3233/shti190423. PMID: 31438122.

[4] Newlin T, McCall T, Ottmar P, Welch B, Khairat S. Assessing the Satisfaction of Citizens Using Teleconsent in Clinical Research. Stud Health Technol Inform. 2018;247:685-9. Epub 2018/04/22. PMID: 29678048.

[5] Khairat S, Haithcoat T, Liu S, Zaman T, Edson B, Gianforcaro R, et al. Advancing health equity and access using telemedicine: a geospatial assessment. J Am Med Inform Assoc. 2019. Epub 2019/07/25. doi: 10.1093/jamia/ocz108. PMID: 31340022.

[6] Khairat S, Pillai M, Edson B, Gianforcaro R. Evaluating the Telehealth Experience of Patients With COVID-19 Symptoms: Recommendations on Best Practices. Journal of Patient Experience. 2020:2374373520952975. doi: 10.1177/2374373520952975.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210946

Inference Control in a Diabetes Data Set Using a Java-Based Prototype of LDH Algorithm

Georgios FERETZAKIS[1], Konstantinos MITROPOULOS, Dimitris KALLES and Vassilios S. VERYKIOS
School of Science and Technology, Hellenic Open University, Patras 263 35, Greece; georgios.feretzakis@ac.eap.gr; kmitrop@otenet.gr; kalles@eap.gr; verykios@eap.gr

Abstract. Data sharing among different entities in the healthcare domain has become an increasingly common practice, where each entity would most likely want to prevent indirect data disclosure via inference channels. The Local Distortion Hiding (LDH) algorithm has been developed to protect sensitive decision tree (DT) rules, which are chosen not to be disclosed when DT construction techniques are applied to the data. This article presents eight experiments using a Java-based prototype that implements the LDH algorithm in a diabetes data set. Our experiments test the ability of the LDH algorithm in two ways, firstly in inference control and secondly in maintaining the structure and the performance metrics of the resulting DT. Our experiments on hiding eight terminal nodes in a diabetes data set using a Java-based prototype that implements the LDH algorithm, yield satisfactory results.

Keywords. Inference control; data security; privacy-preserving; machine learning

1. Introduction and Background

The healthcare sector is being digitally transformed by technological advances in medical information systems, electronic medical records, wearables, and mobile devices. The increase in the amount of global healthcare data and the advancements in the machine learning (ML) and data analytics field allow researchers and clinicians to extract and visualize large-scale medical data in a new spectrum [1]. The Internet facilitates the transfer and the exchange of these data, as well as the delivery of healthcare services and applications, linking this way successfully patients and healthcare providers. While such ecosystems promise a future for widely accessible and more innovative healthcare, the privacy of patients, physicians, nurses, and health care professionals is today more than ever of concern [2]. Data privacy is a critical issue in health informatics, particularly when analyzing datasets collected from various sources, such as health care providers, insurance companies, pharmaceutical companies, and research institutions. Data sharing among different entities in the healthcare domain has become an increasingly common practice, where each entity would most likely want to prevent indirect data disclosure via inference channels. The extraction of knowledge from patients' personal data for research purposes should be made with safety and absolute privacy. Privacy-preserving

[1] Corresponding Author, Georgios FERETZAKIS, PhD; E-mail: georgios.feretzakis@ac.eap.gr

data mining [3] is a research field designed to resolve confidentiality issues arising from data mining. The inference problem in databases occurs when sensitive information can be disclosed via inference channels from non-sensitive data and metadata [4].

1.1. A related to the inference problem scenario

A scenario that could present an inference problem is the following. Let us consider Mary Johnson, who works for a big company. Once a year, the employee association, which is primarily concerned with the welfare and recreational activities, organizes a research study in which all the employees are asked to complete questionnaires n a voluntary basis about their eating habits, their lifestyle, and their health status. Think about the case where the association offers to some diabetic employees coupons for discounted products, given that these individuals have chosen to hide while filling in the questionnaires the fact that they have diabetes. Last year, for the first time, the company's administration asked the employee association for this data set to analyze it, for the benefit of the employees, by using ML techniques to create inference rules that will help accommodate the employees' needs better. The association wonders whether it should provide the data set to the company's administration team mainly because of the indirect disclosure of sensitive data of certain employees through inference channels. One of the paths (rules) of the DT that was deduced from this data set matches Mary's profile, who recently was diagnosed with diabetes, inferring in this way that Mary must be a diabetic. This conclusion was made even if Mary did not provide this information through the questionnaires, mainly because still other individuals who provided this information for themselves share the same habits with Mary. Deletion of the data concerning all the individuals that match the user profile of Mary, which was used to construct the DT would be an insufficient fix to the inference problem since important information that may be needed for other reasons will be lost accidentally. The best approach to address this issue would be to apply an inference control mechanism to hide some specific nodes of the DT without disturbing the overall structure of the original tree and losing the intrinsic value of the remaining rules.

1.2. Background

In articles [5-8], the authors proposed a series of strategies that would effectively protect against the disclosure of the sensitive classification rules. The LDH algorithm [9] was developed on the basis of the concept of preserving sensitive DT rules resulting from the use of data mining techniques. LDH algorithm minimally changes the initial dataset, resulting in the DT generated through hiding being syntactically close to the original one. The modified dataset generated by the LDH algorithm can be shared without any concern for disclosing the sensitive rule. In article [10], the authors presented a Java-based prototype that implements the LDH algorithm.

2. Applying LDH Algorithm for Inference Control Purposes

In our experiments, we used various performance metrics to compare the efficiency of the deduced DT with the original one. One of the most popular algorithms for rule-based classification, the C4.5 algorithm [11], uses the gain ratio as the splitting criterion. In every iteration, the attribute with the highest gain ratio is chosen as the splitting attribute.

Therefore, if we want to suppress a specific attribute test at a node, it would be a reasonable heuristic to change the values of the instances that would enter that node. By this change, the resulting gain ratio (due to that attribute) will be decreased and be equal to zero, where possible. The LDH algorithm locates the parent node of the leaf to be hidden and ensures that the attribute tested at that node will not generate a splitting, which would allow that leaf to re-emerge. A schematic diagram of the LDH algorithm's workflow is shown in Figure 1.

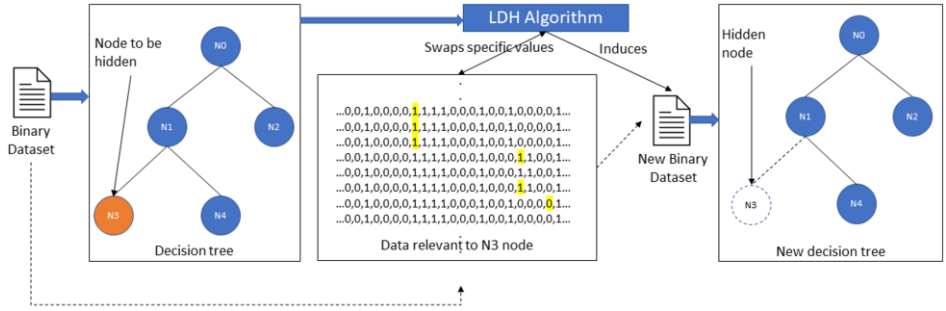


Figure 1. A schematic diagram of the LDH algorithm's workflow

The Java-based prototype [10] uses the Weka library to read an Attribute-Relation File Format (ARFF) data set and visualize the J48 DT with the provided interactive tree visualizer class. Afterward, the user may select an attribute represented with a node in the tree visualizer window, and subsequently, the prototype implements the LDH algorithm. The output of the software is the modified data set and the resulting DT.

3. LDH Implementation on a Diabetes Dataset

This section shows an example regarding an early-stage diabetes risk prediction data set [12] from the UC Irvine Machine Learning Repository [13]. This data set contains reports of 520 persons who have recently become diabetic or are still nondiabetic but have few or many symptoms. The data set includes fifteen binary attributes and one numerical (Age). We modified the original data set by removing Age's numerical variable since the LDH algorithm works only with binary features. The binary variables are Sex, Polyuria, Polydipsia, Sudden weight loss, Weakness, Polyphagia, Genital thrush, Visual blurring, Itching, Irritability, Delayed healing, Partial paresis, Muscle stiffness, Alopecia, and Obesity. The corresponding values for the above binary variables are Yes or No, whereas the class variable represents whether the patient is having a risk of diabetes (positive) or not (negative). We chose for our experiments to use the WEKA [14] framework, an ML software written in Java, and, more specifically, the *J48* classifier, the implementation of the C4.5 algorithm. We performed eight experiments using the Java-based prototype of the LDH algorithm [10] to evaluate the impact of each hiding regarding the DT's performance metrics. Each hiding was applied to every one of the eight terminal nodes of the original DT. Since the DTs mentioned above are too big to fit into a single page, they are available on the website [15]. All dataset files (.arff), before and after hiding, was applied, and the corresponding outputs are also available on the website [15]. Each of the eight terminal nodes was successfully hidden in all experiments, and the corresponding DTs were syntactically close to the original DT. The main performance

metrics of the original DT and the DTs induced from the modified data sets are presented in Table 1.

Table 1. Main performance metrics of the DTs induced from the original and the modified datasets.

	Original	Modified
Kappa statistic	0.9878	[0.9756-0.9878]
Mean absolute error	0.0084	[0.0084-0.0165]
Root-mean-squared error	0.0648	[0.0648-0.0907]
Relative absolute error	1.772%	[1.772%-3.478%]
Root relative squared error	13.314%	[13.314%-18.652%]

4. Conclusion

In this article, several experiments on inference control in a diabetes data set are presented where eight terminal nodes are hidden using a Java-based prototype that implements the LDH algorithm, all of which yielded satisfactory results.

References

[1] Nazir S, Khan S, Khan HU, Ali S, Garcia-Magarino I, et al. A Comprehensive Analysis of Healthcare Big Data Management, Analytics and Scientific Programming. IEEE Access. 2020;8:95714–33.
[2] Iyengar A, Kundu A, Pallis G. Healthcare Informatics and Privacy. IEEE Internet Comp. 2018;22:29–31.
[3] Verykios VS, Bertino E, I. Fovino I, Provenza L, Saygin Y, Theodoridis Y. State-of-the-art in privacy-preserving data mining. ACM SIGMOD Record 2004;33:50.
[4] Farkas C, Jajodia S. 2002. The inference problem: a survey. SIGKDD Explor. Newsl. 2002;4:6–11. doi:https://doi.org/10.1145/772862.772864
[5] Kalles D, Verykios VS, Feretzakis G, Papagelis A. Data set operations to hide decision tree rules. In Proceedings of the Twenty-second European Conference on Artificial Intelligence, Hague, The Netherlands, 29 August–2 September 2016.
[6] Kalles D, Verykios VS, Feretzakis G, Papagelis A, Data set operations to hide decision tree rules. In Proceedings of the 1St International Workshop on AI for Privacy and Security—Praise '16, Hague, The Netherlands, 29–30 August 2016.
[7] Feretzakis G, Kalles D, Verykios VS. On Using Linear Diophantine Equations for Efficient Hiding of Decision Tree Rules. In Proceedings of the 10th Hellenic Conference on Artificial Intelligence—SETN '18, Patras, Greece, 9–12 July 2018. doi:https://doi.org/10.1145/3200947.3201030
[8] Feretzakis G, Kalles D, Verykios VS. On Using Linear Diophantine Equations for in-Parallel Hiding of Decision Tree Rules. Entropy 2019;21(1):66. doi: https://doi.org/10.3390/e21010066
[9] Feretzakis G, Kalles D, Verykios VS. Using Minimum Local Distortion to Hide Decision Tree Rules. Entropy 2019;21:334, doi: https://doi.org/10.3390/e21040334
[10] Feretzakis G, Mitropoulos K, Kalles D, Verykios VS. Local Distortion Hiding (LDH) Algorithm: a Java-based prototype. In 11th Hellenic Conference on Artificial Intelligence (SETN 2020). Association for Computing Machinery, New York, NY, USA, 144–149. doi:https://doi.org/10.1145/3411408.3411419
[11] Quinlan JR, C4.5. Programs for Machine Learning. Morgan Kaufmann: CA, USA, 1993.
[12] Islam MMF, Ferdousi R, Rahman S, Bushra HY. Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques. Computer Vision and Machine Intelligence in Medical Image Analysis Advances in Intelligent Systems and Computing. 2019:113–25.
[13] Dua, D, Graff C. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml] 2019. Irvine, CA: University of California, School of Information and Computer Science.
[14] Frank E, Hall MA, and Witten IE. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.
[15] Feretzakis G. Hiding Decision Tree Rules in a Diabetes Data Set using LDH Algorithm Java-based prototype. Available online: http://www.learningalgorithm.eu/datafiles_ICIMTH2021.html (accessed on 17 July 2021).

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210947

Admission and Discharge Following Ambulance Transport to the Emergency Department

Aikaterini SAKAGIANNI [a,†], Georgios FERETZAKIS[b,c,d,1,†], Georgios KARLIS[e], Evangelos LOUPELIS[d], Lazaros TZELVES[f], Rea CHATZIKYRIAKOU[g], Nikolaos TRAKAS[h], Eugenia KARAKOU[h], Stavroula PETROPOULOU[d], Aikaterini TIKA[i], Zoi RAKOPOULOU[i], Ilias DALAINAS[i] and Vasileios KALDIS[j]

[a] *Intensive Care Unit, Sismanogleio General Hospital, Marousi, Greece*
[b] *School of Science and Technology, Hellenic Open University, Patras, Greece*
[c] *Sismanogleio General Hospital, Department of Quality Control, Research and Continuing Education, Marousi, Greece*
[d] *Sismanogleio General Hospital, IT department, Marousi, Greece*
[e] *Thoracic Diseases General Hospital Sotiria*
[f] *Second Department of Urology, National and Kapodistrian University of Athens, Sismanogleio General Hospital, Athens, Greece*
[g] *Sismanogleio General Hospital, Hematology Laboratory, Marousi, Greece*
[h] *Sismanogleio General Hospital, Biochemistry Department, Marousi, Greece*
[i] *Sismanogleio General Hospital, Administration, Marousi, Greece*
[j] *Sismanogleio General Hospital, Emergency Department, Marousi, Greece*

† Equal contribution (1st authors)

Abstract. Emergency ambulance use is deemed necessary for the transport of acutely ill patients to hospital emergency departments (ED). However, some patients are discharged as they present low acuity or chronic problems and should receive primary healthcare services, while the most severely ill are admitted. In the present study, we examined the descriptive epidemiology of ambulance transports for emergencies in the ED by utilizing the data of the information systems of a public tertiary general hospital in Greece. More than half of the patients transferred to the ED by an ambulance were finally admitted to the hospital (52.25%), whereas only one-third (33.74%) of those transferred by other means. A statistically significant association was detected between ambulance use and hospital admission. Age was also statistically significantly higher in the ambulance group. Higher mean values of creatinine, CRP, LDH, urea, white-blood-cell count, and neutrophils were detected in the ambulance group, in contrast to hemoglobin and lymphocyte count which were higher in the non-ambulance group.

Keywords: emergency medical services, ambulances, hospital information systems

[1] Corresponding Author, Georgios FERETZAKIS, PhD, E-mail address: georgios.feretzakis@ac.eap.gr

1. Introduction and Background

Emergency ambulance use is usually reserved for transporting acutely ill patients to hospital emergency departments (ED). However, a substantial proportion of ambulance users are discharged as they present low acuity or chronic problems and could benefit from primary healthcare services, while the most severely ill are admitted to hospitals [1-3]. A significant increase in ambulance demand has been observed worldwide, reaching 29.2%, in a recent 8-year observational study in Melbourne, Australia [4]. Several articles in the literature have tried to elucidate the phenomenon of increased use of emergent care services by patients with non-acute situations [5-7]. Factors as older age, low socioeconomic status, and homelessness have been suggested [6,7]. Similarly, non-emergent ambulance calls by patients with minor illnesses may have severe implications regarding medical resources adequacy for those who really need them [8]. Moreover, increased ambulance transport of patients with primary care medical problems leads to the overcrowding of hospital emergency departments and potentially to the provision of lower quality healthcare services [5]. The aim of this study was to compare the characteristics of patients visiting the ED in relation to the use of ambulance transport and the ED outcome (admission or discharge).

2. Materials and Methods

This research is an observational retrospective statistical analysis conducted in the ED of a public tertiary care hospital in Greece that has been approved by the Institutional Review Board of Sismanogleio General Hospital (Ref. No 15177/2020, 5969/2021). Between 14 March and 4 May 2019, 2903 ED visits were recorded (missing values of the Ambulance variable were excluded from the statistical analysis). We investigated biochemical markers and coagulation tests that are routinely checked in patients visiting the ED, in relation to the use of ambulance transport and the ED outcome (admission or discharge). Raw data were retrieved from a standard Laboratory Information System (LIS) and a hospital information system (HIS). The data were analyzed by using IBM SPSS Statistics version 27.0 [9]. The data set includes the following variables: ambulance use (Ambulance), serum levels of Urea (UREA), Creatinine (CREA), Lactate Dehydrogenase (LDH), Creatine Kinase (CPK), C-Reactive Protein (CRP), Complete Blood Count with differential, including white blood cells (WBC), neutrophil count (NEUT%), lymphocyte count (LYM%), hemoglobin (HGB), and platelets (PLT), Activated Partial Thromboplastin Time (aPTT), D-Dimer, International Normalized Ratio (INR), age, gender, and triage disposition to ED unit.

The descriptive statistics of the data for the variables Age, Gender, Ambulance use, and Admission are presented in the following table (Table 1).

Table 1. Summary of descriptive statistics for variables Age, Gender, Ambulance use, and Admission

Age		Gender		Ambulance use		Admission	
Mean	61.17	Male	1453	Yes	511	Yes	1074
Standard deviation	20.82	Female	1450	No	2392	No	1829
Range/IQR	86/33	Total	2903	Total	2903	Total	2903

The patient factors were compared between the ambulance and non-ambulance groups using Pearson's Chi-squared test for categorical variables and independent sample t-test for continuous data, both with a 0.05 level of significance.

3. Results

A total of 2903 patients visited the ED from March 14th until May 4th, 2019, with a 1:1 ratio of males to females. The admission rate was equal to 37%, while the rest 63% was discharged. Among the 2903 ED patients, only 17.6% (n=511) used an ambulance. Patients transported by an ambulance had a higher probability of being admitted to the hospital compared to those who did not use emergency medical services (OR 2.15, 95%CI 1.772-2.607). More than half (52.25%) of the patients transported by an ambulance were admitted to the hospital. The rest of them were discharged as their condition was not deemed urgent. Among patients not transferred to the hospital by ambulance, 33.74% were admitted.

The majority of patients that visited ED (93%) were referred to specific units. About 41% (n=1118) were referred to the internal medicine unit, 23.3% (n=630) to the pulmonology unit, and 15.7% (n=321) to the cardiology unit. Among those transferred to the ED by ambulance, 92.3% (472 out of 511) were referred to a specific ED unit. The respective percentages were 43.2% (n=204) with medical problems, 23% (n=109) with respiratory problems, and 21.8% (n=103) with heart-related problems.

Chi-square analysis showed that the use of ambulance and hospital admission had a statistically significant association ($X^2 = 61.911$, $p < 0.001$). Phi and Cramer's V values were 0.146, which indicates a weak association. On the other side, our statistical analysis showed no statistically significant association between gender and ambulance use ($X^2 = 0.170$, $p = 0.680 > 0.05$). Independent samples t-tests were performed between the ambulance and non-ambulance groups and deduced that the means of the following continuous variables were statistically significantly higher in the ambulance group: CREA, CRP, LDH, UREA, WBC, and NEUT%. On the contrary, the means of HGB and LYM% were statistically significantly lower in the ambulance group. No statistically significant difference was observed between the two groups regarding the other continuous variables. A detailed summary of the independent samples t-tests is presented in the following table (Table 2). Age was also statistically significantly higher in the ambulance group, with a mean difference of 13.11 years, compared to patients not using an ambulance.

Table 2. Summary of the independent samples t-tests in laboratory testing data between the ambulance and non-ambulance groups

	CREA	CRP	LDH	UREA	HGB	LYM%	NEUT%	WBC
Mean Difference	-0.2718	-16.625	-45.630	-13.768	0.875	5.229	-6.111	-1.114
t- value	-4.931	-4.139	-3.729	-7.340	7.522	9.364	-9.762	-4.351
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

4. Discussion

The emergency medical service (EMS), worldwide, is ruled from several life-threatening level protocols of time response. This is a useful tool when the demand for services exceeds resource availability, and it is therefore vital that the appropriate transport request is made. In Greece, the public ambulance service covers the transportation of acutely ill patients to hospitals, as well as scheduled local and interregional urgent and non-urgent inter-hospital patient transfers. The present study examined the descriptive epidemiology of EMS ambulance transports in the ED by utilizing the HIS and LIS from a public tertiary general Greek hospital. Demographic inpatient data were merged with patients' laboratory data of the emergency department, elaborating information on whether the patient arrived by ambulance or not. Older patients and those with more deranged laboratory exams had a greater possibility to use an ambulance service as a life-threatening medical emergency code. However, only half of the patients transported by an ambulance were admitted to the hospital, whereas only one-third of those were not transported by an ambulance. Nevertheless, a medical fast-track service is available in the triage room, permitting patients with minor medical issues to avoid unnecessary laboratory or imaging tests, therefore diminishing their discharge time. From the total number of patients visiting the ED, nearly two-thirds were discharged, meaning that they presented with mild problems, which could be initially managed at primary healthcare services. Although ambulance users are more frequently older, with more severe illness and greater need for admission, in approximately half of the cases, EMS ambulance call was actually unnecessary, and patients were discharged, possibly due to minor or chronic problems. Our results are in accordance with similar studies [1,3,5]. The optimal use of public, pro bono emergency ambulance services by the citizens is desirable as it could help alleviate ED crowding and improve the quality of emergency hospital services. An association between poor patient outcomes and ED overcrowding has been suggested [5]. The role of primary healthcare services should also be strengthened as it is fundamental in achieving better health care services in general.

References

[1] Yazaki H, Nishiura H. Ambulance Transport of Patients with Mild Conditions in Hokkaido, Japan. Int J Environ Res Public Health. 2020 Feb 2;17(3):919. doi: 10.3390/ijerph17030919.
[2] Ting JY, Chang AM. Path analysis modeling indicates free transport increases ambulance use for minor indications. Prehosp Emerg Care. 2006 Oct-Dec;10(4):476-81. doi: 10.1080/10903120600885209.
[3] Hegenberg K, et al. Differences between cases admitted to hospital and discharged from the emergency department after emergency medical services transport. BMJ Open. 2019 Sep 27;9(9):e030636.
[4] Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: A systematic review of causes, consequences and solutions. PLoS One. 2018 Aug 30;13(8):e0203316.
[5] Unwin M, Crisp E, Rigby S, Kinsman L. Investigating the referral of patients with non-urgent conditions to a regional Australian emergency department: a study protocol. BMC Health Serv Res. 2018 Aug 20;18(1):647. doi: 10.1186/s12913-018-3411-4.
[6] Andrew E, Nehme Z, Cameron P, Smith K. Drivers of Increasing Emergency Ambulance Demand. Prehosp Emerg Care. 2020 May-Jun;24(3):385. doi: 10.1080/10903127.2019.1635670.
[7] Coster JE, Turner JK, Bradbury D, Cantrell A. Why Do People Choose Emergency and Urgent Care Services? A Rapid Review Utilizing a Systematic Literature Search and Narrative Synthesis. Acad Emerg Med. 2017 Sep;24(9):1137-1149. doi: 10.1111/acem.13220.
[8] Norberg G, Wireklint Sundström B, Christensson L, Nyström M, Herlitz J. Swedish emergency medical services' identification of potential candidates for primary healthcare: Retrospective patient record study. Scand J Prim Health Care. 2015;33(4):311-7. doi: 10.3109/02813432.2015.1114347.
[9] IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210948

Cory COVID-Bot: An Evidence-Based Behavior Change Chatbot for COVID-19

Simon T. VAN BAAL[a], Suong LE[b,c], Farhad FATEHI[d,1], Jakob HOHWY[a,e] and Antonio VERDEJO-GARCIA[d,f]
[a] *Cognition and Philosophy Lab, Monash University, Australia*
[b] *Department of Gastroenterology and Hepatology, Monash Health, Australia*
[c] *School of Clinical Sciences, Monash University, Australia*
[d] *School of Psychological Sciences, Monash University, Australia*
[e] *Centre for Consciousness & Contemplative Studies, Monash University, Australia*
[f] *Turner Institute for Brain and Mental Health, Monash University, Australia*

Abstract. Cory COVID-Bot is an artificial intelligence chatbot designed and built by a multisector collaboration to help people safely step towards COVID normal. Achieving COVID normal and avoiding unnecessary adverse health outcomes requires effective communication to the public regarding COVID safe behaviors, but reaching young, culturally and linguistically diverse members of the community is challenging for government. Cory COVID-Bot was developed to directly engage with difficult to reach populations in English and Vietnamese. In order to resolve public ambiguity and uncertainty about public health guidelines, and to stimulate safe behavior, Cory COVID-Bot provides updated recommendations and behavior change interventions, which emphasize the importance of COVID safe behaviors.

Keywords. Chatbot, COVID-19, Behavior Change, Artificial Intelligence

1. Introduction

Promoting adherence to behavioral changes remains a cornerstone strategy to prevent and contain COVID-19 resurgences. Based on the 2009 H1N1 experience, repeated messaging from government and public health experts often does not impact the likelihood of people routinely engaging in infection prevention behaviors [1]. In particular, there has been a global lack of adherence to COVID-safe behaviors in difficult to reach communities due to age, language or ethnicity during the COVID-19 lockdowns.

There is no scalable solution that can simultaneously educate hard-to-reach members of the community and shift human behavior at a general population level. Chatbots are gaining momentum within mobile health settings for lifestyle risk factor modification and chronic disease management. We proposed to overcome communication problems and bridge the COVID-safe "knowledge-action" gap through the rapid development and validation of a user-friendly, low-cost and scalable chatbot, Cory COVID-Bot.

[1] Corresponding Author, Farhad Fatehi MD PhD, School of Psychological Sciences, Monash University, 18 Innovation Walk, Clayton VIC 3800, Australia; E-mail: farhad.fatehi@monash.edu.

2. Methods

We adopted an agile design approach to build a de novo chatbot as a targeted behavior change intervention tool to combat non-adherence to COVID-19 safe behaviors. The chatbot provides essential information about the prevention, diagnosis and treatment of COVID-19 based on natural language processing that uses training phrases and follow-up questions in order to engage users in a conversation rather than didactic education. Users can interact with Cory through free text, and he can reply through pre-set responses, media, and resource-carousels. Cory COVID-Bot delivers evidence-based behavioral interventions in the form of short animations. The avatar of the chatbot, Cory, is a community librarian in his mid-thirties. He was designed to be a collaborative model citizen (he can demonstrate how he uses a mask in a social situation); resourceful and pragmatic (he can geolocate the closest testing center); informative and accurate (real-time information regarding the latest restrictions relevant to your area) and empathetic (helps to self-manage loneliness, anxiety, and low mood). Cory COVID-Bot has an affable, knowledgeable, and conscientious persona, able to respectfully interact with all demographics, while providing clear, practical, and actionable information (Figure 1).

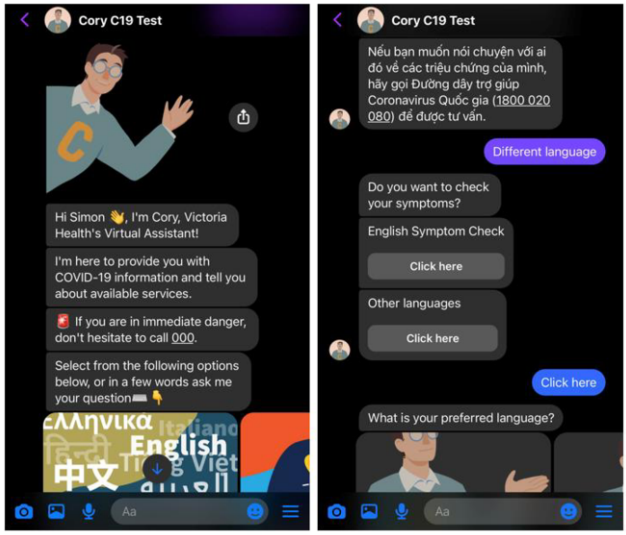


Figure 1. Two screenshots of Cory COVID-Bot deployed in Facebook Messenger app.

The project team developed a series of behavioral interventions to be delivered by Cory for promoting specific actions vital to curtailment of COVID-19. These behavioral interventions were intended to remove mental and practical hurdles to desired COVID safe practices. This is effective because people have an intrinsic tendency to procrastinate decisions if they perceive the decisions as complex, if the consequences seem distant (e.g. present bias), or if they perceive the decision as important [2,3].

The interventions were informed by the literature on behavior change; we focused on two issues with relevance in the pandemic. Altruism (which includes COVID-safe behavior) can be motivated by compassion, which can be trained [4], and that people tend to underestimate exponential disease transmission [5]. The first behavioral

intervention was aimed at diminishing psychological threshold for COVID-19 testing by providing a clear and resonant rationale for why getting tested is essential for helping others including loved ones. The animation shows a family separated when one of the family members contracts COVID-19. This intervention invoked empathy and compassion in the viewer, which has been shown to increase willingness to comply with public health guidelines [6]. The second behavioral intervention addressed exponential growth bias, in this case the tendency to underestimate the increase of case numbers over time, in order to increase the likelihood of getting tested, address personal responsibility and emphasize the exponential nature of viral transmission [5].

3. Results

We built a scalable and practical artificial intelligence (AI) chatbot that can be deployed on various platforms, such as Facebook Messenger, and can therefore be adapted for different purposes as the COVID pandemic evolves. Cory is an AI chatbot capable of delivering targeted behavioral interventions when deemed appropriate in a conversation, thereby increasing its impact as a population-based intervention. In this case, when the individual interacting with Cory asks about COVID-19 symptoms, he would provide them with the necessary information, and prompt them with an animation aimed at eliciting empathy and compassion (see Figure 2), an animation that aims to correct exponential growth bias (see Figure 3), or skips the animations with equal chance.

Figure 2. Empathy manipulation, a man did not wear a mask (left), and caused family separation (right).

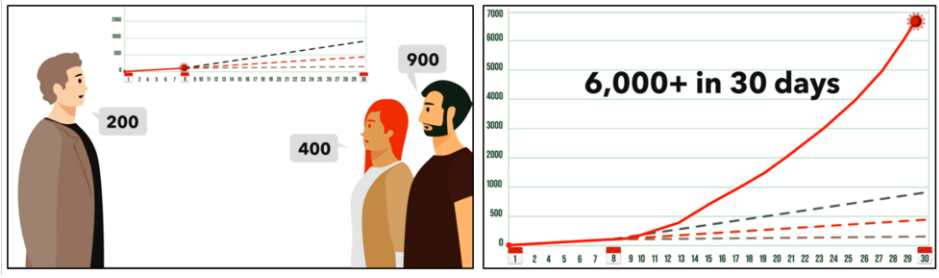


Figure 3. Exponential growth bias intervention, people in the picture on the left guess the trajectory of an outbreak, and an exponential growth curve is traced on the right.

Afterwards, Cory would ask how likely it was the user would get tested. This type of question generates tangible behavior change data, and enables researchers and

developers to analyze whether the intervention is likely to be effective. We thus prepared Cory for conducting a randomized controlled trial, the results of which will be communicated after testing is completed. This setup also allows for continual updating and fine-tuning of the behavioral interventions.

4. Discussion

This project shows that it is possible to develop an artificial intelligence chatbot that informs people about public health guidelines and delivers targeted behavioral interventions. Cory is low-cost, adaptable and agile. Once Cory has engaged with a user, he can maintain a direct line of communication through push notifications. The design allows for translation into different languages, swift adjustments in light of new information, and adaptation to new use potential such as addressing vaccine hesitancy. These are all desirable attributes in a pandemic where information readily changes, misinformation and disinformation are rife on social media channels, and at-risk communities need to selectively target.

We have found that behavioral interventions delivered via animations allow for more degrees of freedom than text-based interventions, which means more effective manipulations can be pursued through iteration. Various strategies can be used to identify whether the chatbot has been effective at achieving the aims set by the project team, which the team can use to update the chatbot content and functionality. Chatbots could play a meaningful role in the pandemic, and could be considered for deployment on platforms that are popular among hard-to-reach communities as both a trusted advisor and an unbiased source of information.

References

[1] Gray L, MacDonald C, Mackie B, Paton D, Johnston D, Baker MG. Community responses to communication campaigns for influenza A (H1N1): a focus group study. BMC Public Health. 2012 Dec;12(1):1-2.

[2] Beshears J, Choi JJ, Laibson D, Madrian BC. The importance of default options for retirement saving outcomes: Evidence from the United States. Social security policy in a changing environment: University of Chicago Press; 2009. p. 167-95.

[3] O'Donoghue T, Rabin M. Choice and procrastination. The Quarterly Journal of Economics. 2001;116(1):121-60.

[4] Weng, HY, Fox, AS, Shackman, AJ, Stodola, DE, Caldwell, JZ, Olson, MC, Rogers, GM, Davidson, RJ. Compassion training alters altruism and neural responses to suffering. Psychological science, 2013;24(7):1171-1180.

[5] Lunn PD, Timmons S, Belton CA, Barjaková M, Julienne H, Lavin C. Motivating social distancing during the Covid-19 pandemic: An online experiment. Social Science & Medicine. 2020;265:113478.

[6] Pfattheicher S, Nockur L, Böhm R, Sassenrath C, Petersen MB. The Emotional Path to Action: Empathy Promotes Physical Distancing and Wearing of Face Masks During the COVID-19 Pandemic. Psychological Science. 2020;31(11):1363-73.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210949

Geographic Information System-Based Mobile Application Design for Health Care in Older Persons in Rural Community by Village Health Volunteers

Kukiat TUDPOR[a,b], Adisorn WONGKONGDECH[a], Ranee WONGKONGDECH[c], Thanyarat CHAIYAKARM[d], Worawit JITSUKKA[d], Kemika SOMBATEYOTHA[a], Songkramchai LEETHONGDEESAKUL[a], Narongsakdi KUBOONYA-ARAGSA[e], Chalard CHANTARSOMBAT[f], Krissanachai KIMCHAI[g], Nutchanun KAMONROEK[h] and Niruwan TURNBULL[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*
[c] *Faculty of Medicine, Mahasarakham University, Thailand*
[d] *Faculty of Humanities and Social Sciences, Mahasarakham University, Thailand*
[e] *Department of Labour Protection and Welfare, Ministry of Labour, Thailand*
[f] *Faculty of Education, North Eastern University, Thailand*
[g] *Northeasthern Regional Center for Primary Health Care Development, Thailand*
[h] *Ubon Ratchathani Provincial Health Office, Thailand*

Abstract. Health care for older persons is important for their well-being. In Thailand, village health volunteers (VHVs) play a crucial role in communicating between the older persons and health care professionals. Traditionally, they record and submit a monthly report as papers. This project was aimed to design a geographical information system (GIS)-based mobile application for health care in older persons in rural community by the VHVs. As a part of the analysis, design, development, implementation, and evaluation (ADDIE) model, health information obtained from 8,348 older persons and 1,125 VHVs living in rural community were analyzed and used for the for iPhone (iOS) and Android devices, and web browser. In summary, this mobile application allows the VHVs locate the older person's addresses and communicate with health care staff online in time-saving manner.

Keywords. Geographic information system, Mobile application, Older persons, Health care

1. Introduction

World's older population is drastically growing. Based on data from an Office of the National Economic and Social Development Council (NESDC), Thailand is considered as an aged society in 2021 since its elderly population has reached 18.7 percent of the

[1]Corresponding author, Niruwan Turnbull, Faculty of Public Health, Mahasarakham University, Kham Rieng, Kantharawichai, Maha Sarakham, Thailand 44150; E-mail: niruwan.o@msu.ac.th

total population (1, 2). Older persons have are prone to multiple age-related changes such as tissue and organ simultaneous degeneration, increased sensitivity to diseases and environmental stressors (3). Therefore, continuous monitoring of health and diseases are crucial for efficient health care services. However, most of older persons in Northeastern part of Thailand live in rural areas faraway from health care facilities. Subdistrict Health Promotion Hospital (SHPH) system has been a key unit providing primary health care for people living in rural areas of Thailand. In close relationship with the SHPH, village health volunteers (VHVs) play a pivotal role in primary care system. Their tasks are to provide health screening, self-care advice, basic health care, and communication between villagers and hospitals. Normally, VHVs have to submit data to the hospital information center on monthly basis by papers or online, depending on policies of the local SHPH. This present project was aimed to demonstrate a designing process of a GIS-based mobile application for health care in rural community-dwelling older persons by the VHVs.

2. Methods

2.1. Informants and data collection

As a part of the analysis, design, development, implementation, and evaluation (ADDIE) model in the project "Development of Mobile Application of Database of Older Persons Using Geographic Information System (GIS) to Detect and Analyze Risks of Chronic Diseases, Quality of Life, and Mental Illness by Village Health Volunteers in 7th Regional Health Office Territory, (VHVA)", all older persons (N = 8,348) and all VHVs (N = 1,125) were from 8 subdistricts in 4 provinces. Inclusion criteria of the older persons were age ≥ 60 years old and signed consent form. The older adults not physically present in the areas during the survey were excluded. All procedures have been conducted in accordance with the Declaration of Helsinki and approved by the Ethical Review Committee for Human Research, Maha Sarakham Provincial Public Health Office (No.6/2564). Partial socio-demographic profiles of the elders are presented in Table 1.

Table 1. Socio-demographic profiles of the older persons (N = 8,348)

Profile	Category	n (%)
Gender	Male	3,587 (42.97)
	Female	4,757 (56.98)
	LGBT	4 (0.05)
Age	60-69	4,597 (55.07)
	≥ 70	3,751 (44.93)
Marital status	Married	5,353 (64.12)
	Single (widowed/divorced/separate)	2,995 (35.88)
Living arrangement	Living with family	7,869 (94.26)
	Living alone	479 (5.74)
Educational level	Higher than primary school	615 (7.37)
	Primary school or no schooling	7,733 (92.63)
Occupational status	Working	5,899 (70.66)
	Not working	2,449 (29.34)
Economic status	Enough income for living	2,912 (34.88)
	Not enough income for living	5,436 (65.12)
Activities of daily living (ADL) level	Independent	8,122 (97.29)
	Dependent	226 (2.71)
Non-communicable disease (NCD)	No hypertension or diabetes mellitus	4,487 (53.74)
	Hypertension or diabetes mellitus	3,861 (46.26)

2.2. Application design

The mobile application was built up with the Apache Cordova™ platform for both iOS and Android devices. The Google Map Application Programming Interface (API) was used by the web administrator to view and share latitude and longitude. After deployment, the app runs in the web browser. Only authorized web administrator had an access to the encrypted server. Data exports and printouts can be obtained by SHPH staff with secured passwords.

3. Results

3.1. User interface

Information of older persons collected by the VHVs on mobile was managed by the web admin. Examples of user interface shown in Figure 1 consists of 1) login page, 2) menus (purple, health assessment; orange, knowledge resources; green, home visit; and red, emergency call), and 3) location search page. SHPH staff were able to double check information and provide feedback.

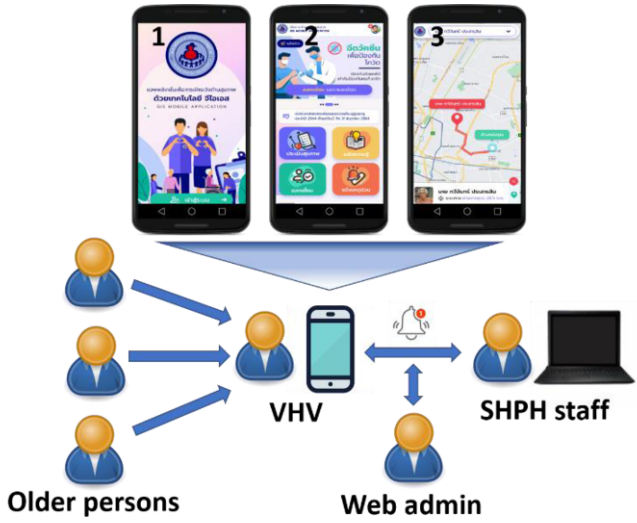


Figure 1. Mobile application information flow and screen displays

Table 2. Outcome measures and data collection tools for the mobile application

Outcome measure	Data collection tools
Geographic coordinate	Global navigation satellite system (GNSS) and GIS programs
Health screening and medical record	Secondary data from SHPH
Psychological stress level	Newly validated questionnaire
Quality of life (QOL)	World Health Organization Quality of Life for Older Persons (WHOQOL-OLD) questionnaire
Activities of daily living (ADLs)	Barthel Index for Activities of Daily Living questionnaire
Knowledge, attitude, and practice (KAP) on health care	Newly validated questionnaires

3.2. Outcome measures and data collection tools

This mobile application contained 6 outcome measures and data collection tools for VHVs and local heal care staff: geographic coordinate, health screening and medical history, psychological stress level, QOL, ADLs, and KAP on health care (Table 2).

4. Discussion

4.1. GIS-based mobile applications for health care in older persons

More than 2,000 mobile health applications for older person were invented, but none were designed for dependent older persons (4). Our application was designed for both dependent and independent groups and was the first GIS-based mobile application for health care in older persons by the VHVs in Thailand. The GIS has been proved as an efficient and travel-time saving tool for people living in remote areas (5).

4.2. User interface and utilities

Our mobile application design was based on user requirements. GIS applications can be divided into 4 themes: health surveillance, support, promotion and prevention; and communication among health care staff (6). Our application served all of these aspects. Limitation of the project was a gap of communication between the design team and the VHVs and the older persons during COVID-19 pandemic; and telecommunication was applied in substitute for face-to-face contact.

5. Conclusion

In conclusion, the GIS-based mobile application for health care in older persons living in rural community of Thailand by the VHVs has been designed. The application will be further developed, implemented, evaluated based on feedback from the users.

References

[1] Khamken PW. Sufficiency Economy Philosophy Towards Poverty Eradication in Thailand. Psychology and Education Journal. 2021;58(1):1406-11.
[2] Lee SB, Oh JH, Park JH, Choi SP, Wee JH. Differences in youngest-old, middle-old, and oldest-old patients who visit the emergency department. Clin Exp Emerg Med. 2018;5(4):249-55.
[3] Nigam Y, Knight J, Bhattacharya S, Bayer A. Physiological changes associated with aging and immobility. J Aging Res. 2012;2012:468469.
[4] Paiva JOV, Andrade RMC, de Oliveira PAM, Duarte P, Santos IS, Evangelista ALP, et al. Mobile applications for elderly healthcare: A systematic mapping. PLoS One. 2020;15(7):e0236091.
[5] Weiss DJ, Nelson A, Vargas-Ruiz CA, Gligoric K, Bavadekar S, Gabrilovich E, et al. Global maps of travel time to healthcare facilities. Nat Med. 2020;26(12):1835-8.
[6] Nhavoto JA, Gronlund A. Mobile technologies and geographic information systems to improve health care systems: a literature review. JMIR Mhealth Uhealth. 2014;2(2):e21.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210950

Exploring the Influence of Users' Characteristics in Continuance Intention to Use Electronic Medical Records (EMR): Extending the Expectation Confirmation Model (ECM)

Zahra MEIDANI[a,b], Hashem MOHAMMADIAN[c], Fatemeh MOBARAKI[a], Walter SWOBODA[d], Gholam Abbass MOOSAVI[e] and Sahar ZARE[a,1]
[a] *Health Information Management Research Center, Kashan University of Medical Sciences, Kashan, Iran*
[b] *Department of Health Information Management & Technology, Kashan University of Medical Sciences, Kashan, Iran*
[c] *Department of Health Education and Health Promotion, Ahvaz University of Medical Sciences, Ahvaz, Iran*
[d] *DigiHealth Institute, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany*
[e] *Department of Biostatistics and Epidemiology Kashan University of Medical Sciences, Kashan, Iran*

Abstract. Even though the "user characteristics" (UC) can influence "acceptance" and "the continuance intention to use" of information systems such as electronic medical records (EMR), the effect of UC has not been adequately evaluated in post-adoption models. This study seeks to examine the effect of UC on post-usage of EMR using the expectation confirmation model (ECM). A total of 450 questionnaires was collected by a survey to extend ECM by integrating UC using structural equation modeling (SEM). Data were analyzed using LISREL through confirmatory factor analysis. A path analysis test was also performed to fit and confirm the model fit. The UC affects the "confirmation of expectations" directly and the "continuance intention to use" indirectly. The findings of the present study can provide a new path for further research on the extending of the ECM model. Moreover, end users' characteristics should be considered in the preparation and training phases of EMR implementation.

Keywords. Information systems, electronic health record, intention, motivation

1. Introduction

EMRs play an important role in the healthcare industry through facilitating timely access to health information, decreasing medical errors, and improving patient care

[1] Corresponding author, Sahar Zare, Health Information Management Research Center, Kashan University of Medical Sciences, Kashan, Iran; Email: zare-s@kaums.ac.ir

effectiveness and efficiency. Despite the potential benefits of such information technologies, developing and deploying these technologies is not enough since the technology should be accepted and used by end-users [1]. In spite of the fact that EMRs are expected to create better healthcare outcomes, the slow rate of its adoption has been a concern in developing countries [2]. Thus, identifying, predicting, and managing users' adoption and post-usage of EMR is one of the key factors in its successful implementation [3]. The expectation confirmation model (ECM) is one of the earliest information system (IS) post-usage and continuance models that integrates major constructs in terms of perceived usefulness, expectation confirmation (EC), and satisfaction (SA) in continuance intention to use (IU) information systems including EMRs [4]. Although ECM is a robust model for understanding the determinants of IS continuance, there is a variety of contributing factors such as the opinions of others, external factors, and user characteristics which have not been addressed by the ECM model. It is shown that user characteristics have significant effect on performance expectancy [5]; in addition, "user characteristics" including self-efficacy, job satisfaction, computer anxiety, age, gender, and computer literacy can influence "the continuance intention to use" which affect job satisfaction and users performance [2-7]. thus, this study aims to extend ECM by incorporating "user's characteristics" into this model.

The research hypotheses were the following:

- User's characteristic (UC) is related to expectation confirmation (EC)
- User's characteristic is related to perceived ease of use (EU)
- User's characteristic is related to perceived usefulness (PF)
- User's characteristic is related to perceived enjoyment (PE)
- Expectation confirmation is related to perceived EU
- Expectation confirmation is related to PF
- Expectation confirmation is related to PE
- Perceived ease of use is related to satisfaction (SA).
- Perceived usefulness is related to SA
- Perceived enjoyment is related to SA
- Users' satisfaction is related to continuance intention to use (IU).

2. Method

A total of 450 questionnaires was collected by a survey to extend ECM by integrating user characteristics using structural equation modeling (SEM). A confirmed questionnaire [7,8] containing 44 questions in three sections including demographic data, users' characteristics, and ECM model constructs were used to collect data. EMR end-users' responses were evaluated based on a five-point Likert strongly agree (5) to strongly disagree (1). Data were analyzed using SPSS and LISREL through confirmatory factor analysis. R-squared coefficients (R2) were used to assess the structural model computed for all endogenous constructs. A path analysis test was also performed to fit and confirm the model fit using LISREL software. The causal relationships of dependent and independent variables were measured through standard coefficient and significance values by LISREL software; consequently, a decision was made to confirm or reject the hypotheses based on the measured values. The current study was supported by the Vice-Chancellorship for Research & Technology at Kashan

University of Medical Sciences [Code# IR.KAUMS.MEDNT.REC.1393.1232] and grant number 9330.

3. Result

72% (n = 324) of the participants in this study were female and 28% (n = 126) were male and most of them (59.9%) had a bachelor's degree. The direct, indirect, and total effect of variables on the dependent variable of "continuance intention to use" in the final model are presented in Table 1.

Table 1. Results of the model validity

Variable	Direct relationship	Indirect relationship	Total effect
User Characteristic (UC)	-	-0.023	-0.023
Expectation confirmation (EC)	-	0.513	0.513
Ease of use (EU)	0.47	0.47	0.47
Perceived usefulness (PF)	0.27	0.27	0.27
Perceived enjoyment (PE)	-	0.061	0.061

(R^2) IU 0.83, (R^2) SA 0.50, (R^2) EU 0.87, (R^2) PF 0.78, (R^2) PE 0.55, (R^2) EC 0.033

As shown in table 1, "Ease of use" with a value of 0.47 had the most impact on the "continuance intention to use". The other factors including "perceived benefits" (0.27), "confirmation of expectations" (0.513), "satisfaction" (0.061), and "user characteristics" (-0.023), affected the "continuance intention to use" respectively after "ease of use" variable. The results of the SEM path analysis for the final model are indicated in Figure 1.

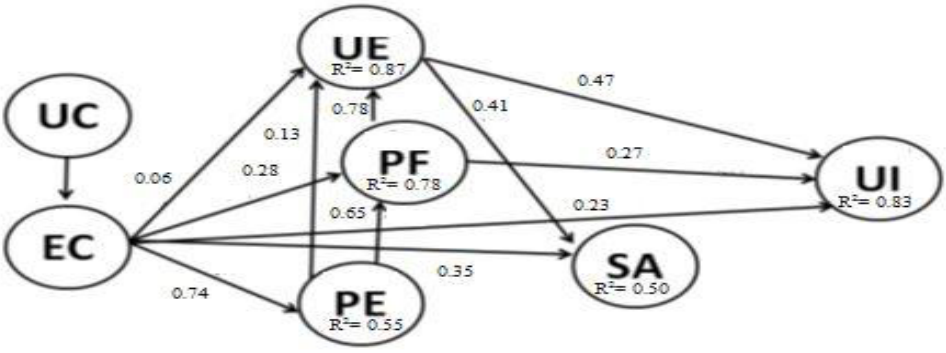


Figure 1. SEM path analysis results Notes: ** $p < 0.01$, * $p < 0.05$

4. Discussion

Although the "user characteristic" has the least effect on the "the continuance intention to use", this construct affects the "confirmation of expectations" directly and the "continuance intention to use" indirectly. Similar to previous studies, "user characteristics" affect users' expectation confirmation [9, 10]; thus, it should be considered in the preparation and training phases of EMR implementation [11]. Considering user's characteristics during EMR implementation and ongoing education not only would ensure the success of adoption and post-usage of EMR but also has a positive impact on end users' satisfaction and efficiency [2]. In contrast to previous studies [12, 13], "user characteristics" had no impact on the ease of use and perceived enjoyment. The findings of the present study can provide a new path for further research on the extending of the ECM model based on the integration of user characteristics and other emerging factors which affect the continuance intention to use information systems.

References

[1] Farzandipour M, Meidani Z, Sadeqi Jabali M, Dehghan Bnadaki R. Designing and evaluating functional laboratory information system requirements integrated to hospital information systems. J Eval Clin Pract. 2019; 25(5):788-799. doi: 10.1111/jep.13074.

[2] Ayanso A, Herath TC, O'Brien N. Understanding continuance intentions of physicians with electronic medical records (EMR): An expectancy-confirmation perspective. Decision Support Systems. 2015;77:112-22

[3] Al-Gahtani SS, King M. Attitudes, satisfaction and usage: factors contributing to each in the acceptance of information technology. Behavior and Information Technology. 1999;18(4):277-97

[4] Bhattacherjee A. Understanding Information Systems Continuance: An Expectation-Confirmation Model. MIS Quarterly. 2001;25(3):351-70

[5] Kim, Sanghyun, Park, Hyun-Sun. An Empirical Study on the Influence of Social Network Services(SNS) and Individual Characteristics on Intention to Continuous Use of SNS2012 Sep 30;11(3):17–38.

[6] Lee H, Choi SY, Kang YS. Formation of e-satisfaction and repurchase intention: Moderating roles of computer self-efficacy and computer anxiety. Expert Systems with Applications. 2009;36(4):7848-59

[7] Arning K, Ziefle M. Understanding age differences in PDA acceptance and performance. Computers in Human Behaviour. 2007;23(6):2904-27

[8] James Y, Thong Se , Kar Y. The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance. int.j.Human-computer Studies 2006:799-810

[9] AL-Khalid M, AL-Jabra I. The Effect of attitudes on computer utilization: new evidence from a developing nation. Computer Attitudes And Utilization 1995

[10] Venkatesh V, Morris MG, Ackerman PL. A Longitudinal Field Investigation of Gender Differences in Individual Technology Adoption Decision Making Processes. Organizational Behavior and Human Decision Processes. 2000;83(1):33-60

[11] Chen YY, Huang H, Hsu Y, al e. Confirmation of Expectations and Satisfaction with the Internet Shopping: The Role of Internet Self-efficacy Computer and Information Science 2010;3(3):14-22

[12] Rangraz Jeddi F, Nabovati E, Bigham R, Khajouei R. "Usability evaluation of a comprehensive national health information system: relationship of quality components to users' characteristics." International journal of medical informatics vol. 133 (2020): 104026. doi:10.1016/j.ijmedinf.2019.104026

[13] Aggelidis VP, Chatzoglou PD. Using a modified technology acceptance model in hospitals. International journal of medical informatics. 2009;78(2):115-26

[14] Hwang Y, Yi MY. Predicting the use of web-based information systems: intrinsic motivation and self-efficacy. ghth Americas Conference on Information Systems. 2002:1076-81

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210951

A Resilience Model for Moderating Outcomes Related to Electronic Medical Record Downtime

Joseph M. WALSH[a,1] and Elizabeth M. BORYCKI[a,b]
[a] *School of Health Information Science, University of Victoria, Victoria, BC, Canada*
[b] *Michael Smith Foundation for Health Research, Canada*

Abstract. The objective of this scoping review is to develop a model to understand the factors that influence clinical downtimes or clinical activities in a healthcare organization**.** To report on the results of searches preformed using seven bibliographic databases, using the logical search criteria of (downtime AND (EMR OR Electronic Medical Record OR EHR OR Electronic Health Record). After a title, abstract and full-text review 26 articles remained. The articles were coded and analyzed for themes. Downtime planning activities mitigate the effects of disasters on patient safety outcomes and clinical delays. A model was developed representing the relationships between disasters, the moderating variable of downtime planning activities and patient safety as well as clinical outcomes. Disasters can have significant impact on patients and health professionals. Downtime planning activities can be enacted when a disaster occurs to moderate the effects of the downtime on patients and clinical activities and can improve safety.

Keywords. Downtime, safety, disasters, resilience

1. Introduction

A recent survey reported that, most North American healthcare institutions experienced an unplanned downtime in the past three years. Nearly 70% of those institutions who completed the survey stated they had experienced a downtime that lasted at least eight hours. Despite those figures, most organizations had not completely implemented a comprehensive downtime plan to address patient safety concerns during Electronic Medical Record (EMR) downtimes [1]. Downtimes can be caused by many different disasters from human error through to natural disasters such as flood, hurricane, and fire [2]. The Intergovernmental Panel on Climate Change (IPCC) refers to these types of environmental events as low-likelihood, high impact events and compound extreme events. Extreme weather events such as these are expected with high confidence to increase in frequency and amplitude according to the recent IPCC report on climate change [3]. During a downtime, some or all functions of an EMR can become unavailable. In both planned and unplanned downtime scenarios, clinical users risk timely access to critical patient information [4]. The objective of this scoping review is

[1] Corresponding author, Joseph M. Walsh, School of Health Information Science, University of Victoria, Victoria BC, Canada; E-mail: joewalsh@uvic.ca

to develop a model to understand the factors that influence clinical downtimes or clinical activities in a healthcare organization.

2. Methodology

To understand the factors influencing clinical downtimes, we performed a scoping study following Arksey and O'Malley's 2005 methodology [5], including the optional consultation stage [6]. In this final stage, subject matter experts were consulted to further refine and interpret the key concepts discovered in the preceding stages. Following the final consultation stage with subject matter experts, a narrative review identified contextual factors and strategies that impact clinical operations during a clinical downtime. A model was created demonstrating the impact of downtime and how planning activities can be used to moderate the effects of a downtime.

3. Results

To capture a broad set of results, searches were preformed using seven different bibliographic databases, using the logical search criteria of (downtime AND (EMR OR Electronic Medical Record OR EHR OR Electronic Health Record)). 672 articles were returned from all sources, with 625 remaining after duplicates were removed.

The remaining results were screened by title, abstract and full text for context. Articles selected must be published in a peer reviewed journal, printed in English, and have abstracts and full text available electronically. One reviewer performed all screening stages, with 30 articles remaining after title and abstract screening. Four additional articles were removed following a full text review as these articles did not discuss technology downtimes in the context of hospital or clinical operations. The remaining 26 articles, published between January 2008 and March 2021, were coded and analyzed for themes.

4. Discussion

The literature shows several patient safety outcomes are affected by downtimes. The theme of patient safety was common in the publications reviewed and included the subthemes of clinical error, patient harm, clinical delays, and data integrity. One study concluded that "downtime-exposed patients had statistically significant longer intraoperative courses and postoperative hospital length of stay" [7]. Many of the included articles directly discuss patient safety during downtime scenarios [8]. Errors occurring during a clinical downtime affect a clinician's ability to access the most relevant patient information. Wang et al. identified "recent surveys of US hospitals have linked downtime to medication errors, increased length of stay and patient harm." [9].

There are several ways that clinical delays affect patient care. Delays can occur in the reporting of test results, length of stay, diagnosis, and treatment. Larsen et al. state that "patients in emergency situations, whose diagnosis depends on timely laboratory results are exposed to significant risks" [10]. Researchers identified the importance of downtime planning such as mitigation in reducing the effects of a downtime on clinical

activities. Mitigation of clinical downtime is considered any action taken before, during, or after a downtime to minimize the impact or duration of a downtime event [11]. Downtime planning is a mitigation strategy commonly discussed in the literature. These downtime plans should be reviewed regularly so that the plans remain current to the evolving clinical workflows and applications [12]. Downtime kits including relevant policy, procedure, and preprinted forms required during a downtime should also be checked and updated regularly. Downtime planning can also include proactive steps taken to build out systems redundancy with backup or clustered systems. One common backup system is the use of uninterruptable power supplies and standby emergency power generators. Other backup systems are online and offline data repositories to use as alternate sources for patient data when the primary system is offline, or for data recovery after data loss or corruption [4, 13]. Clustered servers can also be deployed so that routine maintenance can be performed without affecting application availability. Standby clustered or redundant systems can be staged at the same site, or for greater fault tolerance, at a secondary data center. In some cases, it can take time to bring these standby systems fully online and accessible by clinical users [1, 4, 14]. Routine data backups can be performed to disk, or on removable media such as tape so that it can be taken offline to preserve the historical data integrity for recovery at a later time.

In summary downtimes increase surgical times and postoperative length of stay [7], adverse clinical outcomes, clinical errors [15], medication errors [4] and adverse patient outcomes [9]. We also found that downtime can contribute to delays in patient test results [4], diagnosis [10] and treatment [9]. The scoping review study, revealed that several moderating variables were found to influence the outcomes of a disaster on patients and health professionals. They include contingency planning [16], mitigation [17], systems redundancy [14] and backup systems [1]. Their interactions are shown in figure 1 in the model we created after reviewing the literature. Disasters lead to a number of patient safety outcomes, including adverse patient outcomes such as increased surgical time and length of stay as well as clinical errors such as medication errors and adverse patient outcomes. It was also discovered that downtimes led to delays in patient test results, diagnosis and treatment (see Figure 1). Of note, researchers found that downtime planning (i.e. contingency planning, mitigation, systems redundancy and backup systems) moderates the effect of a downtime on patient safety and clinical errors.

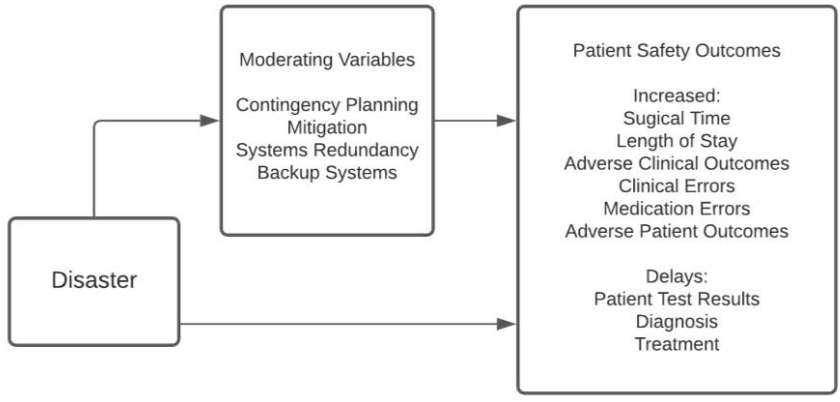


Figure 1. Disaster's effect on patient safety outcomes with moderating variables

5. Conclusion

Downtime planning activities such as contingency planning, mitigation, systems redundancy, and backup systems moderate the effects of a disaster on patient safety outcomes and delays in critical patient care activities. In this scoping review we identified the key findings (or themes) in the literature that need consideration when developing downtime plans and integrating them into healthcare organizational structure. Of note, we created a model that shows the impacts of a downtime and how downtime planning activities can moderate the effects of a downtime on patient safety outcomes and health professional clinical activities. More work needs to be done in the area of studying the effects of downtime planning and preparation and its effect on patient safety outcomes. Future research for our group includes fully testing the models derived from the literature as well as identifying new ways of moderating the effects of downtimes.

Acknowledgements

The authors acknowledge NSERC funding, reference number RGPIN-2020-06950. E. Borycki receives funding from the Michael Smith Foundation for Health Research.

References

[1] Sittig DF, Gonzalez D, Singh H. Contingency planning for electronic health record-based care continuity: a survey of recommended practices. International Journal Of Medical Informatics. 2014;83(11):797-804.
[2] Cleary PA. Lessons learned from unscheduled downtime -- Data inaccessible for up to 10 days. Hospital Access Management. 2016;35(4):40-1.
[3] IPCC. The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Press.: IPCC; 2021.
[4] Larsen E, Fong A, Wernz C, Ratwani RM. Implications of electronic health record downtime: an analysis of patient safety event reports. Journal Of The American Medical Informatics Association: JAMIA. 2017;25(2):187-91.
[5] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology. 2005;8(1):19-32.
[6] Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci. 2010;5(1):69.
[7] Harrison AM, Siwani R, Pickering BW, Herasevich V. Clinical impact of intraoperative electronic health record downtime on surgical patients. Journal Of The American Medical Informatics Association: JAMIA. 2019;26(10):928-33.
[8] Scantlebury A, Sheard L, Fedell C, Wright J. What are the implications for patient safety and experience of a major healthcare IT breakdown? A qualitative study. Digital Health. 2021;7:20552076211010033.
[9] Wang Y, Coiera E, Gallego B, Concha OP, Ong M-S, Tsafnat G, et al. Measuring the effects of computer downtime on hospital pathology processes. Journal Of Biomedical Informatics. 2016;59:308-15.
[10] Larsen E, Haubitz C, Wernz C, Ratwani R, editors. Improving Electronic Health Record Downtime Contingency Plans with Discrete-Event Simulation. 49th Annual Hawaii International Conference on System Science; 2016 2016; Hawaii: IEEE.
[11] Walsh JM, Borycki EM, Kushniruk AW. Effects of Electronic Medical Record Downtime on Patient Safety, Downtime Mitigation, and Downtime Plans. *International Journal of Extreme Automation and Connectivity in Healthcare*. 2020;2(1):161-86.
[12] Fernández MT, Gómez AR, Santojanni AM, Cancio AH, Luna DR, Benítez SE. Electronic Health Record System Contingency Plan Coordination: A Strategy for Continuity of Care Considering Users' Needs. Studies in health technology and informatics. 2015;216:472-6.
[13] Campos F, Luna D, Sittig DF, Bernaldo de Quirós FG. Design, Implementation and Evaluation of an Architecture based on the CDA R2 Document Repository to Provide Support to the Contingency Plan. Studies In Health Technology And Informatics. 2015;216:173-7.

[14] Coffey PS, Postal S, Houston SM, McKeeby JW. Lessons Learned from an Electronic Health Record Downtime. Perspectives in Health Information Management. 2016:1-7.
[15] Fahrenholz CG, Smith LJ, Tucker K, Warner D. Plan B. A practical approach to downtime planning in medical practices. J AHIMA. 2009;80(11):34-8.
[16] Dave K, Boorman RJ, Walker RM. Management of a critical downtime event involving integrated electronic health record. Collegian. 2020;27(5):542-52.
[17] Walsh JM, Borycki EM, Kushniruk AW. Strategies in Electronic Medical Record Downtime Planning: A Scoping Study. Studies In Health Technology And Informatics. 2019;257:449-54.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210952

Inpatient Cost of Stroke Care in Greece: Preliminary Results of the Web-Based "SUN4P" Registry

Olga SISKOU[a,1], Petros GALANIS[a], Olympia KONSTANTAKOPOULOU[a], Iliana KARAGKOUNI[a], Evangelos TSAMPALAS[b], Dafni GAREFOU[b], Helen ALEXOPOULOU[b], Anastasia GAMVROULA[b], Ioannis KALLIONTZAKIS[c], Anastasia FRAGKOULAKI[c], Aspasia KOURIDAKI[c], Argyro TOUNTOPOULOU[d], Ioanna KOUZI[d], Sofia VASSILOPOULOU[d] Efstathios MANIOS[e] Georgios MAVRAGANIS[e], George NTAIOS[f,g], Efstathia KARAGKIOZI[g], Anna Maria LOUKA[g], Christos SAVOPOULOS[h] Gregorios DIMAS[h], Athina MYROU[h], Haralampos MILIONIS[i], Georgios SIOPIS[i], Hara EVAGGELOU[i], Athanasios PROTOGEROU[j], Stamatina SAMARA[j], Asteria KARAPIPERI[j], Nikolaos KAKALETSIS[k], Paris GALLOS[a], Stefanos PAPASTEFANATOS[a], Panayota SOURTZI[a], Kostas VEMMOS[f*], Eleni KOROMPOKI[e,f*] and Daphne KAITELIDOU[a*]

**equal last authors*

[a] *Center for Health Services Management and Evaluation, Department of Nursing National and Kapodistrian University of Athens, Athens, Greece*
[b] *Department of Neurology, Panarkadikon General Hospital, Tripoli, Greece*
[c] *Department of Neurology, General Hospital of Chania, Creta, Greece*
[d] *1st Department of Neurology, Eginition Hospital, National and Kapodistrian University of Athens, Athens, Greece*
[e] *Department of Clinical Therapeutics, Alexandra Hospital, National and Kapodistrian University of Athens, Athens, Greece*
[f] *Hellenic Stroke Organization, Greece*
[g] *Department of Internal Medicine, Faculty of Medicine, School of Health Sciences, University of Thessaly, Larissa, Greece*
[h] *1st Medical Propedeutic Department of Internal Medicine, Aristotle University of Thessaloniki, AHEPA Hospital, Thessaloniki, Greece*
[i] *Department of Internal Medicine, School of Medicine, University of Ioannina, Ioannina, Greece*
[j] *Cardiovascular Prevention & Research Unit in Laiko General Hospital of Athens at the Medical School of the National & Kapodistrian University of Athens, Greece*
[k] *Second Propedeutic Department of Internal Medicine, Aristotle University of Thessaloniki, Hippokrateion General Hospital of Thessaloniki, Greece*

Abstract. The aim of this study was to calculate the average operational cost per sub-type of stroke patient and to investigate cost drivers (e.g. ALoS, NIHSS score, age) correlated to cost. **Methods**: Direct medical costs (diagnostic imaging and clinical laboratory exams, overheads/bed cost, pharmaceuticals, ringers and other

[1] Corresponding author, Olga Siskou; E-mail: olsiskou@nurs.uoa.gr.

non –durables and inpatient rehabilitation) per patient were calculated from the providers' (hospitals') perspective. Resource use data derived from the "SUN4P" web-based registry and unit costs were retrieved from publically available sources and were assigned to resource use. **Results**: The sample comprised 6,282 inpatient days of 750 patients (mean age: 75.5±13.3 years) admitted from July 2019 to July 2021, in nine public hospitals. Mean length of stay was 8.4 ± 7.6 days and mean total operational cost was calculated to €1,239.4 (from which 45% and 35% related to diagnostic exams and overheads/bed cost respectively). Mean cost related to hemorrhagic stroke patients that were discharged alive was calculated significantly higher compared to mean cost related to ischemic stroke patients who didn't undertake thrombolysis and were also discharged alive from the hospital (€2,155.2 vs. €945.2, p<0.001). Linear regression analysis revealed that length of stay was significantly correlated with cost (coefficient beta=232, 95% CI confidence interval = 220 – 243, p<0.001). **Conclusions**: These findings are in accordance with current evidence and should be thoroughly assessed to rationalize inpatient reimbursement rates in order to achieve improved value of care.

Keywords. Stroke, Mean Cost of Hospitalization Inpatient Stroke Care

1. Introduction

Stroke is a major public issue worldwide [1]. Effective management during acute phase (hospitalization) is associated with favorable outcomes, especially for patients suffering from ischemic stroke [2]. Notably, inpatient stroke care accounts for more than 70% of direct stroke health expenses over the cycle of care [3]. To date researchers calculate hospitalization costs by detailed sub-type of stroke [4]. However in Greece, hospitals' charges to National Organization for the Provision of Healthcare (EOPYY) are based on a kind of Diagnostic Related Groups /KEN-DRGs (Gazette 946/27/03/2012) where four generic tariffs (only for operational expenses, wages excluded) were set, without explicit description of diagnosis: *a) N30A=€380 Stroke and other cerebral vessels disorders of a patient who died or was transferred to another hospital in less than 5 days from the admission b) N30X= €900 (Average Length of Stay –ALoS=6) Stroke and other cerebral vessels disorders without catastrophic or severe co-morbidities / complications c) N30Mb (ALoS=9)=€1,625 Stroke and other cerebral vessels disorders with severe co-morbidities /complications d) N30Ma (ALoS=17)= €2,475 Stroke and other cerebral vessels disorders with catastrophic or severe co-morbidities / complications.* However, currently, Greek DRGs are under revision by the Hellenic Institution for DRGs. The aim of this study was to contribute to the efforts of policy makers to rationalize the reimbursement tariffs for inpatient stroke care. Thus, we calculated the average operational cost per sub-group of stroke patients and we investigated cost drivers (e.g. ALoS, NIHSS score, age) correlated to inpatient stroke cost.

2. Methods

Direct medical costs (diagnostic imaging and clinical laboratory exams, overheads/bed cost, pharmaceuticals included ringers and other non –durables and inpatient rehabilitation) per patient were calculated from the providers' (hospitals') perspective. Resource use data per patient derived from the "Improving Stroke Care in Greece in Terms of Management, Costs and Health Outcomes- (SUN4P)" web-based registry

(ClinicalTrials.gov, NCT04109612) [5]. Unit costs, reflecting 2021 prices in euros, were retrieved from publically available sources (Ministry of Health, EOPYY, and National Organization for Drugs) and were assigned to resource use. Descriptive statistics (mean and median) were used in order to present average cost per group of patients and selected patients' characteristics. Student's t-test, ANOVA test, Mann-Whitney test, Kruskal-Wallis test, and Pearson's and Spearman's correlation coefficients were used for bivariate analyses. Finally, multivariate linear regression analysis was also performed with patient cost as the dependent variable and type of stroke, age, baseline NIHSS score and length of stay as independent variables. The SUN4P design was in accordance with the European General Data Protection Regulation (GDPR) and was aligned with the Declaration of Helsinki.

3. Results

The sample comprised 6,282 inpatient days of 750 patients (51.2% males) admitted from July 2019 to July 2021 in nine public hospitals. Only 115 patients were hospitalized in a stroke unit while the majority of patients were admitted either in an internal medicine or in a neurological department. The vast majority of patients (91.5%) underwent the first neuroimaging investigation (CT and/or MRI) during acute phase. In total 310 patients were administered antibiotics (mean cost €350, SD: 2,210) while only 32 patients (30 discharged alive) were administered thrombolysis (mean cost €559.5, SD: 112). Mean cost related to ischemic stroke patients who didn't undertake thrombolysis and were discharged alive was statistically significant lower compared to i) mean cost related to hemorrhagic stroke patients that were also discharged alive ($p<0.001$) and ii) mean cost related to ischemic stroke patients who didn't undertake thrombolysis and discharged died ($p<0.009$) (Table 1). Finally, linear regression analysis revealed that length of stay was significantly correlated with cost (coefficient beta=232, 95% CI confidence interval = 220 – 243, $p<0.001$).

Table 1. Baseline characteristics and average cost per sub group of stroke patients

	Ischemic stroke without Throbolysis-Discharged Alive (N=543)	**Ischemic stroke without Throbolysis-Discharged Died (N=53)**	**Ischemic stroke with Throbolysis Discharged Alive (N=30)**	**Hemorrhagic stroke - Discharged Alive (N=86)**	**Hemorrhagic stroke - Discharged Died (N=36)**	**All Patients (N=750)**
Age*	74.8 (13.4)	85.3 (7.8)	69.2 (12.2)	73.7 (13.7)	81.4 (9.2)	75.5 (13.28)
Baseline NIHSS**	5 (6)	16 (14.8)	10 (7.3)	7 (10.5)	26.5 (13)	7 (8)
Baseline mRS**	0 (1)	1 (4)	0 (0)	0 (1)	0.5 (1.8)	0 (1)
ALoS*	7.16 (5.2)	11.9 (9.8)	9.5 (6.4)	13.4 (13.6)	8.9 (9.7)	8.4 (7.6)
Total* Cost/Patient	**945.2** (652)	**1,957** (1,833)	**1,850.4** (729.1)	**2,155.2** (5,453.2)	**1,897** (3,118)	**1,239.4** (2,150)
Diagnostic Imaging and Biochemical Exams	*504 (301.2)*	*758.3 (543)*	*617 (232)*	*689 (625)*	*633.5 (685)*	*554 (402)*
Bed/ Overheads	*364.5(311)*	*656 (660)*	*585(485)*	*703.6(768.5)*	*478(596)*	*439 (458)*
Pharmaceuticals (included ringers and other non durables)	*70.7(190.6)*	*531.3 (855)*	*636 (157)*	*754.3 (4,353)*	*781(1,978.5)*	*240(1,557)*
Inpatient rehabilitation	*6 (10.5)*	*113 (29.2)*	*12 (17.8)*	*8.2 (13)*	*4.6 (10)*	*6.8(13.3)*

NIHSS-National Institute for Stroke Scale;
mRS- modified Rankin Scale for Neurologic Disability / ALoS- Average Length of Stay in Days * Mean (SD) /** Median (IQR)

4. Discussion

Preliminary results of the SUN4P cohort demonstrate that mean hospitalization cost for the majority of the study patients, those with ischemic stroke who didn't undertake thrombolysis and were alive at the point of discharge, was almost equal to the relevant DRG N30X tariff referring to moderate severity stroke patients (€945 vs €900 /ALoS: 6 vs 7 days). However, as DRGs tariffs are not explicitly assigned with initial diagnosis by type of stroke and health status, misclassification to greater cost DRGs categories may potentially occur. Among inpatient stroke cost components, cost of diagnostic exams is prevalent congruent with previous studies from Greece [6],[7]. The overall mean cost in our study was calculated at less than half of the corresponding cost found in previous studies, as we didn't include wages cost. Those who died in hospital and those suffering from severe hemorrhagic stroke were found to incur remarkably higher total cost as compared to average cost of all patients, a finding that coincides with literature [8], as a result of increased utilization of resources during hospitalization.

5. Conclusions

The costs of inpatient stroke care vary greatly by type of stroke and diagnosis status. These findings demonstrates the necessity for revising the charging tariffs for inpatient stroke care in Greece, taken into consideration the initial explicit diagnosis and the hard outcomes at the point of discharge, and should be incorporated into cost effective strategies for improving value of provided care.

Acknowledgments: This project has received funding from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under grant agreement Nb538.

References

[1] Global Burden of Disease Data Base, 2017.
[2] Sousa D, Martial R, Abilleira RS, et al. Access to and delivery of acute ischemic stroke treatments: A survey of national scientific societies and stroke experts in 44 European countries. European Stroke Journal 2018;1: 13-28.
[3] European Cardiovascular Disease Statistics, 2017.
[4] Wang G, Zhang Z, Ayala C, Dunet D, Fang J.,George M. Costs of Hospitalization for Stroke Patients Aged 18-64 Years in the United States. J Stroke Cerebrovasc Dis. 2014 ; 23(5): 861–868.
[5] Improving Stroke Care in Greece in Terms of Management, Costs and Health Outcomes- SUN4P. Available at: http://www.sun4patients.gr/en/ (09/09/2021).
[6] Kritikou P, Spengos K, Zakopoulos N, et al. Resource utilization and costs for treatment of stroke patients in an acute stroke unit in Greece. Clin Neurol Neurosurg 2016 Mar; 142: 8-14.
[7] Gioldasis G, Taleli P, Chroni E, Daouli J, Papapetropoulos T, Elul J. In hospital direct cost of acute ischemic and hemorrhagic stroke in Greece. Acta Neurol Scand 2008. Oct; 118 (4) 268-274.
[8] Abdo R, Abboud H, Salameh P, et al. Direct Medical Cost of Hospitalization for Acute Stroke in Lebanon: A Prospective Incidence-Based Multicenter Cost-of-Illness Study. INQUIRY: The Journal of Health Care Organization, Provision, and Financing. August, 2018.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210953

Embedding Risk-Based Anonymization into Data Access Control for Providing Individual-Level Health Data in a Secure Way

Murat SARIYAR[a,1] and Jürgen HOLM[a]
[a] *Bern University of Appl. Sciences, Department of Medical Informatics, Switzerland*

Abstract. Especially in biomedical research, individual-level data must be protected due to the sensitivity of the data that is associated with patients. The broad goal of scientific data re-use is to allow many researchers to derive new hypotheses and insights from the data while preserving privacy. Data usage control (DUC) as an attribute-based access mechanism promises to overcome the limitations of traditional access control models achieving that goal. Park and Sandhu provided the usage control (UCON) model as an instance of DUC, which defines policies that evaluate certain attributes. Here, we present an UCON-based architecture, which is augmented with risk-based anonymization as provided by the R package sdcMicro and an extensible Access Control Markup Language (XACML) environment with a core policy decision point as implemented by authzforce.

Keywords. Data Security, Data access control, XACML, Anonymization

1. Introduction

Protecting data in the context of computer-assisted processing of individual-level data requires the usage and/or implementation of security mechanisms. Especially in biomedical research, individual-level data have to be protected due to the sensitivity of the information that is associated with patients, e.g., the propensity to develop breast cancer [1, 2]. Protecting privacy risks requires technical as well as organizational measures, such as the usage of terms & conditions before authorization of data users [3].

One central aim of many scientific data re-use scenarios is allowing many researchers to derive new hypotheses and insights. Such a broad goal requires high flexibility in using as much of the data as possible without compromising data privacy and security. On the one hand, authorization and control of the user activity is often not sufficient for preventing disclosure of sensitive information, as de-anonymization scandals showed [4]. On the other hand, limiting the amount and type of operations on data to ensure high protection decreases the utility of the data. One way to tackle such settings is relying on data-usage-control that considers de-anonymization risks [5].

Data usage control promises to overcome the limitations of traditional access control models. Standard access control protocols regulate the issue of granting access to an

[1] Corresponding Author, Murat Sariyar, Bern University of Applied Sciences, Quellgasse 21, CH2502 Biel/Bienne, Switzerland; E-mail: murat.sariyar@bfh.ch.

object with certain rights (e.g., read only for a certain role, where a role is a named collection of users and relevant permissions). Park and Sandhu [6] provided the usage control (UCON) model for a systematic usage control. Unlike role-based access control, which assigns to certain users pre-defined roles with a set of privileges associated with them, UCON is an attribute-based access control mechanism, which defines policies that evaluate many different attributes. The evaluations are based on three decision factors: authorizations, obligations, and conditions. Authorizations are functional predicates that are evaluated to decide whether a subject is allowed to perform a request on an object (in most cases data). Obligations are requirements a subject must fulfill, and conditions relate to system-related decision factors that are independent of subjects and objects.

There are already works on integrating anonymization procedures into attribute-bases access mechanisms available (see [7]), mainly focusing on the differential privacy (DP) model, which requires that the presence or absence of any individual record must not affect the answer of a query. DP is especially designed for interactive query settings with aggregate outcomes. Even though, there are adaptation to the case of publishing data to allow much more flexibility, the required noise addition is often too high in terms of the resulting data utility [8]. One reason is the difficulty of transforming the output noise addition mechanism into an input-related one for single records, which contradicts the core idea of DP. Hence, k-anonymity models using methods such as shuffling, generalization and micro-aggregation are often more promising in terms of risk-utility trade-offs for dealing with individual-level data in a flexible way [9].

Main motivation for this work is the implementation of a health network platform for pharmacogenetic (PGx) treatments and research. An authorized user of the platform, for example a clinical researcher, should be able to use the platform for SNP-related association analyses. Here, we present the UCON-based architecture and the tools, which is augmented with risk-based anonymization as provided by the R package sdcMicro and an extensible Access Control Markup Language (XACML) environment with a core policy decision point as implemented by authzforce [10].

2. Methods

First, we apply the UCON framework to our pharmacogenetic platform, after which a risk-based anonymization approach is incorporated.

UCON allows to formulate restrictions (rules) on user's access to the data and the operations that are allowed for them. Policies or rules are tuples of the form (subject, action, resource, purpose, system-related condition, user-specific obligation), e.g. (Researcher, machine learning analysis on the data on the client, SNP data, hypothesis generation for research, anytime, ethical statement is signed). The common example for obligations is acceptance of terms of use, being a mandatory requirement that must be met before or during access. Conditions are not directly related to subjects and objects, but to environmental and/or system requirements that must be satisfied, e.g., access is only granted when the system load is under some threshold. Based on the associated attributes of the subject and the resources, the system evaluates these policies for decisions on data access. Hence, defining these attributes is central for the implementation of UCON. For implementing the UCON components, the natural choice is the extensible Access Control Markup Language (XACML).

For anonymizing data, it is crucial to determine the quasi-identifiers (QIDs), which are those attributes that have discriminatory value (i.e., they increase the probability of

re-identification), can be obtained from external resources, and are potentially useful for the data user. Examples of QIDs are gender, age, postal codes, race, ethnicity, etc. These QIDs must be protected from being used for disclosing information of identifiable individuals, which is done by using data perturbation techniques, such as shuffling, generalization and micro-aggregation. From a statistical perspective, a procedure for protecting sensitive data should be based on a disclose scenario, dealing with risks and utility at the same time. Typical thresholds for maximal accepted risks lie between 0.005 and 0.01, while thresholds for utility are use-case dependent. As a general utility measure, usually the entropy measure of the loss of information is used. While definition of QIDs and risk threshold will be made together with the implementation of UCON, risk and utility estimation are part of the policy evaluation during access request.

3. Results

The relevant subject (s) and resource (r) attributes of the privacy aware UCON system are given in Table 1. Subjects are assigned to different roles according to their clearance level, which are associated with certain risk thresholds (the concrete numbers are omitted here), general data utility properties, and the related data perturbation. For example, clinical users can access data with high-risk thresholds and high utility by changing nothing of the QIDs. For researchers, the necessary changes for a risk-utility balance are dependent on the origin of the researcher. Trade-off for data utility means, that an iterative process is allowed, in which the allowed risk threshold can be changed through additional security measures (e.g., data usage is continuously monitored), if the researcher is not satisfied with the anonymization result. For public use, only highly aggregated data is provided without allowing any compromise.

Table 1. Core attributes for the subjects (s) and the resources (r) that are used by our UCON system.

Roles=Purpose (s)	Security Level (r)	Data utility (r)	Data perturbation (r)
Clinical use	High risk threshold	High	Raw or pseudonymized
Research intern	Medium risk threshold	Trade-off	1st level anonymization
Research extern	Low risk threshold	Trade-off	2nd level anonymization
Public use	Very low risk threshold	Low	Highly aggregated

For implementing UCON with risk-based anonymization, an authzforce server was installed that provides a multi-tenant RESTful API to policy administration points (PAP) and policy decision points (PDP). A web service wraps the data access request of a subject, which is sent to the server. The REST request has four parts: subject, resource, action, and an environment. In our use case, the subject is an internal researcher (role), the resources are PGx data sets, actions are they ways to access the data (e.g., reading, or on-site analysis) and via the environment part the purpose, obligation statements as well as the required data utility is specified (see Section 2 on the expected 6-element tuples). Credentials are provided by SAML tokens, which contain user IDs and roles. On the client, the request is processed by an authzforce module that extracts the various attributes to request a decision from the remote authzforce server, which enforces PDP decisions.

The PDP decisions are implemented with respect to the attributes in Table 1. If an internal researcher requires data with maximum allowed information loss in the environment attribute that cannot be achieved by anonymization for his security level, the request is denied with a proposal to adjust his required entropy value. To compute

the best achievable entropy value, the sdcMicro package is run manually before any requests and the results are stored on the PDP as resources for the responses. We used minimal sample uniques for the risk estimation (SUDA) and the following four anonymizing techniques: generalization by recoding, post-randomization, micro-aggregation, and shuffling. The system is just a prototype and not productive yet, due to the lack of certification of the whole PGx platform as a medicinal product.

4. Discussion

Even though, our proposed system seems feasible for the practice, we advise to update the risk model and the risk thresholds in certain time intervals, since risks and the status of anonymity change with time. If none of the allowed risk thresholds is compatible with the required utilities, additional measures such as highly secure analysis environment could be established to allow an increase of the allowed risk threshold. In other words, the whole system cannot be assessed by certain results of performing analysis on the perturbed data, as this only evaluates the anonymization component, not the attribute-based data access system. For assessing the system in practice, it will be important to record the number of users, the number of rejections as well as trade-off rounds, and the satisfaction of the data users with the data quality.

There are many options for concretizing authorizations, obligations, and conditions. We emphasize, that data security is not only a technical issue. Should the platform be open for all kind of analysis? What kind of restrictions seems necessary to guarantee data protection? What kinds of restrictions have small and high impact on the utility of the data? What is the standard use case? Who is the typical user? Based on answers to these and related questions, the system should be adapted to different scenarios. The flexibility of the UCON model is conducive for such adaptations, even though it is not necessary to stick totally to it if the system has a limited scope as in our scenario.

References

[1] Castellani J. Are clinical trial data shared sufficiently today? Yes. BMJ 357 (2013), f1881.
[2] Eichler H-G, Pétavy F, Pignatti F, et al. Access to Patient-Level Trial Data — A Boon to Drug Developers. N Engl J Med 369 (2013), 1577–1579.
[3] Peltier TR. Information Security Policies, Procedures, and Standards: Guidelines for Effective Information Security Management. CRC Press, 2016.
[4] Janmey V, Elkin PL. Re-Identification Risk in HIPAA De-Identified Datasets: The MVA Attack. AMIA Annu Symp Proc 2018 (2018), 1329–1337.
[5] Dimitrakos T, Martinelli F, Pretschner A, et al. Proceedings of the 4th International Workshop on Security and Trust Management (STM 2008): Policy Evolution in Distributed Usage Control. Electron Notes Theor Comput Sci 244 (2009), 109–123.
[6] Park J, Sandhu R. Towards Usage Control Models: Beyond Traditional Access Control. In: Proceedings of the Seventh ACM Symposium on Access Control Models and Technologies. New York, NY, USA: ACM, pp. 57–64.
[7] Metoui N, Bezzi M. Differential Privacy Based Access Control. In: Debruyne C, Panetto H, Meersman R, et al. (eds) On the Move to Meaningful Internet Systems: OTM 2016 Conferences. Cham: Springer International Publishing, 2016, pp. 962–974.
[8] Domingo-Ferrer J, Sánchez D, Blanco-Justicia A. The limits of differential privacy (and its misuse in data release and machine learning). Commun ACM 64 (2021), 33–35.
[9] Templ M, Kowarik A, Meindl B. Statistical Disclosure Control for Micro-Data Using the R Package sdcMicro. J Stat Softw 67 (2015), 1–36.
[10] AuthzForce (Community Edition) - XWiki, https://authzforce.ow2.org/ (accessed 23 August 2021).

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210954

A Safety Maturity Model for Technology-Induced Errors

Elizabeth M. BORYCKI[a,b,1] and Andre W. KUSHNIRUK[a]
[a] *School of Health Information Science, University of Victoria, Canada*
[b] *Michael Smith Foundation for Health Research, Canada*

Abstract. In this paper we describe a review of the literature and the development of an initial maturity model for the research literature in the area of technology-induced errors and health technology safety. The capability maturity model provides a way forward for organizations to formalize their health technology safety processes and also compare their efforts in this area to other organizations to support organizational learning. The application of maturity models to technology-induced error is described.

Keywords. Safety, health technology safety, technology-induced errors, maturity models, risk management

1. Introduction

Maturity models (MM) help organizations to assess their capabilities or competencies in a domain area. Organizational units or teams use maturity models to determine what capabilities their organizations need to acquire in order to reach the next level of organizational performance in a given business context [1]. Most MMs rate performance using levels (i.e. level 1 through to 5 in maturity) [2]. MMs are used to improve electronic medical record (EMR) adoption, implementation and agility in undertaking these endeavors [2]. These adoption and implementation models are critical to organizations as they help leadership teams to assess their organization's current state, compare the organization's performance to other similar organizations, chart the organization's progression and accomplishments through the model and identify what needs to be implemented for the organization to reach the next level of maturity [2]. MMs have advanced in varying organizational domains such as digital health. In a recent review [3], the focus of these models has been on technology adoption, implementation, interoperability, analytics, and usability in the domain area of health information systems. To date, few of these models have exclusively focused on health technology safety, in particular, no MMs have focused on implementing protections against technology-induced errors. In this paper, the authors describe an initial MM developed with a primary focus on addressing health technology safety.

Health technology safety has been recognized as important public health issue that has led to the study of technology-induced errors. This emerging area of health informatics research aims to address those sources of error that arise from: "a) the design

[1] Corresponding Author, Elizabeth Borycki, School of Health Information Science, University of Victoria, Canada; E-mail: emb@uvic.ca.

and development of technology, b) the implementation and customization of a technology, and c) the interactions between the operation of a technology and the new work processes that arise from a technology's use" [4]. Much of the research in this area has focused on identifying incidents where a technology-induced error had occurred. Here, researchers have focused on methods that can be used throughout the software development lifecycle and after implementation to identify and then prevent future errors. Such research is critical to preventing death, harm and injury and the significant costs to the individual and the healthcare system associated with such events [4,5]. Over the past few years, there have been several attempts to integrate newly developed approaches to preventing such errors throughout the health technology industry in vendor as well as healthcare settings with some success [6,7]. Yet, there remains a patchwork approach to integrating innovative approaches to improving health technology safety.

Maturity models offer an opportunity to aid healthcare organizations in formalizing activities and processes aimed at preventing, mitigating and monitoring for technology-induced errors. Such models are used in the general organizational literature to help organizations develop and formalize internal processes and they are used across organizations to allow for comparison, benchmarking and learning between organizations to enhance systems across an industry. Some researchers and professional organization have developed MMs to support internal organization improvements and industry wide comparisons across organizations [1,2]. One of the most commonly used MMs is the HIMSS electronic medical record adoption model which helps organizations to identify which aspects of the electronic medical record to implement and support their adoption [2,3]. This model supports inter-organizational learning as organizations can look at where they fit in the model in terms of the implementation and adoption process. Such models have also been developed for decision support systems and personal health record implementation. None of the models exclusively focus on technology-induced errors or improving technology safety within and across organizations.

2. Method

The following method was used to develop an initial MM described in this paper. The research was conducted in three phases described below.

2.1 Phase 1

In phase one we reviewed the literature on information system MMs in healthcare to identify MMs specific to strategies for addressing "technology-induced errors" in healthcare organizations. The review began with two researchers identifying keywords and conducting searches of PubMed (see Table 1).

Table 1. Model Search Terms

Keyword Terms
"maturity model" AND "system"
"maturity model" AND "safety"
"maturity model" AND "technology"
"maturity model" AND "error"
"maturity model" AND "error" AND "system"
"maturity model" AND "error" AND "technology
"maturity model" AND "error" AND "safety"

The titles and abstracts of the identified articles were then reviewed by the researchers for their possible inclusion in: (1) a MM focused on technology-induced errors or health technology safety, and (2) help with identifying gaps in the literature relevant to development of MMs focused on technology-induced errors.

2.2 Phase 2

In Phase 2 the results of the literature search were reviewed by the research team and gaps were identified. Preliminary levels for an initial MM were developed. Then the research team met to discuss the levels and disagreements were resolved through review of the articles and discussions.

2.3 Phase 3

An initial MM was created based on the results of phase 1 and 2. This model was developed specifically in the context of technology-induced errors.

3. Results

The review of the literature conducted in Phase 1 revealed that 69 articles were returned (see Table 2). Following a title, abstract and full text review, the researchers found that the MMs focused on either system implementation maturity (e.g. electronic health record features and function implementation) or on specific process implementation such as health information system adoption. As Table 2 shows there were very few articles returned for MMs in relation to technology-induced errors. We then focused our review on the 15 articles at the intersection of "maturity model" and "safety" to develop a new initial MM that incorporated some of the levels described from these articles in the new context of technology-induced errors. The layers in the newly proposed MM for technology-induced error were defined and refined through further discussion by the research team.

Table 2. Maturity Model Search Results

Keyword Terms	Resulting Number of Papers Returned
"maturity model" AND "system"	29
"maturity model" AND "safety"	15
"maturity model" AND "technology"	24
"maturity model" AND "error"	1
"maturity model" AND "error" AND "system"	0
"maturity model" AND "error" AND "technology	0
"maturity model" AND "error" AND "safety"	0
Total Number of Papers Returned	69

In our first discussions, we identified that MMs from the health informatics and health information technology literature. We found that they typically consist of five or more different levels. The levels draw on an organization's ability to address a particular problem, event or issue. In Level 1 organizations employ ad hoc responses. In Level 2 organizations employ repeatable activities in response to problems, events or issues. In Level 3 organizations fully define and respond to problems or issues. In Level 4

organizations create formal structures to manage these processes and in Level 5 organizations have formal managed processes to respond to problems, issues and events, continually monitor and focus on optimization and improved safety (see Figure 1).

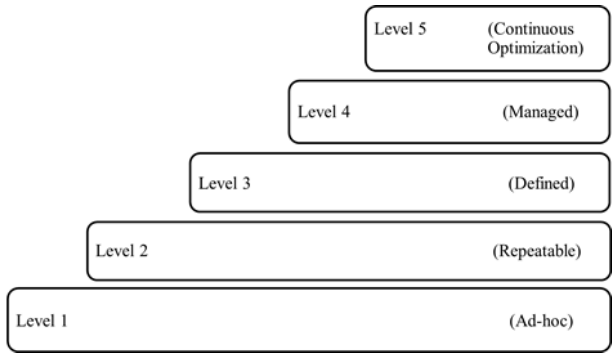


Figure 1. Generic Maturity Model Levels.

In Figure 2, we map the findings from our literature review and discussions on to generic layers of a capability MM (described in Figure 1), specifically in the context of technology-induced errors to arrive at a new model in the context of technology-induced error (TIE).

Level	Description	Practice Area	Key Safety Practices	Desired Outcomes
1	Ad-hoc	• Planning for Requirements Gathering	• Interviews • Focus Groups	
2	Repeatable	• Requirements Gathering • Project Planning • Project Tracking an Oversight • Contract (and Subcontract) Management • Quality Assurance • Configuration Management	• Interviews • Focus Groups • Team meetings and Review	• Requirements are free of TIE.
3	Defined	• Organizational Process Definition • Organizational Process Description • Peer Reviews • Software Management • Intergroup Coordination • Training	• Systems Modelling • White Box Testing • Black Box Testing • Usability Testing • Workflow Testing • Clinical Simulation Testing	• Models that are free of TIE. • Programming code that is free of TIE. • User interface features and functions that are free of TIEs. • Workflows that are free of TIEs. • Technology savvy and resilient users.
4	Managed	• Process Management • Quality Management	• Control Charts • Run Charts	• Software that is free of TIE. • Services that are free of TIE.
5	Optimization	• Learning from Technology-induced Errors • Technology Change Management • Process Change Management	• Qualitative Coding of Individual Incident Reports for Technology-induced Errors • Root Cause Analyses • Usability Testing • Simulation Testing • Data Analytics of Technology-induced Errors	• Ensuring ongoing continuous improvement (CI) to address the root causes of defects that are inherent in processes, tools, and designs and that have a significant impact.

Figure 2. Maturity Model for Technology-induced Error

In the context of the research literature we focused on methodologies as applied to technology-induced errors. Level 1 represents organizations where the practices and activities surrounding a technology are ad-hoc or reactive to events in an organization. In Level 2 we identified that organizations identified activities associated with managing and repeated activities for events related to technology-induced error. In Level 3 safety activities focused on technology-induced errors were defined and enacted when errors occurred. In Level 4 activities were not only defined and enacted specific to technology-induced errors, but they are part of organizational processes and technology-induced errors were actively managed. Lastly, Level 5 represents organizations where there is a formalized program of activities aimed at identifying, diagnosing, addressing, mitigating and monitoring for technology-induced errors with a focus on improving safety through optimization (see Figure 1). Figure 2 also shows how the five generic MM levels can be mapped to: (1) practice areas, (2) key safety practices and (3) desired outcomes.

4. Conclusion

Maturity models help organizations to assess their capabilities or competencies in a given area. They are used extensively in the general organizational literature. In the health informatics research literature, there have emerged several MMs that have helped healthcare organizations and systems with specific health information technology activities such as the adoption of electronic health records, personal health records and decision support systems. None of this research has focused specifically on technology-induced errors. In the absence of a MM that allows for and supports the development of within and between organizational processes and activities, we have developed an initial MM that focuses on processes, methods and monitoring approaches for technology-induced errors. The MM helps to diagnose, mitigate and monitor for these safety issues with the goal to improve the safety of technology systems and overall patient care. Our ongoing work is focusing on the validation of the initial model.

References

[1] Carvalho JV, Rocha Á, Abreu A. Maturity models of healthcare information systems and technologies: A literature review. J Med Syst. 2016 Jun;40(6):131.
[2] Carvalho JV, Rocha Á, Abreu A. Maturity assessment methodology for HISMM: Hospital information system maturity model. J Med Syst. 2019 Jan 7;43(2):35.
[3] Kazley AS, Ozcan YA. Organizational and environmental determinants of hospital EMR adoption: A national study. J Med Syst. 2007 Oct;31(5):375-84.
[4] Borycki EM, Kushniruk AW, Bellwood P, Brender J. Technology-induced errors. The current use of frameworks and models from the biomedical and life sciences literatures. Methods Inf Med. 2012;51(2):95-103.
[5] Borycki E, Keay E. Methods to assess the safety of health information systems. Healthc Q. 2010;13 Spec No:47-52.
[6] Borycki E, Dexheimer JW, Hullin Lucay Cossio C, Gong Y, Jensen S, Kaipio J, Kennebeck S, Kirkendall E, Kushniruk AW, Kuziemsky C, Marcilly R, Röhrig R, Saranto K, Senathirajah Y, Weber J, Takeda H. Methods for addressing technology-induced errors: The current state. Yearb Med Inform. 2016 Nov 10;(1):30-40.
[7] Borycki E, Kushniruk A, Nohr C, Takeda H, Kuwata S, Carvalho C, Bainbridge M, Kannry J. Usability methods for ensuring health information technology safety: Evidence-based approaches contribution of the IMIA working group health informatics for patient safety. Yearb Med Inform. 2013;22(1):20-7.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210955

The Wellness Center in Public Hospitals in Thailand: A Qualitative Assessment Using RE-AIM Model

Utoomporn WONGSIN[a,1], Arnat WANNASRI[b,c], Kwanpracha CHIANGCHAISAKULTHAI[d] and Tuo-Yu CHEN[a,e]
[a] *PhD Program in Global Health and Health Security, College of Public Health, Taipei Medical University, Taiwan*
[b] *Nopparat Rajathanee Hospital, Department of Medical Services, Ministry of Public Health, Thailand*
[c] *National Yang Ming Chiao Tung University, Taiwan*
[d] *International Health Policy Program (IHPP), Thailand*
[e] *Master Program in Global Health and Development, College of Public Health, Taipei Medical University, Taiwan*

Abstract. Non-communicable diseases (NCDs) become one of serious health issues globally including Thailand. The Wellness Center was introduced by the Ministry of Public Health (MOPH) in public hospitals since 2018 for both healthy people and people who at risk for NCDs. It incorporates risk evaluation, health assessment, health behaviors changing, health promotion and prevention, and health record system to follow participant's health status. This study aimed to evaluate the Wellness Centers in three public hospitals in Thailand using RE-AIM framework. Findings indicated that the Wellness Centers successful reached its target groups. Although sample size was small, the findings provided a positive outcome. Besides, the average weight and Body Mass Index (BMI) had reduced significantly after participants attending both services of Wellness Centers. However, there were some challenges from this program such as financial sustainability, performance indicator, and data system.

Keywords. Wellness Center, Qualitative assessment, NCDs, Prevention and promotion program, RE-AIM model

1. Introduction

Non-communicable diseases (NCDs) are serious health issues in global level. In Thailand, of 75% of all death in 2009 attributed to NCDs. Stroke, diabetes and ischemic heart disease were top three of NCDs [1]. The economics loss from NCDs was also high in Thailand. It was estimated that the costs of NCDs were approximately USD 6,158 million in 2013 [2]. In 2011, the Thai cabinet introduced the Thailand Healthy Lifestyle Strategic Plan for 2011-2020 [3] to encourage the ministries to collaborate in controlling NCDs. The Ministry of Public Health (MOPH) – as a main organization of this plan –

[1] Corresponding Author, Utoomporn Wongsin, PhD Program in Global Health Department, College of Public Health, Taipei Medical University, Taiwan; Email: d537108007@tmu.edu.tw.

introduced NCDs and health risks with focuses on promotion and prevention program to improve health issues. Four main diseases included in this plan are cardiovascular disease, diabetes, chronic obstructive pulmonary disease, and cancer. Risk factors included in this plan are smoking, drinking, unhealthy diet, and lack of physical exercise. Several kinds of promotion and prevention programs were introduced and implemented in public hospitals by MOHP such as the NCDs Clinic Plus 2017. The Wellness Center was also introduced by the MOPH in public hospitals in 2018 for both healthy people and people who at risk of NCDs. A pilot of Wellness Centers implemented in 16 public hospitals in the same year. The Wellness Center incorporates risk evaluation, health assessment, health behaviors modifications, health promotion and prevention, and health record system to follow participant's health status. However, only Wellness Centers from three public hospitals had available data for us to do the evaluation. This study used RE-AIM model to evaluate Wellness Center program of the three public hospitals.

2. Methods

We used the Reach, Effectiveness, Adoption, Implementation and Maintenance (RE-AIM) framework to evaluate program. This study conducted in three public hospitals in Thailand selected by purposive sampling. Data were retrospectively collected from hospitals' database in fiscal year 2019. Inclusion criteria were hospitals have prevention and promotion services, and an adequate data to evaluate the program. The Wellness center had two services including services and fitness service. There were three service process in this program, including assessment, empowerment, and coaching. Each participant received services once a month for six months. Besides, individual fitness services were planned by a sport science.

3. Results

Two hundred and thirty-five participants from three public hospitals in 2019 were analyzed. Their ages were from 14 to 67 years old, with an average age of 43 years old. Most participants were female (69.1%). Participants were general people (NCDs patients and walk-in). Over 50% of the participants had Body Mass Index (BMI) at least 25. And 57 participants had Hypertension, Diabetes and Dyslipidemia. A total of 132 participants attended NCD programs for more than two months

Overall, the results showed that weight and BMI of participants changed significantly. Considering in each hospital, participants from hospital B had the highest weight changed (5.77) after attending a program followed by hospital C (3.97) and A (1.89), respectively. For BMI, participants from hospital B had the highest BMI changed (2.4) after attending a program followed by hospital C (1.63). However, there was not BMI changed from participants from hospital A.

Table 1. The Wellness Centers evaluation using RE-AIM model

RE-AIM Framework	Wellness Center	Notes
Reach (Individual)	There were 4 target groups. 1. People at risk of NCDs who has annual health checks from hospital.	1.Target groups covered healthy people and people who are at risk of NCDs following The Wellness Center policy.

	2.People who willing to attend program. 3.Patients from NCDs clinics 4. Hospital staff.	2.However, sample sizes were low and they were not good representative sample but the result showed a positive outcome.
Efficacy (Individual)	An average of weight and Body Mass Index (BMI) had reduced significantly after participants attending Wellness center program.	Motivation process plays a crucial role in behavior changing.
Adoption (Setting level)	1.A pilot study in 16 public hospitals during year 2018 but there were only 3 public hospitals had enough data to evaluate the program. 2.A total 253 participants from 3 public hospitals attended the program (using data 2019)	One Wellness Center provided services to hospitals employees only.
Implementation (setting/staff level)	Only 132 from 253 participants attended programs for more than two months. (52%)	48% of total participants attended program less than two months. 2.They were unwillingness to attend the program.
Maintenance (Individual and Setting level)	1. This program still continues providing services in hospitals and expanding to other public hospitals. 2. It is included in P&P package which has annually budget from three main insurance schemes in Thailand. 3. It is collaboration between Department of Disease Control, Department of NCDs and Department of Health, MOPH.	1.Financial sustainability (many projects include in P&P package). 2.Program (service and fitness).

4. Discussion and Conclusion

The results show that an average of weight and Body Mass Index (BMI) had reduced significantly after participants attending Wellness Center both service and fitness program. This is in line with previous studies indicated that lifestyle modification intervention, weight reduction can be reduced the incidence of people at high risk living with diabetes [5], [6] improving NCDs risk factors [7],[8], [9] and it also increase quality of life [10].

The Wellness Centers from Three public hospitals in Thailand successful reached its target groups needed both healthy people and people who at risk for non-communicable diseases (NCDs) which contributes in strengthening public health.

However, there are some challenges from this program

1. Data system should be managed to provide adequate data, we found that some information was not completed. Hospital uses paper-based to record participant's information.
2. The Wellness Center program was based on hospital financing, and some hospitals collect a small money from participants. Therefore, long term financing should be prepared by hospital and governments to ensure financial sustainable.

3. There was a difference in services process between three hospitals which may impact to the results. Diet Coaching, for example, one hospital taught participants by conducting workshops and food model. Also, they gave advice to participants via line application. Whereas, two hospital provide knowledge by using food model. Thus, a standard guideline should be preparing to ensure that program is more effective.
4. There was not an evaluation process/ a standard performance indicators for the wellness center program which are necessary improving program. To ensure that this program is much more effective to target groups, a set of indicators and evaluation process should be regulated.

Further studies are recommended to involve more Wellness centers providing services at hospitals to improve generalizability of our findings.

Acknowledgments

The authors would like to acknowledge hospital directors and staff from three public hospitals, National Health Foundation, International Health Program, Thailand, and College of Public health, Taipei Medical University, Taiwan.

References

[1] International health program,Thailand. Burden of Disease and injury of the Thai population in Thailand 2014 [Internet]. [cited 2018 September 25]. Available from: http://bodthai.net/en/download/burden-of-diseases-and-injuries-of-thai-population-in-2014-english-report-151118-np_final/
[2] NCDs - Thaihealth. Thai Health Promotion Foundation. [Internet]. [cited 2018 September 25]. Available from: https://www.thaihealth.or.th/microsite/categories/5/ncds/2/173/176-+NCDs.html. (In Thai)
[3] Ministry of public health. Thailand healthy lifestyle strategic plan B.E 2011-2020 [Internet]. [cited 2018 September 25]. Available from: http://wops.moph.go.th/ops/oic/data/20110316100703_1_.pdf(In Thai)
[4] Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, et al. RE-AIM Planning and Evaluation Framework: Adapting to New Science and Practice with a 20-Year Review. Front Public Health. 2019 Mar 29; 7:64.
[5] Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med. 2002 Feb 7;346(6):393–403.
[6] Hamman RF, Wing RR, Edelstein SL, Lachin JM, Bray GA, Delahanty L, et al. Effect of weight loss with lifestyle intervention on risk of diabetes. Diabetes Care. 2006 Sep;29(9):2102–7.
[7] Wing RR, Lang W, Wadden TA, Safford M, Knowler WC, Bertoni AG, et al. Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. Diabetes Care. 2011 Jul;34(7):1481–6.
[8] Salas-Salvadó J, Díaz-López A, Ruiz-Canela M, Basora J, Fitó M, Corella D, et al. Effect of a Lifestyle Intervention Program with Energy-Restricted Mediterranean Diet and Exercise on Weight Loss and Cardiovascular Risk Factors: One-Year Results of the PREDIMED-Plus Trial. Diabetes Care. 2019 May 1;42(5):777–88.
[9] Gregg EW, Lin J, Bardenheier B, Chen H, Rejeski WJ, Zhuo X, et al. Impact of Intensive Lifestyle Intervention on Disability-Free Life Expectancy: The Look AHEAD Study. Diabetes Care. 2018 May;41(5):1040–8..
[10] Langan A, Bader A, Goedkoop S, Cummings S, Tsikitas M, Nogueira I, et al. A longitudinal study to investigate the effects of a 12-week comprehensive lifestyle weight management program on body weight and quality of life. Clin Nutr ESPEN. 2020 Dec; 40:125–32.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210956

The Contribution of Informatics to Overcoming the Covid-19 Fake News Outbreak by Learning to Navigate the Infodemic

Marianna ISAAKIDOU[a,1] and Marianna DIOMIDOUS[a]
[a] *Faculty of Nursing, National and Kapodistrian University of Athens, Athens, Greece*

Abstract. Daily, people are being exposed to an enormous amount of Covid-19 information, but not all of it is reliable. This phenomenon seriously affects the public health policy effectiveness, because there is a lot of misleading or inaccurate information, which is spreading rapidly and makes it more difficult to restrict the pandemic. Healthcare informatics has emerged as a diverse and important new field. Healthcare informatics applications are becoming more and more popular and are providing easy access to new sources of knowledge. This way, the quality of patient care will improve and productivity will increase. However, people should also learn how to navigate this infodemic properly.

Keywords. Healthcare Informatics, fake news, misinformation, Covid-19

1. Introduction

The World Health Organisation (WHO) has declared that false allegations are spreading faster than the virus itself and has already defined the situation as a "global information pandemic" [1]. The word "Infodemic" is consisted of the words "information" and "epidemic" or "pandemic", which means "epidemic/pandemic of information" [2]. Due to Covid-19, another word has been added to people's vocabulary daily, called: "epidemiology", which is the branch of medical science that deals with how diseases are transmitted and can be controlled in populations [1]. So, during the current pandemic, a new word has been aroused, called: "infodemiology". The term "Infodemiology" is defined as "the science of distribution and determinants of information in an electronic medium, specifically the Internet or in a population, with the ultimate aim to inform public health and public policy" [3].

The great amount and multimodality of data produced in health, give a key role of data analytics methods as a speciality of health informatics. As a result, an increased interest appeared in development of Machine Learning. Furthermore, techniques with their foundations in artificial neural networks, like Deep Learning, have emerged in recent years as a powerful tool and promise to reshape the future of Artificial

[1] Corresponding Author, Marianna Isaakidou, PhD cand. Faculty of Nursing, School of Health Sciences, National and Kapodistrian University of Athens, Athens, Greece; E-mail: marianis@nurs.uoa.gr.

Intelligence. Newly appeared online platforms are already operating to reduce the effects of fake news [4].

2. Methods

A systematic review of academic literature was conducted, aiming to identify strategies in tackling online misinformation and fake news about public health and especially the Covid-19 pandemic. Other articles were reviewed, involving vaccine misinformation, as well as online and open provided information by health-related authorities, regarding the phenomenon of fake news in the healthcare sector. A systematic review of academic literature was performed using mainly PubMed and ScienceDirect databases in July and August 2021 using terms like "infodemic", "healthcare informatics", "fake news", "Covid-19". In total, 58 academic publications identified and extracted for further assessment. Other than the title, abstract and full-text assessment, 16 articles were selected for inclusion in this literature review. In addition to the systematic review of academic literature, research has also been contacted on National Websites dealing with Public Health and an overview of where people can find accurate and proper information on health issues is presented. This article also attempts to report the sources of Covid-19 misinformation and to find ways to navigate with safety the various websites dealing with information on the pandemic.

3. Results

The Covid-19 infodemic is full of false claims, half-backed conspiracy theories and pseudoscientific therapies, regarding the diagnosis, treatment, prevention, origin, and spread of the virus. Reliable and authorised "places" to search for accurate worldwide data about Covid-19 infodemic and all useful information needed, are the websites of every national Ministry of Health and any other national public health authorities, as well as other global organisations, such as the World Health Organisation (WHO), the European Centre for Disease Prevention and Control (ECDC) website, and the Centre for Disease Prevention and Control (CDC) website. It is worth mentioning that as we learn more about the virus, information may change or vary [5, 6].

Worldwide data sources for Covid-19 and healthcare issues related to it (e.g. Covid-19 vaccines, behaviours against them, etc) are already starting to increase and healthcare informatics applications are becoming a very useful tool of defence [7]. Previous works proved that "Machine Learning" application seems very useful in recognising and striking out fake news and other disinformation [8, 9, 10].

The WHO has developed guidelines to help individuals, community leaders, governments and the private sector understand some key actions they can take to manage the Covid-19 infodemic. A very important intervention by the WHO is working closely with more than 50 digital companies and social media platforms, including the most popular ones such as Facebook, Twitter, Viber, and others, to ensure that science-based health messages from the organisation itself or other official sources appear first when people search for health information. The WHO has also collaborated with the Government of the United Kingdom on a digital campaign to raise awareness on misinformation concerning the coronavirus and to encourage individuals to report false or misleading online content [1]. Another awareness campaign by the WHO

collaborated with the UK is called: "Stop The Spread", which clearly urged people to double-check all information. An additional effort to limit the infodemic was the campaign called "Reporting Misinformation", which urged people to not only verify information, but also report misinformation in various social media platforms. The WHO, also, created an innovative online game called "Go Viral!". Research has shown that even one game can reduce perceived reliability of fake news by an average of 21% [11]. Lastly, the WHO has created the "EPI-WIN", its own Information Network for Epidemics that seeks to give everyone access to accurate and valid advice or information from trusted sources on public health emergencies [12].

The ECDC website is also ensuring a maximum level of transparency. It screens up to 500 sources to collect Covid-19 figures from 196 countries. It screens official social media accounts (e.g. Facebook) from national authorities. Social media sources are screened to gather additional information, which can be validated by the official sources previously mentioned. This process is being refined constantly, as to ensure the validity and reliability of the data and it is very useful for monitoring and interpreting the dynamics of the Covid-19 pandemic not only in the EU, but also worldwide [13].

The CDC website is also exploring and monitoring similar activities to support the global effort against SARS-CoV-2. The "COVID Data Tracker" finds maps and tracks cases, deaths, and trends of COVID-19 in the United States and it is being updated daily [14]. Additionally, the CDC Health Alert Network (HAN) updates a huge number of guidelines and recommendations on how to evaluate and identify patients who should be tested for Covid-19. The CDC has, also, updated travel guidance to closely monitor and respond to the Covid-19 outbreak [15].

4. Discussion

People don't usually give it a second thought before sharing information on social media and forget to verify if it is fake or true. It is extremely useful to teach people how to identify misinformation and how to navigate this infodemic. First of all, they have to assess the source and find out who shared the information and where they got it from. Attractive and catchy headlines are being used to mislead people intentionally. Therefore, people should read the entire article carefully and not just the headline. In addition, they should identify the author's name to ensure they are real and credible. It would be extremely helpful if people were to check the date of the article and any other biases. Examination of the supporting evidence with fact-checkers to debunk misinformation could also be enlightening [1,8]. There are different ways to identify fake news and stay accurately informed. People can verify fake news either by traditional verification methods (e.g. getting on to the primary source and verifying it) or online verification tools (e.g. Machine Learning).

The pandemic has been the breeding ground for the germination of misinformation, conspiracy theories and, in general, a climate of suspicion and uncertainty in the face of any message about the situation. Through today's different communication platforms, information flows unfiltered, generated by any individual without credentials to report on specialised topics, giving way to the invention of myths and stories that easily go viral, activating waves of opinion and diffuse beliefs that deepen the climate of misinformation and mistrust [6]. Scientists, health professionals, as well as journalists, should take serious steps to help the public identify and recognise fake news.

5. Conclusions

The fake news crisis is putting the public health at risk as it is taking over social media. If we want to win the war against fake news and misinformation, there is an urgent need to train people on the nature of social media and how to use them effectively. However, personal responsibility is the first and fundamental step to protecting our society from the dangerous phenomenon of fake news. Future steps to combat fake news could be to further apply machine-based classification methods on different types of "fake news". The goal is to develop algorithms that make predictions about future instances. For example, an algorithm could be trained with pre-established indicators or features from a training set to predict its corresponding category for unseen instances by a set of learned rules (e.g. Decision Trees) [16]. A new era of AI is inevitable.

References

[1] WHO (2021), Let's flatten the infodemic curve. Accessed in 12 July 2021. Available from: https://www.who.int/news-room/spotlight/let-s-flatten-the-infodemic-curve.
[2] Isaakidou M, Zoulias E, Diomidous M. Investigating the Impact of Misinformation Sources on Health Issues: Implications for Public Health. Stud Health Technol Inform. 2021 May 27;281:494-495. doi: 10.3233/SHTI210211.
[3] Eysenbach G. Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet. J Med Internet Res 2009;11(1): e11. doi: 10.2196/jmir.1157.
[4] Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ. Deep Learning for Health Informatics. IEEE J Biomed Health Inform. 2017 Jan;21(1):4-21.
[5] European Centre for Disease Prevention and Control (ECDC). Countering online vaccine misinformation in the EU/EEA. Stockholm: ECDC; 2021, ISBN: 978-92-9498-542-2 doi: 10.2900/329304. Accessed in 23 July 2021. https://www.ecdc.europa.eu/sites/default/files/documents/Countering-online-vaccine-misinformation-in-the-EU-EEA.pdf.
[6] Bin Naeem S, Bhatti R, and Khan A. An exploration of how fake news is taking over social media and putting public health at risk. Health Info Libr J. 2021 Jun;38(2):143-149. doi: 10.1111/hir.12320.
[7] Darvish A, Bahramnezhad F, Keyhanian S, Navidhamidi M. The Role of Nursing Informatics on Promoting Quality of Health Care and the Need for Appropriate Education. Glob J Health Sci. 2014 Jun 25; 6(6):11-8. doi: 10.5539/gjhs.v6n6p11.
[8] Isaakidou M, Zoulias E, Diomidous M. Machine Learning to Identify Fake News for COVID-19. Stud Health Technol Inform. 2021 May 27;281:108-112. doi: 10.3233/SHTI210130.
[9] Raju V, Mohd J, Ibrahim HK, Abid H. Artificial Intelligence (AI) applications for COVID-19 pandemic, Diabetes & Metabolic Syndrome. Clinical Research & Reviews. 2020;14(4):337-339.
[10] Hunt K, Agarwal P, Zhuang J. Monitoring Misinformation on Twitter During Crisis Events: A Machine Learning Approach. Risk Anal. 2020 Nov 14. doi: 10.1111/risa.13634.
[11] WHO (2021), Fighting misinformation in the time of COVID-19, one click at a time, accessed in 12 Aug 2021, https://www.who.int/news-room/feature-stories/detail/fighting-misinformation-in-the-time-of-covid-19-one-click-at-a-time.
[12] WHO (2021), Risk communication-EPI-WIN: WHO Information Network for Epidemics, accessed in 8 Aug 2021, https://www.who.int/teams/risk-communication.
[13] European Centre for Disease Prevention and Control (2021), How ECDC collects and processes COVID-19 data L, accessed in 24 Aug 2021, https://www.ecdc.europa.eu/en/covid-19/data-collection.
[14] Taveras J, León A, Tejeda F, Infodemic and disinformation of the COVID-19: A reflect of the systemic crisis in the post-truth era, accessed in 29 August 2021, citibeats.com/infodemic-and-disinformation-of-the-covid-19-a-reflect-of-the-systemic-crisis-in-the-post-truth-era/.
[15] Centre for Disease Prevention and Control (CDC). Updated Guidance on Evaluating and Testing Persons for Coronavirus Disease 2019 (COVID-19), accessed in 30 Aug 2021, https://emergency.cdc.gov/han/2020/han00429.asp.
[16] Maria D. Molina, S. Shyam Sundar, Thai Le, Dongwon Lee. "Fake News" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. American Behavioral Scientist 2021;65(2):180–212. https://doi.org/10.1177/0002764219878224.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210957

A GDPR-Compliant Partner Notification Service

Christine KAKALOU[a1], Eleftheria POLYCHRONIDOU[b], Vicky DROSOU[b], Vlasios K. DIMITRIADIS[a], Thomas DERMARIS[a], Rafael KORDONIAS[b], Aris PAPAPRODROMOU[b], Triantafillos TSIRELIS[b], Christos MARAMIS[a], Konstantinos VOTIS[b], Dimitrios TZOVARAS[b], Domenico SAVARINO[c], Manuel MAFFEO[d], Nedim JASIC[e], Tatjana NEMETH-BLAŽIĆ[f], Zoran DOMINKOVIĆ[g], Dubravko POGLEDIĆ[g], Iva JOVOVIC[h], Agne SIMKUNAITE-ZAZECKE[i], Loreta STONIENE[j], Antonella SAMMUT[k], Lella COSMARO[c] and Pantelis NATSIAVAS[a]

[a] *Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece*
[b] *Information Technologies Institute, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece*
[c] *Fondazione LILA Milano - Italian League for Fighting AIDS, Milan, Italy*
[d] *Arcigay, Bologna, Italy*
[e] *Croce Rossa Italiana, Rome, Italy*
[f] *CIPH Croatian Institute of Public Health, Zagreb, Croatia*
[g] *Iskorak, Zagreb, Croatia*
[h] *Life Quality Improvement Organisation "Flight", Zagreb, Croatia*
[i] *ULAC/CCDA Centre for Communicable Diseases and AIDS, Vilnius, Lithuania*
[j] *RPLC Republican Center for Addictive Disorders, Vilnius, Lithuania*
[k] *Public Mental Health Services. Ministry for Health, Valletta, Malta*

Abstract. Partner Notification processes focus on the notification of sexual partners to prevent the transmission of Sexually Transmitted Infections (STIs). The INTEGRATE Joint Action provides an integrated platform called RiskRadar, for combination prevention activities targeting STIs, including an anonymous, free and voluntary Partner Notification service. The presented service information flow ensures privacy, security and GDPR compliance which were identified as vital with similar tools. The service is available via web and mobile interfaces using a unique random code provided from authorised healthcare professionals to support privacy.

Keywords. HIV Infections, Sexually Transmitted Diseases, Contact Tracing

1. Introduction

Contact Tracing (CT) is the process of identifying potential infection sources and/or newly infected persons, aiming to reduce further transmission by offering testing or other preventive interventions, along with potential treatment options, depending on the specific infection per se [1]. Partner Notification (PN) is a process tightly related with

[1] Christine Kakalou, Institute of Applied Biosciences, Centre for Research & Technology Hellas, 6th Km. Charilaou-Thermi Road, Thermi, Thessaloniki; E-mail: ckakalou@certh.gr.

CT which involves voluntary notification of sexual partners of a diagnosed patient, in order to prevent further transmission of Sexually Transmittable Infections (STIs), as high prevalence rates among notified partners have been noted compared with unselectively screened individuals [2]. Obviously, PN/CT processes come with many ethical, privacy and legally related challenges, while they also come with high cost [3], [4]. Thus, digital technologies are identified as a way to improve them in a cost-effective manner [5].

The INTEGRATE Joint Action has implemented integrated activities to improve prevention, early diagnosis and linkage to care for HIV, viral hepatitis, STIs and tuberculosis in Europe. In this context, RiskRadar a publicly available web and mobile toolkit has been developed to support combination prevention, testing and linkage to care activities for the four disease areas[2]. This paper presents a technical solution to support the PN process in a GDPR complaint, secure and privacy-preserving manner.

2. Methods

The development of the PN service started with an extensive desk review to evaluate existing relevant tools and propose a prototype based on the resulting insights. Through an iterative process, the PN service was developed and tailored to identified priority groups (men who have sex with men, people who use drugs, migrants, young adults and prisoners) while ensuring its correct adaptation to all STIs. Furthermore the effective communication strategies proposed by ECDC and WHO [6], [7] were adopted and a Data Protection Impact Assessment was conducted to mitigate information security risks (e.g. service misuse, sensitive data exposure etc.) to ensure GDPR compliance.

3. Results

The architecture of the PN service is composed of 4 main technical components (Figure 1): a) An interface for HCPs dedicated to unique passcode generation, b) The RiskRadar backend that handles the unique random passcode generation and RiskRadar's PN service requests, c) RiskRadar's PN service frontend for index patients that can be accessed only via the generated passcodes and d) an SMS sending mechanism. To maintain privacy, log encryption and pseudonymisation of the phone numbers were applied; in addition, the phone numbers are automatically and permanently deleted from the system once the notification messages are delivered to the recipients.

RiskRadar is available to download from the official app stores[3,4]. Figure 2 presents the PN service information flow; healthcare professionals (HCPs) receive login credentials to use the code generator web interface in order to issue and deactivate passcodes. Once a patient receives the diagnosis, the HCP encourages the index patient to notify their past sexual partners so that they can get tested. The index patient receives a printout with the generated passcode in two formats: a serial passcode for the RiskRadar web app and a QR passcode to be scanned through the RiskRadar mobile app, along with printed instructions on how to use the PN service. The patient can then notify past sexual partners in their own time, outside of the clinical setting if they wish –

[2] https://integrateja.eu/riskradar/

[3] https://play.google.com/store/apps/details?id=com.tdermaris.integratetoolkit

[4] https://apps.apple.com/app/id1469764662

although contact information to reach affiliated organisations is offered should they need further support.

To use the passcode, the index patient can type/scan the passcode and then enter the mobile number(s) of the partners to be notified. They will then receive an SMS informing of potential STI exposure with information on where they can get tested. To prevent malicious use, the passcode is valid for 30 days and can be used up to 10 times - however the user is able to notify multiple partners in each round. Furthermore, every time a misuse/abuse incident is identified by the system through the auditing of the code generator database (e.g. a bot attempting to generate codes and use them to send PN messages), the healthcare professional is notified to deactivate the passcode.

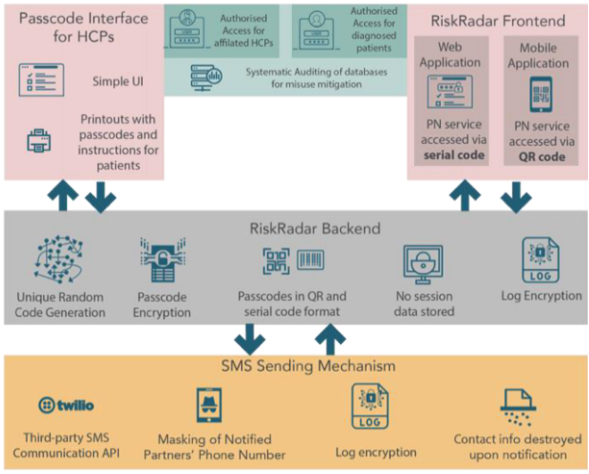


Figure 1. Partner Notification Service Architecture

4. Discussion

In the context of combination prevention efforts, the proposed service offers free, voluntary, anonymous Partner Notification and thus promotes patient empowerment and reduces stigmatisation. Concerning the service's strengths, its architecture ensures the security, privacy and confidentiality of all stakeholders involved; Furthermore, the HCP-authorised access hinders misuse attempts. We expect that the enhanced security and privacy of the service will boost its acceptance especially among vulnerable populations. Regarding limitations of this work, providing a printout of the instructions and passcodes is an obvious liability that could expose a patient's diagnosis. Another limitation is the prerequisite of a patient knowing the phone number of past sexual partners, which may be problematic especially for people engaging in casual sex. A possible workaround would be the integration of the PN service to popular dating mobile apps. Additionally, detecting a misuse of the PN service could be challenging, even with the authorisation process that is currently in place. Finally, the acceptability, usability and increased preference to the proposed PN tool over other solutions needs to be validated through a robust evaluation phase by engaging members of the community.

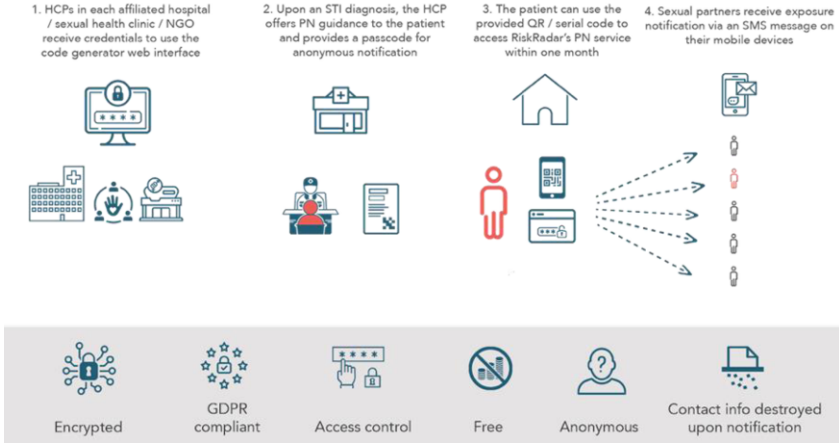


Figure 2. Information flow for Partner Notification service.

5. Conclusions

The presented PN service emphasizes the need to efficiently and securely notify sexual partners for a potential transmission of STIs. While the presented platform has not yet been pilot tested, we argue that such an approach could provide an alternative to traditional costly and ethically, legally and privacy-challenging PN approaches.

Acknowledgements

This work has received funding from the European Union for the INTEGRATE Joint Action within the framework of the Third Health Programme (grant number 761319).

References

[1] European Centre for Disease Prevention and Control. Public health guidance on HIV, hepatitis B and C testing in the EU/EEA –. 2018.
[2] Wang AL, Peng RR, Tucker JD, Cohen MS, Chen XS. Partner notification uptake for sexually transmitted infections in China: A systematic literature review. Sexually Transmitted Infections. NIH Public Access. 2012;88:386–93.
[3] Johnson BL, Tesoriero J, Feng W, Qian F, Martin EG. Cost Analysis and Performance Assessment of Partner Services for Human Immunodeficiency Virus and Sexually Transmitted Diseases, New York State, 2014. Health Serv Res. 2017 Dec;52(2):2331–42.
[4] Gibbs J, Gkatzidou V, Tickle L, Manning SR, Tilakkumar T, Hone K, Ashcroft RE, Sonnenberg P, Sadiq ST, Estcourt CS. 'Can you recommend any good STI apps?'A review of content, accuracy and comprehensiveness of current mobile medical applications for STIs and related genital infections. Sexually transmitted infections. 2017 Jun 1;93(4):234-5.
[5] Althaus CL, Turner KM, Mercer CH, Auguste P, Roberts TE, Bell G, Herzog SA, Cassell JA, Edmunds WJ, White PJ, Ward H. Effectiveness and cost-effectiveness of traditional and new partner notification technologies for curable sexually transmitted infections: observational study, systematic reviews and mathematical modelling. Health technology assessment (Winchester, England). 2014;18(2):1-viii.
[6] European Centre for Disease Prevention and Control. Utilising social media to support HIV/STI prevention: evidence to inform a handbook for public health programme managers. 2017.

[7] WHO. Guidelines on HIV self-testing and partner notification. Supllement to Consolidated Guildelines on HIV Testing Services. World Health Organization. 2016.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210958

Quantifying Citizens' Well-Being in Areas with Natural Based Solutions Using Mobile Computing

Parisis GALLOS[a,b,1], Andreas MENYCHTAS[a,b], Christos PANAGOPOULOS[a], Manthos BIMPAS[c] and Ilias MAGLOGIANNIS[b]
[a] *BioAssist S.A., Athens, Greece*
[b] *Computational Biomedicine Research Lab, Department of Digital Systems, University of Piraeus, Piraeus, Greece*
[c] *National Technical University of Athens, Athens, Greece*

Abstract. Urban planners, architects and civil engineers are integrating Nature-Based Solutions (NBS) to address contemporary environmental, social, health and economic challenges. Many studies claim that NBS are poised to improve citizens' well-being in urban areas. NBS can also benefit Public Health, as they can contribute to optimising environmental parameters (such as urban heat island effects, floods, etc.), as well as to the reduction of diseases, as for example cardiovascular ones and the overall mortality rate. In addition, the usage of mobile health (mHealth) solutions has been broadly applied to support citizens' well-being as they can offer monitoring of their physical and physiological status and promote a healthier lifestyle. The aim of this paper is to present the specifications, the design and the development of a mobile app for monitoring citizens' well-being in areas where NBS have been applied. The users' physical activity and vital signs are recorded by wearable devices and the users' locations are recorded by the proposed mobile application. All collected data are transferred to the cloud platform where data management mechanisms aggregate data from different sources for combined analysis. The mobile application is currently available for Android and iOS devices and it is compatible with most smart devices and wearables. The "euPOLIS by BioAssist" application can be used as a health and other data collection tool to investigate citizen's well-being improvement in areas with NBS.

Keywords. Natural Based Solutions (NBS), Mobile Systems, Well-Being, Pervasive Computing

1. Introduction

Urban planners, architects and civil engineers are integrating Nature-Based Solutions (NBS) to address contemporary environmental, social, health and economic challenges [1-2]. NBS can be characterized as a new way for environmental management by applying solutions such as "green" roofs or walls to cool down urban areas during summer, to capture and save rain water, to reduce pollution, and to increase human well-being while enhancing biodiversity [2]. Apart from the contribution of NBS to the

[1] Corresponding Author, Parisis Gallos, Computational Biomedicine Research Lab, Department of Digital Systems, University of Piraeus, Piraeus, Greece; E-mail: parisgallos@bioassist.gr.

environment, many relative studies claim that NBS are poised to improve citizens' well-being in urban areas. Specifically, citizens' stress levels have been reduced, while physical activity levels are increased within NBS areas. NBS can also bolster Public Health, by contributing to the optimisation of relevant environmental parameters (e.g., urban heat island effects, floods, etc.), thus reducing the risk of diseases, such as cardiovascular ones, and, consequently, mortality rates [3].

Meanwhile, the usage of mobile health (mHealth) solutions has been broadly applied to support people's well-being [4-6]. mHealth applications have been used to support patients with cardiovascular diseases [7] with very promising results. Many studies also present the value of mHealth in the well-being of mental health patients [8,9]. The usage of mHealth technologies has been also linked to highly positive impact on public health [10,11]. By measuring and continuously monitoring the user's physical activity and vital signs, mHealth applications can be used to collect and analyse biosignals to ensure the users' good health status [12-17]. Finally, mHealth solutions can encourage a healthier lifestyle through coaching mechanisms that promote physical activity [18,19].

The aim of this paper is to present the specifications, the design and the development of an mHealth app for monitoring citizen well-being in areas where NBS have been applied in the frame of the euPOLIS project.

2. Methods

In order to achieve the aim of the study, requirement analysis took place, based on the potential user needs, as these were described by urban planners in the municipalities of the test areas (demo sites of the euPOLIS project) and medical experts. User requirements were collected through interviews from November 2020 until May 2021. The specifications of the mobile application for monitoring citizen well-being in areas where NBS are applied include location tracking, the timestamp of the visits in the areas, as well as recording of specific biomarkers, such as physical activity (number of steps, daily exercise, walking/running etc.), heartrate, SpO_2, sleep quality, and stress levels.

To collect the aforementioned data, citizens need to use a mobile phone and a wearable device, such as a smartwatch or a smart-band. Figure 1 (a) presents the recorded data flow. The user's activity data and vital signs are recorded by the wearable devices and the user's location is recorded by the mobile application. All data are transferred to the cloud platform, where information from different sources is aggregated for combined analysis. The final recorded data will be further analysed to examine the impact of the NBS on citizens' well-being.

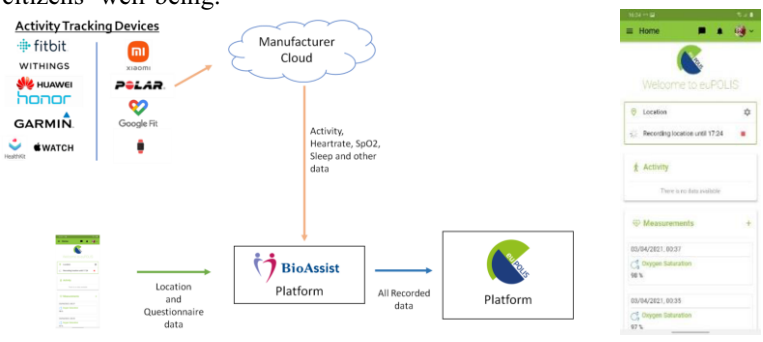


Figure 1. Collected Data Flow (a) and Mobile Application Screenshot (b)

3. Results and Discussion

A mobile application has been developed (Figure 1 (b)) to monitor the users' daily physical activity and other variables related to citizens' well-being in areas with NBS. The mobile application is currently available for Android and iOS devices and it is compatible with most commercially available smart devices and wearables. The proposed solution supports integration on the platform level, where the acquisition of measurements takes place via services supported by the cloud platforms of the device manufacturers (i.e., Apple Watch, Fitbit, Garmin, Polar, Withings, Xiaomi, Huawei, etc.). For data storage and data exchange, HL7-FAIR standards have been adopted to ensure the validity and the reliability of the system.

Users can activate the location tracking feature any time they visit the NBS pilot sites and they can also respond to related questionnaires, which are available through the mobile application. In addition, the application offers a timeline of data related to the user's physical activity and vital signs measurements. Users have access to their historical data through the timeline and can also apply filters and view the information if graph format.

The "euPOLIS by BioAsssist" application can be used as a health and other data collection tool to investigate the potential improvement of citizen's well-being in areas with NBS. Relative studies have used comparable approaches to monitor people's well-being and public health using similar mHealth applications [6,11]. The expected value of the proposed application is to collect reliable and accurate data to examine citizens' well-being in an "easy-to-use" and relatively non-intrusive manner.

4. Conclusions

The aforementioned application is developed in the context of the EU-funded project "EuPOLIS - Integrated NBS-based Urban Planning Methodology for Enhancing the Health and Well-being of Citizens: the euPOLIS Approach", aiming to deploy natural systems to enhance public health and well-being and create resilient urban ecosystems. The project's solutions will be tested in four cities: Belgrade, Lodz, Piraeus and Gladsaxe.

Future work includes the analysis of the collected data using specific indicators to measure the improvement of the participating in the pilot study citizens' well-being, as well as the enhancement of the current mobile application with additional features for achieving more comprehensive and pervasive monitoring.

Acknowledgements

This work is co-funded by the Horizon 2020 Programme of the European Commission Grant Agreement number: 869448 - Integrated NBS-based Urban Planning Methodology for Enhancing the Health and Well-being of Citizens: the euPOLIS Approach.

References

[1] Cohen-Shacham E, Walters G, Janzen C, Maginnis S. Nature-based solutions to address global societal challenges. IUCN: Gland, Switzerland. 2016;97.

[2] Eggermont H, Balian E, Azevedo JM, Beumer V, Brodin T, Claudet J, Fady B, Grube M, Keune H, Lamarque P, Reuter K. Nature-based solutions: new influence for environmental management and research in Europe. GAIA-Ecological Perspectives for Science and Society. 2015 Jan 1;24(4):243-8.

[3] Van den Bosch M, Sang ÅO. Urban natural environments as nature-based solutions for improved public health–A systematic review of reviews. Environmental research. 2017 Oct 1;158:373-84.

[4] Khorakhun C, Bhatti SN. Wellbeing as a proxy for a mHealth study. In2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2014 Nov 2 (pp. 32-39). IEEE.

[5] Saxena M, Saxena A. Evolution of mHealth eco-system: a step towards personalized medicine. Adv Intell Syst Comput. 2020 Feb 28;1087:351-70.

[6] Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res. 2016 Sep 5;16 Suppl 5(Suppl 5):290. doi: 10.1186/s12913-016-1522-3. PMID: 27608677; PMCID: PMC5016733.

[7] Chow CK, Ariyarathna N, Islam SM, Thiagalingam A, Redfern J. mHealth in Cardiovascular Health Care. Heart Lung Circ. 2016 Aug;25(8):802-7. doi: 10.1016/j.hlc.2016.04.009. Epub 2016 May 11. PMID: 27262389.

[8] Hensel JM, Ellard K, Koltek M, Wilson G, Sareen J. Digital Health Solutions for Indigenous Mental Well-Being. Curr Psychiatry Rep. 2019 Jul 1;21(8):68. doi: 10.1007/s11920-019-1056-6. PMID: 31263971; PMCID: PMC6602981.

[9] Patel S, Saunders KE. Apps and wearables in the monitoring of mental health disorders. Br J Hosp Med (Lond). 2018 Dec 2;79(12):672-675. doi: 10.12968/hmed.2018.79.12.672. PMID: 30526097.

[10] Grady A, Yoong S, Sutherland R, Lee H, Nathan N, Wolfenden L. Improving the public health impact of eHealth and mHealth interventions. Australian and New Zealand journal of public health. 2018 Apr 1;42(2):118-9.

[11] Freudenberg N. Assessing the Public Health Impact of the mHealth App Business. Am J Public Health. 2017 Nov;107(11):1694-1696. doi: 10.2105/AJPH.2017.304083. PMID: 29019765; PMCID: PMC5637691.

[12] Majumder S, Mondal T, Deen MJ. Wearable Sensors for Remote Health Monitoring. Sensors (Basel). 2017 Jan 12;17(1):130. doi: 10.3390/s17010130. PMID: 28085085; PMCID: PMC5298703.

[13] Block VA, Pitsch E, Tahir P, Cree BA, Allen DD, Gelfand JM. Remote Physical Activity Monitoring in Neurological Disease: A Systematic Review. PLoS One. 2016 Apr 28;11(4):e0154335. doi: 10.1371/journal.pone.0154335. PMID: 27124611

[14] Kallipolitis A, Galliakis M, Menychtas A, Maglogiannis I. Affective analysis of patients in homecare video-assisted telemedicine using computational intelligence. Neural Computing and Applications. 2020 Dec;32(23):17125-36.

[15] Panagopoulos C, Malli F, Menychtas A, Smyrli EP, Georgountzou A, Daniil Z, Gourgoulianis KI, Tsanakas P, Maglogiannis I. Utilizing a homecare platform for remote monitoring of patients with idiopathic pulmonary fibrosis. InGeNeDis 2016 2017 (pp. 177-187). Springer, Cham.

[16] Menychtas A, Doukas C, Tsanakas P, Maglogiannis I. A versatile architecture for building IoT quantified-self applications. In2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS) 2017 Jun 22 (pp. 500-505). IEEE.

[17] Panagopoulos C, Menychtas A, Tsanakas P, Maglogiannis I. Increasing Usability of Homecare Applications for Older Adults: A Case Study. Designs. 2019; 3(2):23

[18] McGarrigle L, Todd C. Promotion of Physical Activity in Older People Using mHealth and eHealth Technologies: Rapid Review of Reviews. J Med Internet Res. 2020 Dec 29;22(12):e22201. doi: 10.2196/22201. PMID: 33372894; PMCID: PMC7803474.

[19] Dobkin BH, Dorsch A. The promise of mHealth: daily activity monitoring and outcome assessments by wearable sensors. Neurorehabil Neural Repair. 2011 Nov-Dec;25(9):788-98. doi: 10.1177/1545968311425908. PMID: 21989632; PMCID: PMC4098920.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210959

Perform the Regulatory Process: A National Approach for Evaluation of the Pre-Hospital Emergency Information System in I.R. Iran

Somayeh ABEDIAN[a,b,1], Pirhossein KOLIVAND[a], Hamid Reza LORNEJAD[a]
[a] *National Emergency Medical Services Organization, Ministry of Health and Medical Education, I.R. Iran*
[b] *Department of Information Technology Management, Islamic Azad University, Qazvin, Iran*

Abstract. Evaluation of Emergency Medical Services Management Information System (EMSIS) makes it possible to assess the extent to which the objectives of supporting of healthcare delivery services. This paper presents an overview of the regulatory process in prehospital EMS electronic data registration and provides a minimum data set for the purpose of developing such a care system on a national scale. It further offers an evaluation framework for such systems.

Keywords. Pre-Hospital Emergency Medical Services Management Information System (EMSIS), operations leadership center (OLC), Key performance indicators, Pre-Hospital Care Report (PCR)

1. Introduction

Prehospital care is provided by Emergency Medical Services (EMS) organization staff, the initial health care providers at the disaster scene [1]. EMS personnel are often the first to recognize the nature of a disaster, immediately evaluate the situation, determine the need for resources and perform emergency medical procedures [2]. The Emergency Medical Services Management Information System (EMSIS) has been developed with the aim of improving the follow-up process, monitoring the calls for first aid procedures of pre-hospital emergency (phone number: 115 in Iran). In this system, the information of a mission is recorded from the moment of receiving a phone call and all information received from the caller by the call center operators (trained nurses) is added to the mission Electronic Medical Record (EMR) in order to meet the information needs and support the delivery of emergency medical services [3]. In case of the need to send a rescue vehicle such as a motorlance, ambulance or helicopter, this operation is performed by the Operations Leadership Center (OLC) and dispatch information is also registered in the patient's EMR. As soon as the paramedics or Emergency Medical Technicians (EMTs) arrive at the patient bedside, the information of the patient observations and vital signs are added to the mission information. If the EMT needs any advice from specialists present at OLC (under the supervision of National EMS organization), their diagnosis and prescribed item are added to the mission information. In addition, all telephone calls of a mission are recorded for future reference (for example, in case of legal proceedings or complaints) [3,4]. After transferring patient to the hospital and hospitalization, all the information about administration data or medical procedures performed for the patient in the hospital are registered in the mission EMR. To perform the whole EMS processes through a comprehensive information system, the existence of special capabilities and

[1] Corresponding and first author: Somayeh Abedian, Ministry of Health and Medical Education, Tehran, Iran, E-mail: S.abedian@behdasht.gov.ir, So.abedian@gmail.com

key indicators is crucial [5]. Therefore, the Iranian EMS organization and the Statistics and Information Technology Management Center of the Ministry of Health and Medical Education (ITMC-MOHME) determined and developed the main indicators of a capable system to meet EMS requirements. Due to the vitality of providing pre-hospital services and direct link with human lives, the high efficiency of this system in both technical and functional dimensions was studied. This paper provides the definitions of a number of Key Performance Indicators (KPI) to assess the quality, functionality and reliability of such a comprehensive system.

2. Methods

In this project, a model for evaluation of Emergency Medical Services Management Information Systems (EMSIS) was developed. The model contains a number of indicators in two categories that evaluate in line with weight-based decision tree framework [6,7]. In order to identify and classify the indicators, help was enlisted from emergency medicine specialists and experts in EMS affairs alongside health informatics professionals, IT experts and other involved people [9]. Also, the collaboration of ISO corresponding technical committees in Health informatics (ISO/TC 215) members was utilized. EMSIS is divided into different parts. According to experts, the categories fall in two groups of Functional and Technical, each supporting some primary sub-groups. More than 25 separate parts with unique features were defined. After the first round of consulting experts, they were decreased to 10 subcategories, and then final KPIs were reallocated. Thus, in this model attempts were made to combine all similar features and define some categories to support all KPIs. The criteria identified in the framework were given expert ratings. In other words, they weighted to qualitative indicators were converted to quantitative and metric indices. Then, the system was assessed in terms of supporting the functional criteria.

3. Results

3.1 Evaluation components

The indicators are used for the government to guarantee the reliability, responsibility and accessibility of this critical information system. The system is a live operational information system at the level of Transaction Processing Systems (TPS), and uninterrupted activity is inevitable [5]. The first step constituted determining clear objectives for the system. For every goal, sub-groups were defined under functional or technical groups. The framework includes more than 160 criteria assigned to 10 categories in two main branch of functional and technical components. The technical components are typically adapted with the infrastructure of software and technical support. The functional components are related to the operational requirements necessary to be registered electronically. These components are not specific to parts of the procedure. They may be utilizing in different parts of the phone-calls to discharge patient from EMS. For instance, vital signs that are recorded in EMT's part in a cellphone application will be used by physicians at medical advice department to observe the patient's clinical condition and prescribe medicines or a treatment course. In the technical aspect, the security of system has to be observed in all sub-systems. The functional components include the following:

1. Telephone triage support,
2. Electronic registration of all dispatch and mission information,

3. Medical Advice Unit Support,
4. Ability to record admission unit information (pre-hospital and hospital),
5. Support for electronic and automatic recording of clinical and non-clinical data by EMT or EMS workers (completion of the PCR electronic form and other critical data at the patient's bedside).

The functional indicators include more than 130 items of which a few of Medical Advice Unit Support criteria are shown in Table 1.

Table 1. A number of key performance indicators (KPI) for the evaluation of Medical Advice Unit Support (Functional type)

Indicator(s)	Weight in category
Fields to insert the doctor's diagnosis in the form of checkbox menus or optional and free-text field for descriptions	5
Ability to automatically register and attach all conversations with the consulting physician to EMR	8
Text fields to register suspicious cases	4
The possibility to attach a new medical record to the original record with the insertion of time (connection to last medical record) in case of re-consultation for a patient and opening a new consultation form	4
Visibility of the number of EMTs and callers in the waiting queue with separate alarms (mechanized, audible, colored, etc.) for the physician	8

The technical components include the following:

6. Security considerations (according to security certification from ICT Ministry)
7. Technical performance
8. Standard documentation of software
9. System performance and efficiency
10. Software flexibility

Technical type encompasses more than 35 indicators of which a few standard documentation criteria of software category are presented in Table 2.

Table 2. A number of key performance indicators (KPI) for evaluation of standard documentation of software category (Technical aspect)

Indicator(s)	Weight in category
Standard software development document	4
A competency certificate issuance from EMS organization	4
Ability to export dynamic reports as Excel, Microsoft Word, XML, etc.	3
Software roadmap document	2
Software application document	3
Software architecture document	3
Software data model and database design document	3
Existing process models and roles document	3
Provision of educational documentation depending on user levels	3
Data dictionary document	3

3.2 Indicator weighting

In this model, each component comprises some main subgroups and also their subsidiary subgroups. Functional and technical component each include 5 subsets. The functional and technical components have more than 130 and 35 criteria, respectively. The criteria were weighted according to experts' comments, and after finalization, were assessed by private companies developing EMSIS product based on published criteria via a self-evaluation process. Finally, a some time is allocated to each company to present their product and defend its potentials and capability of their system within the context. Finally, based on the evaluation results, a national certificate was issued to represent

capability level of the software. This certificate indicates a product's quantitative ranking in two categories and ten subgroups. The quantitative results of the certification involve three levels which are shown in Table 3.

Table 3. Status of evaluation results in sub-groups

Status	Description	Score
Optimal	In this status, the system ideally meets the requirements.	Above 80%
acceptable	In this status, the system satisfies minimum acceptable functionality and the software needs further development.	Between 60 and 80%
Rejected	This level is rejected and the technical and functional capabilities are not achieved.	Less than 60%

4. Discussion

An assessment framework is designed, implemented and would annually improve the criteria based on professional advice and emerging requirements. To this aim, a committee consisting of all relevant experts was formed. In this process, sponsored by the National EMS organization, ITMC-MOHME, Computer Union Organization (Non-Governmental) and Supreme Council of Cyberspace (under presidential supervision), attempts were made to improve the capabilities of EMSIS based on national and international standards. This paper seeks to serve the function of evaluating and achieving more efficient systems, effective KPIs in service delivery (aspects evaluation) and developing evaluation methods for such systems. Moreover, the first purpose of this work was improving the quality of EMS services in response time, acceptability, continuity/sustainability, patient safety, clinical effectiveness and welfare for citizens. In this regard, a set of criteria and indicators have been published, according to which relevant software is assessed, and certificates are issued for each product in ten classifications.

References

[1] Washington DC. Committee on Guidance for Establishing Crisis Standards of Care for Use in Disaster Situations, Institute of Medicine. Crisis Standards of Care: A Systems Framework for Catastrophic Disaster Response. Washington (DC): National Academies Press (US). DC: National Academies Press; 2012.

[2] Marks MR, Bowers C, DePesa NS, Trachik B, Deavers FE, James NT. REACT: A paraprofessional training program for first responders—A pilot study. Bulletin of the Menninger Clinic. 2017 Jun;81(2):150-66.

[3] Ramadanov N, Klein R, Ramadanova N, Behringer W. Influence of Time of Mission on Correct Diagnosis by the Prehospital Emergency Physician: A Retrospective Study. Emergency Medicine International [Internet]. 2019 Apr 2;2019. Available from: https://www.hindawi.com/journals/emi/2019/3727081.

[4] EMS1 Academy. EMS protocols all paramedics should follow [Internet]. https://www.ems1.com. 2019 [cited 2019 Aug 16]. Available from: https://www.ems1.com/ems-education/articles/ems-protocols-all-paramedics-should-follow-7c1zyNEC6v1SjDQM/.

[5] Pap R, Lockwood C, Stephenson M, Simpson P. Indicators to measure prehospital care quality: a scoping review. JBI Database of Systematic Reviews and Implementation Reports. 2018 Nov;16(11):2192–223.

[6] Doğan SZ, Arditi D, Murat Günaydin H. Using decision trees for determining attribute weights in a case-based model of early cost prediction. Journal of Construction Engineering and Management. 2008 Feb;134(2):146-52.

[7] Mark H. A Decision Tree-Based Attribute Weighting Filter for Naive Bayes. Research and Development in Intelligent Systems XXIII. New Zealand: Springer; 2007;59–70.

[8] Kass GV. An Exploratory Technique for Investigating Large Quantities of Categorical Data. Journal of the Royal Statistical Society. 1980;29(2):119–27.

[9] McSherry D. Strategic induction of decision trees. Knowledge-Based Systems. 1990;12(5–6):269–75.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210960

Pitfalls of Social Media in the Era of COVID-19 Pandemics

Abdullah ALANAZI[a], Hanan ALDOSARI[a] and Bakheet ALDOSARI[a,1]
[a] *College of Public Health and Health Informatics, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia*

Abstract. In December 2019, Wuhan, China, reported an outbreak of nSARS-CoV2 that caused viral pneumonia, COVID-19. Li Wenliang, a Chinese ophthalmologist, first communicated on Chinese social media about the existence and spread of the unknown viral pneumonia in Wuhan, China. By the end of March 2020, the virus had spread worldwide. However, non-scientific information related to the viral outbreak, disease, and mortality spread even faster on social media. This study performed literature searches among different databases, i.e., PubMed, PubMed Central, and Web of Science, to understand the pitfall of social media during the COVID-19 pandemic and the impact of non-scientific information on public health. Social media not only shared information regarding the outbreak of nSARS-CoV2 and COVID-19 disease but also misinformation regarding epidemiology, government policies, prevention, cure, and vaccination. Thus, strict regulation is required to control the spread of misleading information.

Keywords. COVID-19, nSARS-CoV2, social media, pitfalls, and falsification

1. Introduction

The outbreak of beta coronavirus in Wuhan, China, in December 2019, resulted in a global pandemic named COVID-19, affecting more than 150 million people with more than 3 million deaths so far and the world is still fighting to find a cure. COVID-19 vaccines received mixed responses resulting in vaccine hesitancy with social media playing a critical role in spreading information (generally fabricated) towards safety and efficacy.

Last year, several regions reported multiple waves of infection, and India being the worst hit. Platforms such as Facebook, YouTube, WhatsApp, Instagram, and Twitter, are the most widely used, and never before in human history, the sharing of information was so easily possible [1]. However, these social media platforms have shown a dual effect on human life. On the one hand, they have spread awareness and shared correct information, enabling many populations to adopt recommended precautions against the current infection pandemic [2]. On the other hand, there is the dark side of social media as well, and in the present pandemic, it has created a panic situation in every aspect of COVID-19 related to cause, spread, treatment, and vaccine.

It is essential to understand that more than 3 billion social media users worldwide remain active in sharing information, particularly non-scientific knowledge [3]. In India

[1] Corresponding Author, Bakheet ALDOSARI, College of Public Health & Health Informatics, King Saud Bin Abdul-Aziz University for Health Sciences, Saudi Arabia. Email: dosarib@gmail.com.

alone, nearly 375 million social media users represent a mass population who indulge in sharing information. More than 75% of social media users in India and the rest of the world rely on information shared on WhatsApp and other social media platforms [4]. This review focuses on the effects of social media on sharing information on nSARS-CoV2 spread and COVID-19. It also highlights how false, fabricated, and manipulated information shared on different social media platforms affected lives during the pandemic.

2. Methodology

To address the pitfall of social media during the COVID-19 pandemic, various databases such as PubMed, PubMed Central, Web of Science, and Scientific blogs were searched for studies reporting a negative role of social media in the present COVID-19 pandemic. Different keywords such as "Pitfall", "Impact of Social Media", "Pros and Cons of Social Media", and "Scientific relevance of Social Media', in context with COVID-19 pandemic, were used. We selected articles associated with information on COVID-19 and social media as sharing platforms in each search. These articles were examined for social media issues on the COVID-19 pandemic and how it affected human lives.

3. Results

We found that 18 most relevant articles highlighted the negative impact of social media in sharing rumors on the cause/source of infection, disease, etiology and symptoms, preventive measures (restrictions and lockdown), guidelines for appropriate covid behavior, treatment, and vaccination.

3.1. Social Media versus the nSARS-CoV2 spread, restrictions, and containment

In early 2020, most of the countries across the globe were in a strict lockdown to contain the spread of the virus. In many instances, social media platforms were reported with misleading information and manipulated content. Brennen et al. demonstrated social media as a potential enemy in the current pandemic by sharing information via different types, sources, and claims of misinformation in COVID-19 [5]. It is evident that information on the global internet travel even faster than a virus and affects millions of individuals mentally, creating a panic situation. In fact, in 2020, The Lancet published an editorial "COVID- 19; fighting panic with information" [6]. The regulations associated with social media for sharing information are fragile, and in most cases, users do not abide by such regulations. The repeated and bulk amount of misleading/manipulated information creates a devastating consequence in the public domain, as seen in the current COVID- 19 pandemic [2]. The low low-quality research work often promoted via these social media platforms becomes an example of self-validation of shared information [7].

3.2. Social media versus COVID-19 management

The most challenging aspect of the current pandemic is the lack of treatment and an authentic source of information. In such a panic situation, a large percentage of the population relies on self-medication and seeks information on the internet. Gottlieb and Dyer demonstrated that more than 1.5 million Twitter users targeted malicious and manipulative behavior to the large population seeking self-medication in the COVID-19 pandemic [7]. Malik et al. reported that social media offered non-authenticated remedies for self-medication during COVID-19 [8]. Various social media pages are filled with overwhelming information (non-authenticated scientifically) on herbal and traditional medicine [9]. More than benefits, these remedies remain associated with negative, harmful health impacts. Devastating consequences are reported in many developing countries, such as India, where a prescription is not mandatory to buy several therapeutics such as chloroquine and hydroxychloroquine [10]. The use of alternative medicine to cure disease is known since ancient times, and during COVID-19, various herbs are being used on a large scale. Here, social media platforms are the promoter for such uses. In addition, the lack of medicine and treatment has caused a large percentage of the population to find an appropriate treatment remedy not only in India but also in the rest of the world as well [11]. Several life-saving therapeutics in the case of COVID-19 are not available adequately; however, social media provides details of suppliers and sellers who are extremely unethical. There are reports that Remdesivir, a potential therapeutic for chronic COVID-19, is available for sale without prescription [12]. As we are in the middle of the pandemic, the details of life support devices such as oxygen cylinders and ventilators are readily available on social media [1]. However, in most of these social media reports, the information is either false or manipulated [13]. Additionally, there is a growing debate on the safety and efficacy of vaccines across the globe among the scientific community. There is even a new trend in social media that debates whether to take vaccines or not without any scientific basis [14].

4. Discussion and Conclusions

COVID-19 is the second most and worst pandemic in human history after the Spanish Flu in 1919. This pandemic is not over yet and is still affecting millions of lives daily [15,16]. The lack of therapeutics, life support care, and facilities has forced people to opt for alternative approaches to fight against the disease [17]. Here, social media becomes a key ally but largely an enemy as it offers non-scientific information about the diagnosis, precautions, and treatments of COVID-19 [18]. COVID-19 is not only a pandemic but also represents an unprecedented setting for the spread of misleading/manipulated information [19]. Though the first report on COVID-19 was shared on social media (Chinese social media), the role of social media has since then changed completely [15]. Different social media pages such as Facebook, YouTube, WhatsApp, Instagram, and Twitter share overwhelming information on COVID-19.

In most cases, such information is non-scientific and misleading, affecting the mental and physical health of many populations [20]. Social media is a vital tool for sharing information across the globe, but in the case of COVID-19, a large percentage of such information remains misleading and manipulated [21]. Most social media platforms have regulations related to information sharing, but these regulations are

fragile, and users find an opportunity to violate such regulations. Therefore, there is an urgent need to find a measure to control the spread of misleading/manipulated information to help us fight against COVID-19.

References

[1] González-Padilla DA, Tortolero-Blanco L. Social media influence in the COVID-19 Pandemic. Braz J Urol. 2020 Jul 27;46:120-4.
[2] Venegas-Vera AV, Colbert GB, Lerma EV. Positive and negative impact of social media in the COVID-19 era. Rev Cardiovasc Med. 2020 Dec 30;21(4):561-564.
[3] Singh S, Dixit A, Joshi G. Is compulsive social media use amid COVID-19 pandemic addictive behavior or coping mechanism? Asian J Psychiatr. 2020 Dec;54:102290.
[4] Brennen JS, Simon F, Howard PN, Nielsen RK. Types, sources, and claims of COVID-19 misinformation. Reuters Institute. 2020 Apr 7;7(3):1.
[5] The Lancet. COVID-19: fighting panic with information. Lancet (London, England). 2020 Feb 22;395(10224):537.
[6] Gottlieb M, Dyer S. Information and Disinformation: Social Media in the COVID‐19 Crisis. Acad Emerg Med. 2020 Jul 1.
[7] Malik M, Tahir MJ, Jabbar R, Ahmed A, Hussain R. Self-medication during Covid-19 pandemic: challenges and opportunities. Drugs Ther Perspect. 2020 Dec;36(12):565-7.
[8] Chan AKM, Nickson CP, Rudolph JW, Lee A, Joynt GM. Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. Anaesthesia. 2020 Epub ahead of print.
[9] Schillinger D, Chittamuru D, Ramírez AS. From "infodemics" to health promotion: a novel framework for the role of social media in public health. Am J Public Health. 2020 Sep;110(9):1393-6.
[10] Badell-Grau RA, Cuff JP, Kelly BP, Waller-Evans H, Lloyd-Evans E. Investigating the Prevalence of Reactive Online Searching in the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res. 2020;22(10):e19791.
[11] Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. Lancet Digital Health. 2021 Jan 28.
[12] Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res. 2020 Apr 21;22(4):e19016.
[13] Sumbaly R, Miller M, Shah H, Xie Y, Culatana SC, Khatkevich T, Luo E, Strauss E, Szilvasy G, Puri M, Manadhata P, Graham B, Douze M, Yalniz Z, Jegou H. Using AI to detect COVID19 misinformation and exploitative content. Facebook Artificial Intelligence. Available at: https://ai.facebook.com/blog/using-ai-to-detect-covid-19-misinformation-and-exploitative-content/.
[14] Ashton J. COVID-19 and the 'Spanish'flu. J Royal Soc Med. 2020 May;113(5):197-8.
[15] Franchini AF, Auxilia F, Galimberti PM, Piga MA, Castaldi S, Porro A. COVID 19 and Spanish flu pandemics: All it changes, nothing changes. Acta Biomed. 2020;91(2):245.
[16] Williams OD. COVID-19 and Private Health: Market and Governance Failure. Development. 2020 Dec;63(2):181-90.
[17] Al-Dmour H, Masa'deh R, Salman A, Abuhashesh M, Al-Dmour R. Influence of Social Media Platforms on Public Health Protection Against the COVID-19 Pandemic via the Mediating Effects of Public Health Awareness and Behavioral Changes: Integrated Model. J Med Internet Res. 2020;22(8):e19996.
[18] Chen E, Lerman K, Ferrara E. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. JMIR Public Health Surveill. 2020 May 29;6(2):e19273.
[19] Obiała J, Obiała K, Mańczak M, Owoc J, Olszewski R. COVID-19 misinformation: Accuracy of articles about coronavirus prevention mostly shared on social media. Health Policy Technol. 2021 Mar 1;10(1):182-6.
[20] Naeem M. Do social media platforms develop consumer panic buying during the fear of Covid-19 pandemic. J Retail Consum Serv. 2021 Jan 1;58:102226.
[21] Anwar A, Malik M, Raees V, Anwar A. Role of Mass Media and Public Health Communications in the COVID-19 Pandemic. Cureus. 2020 Sep 14;12(9):e10453.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210961

The Impact of Virtual Reality in Enhancing the Quality of Life of Pediatric Oncology Patients

Abdullah ALANAZI[a], Faisal ASHOUR[a], Hanan ALDOSARI[a] and Bakheet ALDOSARI[a,1]
[a] *College of Public Health and Health Informatics, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia*

Abstract. Pediatric oncology interventions involve many challenges, such as multiple hospitalizations, invasive procedures, and adverse physical and psychological treatment side effects. Treatment burden, including administration of general anesthesia, contributes to high levels of psychological distress among pediatric patients and their families. Virtual reality (VR) is a distraction method, which offers an extremely realistic and interactive virtual environment and helps reduce pain and distress by means of a head-mounted display and headphones. VR is based on two crucial dimensions: immersion and presence, which results in the complete suspension of disbelief that the experience is artificial and allows a greater degree of presence and reaction to the stimulations. The VR technology has become a common practice in scientific and clinical research due to its affordability and ease of use. In pediatric settings, the most widely researched clinical application of VR has focused on the effectiveness of VR distraction therapy in the attenuation of acute pain, anxiety, and distress during invasive medical procedures. It has also been hypothesized to be a nonpharmacological form of analgesia that positively influences the body's intricate pain modulation system during painful medical procedures. In this review, we showed the potential benefits of VR technology during radiotherapy and intrathecal procedures on pediatric oncology patients and its involvement in enhancing their quality of life during and after the treatment. Therefore, a collaboration between researchers, clinicians, and programmers is crucial for the inclusion of VR technology in more clinical procedures, which would consequently enhance the patient's quality of life.

Keywords. Virtual Reality, Healthcare, Challenges, Oncology, Pediatrics

1. Introduction

Pediatric oncology patients face tremendous challenges during the treatment process since most of the procedures require the administration of general anesthesia (GA), either to reduce anxiety during radiotherapy sessions or for pain management during intrathecal chemotherapy procedures. General anesthesia exerts a tremendous burden on patients as well as the healthcare system. The various side effects of GA include nausea and vomiting, sore throat, dental injury, and confusion. Moreover, it can cause serious complications, such as allergic reactions, respiratory compromise, and death [1].

[1] Corresponding Author, Bakheet Aldosari, College of Public Health & Health Informatics, King Saud Bin Abdul-Aziz University for Health Sciences, Saudi Arabia; Email: dosarib@gmail.com.

In addition to the complications, another challenge of delivering radiotherapy to pediatric patients is the recovery process after GA. The recovery process includes three phases: immediate recovery, intermediate recovery, and long-term recovery, which depends on many factors, such as the type of anesthesia, drug factors, and patient characteristics. The recovery should occur within 60-90 min after GA, which is a considerable amount of time for pediatric patients [2]. Furthermore, there is a need for a larger team comprising anesthesiologists, physicians, nurses, and technicians. Problems are also encountered during anesthesia scheduling owing to the large number of patients and paucity of time. Each patient requires multiple sessions, like five days a week. Therefore, there is a need for targeted, effective interventions that will support pediatric patients in regaining or maintaining their emotional wellbeing and quality of life.

Virtual reality (VR) is an emerging therapeutic tool that engages individuals in a made-up artificial environment by means of a head-mounted display and headphones. The head-mounted display helmet obscures the user's entire real-life visual field and replaces it with a virtual world of visual stimulation, while the headphones replace external auditory sounds with virtual-environment sounds [3]. The factors that contribute to the VR experience are immersion and presence. Immersion considers the action and movement in the virtual space, while presence reflects the involvement of perception in that space. Together, they allow the individual to suspend any disbelief that the experience is artificial [4]. Recently, VR has been used in some areas of clinical practice. The reports from multiple studies conducted in the field of pain management and palliative care are promising [5]. Moreover, VR effectively distracts the patients undergoing wound debridement and reduces pain intensity [6]. VR distraction has also been shown to reduce preoperative anxiety and improve pediatric intra-operative compliance [1]. Although VR was found to significantly reduce pain and distress scores, the effect on anxiety, depression and fear was not significantly attained in pediatric cancer patients [7]. The use of such a technology may decrease the burden of using general anesthesia on patients during medical procedures.

The aim of this review was to determine the potential benefits of VR technology while delivering radiotherapy or intrathecal chemotherapy to pediatric patients. We assessed the delivery time, side effects, and patients' acceptance and satisfaction, which will ultimately enhance their quality of life.

2. Methodology

Articles related to VR technology were retrieved from and PubMed databases as well as from Google Scholar. The search strategy included the following MeSH terms ("Virtual Reality"[Mesh]) AND "Quality of Life"[Mesh]), ("Neoplasms"[Mesh] AND "Virtual Reality"[Mesh]). Furthermore, keywords, such as "virtual reality", "healthcare", "oncology", "pediatrics", and "challenges". All articles with these keywords in their title or abstract were screened, with the focus on the limitations and challenges of using VR in medicine. References were hand-searched to identify additional studies not covered by the literature search. Relevant information from each article was extracted, analyzed, and included in the review.

3. Results

Numerous relevant articles on the challenges around the use of VR in medicine were retrieved. In this review, we summarize our findings on VR under two themes: VR as a distraction method and VR in pain management.

3.1. Virtual Reality as a Distraction method

Patients typically interacted with different distraction scenarios based on gender and patient age. Distraction methods can be either a game-based software like "Snow World" [2] or an observation-based environment for the user to be relaxed like "Beach or Quit Jungle". The intent of distracting the patient is to have an illusion of going into a different environmental dimension and explore it while they are in their safe zone "next to their parents". The most successful method of using VR relies on the complete suspension of disbelief of being in an artificial/virtual environment so as to achieve a greater degree of presence and reaction to the stimulations. Prior to the initiation of the radiotherapy session, the patients are allowed some time with the VR headsets to adapt to the virtual environment while being in their safe zone environment, which helps in reducing their anxiety level. In general, the critical part is to convince the patients, especially pediatric patients, to wear the goggles before starting the radiotherapy session and then transfer them back to their safe zone without gaining the feeling that they actually went anywhere.

3.2. Virtual Reality for Pain-related Movement Patterns

Pediatric patients typically experience increased pain while moving the injured body part, which makes pain management incredibly challenging. The pediatric patients have difficulty accepting treatments like intrathecal chemotherapy procedure [8]. VR has the potential to eliminate that fear by distracting the patient from the pain. Intrathecal chemotherapy is a procedure where the physician will inject a drug into the spinal canal so that it directly reaches the cerebrospinal fluid [8]. Typically, the patients require multiple sessions of drug administration, for example, 3 times a month. Currently, 60% of the sessions include general anesthesia for patients under 10 years of age, which brings about behavioral changes and is generally not accepted by the patients. The inclusion of the VR technique, considered as an entertainment method, will generate trust and eliminate the fear and pain associated with intrathecal chemotherapy procedure. VR can also be used to modulate behavioral responses during the treatment process. For instance, during the process of locating the spinal canal to inject the drug, the physician will ask the patient to look at the long tree in the artificial environment and to take a deep breath to smell the fresh air. Thus, VR will help the physician in injecting the drug without any difficulties or movement from the patient. In addition, such a procedure will be beneficial for medical institutes in reducing the cost and time associated with the management of a general anesthesia team and for the patients in enhancing their quality of life [10].

4. Discussion and Conclusions

VR has been applied to different fields, such as dental procedures to increase patient trust, physical therapy to enhance patient recovery, and psychology to enhance the psychological behavior of the patient [11]. However, the medical procedures where VR has proven to be the most effective are radiotherapy and intrathecal procedures. In each case, numerous studies have shown the effectiveness of VR, either as a distraction method or pain management, in enhancing the quality of life of the patient during treatment process. In addition, many studies evaluating the advantages and disadvantages of using VR technologies in pediatric assessments and interventions have demonstrated them to be safe, repeatable, and diversifiable interventions that can be of immense help to healthcare providers. Moreover, VR reduces the anxiety level as well as the experience of aversive stimuli in children. Research on VR interventions for pain and anxiety level holds considerable promise in the medical field. A collaboration between researchers, clinicians, and programmers is crucial for the inclusion of VR technology in more clinical procedures, which would consequently enhance the patient's quality of life.

References

[1] Hoffman HG, Chambers GT, Meyer WJ, 3rd, Arceneaux LL, Russell WJ, Seibel EJ, Richards TL, Sharar SR, Patterson DR. Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. Ann Behav Med. 2011 Apr;41(2):183–91.

[2] Maani C, Hoffman HG, DeSocio PA, Morrow M, Galin C, Magula J, Maiers A, Gaylord K. Pain control during wound care for combat-related burn injuries using custom articulated arm mounted virtual reality goggles. Journal of CyberTherapy and Rehabilitation. 2008;1:193–198.

[3] Riener R, Harders M. Virtual Reality in Medicine. London: Springer; 2012:1–2.

[4] Cornick JE, Blascovich J. Are virtual environments the new frontier in obesity management? Soc Personal Psychol Compass. 2014;8:650–658.

[5] Kenny MP, Milling LS. The Effectiveness of Virtual Reality Distraction for Reducing Pain: A Meta-analysis. Psychology of Consciousness: Theory, Research and Practice. Advance online publication. 2016.

[6] Keefe FJ, Huling DA, Coggins MJ, Keefe DF, Zachary RM, Herr NR, Hoffman HG. Virtual reality for persistent pain: a new direction for behavioral pain management. Pain, 2012;153(11):2163–2166. https://doi.org/10.1016/j.pain.2012.05.030.

[7] Yap KY, Koh DW, Lee VS, Wong LL. Use of virtual reality in the supportive care management of paediatric patients with cancer. The Lancet Child & Adolescent Health. 2020 Dec 1;4(12):899-908.

[8] Hoffman HG, Patterson DR, Magula J, Carrougher GJ, Zeltzer K, Dagadakis S, Sharar SR. Water-friendly virtual reality pain control during wound care. J Clin Psychol. 2004;60:95.

[9] Pullen J, Boyett J, Shuster J, Crist W, Land V, Frankel L, Iyer R, Backstrom L, van Eys J, Harris M. Extended triple intrathecal chemotherapy trial for prevention of CNS relapse in good-risk and poor-risk patients with B-progenitor acute lymphoblastic leukemia: a Pediatric Oncology Group study. J Clin Oncol. 1993 May;11(5):839-49. doi: 10.1200/JCO.1993.11.5.839. PMID: 8487048.

[10] McQuay H, Moore A. An Evidence-Based Resource for Pain Relief. Oxford, UK: Oxford University; 1998.

[11] Pullen J, Boyett J, Shuster J, Crist W, Land V, Frankel L, Iyer R, Backstrom L, van Eys J, Harris M. Journal of Clinical Oncology 1993 11:5, 839-849.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210962

Electronic Health Records and Physician Burnout: A Scoping Review

Raghad MUHIYADDIN[a], Asma ELFADL[a], Ebtehag MOHAMED[a], Zubair SHAH[a], Tanvir ALAM[a], Alaa ABD-ALRAZAQ[a] and Mowafa HOUSEH[a,1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*

Abstract. This scoping review aims to identify the causes and consequences of physician burnout resulting from using Electronic Health Records (EHRs), as reported by current literature. We identified studies by searching PubMed, Wiley Online Library, and Google Scholar. Study selection and data extraction were conducted by three reviewers independently. Extracted data was then synthesized narratively. Out of 500 references retrieved, 30 studies met all eligibility criteria. We identified six main causes that lead to physician burnout related to the use of EHRs: EHRs' documentation and related tasks, EHRs' poor design, workload, overtime work, inbox alerts, and alert fatigue. We also identified the following consequences of physician burnout: low-quality care, behavioral issues, mental health complications, substance abuse, career dissatisfaction, costly turnover, and a decline in patient safety and satisfaction.

Keywords. Physicians, Electronic Health Records, Electronic Medical Records, burnout.

1. Introduction

The Electronic Health Record (EHR) is an advanced version of paper-based health records wherein a patient's information is collected and stored digitally [1]. EHRs are an essential component in the e-health domain, and the demand for interaction with electronic systems is increasing. Although using EHRs is associated with many advantages, they have now been identified to be one of the main reasons for physician burnout [2]. Burnout, which often occurs when physicians are faced with long-term stress, may cause depersonalization, emotional exhaustion, and the absence of a sense of self-achievement [3]. The impact of EHRs on physician burnout is an important topic, as it affects many physicians. This phenomenon and subsequently may endanger both their lives as well as the lives of the patients they are treating. Many studies examined how EHRs lead to physician burnout, but only a few reviews thoroughly covered all aspects of this topic. Specifically, the existing reviews mainly focused on physician burnout in general or gave an overview of the negative aspects of EHRs. Thus, the objective of this scoping review is to identify the causes and consequences of physician burnout as a result of using EHRs as reported by the current literature.

[1] Corresponding Author, Dr. Mowafa Househ, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar; Email: mhouseh@hbku.edu.qa.

2. Methods

This scoping review was conducted in line with the Joanna Briggs Institute guidelines. We searched three different databases: PubMed, Wiley Online Library, and Google Scholar. The study selection process consisted of three steps: removing the duplicates in the 500 retrieved studies, screening titles and abstracts of the studies, and reading the full text of the remaining studies (Appendix 1). We then extracted data into an Excel sheet, which was synthesized using a narrative approach. Then, all included papers were summarized into simplified text and tables.

3. Results

A total of 30 studies were included out of the 500 retrieved studies. The included studies reported six main causes of how EHRs lead to physician burnout. These causes were EHRs' documentation and related tasks, EHRs' poor design, workload, work overtime, inbox alerts, and alert fatigue. Eight studies addressed burnout factors stemming from indirect patient care tasks such as documentation, note-taking, and electronic communication in EHRs [4-11]. Nine studies found that EHRs' difficult use and poor design were associated with physician stress and burnout [6,12-19], where EHRs with a high number of functions are likely to have more stress on them than those with fewer functions [19]. Fourteen studies discussed how the increase in workload in EHRs increased burnout [4,10,15,18-28]. Four studies examined how physicians felt burned out due to the workload that results from using EHRs, which led to working overtime during evenings and weekends at home [8,13,25,29]. Six studies explored how inbox alerts from EHRs, including medication refills and test results, caused alert fatigue and eventual physician burnout [4,5,23,24,29,30].

Nine studies discussed the consequences of EHR burnout; namely, that physicians could not provide quality care, which ultimately results in global degradation of the quality of care [4,5,13,19,20,25,31-33]. Burnout affects physicians' behavior [31], including that some burned-out physicians may behave unprofessionally towards their colleagues [11,28]. Additionally, burnout affects the health of the physician [31].Physicians experiencing burnout are more likely to have mental problems, including depression, which can lead to suicide [25,28,33]. They are also more prone to substance abuse disorders [11,32]. A physician's burnout also affects patients' safety and satisfaction [31]. Patients treated by burned-out physicians were dissatisfied with their care, as they experience medical errors and unneeded assessments [11,28,30,31,33]. In general, burned-out physicians are not satisfied with their careers [31]. Lastly, burnout leads to costly turnover if the physician decides to leave as it is costly to replace them [11,30].

4. Discussion

This scoping review summarizes the studies that focus on physician burnout related to EHRs, as well as the causes and consequences of burnout. Two reviews discussed how the United States' physicians now spend more time on their computers than they do interacting with their patients [16,32,33]. Our findings align with previous reviews, which discuss that physicians are now communicating through portals about their

patients' concerns. They describe this as “doctoring while typing” clinical workflow, which leads to an overload of information and extra time spent on documentation; all worsening the physician’s satisfaction [33-36]. Another similar review also found that physicians suffering from alert fatigue and must ignore reminders that pop-up on their screen to combat this [37].

5. Conclusion

The paper’s cited reviews have found that EHRs are the leading cause of physician burnout. Many causes lead to physicians who use EHRs feeling burnt out, including as the necessary documentation and related tasks, poor design of EHRs, extra workload and working overtime, constant inbox alerts, and alert fatigue. There were also consequences to burnout such as low-quality care, physician dissatisfaction, turnover, mental health issues, increased substance abuse, behavioral problems, and patient dissatisfaction. We recommend researchers conduct this research to increase efforts in finding ways to relieve physician burnout. We advise them to identify the causes and consequences of burnout resulting from using EHRs by other health providers such as nurses, pharmacists, and technicians. Appendix files are available at Github: https://github.com/rmuhiyaddin/Appendix...git.

References

[1] HealthIT.gov. What Is an Electronic Health Record (EHR)?. 10 Sept. 2019. Available from: https://www.healthit.gov/faq/what-electronic-health-record-ehr.
[2] Beresford, Larry. Research Shows Link Between EHR and Physician Burnout. The Hospitalist.
[3] AHRQ. Physician Burnout. July 2017. Available from: https://www.ahrq.gov/prevention/clinician/ahrq-works/burnout/index.html.
[4] Tran, Brian, et al. Burnout and EHR Use among Academic Primary Care Physicians with Varied Clinical Workloads. AMIA Joint Summits on Translational Science Proceedings. AMIA Joint Summits on Translational Science, 2019;2019:136–144.
[5] Wang, Jason K., et al. Characterizing Electronic Health Record Usage Patterns of Inpatient Medicine Residents Using Event Log Data. PLOS ONE. 6 Feb. 2019;14(2):e0205379.
[6] Hauer A. et al. Physician Burnout in Wisconsin: An Alarming Trend Affecting Physician Wellness. WMJ : Official Publication of the State Medical Society of Wisconsin, 2018;117(5):194–200.
[7] Tutty, Michael A, et al. The Complex Case of EHRs: Examining the Factors Impacting the EHR User Experience. Journal of the American Medical Informatics Association. 2 Apr. 2019;26(7):673–677.
[8] Robertson SL, Robinson MD, Reid A. Electronic health record effects on work-life balance and burnout within the I3 population collaborative. Journal of graduate medical education. 2017 Aug;9(4):479-84.
[9] Chen L. et al. Racing Against the Clock: Internal Medicine Residents’ Time Spent On Electronic Health Records. Journal of Graduate Medical Education. Feb. 2016;8(1):39–44, 10.4300/jgme-d-15-00240.1.
[10] Reisman M. EHRs: The Challenge of Making Electronic Data Usable and Interoperable. P & T : A Peer-Reviewed Journal for Formulary Management. 2017;42(9):572–575.
[11] Gardner RL, et al. Physician Stress and Burnout: The Impact of Health Information Technology. Journal of the American Medical Informatics Association. 5 Dec. 2018;26(2):106–114, 10.1093/jamia/ocy145.
[12] Friedberg MW. et al. Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy. Rand Health Quarterly. 2014;3(4):1.
[13] Kroth PJ, et al. The Electronic Elephant in the Room: Physicians and the Electronic Health Record. JAMIA Open. 11 June 2018;1(1):49–56.
[14] Fred HL, Scheid MS. Physician burnout: causes, consequences, and (?) cures. Texas Heart Institute Journal. 2018 Aug;45(4):198.
[15] Domaney NM, et al. Exploring the Association Between Electronic Health Record Use and Burnout Among Psychiatry Residents and Faculty: A Pilot Survey Study. Academic Psychiatry. 21 May 2018;42(5):648–652.

[16] Alexander AG, Ballou KA. Work–life balance, burnout, and the electronic health record. The American journal of medicine. 2018 Aug 1;131(8):857-8.
[17] Kroth PJ, et al. Association of Electronic Health Record Design and Use Factors With Clinician Stress and Burnout. JAMA Network Open. 16 Aug. 2019;2(8):e199609.
[18] Heponiemi T, et al. Predictors of Physicians' Stress Related to Information Systems: A Nine-Year Follow-up Survey Study. BMC Health Services Research. 13 Apr. 2018;18(1).
[19] Babbott S, et al. Electronic Medical Records and Physician Stress in Primary Care: Results from the MEMO Study. Journal of the American Medical Informatics Association. Feb. 2014;12(e1):e100–e106. 10.1136/amiajnl-2013-001875.
[20] Arndt BG, et al. Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations. The Annals of Family Medicine. Sept. 2017;15(5):419–426. 10.1370/afm.2121.
[21] Robinson KE, Kersey JA. Novel electronic health record (EHR) education intervention in large healthcare organization improves quality, efficiency, time, and impact on burnout. Medicine. 2018 Sep;97(38).
[22] Roberts DL, Shanafelt TD, Dyrbye LN, West CP. A national comparison of burnout and work‐life balance among internal medicine hospitalists and outpatient general internists. Journal of hospital medicine. 2014 Mar;9(3):176-81.
[23] Gilleland M, Komis K, Chawla S, Fernandez S, Fishman M, Adams M. Resident duty hours in the outpatient electronic health record era: inaccuracies and implications. Journal of graduate medical education. 2014 Mar;6(1):151-4.
[24] Khairat S, et al. Focus Section on Health IT Usability: Perceived Burden of EHRs on Physicians at Different Stages of Their Career. Applied Clinical Informatics. Apr. 2018;9(02):336–347.
[25] Khairat S, et al. Physicians' Gender and Their Use of Electronic Health Records: Findings from a Mixed-Methods Usability Study. Journal of the American Medical Informatics Association 28 Aug. 2019;26(12). 10.1093/jamia/ocz126.
[26] Rassolian M, Peterson LE, Fang B, Knight HC, Peabody MR, Baxley EG, Mainous AG. Workplace factors associated with burnout of family physicians. JAMA internal medicine. 2017 Jul 1;177(7):1036-8.
[27] Privitera MR, et al. Physicians' Electronic Health Records Use at Home, Job Satisfaction, Job Stress and Burnout. Journal of Hospital Administration 20 June 2018;7(4): 52.
[28] Ripp JA, et al. Well-Being in Graduate Medical Education. Academic Medicine. July 2017:92(7):914–917.
[29] Russo E, et al. Electronic Health Record Alert-Related Workload as a Predictor of Burnout in Primary Care Providers. Applied Clinical Informatics. 2017;08(03):686–697.
[30] Meigs SL, Solomon M. Electronic health record use a bitter pill for many physicians. Perspectives in health information management. 2016;13(Winter).
[31] Marckini DN, et al. Electronic Health Record Associated Stress: A Survey Study of Adult Congenital Heart Disease Specialists. Congenital Heart Disease. Mar. 2019;14(3):356–361.
[32] Stehman C, et al. Burnout, Drop Out, Suicide: Physician Loss in Emergency Medicine, Part I. Western Journal of Emergency Medicine. 23 Apr. 2019;20(3):485–494.
[33] Kuhn CM, Flanagan EM. Self-Care as a Professional Imperative: Physician Burnout, Depression, and Suicide. Canadian Journal of Anesthesia/Journal Canadien d'anesthésie. 1 Dec. 2016;64(2):158–168. 10.1007/s12630-016-0781-0.
[34] Ratanawongsa N, et al. Reducing Misses and Near Misses Related to Multitasking on the Electronic Health Record: Observational Study and Qualitative Analysis. JMIR Human Factors. 6 Feb. 2018;5(1):e4.
[35] Bodenheimer T, Sinsky C. From Triple to Quadruple Aim: Care of the Patient R7quires Care of the Provider. The Annals of Family Medicine. 1 Nov. 2014;12(6):573–576. 10.1370/afm.1713.
[36] Houkes I, Winants Y, Twellaar M, Verdonk P. Development of burnout over time and the causal order of the three dimensions of burnout among male and female GPs. A three-wave panel study. BMC Public health. 2011 Dec;11(1):1-3.
[37] Backman R, et al. Clinical Reminder Alert Fatigue in Healthcare: A Systematic Literature Review Protocol Using Qualitative Evidence. Systematic Reviews. Dec. 2017;6(1). 10.1186/s13643-017-0627-z. Accessed 12 Dec. 2019.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210963

openEHR-to-FHIR: Converting openEHR Compositions to Fast Healthcare Interoperability Resources (FHIR) for the German Corona Consensus Dataset (GECCO)

Nektarios LADAS[a,1], Stefan FRANZ[a], Birger HAARBRANDT[a], Kim Katrin SOMMER[a], Severin KOHLER[b], Sarah BALLOUT[a], Johanna FIEBECK[c], Michael MARSCHOLLEK[a] and Matthias GIETZELT[a]
[a] *Peter L. Reichertz Institute for Medical Informatics, TU Braunschweig and Hannover Medical School, Germany*
[b] *Digital Health Charite, Berlin, Germany*
[c] *Center for Information Management, Hannover Medical School, Germany*

Abstract. The German Corona Consensus (GECCO) established a uniform dataset in FHIR format for exchanging and sharing interoperable COVID-19 patient specific data between health information systems (HIS) for universities. For sharing the COVID-19 information with other locations that use openEHR, the data are to be converted in FHIR format. In this paper, we introduce our solution through a web-tool named "*openEHR-to-FHIR*" that converts compositions from an openEHR repository and stores in their respective GECCO FHIR profiles. The tool provides a REST web service for ad hoc conversion of openEHR compositions to FHIR profiles.

Keywords. GECCO: German Corona Consensus Data Set, openEHR, FHIR, COVID-19, interoperability

1. Introduction

In Hannover Medical School, the GECCO dataset is stored in the openEHR format, which is modeled based on the GECCO FHIR profile [1, 2]. To exchange the data, it has to be converted to the FHIR R4 format.

In this paper, we introduce our solution through a tool named "*openEHR-to-FHIR*" for converting and sharing the COVID-19 patient information.

[1] Corresponding Author, Nektarios Ladas, Hannover Medical School, PLRI-OE 8420, Carl-Neuberg-Straße 1, 30625 Hannover, Germany; E-mail: ladas.nektarios@mh-hannover.de.

2. Methods

The data model is based on two main compositions from the openEHR Clinical Knowledge Manager templates [3, 4].

Each GECCO composition is mapped with a Java class that contains the JSON code equivalent to an FHIR resource based on the GECCO – Implementation Guide [5].

Through the application of regular expressions, each element in FHIR is matched with the corresponding element in composition flat format [6] of the openEHR.

3. Results

From a total of 4 compositions, data for a total of 30 compositions are selected and separated. The GECCO openEHR compositions are then converted to 66 datasets to match the definition in FHIR R4 format and stored on HAPI FHIR server.

The converted dataset is shared through the HiGHmed Data Sharing Framework for exchanging COVID-19 specific information between university hospitals in Germany [7].

4. Discussion

Our approach provides a solution for converting compositions from an openEHR repository to FHIR format directly. The manual mapping for generating FHIR format datasets helps to easily modify and update changes on the data structure.

This approach can be used for converting any openEHR data to FHIR data. The source code for openEHR-to-FHIR is available at https:// gitlab.plri.de/NektariosLadas/openehr-to-fhir.

References

[1] Sass J, Bartschke A, Lehne M, Essenwanger A, Rinaldi E, Rudolph S, Heitmann KU, Vehreschild JJ, von Kalle C, Thun S. The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. BMC Medical Informatics and Decision Making. 2020 Dec;20(1):1-7.

[2] German Corona Consensus Data Set - Implementation Guide [Internet]. Simplifier.net. 2021 [cited 20 August 2021]. Available from: https://simplifier.net/guide/GermanCoronaConsensusDataSet-ImplementationGuide/GECCOCore.

[3] Clinical Knowledge Manager 1 [Internet]. Ckm.highmed.org. 2021 [cited 19 August 2021]. Available from: https://ckm.highmed.org/ckm/templates/1246.169.1443

[4] Clinical Knowledge Manager 2 [Internet]. Ckm.highmed.org. 2021 [cited 19 August 2021]. Available from: https://ckm.highmed.org/ckm/templates/1246.169.1440

[5] GECCO - Implementation Guide [Internet]. Simplifier.net. 2021 [cited 25 August 2021]. Available from: https://simplifier.net/guide/germancoronaconsensusdataset-implementationguide/home

[6] Simplified Data Template (SDT) [Internet]. Specifications.openehr.org. 2021 [cited 25 August 2021]. Available from: https://specifications.openehr.org/releases/ITS-REST/latest/simplified_data_template.html

[7] Hund H, Wettstein R, Heidt CM, Fegeler C. Executing Distributed Healthcare and Research Processes–The HiGHmed Data Sharing Framework. InGerman Medical Data Sciences: Bringing Data to Life 2021 (pp. 126-133). IOS Press.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210964

Healthcare Professionals' Perceptions of Adoption of EHRs in Primary Care in Saudi Arabia

Bander Dowahi ALANAZI[a,1] and Bella NAPTSOK[b]
[a] *College of Health and Medicine, University of Tasmania, Launceston, Australia*
[b] *Department of Literature, Adyghe State University, Maykop City, Russia*

Abstract. EHRs are crucial for the provision of high-quality healthcare. However, their adoption and utilization are influenced by several factors, including users' perceptions. This study evaluated the perceptions of Saudi healthcare professionals towards the adoption of EHRs in primary care. All healthcare professionals working in primary care centers in Riyadh city were surveyed, resulting in a 65.9% response rate. Overall, the respondents had a positive perception of EHRs, as demonstrated by high agreement levels across all benefits and low with obstacles. There was also a positive relationship between perceived benefits and satisfaction with EHRs but negative with obstacles to adoption.

Keywords. electronic health records, healthcare professionals, perceptions

1. Introduction

Electronic health records (EHRs) provide several benefits that improve the quality of health care service provision [1]. These may include improving communication between healthcare providers and patients, clinical data documentation, and access to patient records [2, 3]. However, concerns with the system's limitations, including but not limited to increased risk of medical errors, complexity in use, and increased workload, also persist [4]. Users' perceptions of these factors can influence the acceptance, adoption, and use of EHRs [5, 6]. This study evaluates perceptions of healthcare professionals towards the adoption of EHRs in primary care in order to inform adoption in such settings.

2. Methods

All (1,710) healthcare professionals working in primary care centers in Riyadh city, Saudi Arabia, were surveyed to explore their perceptions regarding benefits and obstacles to adopting EHRs, as well as satisfaction with the systems. A 33-item survey tool adopted from a Turkish study [7] and validated for the Saudi context [8] was deployed online using REDCap to collect data between 11/30/2017 and 01/30/2018. Prior consent was obtained, and all the responses were anonymous.

[1] Corresponding Author, Bander Dowahi Alanazi, College of Health Science, University of Tasmania, Launceston, Australia; E-mail: bander12@hotmail.com.

3. Results

A total of 1127 participants (65.9%) completed the survey. The perceptions varied across the items for benefits, obstacles, and satisfaction with EHRs in primary care. The agreement with benefit statements ranged between 63.5% for an EHR 'reduces medical errors' and 77.1% for an EHR 'decreases paper-based documentation' [8]. Similarly, the agreement levels for satisfaction items were also high (65.8% to 78.8%). Conversely, the obstacle statements had low agreement levels except for an EHR system 'needs frequent revisions due to technological developments', with a 45.3% agreement. In addition, there was a strong positive association with perceived benefits (canonical correlation coefficient, r = 0.91) and medium association with perceived obstacles (r = 0.45).

4. Discussion

The results show that primary healthcare professionals in Saudi Arabia generally have a positive perception of EHRs, which is consistent with previous studies [9, 10]. The positive perceptions could be related to the benefits of EHRs, as demonstrated by the results of a positive association between perceptions of benefits and satisfaction with EHRs. These findings suggest that primary care professionals in Saudi Arabia are more likely to adopt EHRs mainly due to their perceived usefulness as opposed to perceived obstacles. However, technological challenges could remain a common barrier to adopting EHRs and must be adequately addressed to improve perception and adoption.

References

[1] Manca DP. Do electronic medical records improve quality of care?: Yes. Canadian Family Physician. 2015;61(10):846-7.
[2] Hoover R. Benefits of using an electronic health record. Nursing2018. 2016;46(7):21-2.
[3] Schopf TR, Nedrebø B, Hufthammer KO, Daphu IK, Lærum H. How well is the electronic health record supporting the clinical tasks of hospital physicians? A survey of physicians at three Norwegian hospitals. BMC health services research. 2019;19(1):1-9.
[4] Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. Risk management and healthcare policy. 2011;4:47.
[5] Kruse CS, Kristof C, Jones B, Mitchell E, Martinez A. Barriers to electronic health record adoption: a systematic literature review. Journal of medical systems. 2016;40(12):252.
[6] Garavand A, Mohseni M, Asadi H, Etemadi M, Moradi-Joo M, Moosavi A. Factors influencing the adoption of health information technologies: a systematic review. Electronic physician. 2016;8(8):2713-8. Epub 2016/10/21. doi: 10.19082/2713. PubMed PMID: 27757179; PubMed Central PMCID: PMCPmc5053450.
[7] Secginli S, Erdogan S, Monsen KA. Attitudes of health professionals towards electronic health records in primary health care settings: a questionnaire survey. Informatics for Health and Social Care. 2014;39(1):15-32.
[8] Alanazi B. Evaluating the healthcare professionals' perceptions about the adoption of electronic health records in primary care centres in Riyadh City, Saudi Arabia: University of Tasmania; 2020.
[9] Al-Harbi A. Healthcare providers' perceptions towards health information applications at King Abdul-Aziz Medical City, Saudi Arabia. International Journal of Advanced Computer Science and Applications. 2011;2(10):10-3.
[10] Shaker HA, Farooq MU, Dhafar KO. Physicians' perception about electronic medical record system in Makkah Region, Saudi Arabia. Avicenna journal of medicine. 2015;5(1):1-5. Epub 2015/01/28. doi: 10.4103/2231-0770.148499. PubMed PMID: 25625082; PubMed Central PMCID: PMCPmc4296390.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210965

Combining Machine Learning and Network Analysis Pipelines: The Case of Microbiome and Metabolomics Data in Colorectal Cancer

Eleni NTZIONI[a] and Ioanna CHOUVARDA[a,1]
[a] *Laboratory of Computing, Medical Informatics and Biomedical Imaging Technologies, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece*

Abstract. This study analyzes samples of intestinal microbiome and metabolites, from healthy individuals (HE) and patients with adenomas (AD) or colorectal carcinomas (CRC). A network analysis (NetAn) method was applied to the data, to identify the metabolites and microbial genera associated with the 3 classes and then 7 classification models were used. The models were initially trained with classic feature selection vs features resulting from NetAn. The distinction of HE and AD is successful, while CRC distinction presented lower success.

Keywords. Machine learning, network analysis, metabolomics, microbiome

1. Introduction

The primary aim of this study is to examine the correlations of specific metabolites and microbes with the occurrence of cancer via NetAn methods, so as to select informative features for a predictive model. Consequently, predictive models are explored using machine learning methods. The results obtained with a classic feature selection method are compared with the features selected by the NetAn method.

2. Methodology

The analysis of microbiome (MI) and metabolomic (ME) data in the present work is based on the study carried out by Kim et al. [1], which included 102 HE samples, 102 AD samples and 36 CRC samples, normalized and openly available. ME samples include data for 462 metabolites and MI samples include data for 70 genera. NetAn was performed with the R package NetCoMi [2], employing Pearson correlation after normalization and zero-handling. NetAn was applied to each of the 3 classes separately, and then network comparisons of the classes took place. Based on this process, 30 features were selected for MI and 65 for ME data. The machine learning methodology used for prediction is based on the work of Topçuoğlu et al. [3]. The open-source script

[1] Corresponding Author, Ioanna Chouvarda; E-mail: ioannach@auth.gr.

was adjusted as needed and 7 multi-class classification models were applied to the 2 datasets (MI/ME). Initially, the 7 models were applied to the complete dataset and permutation analysis was used for feature selection. In a second iteration, the 7 models were applied to the features selected via NetAn, and in a third iteration they were applied to the subset produced with permutation analysis. All methods were compared to determine whether the feature selection with NetAn produced better prediction results.

3. Results

The comparisons indicated that by a large majority, the models produced better results in the analysis of the subsets compared to the complete set of features. The distinction of HE and AD is successful (average values MI: HE~77%, AD~78% and ME: HE~80%, AD~81%), but the distinction of CRC presented lower success regardless of the model used, the dataset or the subset applied. In both the MI and ME datasets, the prediction using features from NetAn showed better results in distinguishing HE from AD or CRC.

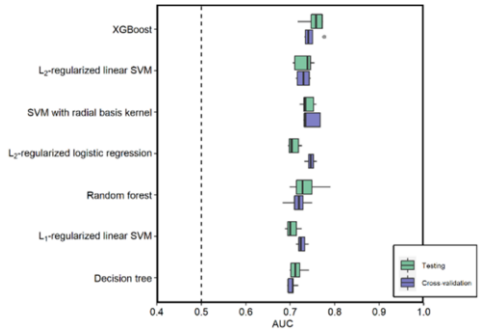


Figure 1. AUC values of the 7 models for the ME subset with the 30 features from NetAn.

4. Conclusions

The results show that the fewer samples of the carcinoma class had a strong negative effect on the performance of the models. For a set of samples where the number of samples will be equal to the other classes, we consider that the features selected with NetAn may lead to better predictions than a classic feature selection method.

Acknowledgement: This work has received funding from H2020 programme under grant agreement 952179.

References

[1] Kim M, Vogtmann E, Ahlquist D, Devens M, Kisiel J, Taylor W et al. Fecal Metabolomic Signatures in Colorectal Adenoma Patients Are Associated with Gut Microbiota and Early Events of Colorectal Cancer Pathogenesis. mBio. 2020;11(1).

[2] Peschel S, Müller C, von Mutius E, Boulesteix A, Depner M. NetCoMi: network construction and comparison for microbiome data in R. Briefings in Bioinformatics. 2020;22(4).

[3] Topçuoğlu B, Lesniak N, Ruffin M, Wiens J, Schloss P. A Framework for Effective Application of Machine Learning to Microbiome-Based Classification Problems. mBio. 2020;11(3).

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210966

Digital Entrepreneurial Nest: Supporting Digitization of Healthcare at National Level in Montenegro

Milovan ROGANOVIĆ[a,1], Ivana OGNJANOVIĆ[b], Ramo ŠENDELJ[b], Christoph REICH[c], Laszlo BOKOR[d], John MANTAS[e], Martina GOLOB[f], Nataša ŠIMŠIĆ[g], Irena OROVIĆ[h], Tanja RADUSINOVIĆ[i], Jevto ERAKOVIĆ[a], and Nataša ŽUGIĆ[j]
[a]*Clinic for Neurology, Clinical Centre of Montenegro, Montenegro*
[b]*University of Donja Gorica, Montenegro*
[c]*Hochschule Furtwangen University, Germany*
[d]*Budapest University of Technology and Economics, Hungary*
[e]*National and Kapodistrian University of Athens, Greece*
[f]*Arctur, Slovenia*
[g]*Comtrade System Integration, Serbia*
[h]*University of Montenegro*
[i]*Chamber of economy of Montenegro*
[j]*Ministry of Health, Montenegro*

Abstract. Health and health systems are not excluded from the influence of digitalization. In Montenegro, regarding the digitization process, when compared to other sectors, the health sector is lagging. In this poster presentation, we present an ambitious Erasmus+ DigN€ST project aimed on modernization of digitalization of healthcare system in Montenegro, as one of priority fields at national level.

Keywords. digitization of healthcare, eHealth, national level

1. Introduction

The strategic vision of development of Montenegro (MNE) is focused on increasing the economy's competitiveness. Montenegro is the only country in the Western Balkans that has adopted a Smart Specialisation Strategy. Amongst four priority areas that have been defined (energy and sustainable environment, sustainable agriculture and food value chain, sustainable and health tourism) ICT is a horizontal sector as it provides business and technological support to other priority sectors [1]. The aging of population, the rising burden of noncommunicable diseases, the influence of vested interests on behavioral risk factors and the sustainability of national health service are main problems in healthcare systems worldwide. Potentials of digital technologies should be used in the process of overcoming these problems [2].

[1] Corresponding author, Milovan Roganović, Clinic for Neurology, Clinical Centre of Montenegro, Montenegro, Ljubljanska bb, 81000 Podgorica, Montenegro; E-mail: milovan.roganovic@kccg.me.

2. Methods

Erasmus+ project Digital Entrepreneurial Nest and Industry 4.0 in Montenegro (DigN€st) (https://dignest.me/) aims to modernize higher education institutions (HEIs) in terms of industry digitization. With the complexity of such a process in mind, as well as current standards and best practices at the EU level, and clearly identified priorities and initiatives at the national level, we decided to take a step-by-step approach of 'learning from others' (via peer review processes), 'making adopted to own HE system and S3 Strategy' (via review and benchmarking), and final implementation and evaluation. All these steps are planned to be implemented in 3 dimensions by several coherent actions: (i) establishment of formal links between HEIs and business/industry at national level in MNE with strong connections to EU leaders in digitization; (ii) modernization of HEIs by development of practical training schemes; (ii) raising awareness, promotion of digitization process and digitized technologies. The proposed methodology is clearly rooted in the Digitising European Industry (DEI) Initiative (2016) [3], which identifies the establishment of Digital Innovation Hubs (DIHs) as a critical component for its implementation, with clear recommendations within DEI Working Group 1 - DIHs: Mainstreaming Digital Innovation Across All Sectors (in June 2017) identifying priority actions required to move the initiative forward.

3. Results

The main outcome of the proposed approach is a newly established MNE ACADEMIC DIH with an innovative topology and structure: it has two first layer nodes: the Digital Entrepreneurial Nest at the University of Donja Gorica and the Digital Support Centre at the University of Montenegro. Both HEIs implement a set of services titled Digital Business Support, while the following two sets of services are shared by both HEIs: Digital Education, as well as Innovation and Collaboration. Strong collaboration with relevant healthcare system institutions will maximize the use of existing capacities and resources at HEIs. This will lead to the development of high-level digital solutions that will improve the quality of the MNE health system.

4. Discussion

The services available through MNE ACADEMIC DIH will enable healthcare system to access the latest knowledge, expertise and technology for testing and experimenting with digital innovations relevant to its products, processes and services. Key characteristics of this model are the following: maximisation of available resources, cooperation between key stakeholders at national level, sustainable support to innovations and use of best practices and relevant experience at global level, from both academia and health industry. The model will be implemented and evaluated during funding period2021-2024.

References

[1] S3 in Montenegro [Internet]. S3. 2020 [cited 2021Aug27]. Available from: https://www.s3.me/en/s3-montenegro

[2] Odone A, Buttigieg S, Ricciardi W, Azzopardi-Muscat N, Staines A. Public health Digitalization in Europe. European Journal of Public Health. 2019;29(Supplement_3):28–35.
[3] The digitising European industry initiative in a nutshell [Internet]. The Digitising European Industry initiative in a nutshell | Shaping Europe's digital future. [cited 2021Aug27]. Available from: https://digital-strategy.ec.europa.eu/en/library/digitising-european-industry-initiative-nutshell

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210967

Co-Designing Smartphone Notifications According to the Clinical Routine of Cancer Patients

Panos BONOTIS[a,1], Fatima SCHERA[b], Christina KARAMANIDOU[a], Lydia SCARFO[c], Charalambos PONTIKOGLOU[d] and Christos MARAMIS[a]
[a] *Institute of Applied Biosciences, Centre for Research & Development Hellas, Thermi Thessaloniki, Greece*
[b] *Institute for Biomedical Engineering, Fraunhofer, Germany*
[c] *Universita Vita-Salute San Raffaele and IRCCS Ospedale San Raffaele, Milan, Italy*
[d] *Department of Hematology and Oncology, University Hospital of Heraklion, Greece*

Abstract. Although smartphone-based notifications offer a promising tool to support patient engagement and data collection via mobile health apps, attention must be given to the burden caused by frequent notifications and notification timing. This study presents a personalized mobile notification scheme, designed and developed to optimize reachability, and thus data collection from patients. Engineers, psychologists, oncologists, and patients were involved in various stages of a co-design approach and the presented implementation is currently used in the context of a clinical study.

Keywords. mHealth, palliative care, notification fatigue, eHealth, cancer

1. Introduction

Smartphone notifications are typically used to remind tasks and raise data collection prompts to end-users/patients. Considering the fact that users can receive in excess of 50 notifications per day from a variety of apps, and that different message formats (i.e. email, SMS, "push" notifications etc.) may have adverse effects on desired behavior compared to the use of just one message type, it is evident that notification delivery should be carefully designed [1] to reduce burden, a.k.a. notification fatigue. This becomes more clear considering the extensive research indicating that apps may be quickly discarded if notifications are perceived to be irritating or intrusive [2].

In this paper, we present a mobile notification scheme that was conceptualized, designed, and deployed as part of a mobile app used in the context of the MyPal project's[2] adult study [3] and is used to prompt patient input along the paradigm of electronic Patient Reported Outcomes (ePROs) for palliative care [4] and the delivery of personalized motivational messages.

[1] Corresponding Author, Panos Bonotis, eHealth Lab, INAB-CERTH, 6th Km. Charilaou-Thermi Road, Thermi, Thessaloniki, Greece; E-mail: pbonotis@certh.gr.

[2] The MyPal project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825872, https://mypal-project.eu/.

2. Methods & Results

In order to align ePRO notifications timing with the clinical routine of patients and reduce "notification fatigue", a participatory design approach was applied, organized in 3 distinct design phases: (a) a panel discussion and focus group session with stakeholders (healthcare professionals, patients) during which the notification scheme requirements were identified, (b) the initial design phase (2 engineers and a psychologist), where it was decided to merge each day's notifications (ePRO prompts, motivational messages) as one notification and the frequency of the notifications was defined as 4-5 per week, and (c) the definition of the in-day timeslots for the notifications to be presented. Considering the specific disposition of cancer patients during or after treatments, in order to define the in-day slots, oncologists were actively involved to identify patient communication patterns: interviews were performed, with chronic lymphocytic leukemia (CLL) and a myelodysplastic syndrome (MDS) oncologists to initially document the clinical routine of patients and these findings were confirmed via a survey by 6 independent physicians. The survey responses served in the design of a set of priority rules for presenting a notification in a day. To this end, a day was represented by a 24-size matrix (24 hours in a day) depicting the users' "quiet hours" (default 23:00-7:00) and "preferred notification time" (user-entered), also aligned with the patient's appointment and treatment schedule. The app calculates the optimum slot based on a decision tree defined by oncologists which was later also validated by patient advocates.

3. Conclusions

Currently, the presented notification scheme is operating as part of the MyPal adult study and is currently being evaluated in the context of the MyPal clinical trials.

Acknowledgments

The authors would like to thank Dr. Pantelis Natsiavas, the technical coordinator of the MyPal project for his contribution in the presented poster manuscript.

References

[1] M. Czerwinski, et. al, Challenges for designing notifications for affective computing systems, UbiComp 2016 Adjun. - Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. (2016) 1554–1559. doi:10.1145/2968219.2968548.

[2] L.G. Morrison, et. al, The Effect of Timing and Frequency of Push Notifications on Usage of a Smartphone-Based Stress Management Intervention: An Exploratory Trial, PLoS One. 12 (2017). doi:10.1371/JOURNAL.PONE.0169162.

[3] C. Karamanidou, et. al, Electronic Patient-Reported Outcome-Based Interventions for Palliative Cancer Care: A Systematic and Mapping Review., JCO Clin. Cancer Informatics. 4 (2020) 647–656. doi:10.1200/CCI.20.00015.

[4] C. Maramis, et. al, Using Electronic Patient Reported Outcomes to Foster Palliative Cancer Care: The MyPal Approach, in: Proc. - 2019 IEEE 19th Int. Conf. Bioinforma. Bioeng. BIBE 2019, Institute of Electrical and Electronics Engineers Inc., 2019: pp. 405–409. doi:10.1109/BIBE.2019.00079.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210968

Range-Measurement Sensor to Improve the Authentication Workflow for Users of a Hospital Information System: A Proof-of-Concept Study

Kei TERAMOTO[a,1], Shigeki KUWATA[b], Masashi HAMAMOTO[c] and Hiroshi KONDOH[a]
[a]*Division of Medical Informatics, Tottori University Hospital, Japan*
[b]*Department of Clinical Information Management, Nara City Hospital, Japan*
[c]*Ricoh Japan Co., Ltd*

Abstract. In this study, we developed an authentication rangefinder (AR) system for hospital information system (HIS) terminals to support the user authentication workflows. The logoff process of the AR system is triggered if no object is placed at least 90 cm in front of the HIS terminal laptop for ≥5 s. We conducted an anonymous survey of medical staff who used the AR system. 33/42(78%) respondents acknowledged an improvement in the logoff process. This study indicates that the AR system improves the user authentication workflow.

Keywords. Authentication system, hospital information system, medical staff

1. Introduction

Medical staff assigned to wards frequently log in and log off from hospital information system (HIS) terminals as they alternate between treating patients and referring to recording in patients' electronic medical records [1-3]. We surveyed the average number of logins per day made by doctors and nurses at Tottori University Hospital and found that doctors and nurses logged in 22 and 41 times, respectively. If the authentication process takes a long time, operational efficiency can be significantly reduced. If staff do not log out of terminals to save the time and effort of the HIS authentication process, not only is there an increased risk of data breaches resulting from shoulder surfing by third party applications and unauthorized logins but the operational efficiency of HIS terminals will also decrease. Therefore, simplifying login and logoff for HIS terminals is very important for the staff of medical institutions that have implemented HIS to improve workflow efficiency [4]. This study was aimed at developing an authentication rangefinder system (AR system) for to achieve an efficient HIS-terminal authentication workflow required by medical staff, and it was evaluated using a user survey.

[1] Corresponding Author, Kei Teramoto, Division of Medical Informatics, Tottori University Hospital, 36-1 Nishi-cho, Yonago, Tottori, 683-8504, Japan; E-mail: kei@tottori-u.ac.jp.

2. Method

We describe a method for authorizing login and logoff of HIS terminals by using an AR system. Login authentication using the AR system identifies users according to two elements: the BLE staff card worn by the user and a vein biometric authentication device. BLE staff authentication is equipped with a function that transmits a beacon signal to the staff ID cards, by which the Bluetooth functionality of the HIS terminal detects that a user is in the vicinity of a PC. The laser rangefinder module was detected when the medical staff are at least 90 cm away from the HIS terminal, and if ≥ 5 s have passed, the screen saver screen is activated. The screen saver was configured to be canceled if the user returns to the PC within 30 min. If a beacon signal cannot be detected for at least 30 min, the system will log off. The AR system was introduced into a general ward of Tottori University Hospital for one year. After that, the system was evaluated by the nurse using the paper-based questionnaire.

3. Results

For the login process using the AR system, 21 of 44 respondents acknowledged an improvement in the workflow; 16 were indifferent, while 7 responded that the workflow was worse. 5 of 7 answered that they experienced frequent login failures. 33 of 42 responded that the AR system improved the logoff process workflow. Negative feedback concerning operational improvements was that the screen saver did not disappear promptly and the distance required before the screen cleared felt long; this could have been caused by the values set for the rangefinder.

4. Discussion

The AR system was developed to support HIS authentication workflows and simplify the sequence of HIS terminal operations from login to logoff. In the login process, the accuracy of login authentication was an issue, with only 47.7% of respondents feeling that the workflow was improved. The logoff process could be further improved by adjusting the settings for the startup time of the screen saver and the laser rangefinder distance. The introduction of the AR system could help medical institutions improve the security of user authentication as well as the efficiency of their operations.

References

[1] Andersen P, Lindgaard AM, Prgomet M, Creswick N, Westbrook JI. Mobile and fixed computer use by doctors and nurses on hospital wards: multi-method study on the relationships between clinician role, clinical task, and device choice. J Med Internet Res. 2009;11(3):e32.

[2] Pearce C, Walker H, O'Shea C. A visual study of computers on doctors' desks. Inform Prim Care. 2008;16(2):111-7.

[3] Badger SL, Bosch RG, Toteja P. Rapid implementation of an electronic health record in an academic setting. J Healthc Inf Manag. 2005;19(2):34-40.

[4] Fontaine J, Zheng K, Van De Ven C, Li H, Hiner J, Mitchell K, et al. Evaluation of a proximity card authentication system for health care settings. Int J Med Inform. 2016; 92:1-7.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210969

Application Using Standard Communication Between Medical Facilities

Mihaela CRISAN-VIDA[a,1], Iulia GOLEA[a], Răzvan BOGDAN[a] and Lăcrămioara STOICU-TIVADAR[a]
[a] *Faculty of Automation and Computers, University Politehnica Timisoara, Romania*

Abstract. The web-based application described in this paper will support the patient to receive the treatment quicker and the physician to generate the prescription easier. The patient will have a real time information if their treatment/prescription is available in the pharmacy. Using a cloud solution will have all the information always available and without delays, the only requirement is the Internet connectivity. Using standardized communication as HL7 FHIR, the information exchanged is easier understanded by different medical units, and in the future other medical units can access the patient treatment/prescription and have a medical history, in this way the patient will receive better quality in treatment and health care.

Keywords. Web-based application, cloud computing, interoperability, HL7 FHIR

1. Introduction

In the current healthcare context one of the important things in digital health is interoperability. The interoperability in healthcare domain is defined as: "exchange of information between different entities and the received information to be understand by all the entities" [1]. This interoperability can be achieved by using a standardized communication, as HL7 FHIR (Fast Healthcare Interoperability). FHIR [2] is a standard for healthcare data exchange published by HL7 organization. FHIR specification describes a set of base resources, frameworks and APIs that can be use in different contexts in digital healthcare. For having the web-based application always available we provide a cloud computing solution, based in our case on Microsoft Azure [3] different services, hosting the web-based application and API for FHIR.

This paper presents the communication between the medical unit and pharmacies using HL7 FHIR. One of the main issues in Romania is that the medical unit generates a prescription with electronic support, but only for compensated or free prescription, and the physician gives it to the patient in paper format. The natural process isthe electronic support, where the medical entity sends in a standardized way the prescription to the pharmacy. After an analysis of applications on the market, as Dataklas Pharmaceutical [4], Pharmec [4], BizPharma [4] the conclusion was that they could not receive prescriptions in a standardized way.

[1] Corresponding Author, Mihaela Crișan-Vida, Faculty of Automation and Computers, University Politehnica Timisoara, Romania, Bd. Vasile Pârvan no. 2, Timisoara, Romania; E-mail: mihaela.vida@upt.ro.

2. Materials and methods

Figure 1 presents the system architecture where a medical unit communicates with the pharmacy in a standardized way. The cloud computing facility is useful ensuring to have the communication and the information always available.

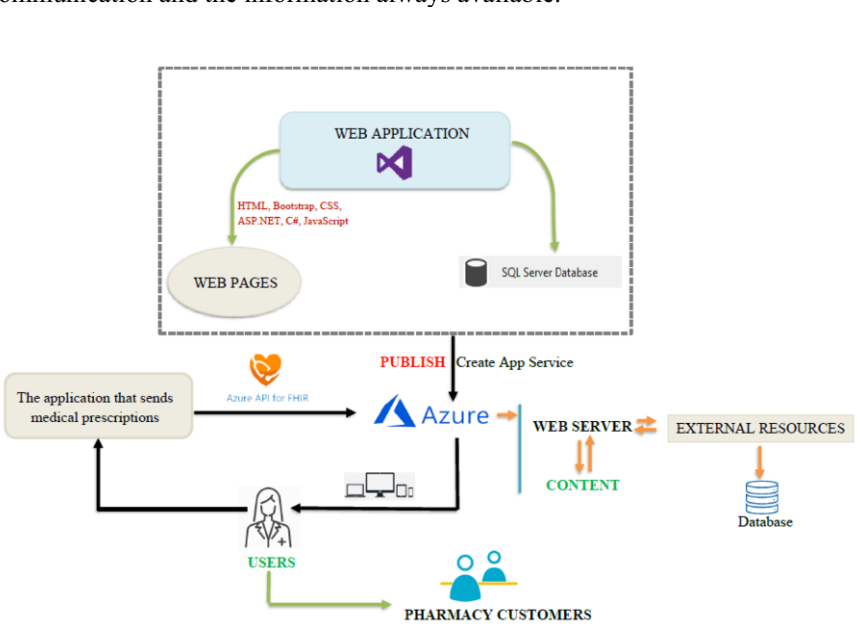


Figure 1. System architecture

The web-based application developed for pharmacies gives the possibility to receive prescriptions in the HL7 FHIR format. The web application supports prediction, and alerts if the medicine is not available in the pharmacy. It is developed using Visual Studio .NET 2017, using ASP.net pages and C# language. The standardized communication uses the Microsoft Azure cloud facilities, with an API facilitating the communication in the standardized way.

3. Conclusions

It is important to have all the medical information in real time; this can be achieved using cloud computing and standardized communication. This paper presents initial work for supporting a medical unit to send a prescription using a standardized communication, with HL7 FHIR standards.

References

[1] IEEE Standard Glossary of Software Engineering Terminology. IEEE Std 610.12-1990.
[2] HL7 FHIR standard. Available from: https://www.hl7.org/fhir/. Accessed on 07.09.2021.
[3] Microsoft Azure. Available from: https://azure.microsoft.com/. Accessed on 06.09.2021.
[4] Pharmacy applications. Available from: https://softlead.ro/aplicatii-software/farmacii. Accessed on 08.09.2021.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210970

Scientific Model Application "Smart Personalized Telecare Approach in Primary Healthcare"

Danae ALEXANDRI[a,1] and Maria TSIRINTANI[a,2]
[a] *University of West Attica, Athens, Greece*

Abstract. This research endeavor is based on the theoretical model of smart personalized telecare approach. As a research hypothesis, the model has the potential to develop utility during the mass crisis effect of Covid-19 pandemic. It will be tested for its applicability in two parallel dimensions for the utilization of the properties of the Delphi research method by tele-monitoring mild cases of Covid-19, because during the pandemic the meaning of holistic health has been shown as a supplementary good to every human activity and the technology of information and communication converts the healthcare into a generally sufficient service with the appropriate cost-management.

Keywords. Smart personalized telecare approach

1. Introduction

This research endeavor is based on an existing paper in which a theoretical model for smart personalized telecare approach and management indexes' sustainable enhancement was proposed. The primary qualitative research on the four (4) groups designated by the scientific bibliography (individual in need, caregivers – optionally, health official, information-communication και electronic systems administrator), has proven the applicability towards prevention or social protection through the dimensions of need / holistic health (physical, psychological, social and self efficiency level) and technological utility (health status notifications, accidents prevention, synchronized communication, (advising) decision-making systems [1]. Smart technology was featured as a powerful tool of personalized approach which functions as an evolution driver for the primary healthcare and the interdisciplinary service providing, that in turn result in the secondary healthcare decongestion via the hospital (re)admission indexes' improvement.

[1] Corresponding Author, Danae Alexandri, MSc in Health & Social Care Management, Dept of Business Administration, University of West Attica, Athens, Greece; E-mail: danah_alex@hotmail.com.

[2] Author: Maria Tsirintani, Professor, Health Information Systems, Dept of Business Administration, University of West Attica, Athens, Greece; E-mail: martsi@uniwa.gr.

2. Methods

As a research hypothesis, the administrative modelization of the interrelation of smart technology to the personalized primary healthcare has the potential to develop utility during the Covid-19 pandemic. It surpasses the obstacles of local (geographical, social) constraints by achieving interconnectiveness in real time [2]. The interface has a limitless amount of levels that range from the simple connection of systems to the interoperability via the holistic health care record [3]. The smart technology in healthcare and social care constitutes a dynamic factor in circumstances of crisis management [4]. The current theoretical model will be tested for its applicability in two parallel dimensions for the utilization of the properties of the Delphi research method [5]: care through tele-monitoring the symptomology of a typical quarantine or verified case (without the obligation, but having the option to resort to the secondary healthcare level if there is medical need) and smart technology amelioration progress through open-source software (for increasing the speed and validity). The population of the pilot model application will include mild cases of Covid-19, which characterize the accessibility needs coupled with the co-ordination for phenomena of mass crisis by competent authorities (National Public Health Organization and General Secretariat for Civil Protection).

3. Limitations

a. The launch of the application with simple data entry of bio-signals due to funding factors.
b. The initial focus on the private sector dynamics aiming at the for-profit restructuring of the system's provided services.

4. Discussion and Conclusions

During the pandemic, the meaning of holistic health has been shown as a supplementary good to every human activity. The technology of information and communication converts the healthcare into a purely public service, generally sufficient with the appropriate cost-management. The utilization of good practices increases the employability and thusly the social fare.

References

[1] Alexandri D, Tsirintani T. Smart Telecare Technology in Health and Social Primary Care Management for Personalized Approach in Greece, Informatics Empowers Healthcare Transformation, Studies in Health Technology & Informatics. 2017;238:177-180.
[2] Wang F, Docherty LS, Turner KJ, Kolberg M, Magill EH. Services & Policies for Care at Home, Pervasive Health Conference & Workshops IEEE. 2006;1-10.
[3] De Silva LC, Morikawa C, Petra IM. State of the Art of Smart Homes, Engineering Applications of Artificial Intelligence. 2012;25:1313-1321.
[4] Tsirintani M, Databases & Multimedia in Health – Distance Learning & Telemedicine, Paschalidis, Broken Hill, Publications, Athens, 2012.
[5] Ramalho R, Adiukwu F, et al. Telepsychiatry during the COVID-19 pandemic: development of a protocol for telemental health care. Frontiers in Psychiatry. 2020;11.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210971

Accreditation on Biomedical and Health Informatics in Europe: Guidelines and Procedures

Emmanouil ZOULIAS[a] and John MANTAS[a,1]
[a] *Health Informatics Laboratory, Faculty of Nursing, National and Kapodistrian University of Athens, Athens, Greece*

Abstract. Accreditation is the evaluation and monitoring process assuring that educational programs and institutions meet academic standards and operational integrity and quality. The number of programs in the field of Biomedical and Health Informatics is rapidly increasing. In this paper we briefly introduce the accreditation procedure that has been established by Accreditation Committee (AC2) of European Federation of Medical Informatics according to international standards.

Keywords. Biomedical and Health Informatics, accreditation, education

1. Introduction

Accreditation is a diligent evaluation and monitoring peer review process assuring that educational programs and institutions meet academic standards and operational integrity and quality [1]. There is a great number of European Universities and Institutions implementing and having established programs (undergraduate and postgraduate) in the field of Biomedical and Health Informatics [2]. A mechanism already applied by EFMI to accredit those programs, since EFMI is the European scientific body of Biomedical and Health Informatics, where all national associations have joined in EFMI. This accreditation provides European added value to the programs, be supportive of cross-national mobility and be complimentary to the required national accreditations processes [3]. The number of programs in the field of Health Informatics increased in Europe. On the other hand, a small number of Universities/Institutions with an accredited program in the field of Health Informatics reveals that accreditation activities have low effect in Europe.

EFMI is represented by the AC2 Committee which establishes an Accreditation and Certification process across. The aim of EFMI AC2 initiative is the promotion and provision awareness of the educational initiative to the wider biomedical and health informatics community in Europe.

[1] Corresponding Author, John Mantas, Health Informatics Laboratory, Faculty of Nursing, National and Kapodistrian University of Athens, Athens, Greece; E-mail: jmantas@nurs.uoa.gr.

2. Methods

The EFMI AC2 Accreditation assessment do not intend to replace or overcome any National or International accreditation or law. The accreditation process is robust and smooth succession of well-established steps and has already applied throughout Europe. The aim of EFMI AC2 Accreditation is to act as a complimentary judgement assisting and supporting to reach high quality academical educational programs in the field of Biomedical and Health Informatics.

The assessment procedure for successful accreditation examines five criteria. The criteria are based on the Dublin descriptors. Dublin Descriptors are uniform requirements for higher education under the Bologna Process acting as a system of qualifications frameworks for evaluating students for bachelor's, master's, doctoral degrees [4].

The EFMI AC2 Accreditation criteria are the following:

1. Needs and relevance
2. Intended learning outcomes
3. Academic/Teaching-learning environment
4. Organization and implementation
5. Internal quality assurance and development

The results of the assessment for each criterion are either satisfactory or unsatisfactory. In case that all five (5) criteria are satisfactory, according to the site-visit panel, the result of accreditation will be positive for the Universities/Institutions. The final decision of EFMI AC2 Committee is based on the evaluation report of site-visit panel. The positive decision of EFMI AC2 Committee gives the grant of EFMI Accreditation to the programme for three years. A re-accreditation process is required at the end of this period [5]-[8].

In case that only one (1) out of five (5) criteria measured as unsatisfactory then the programme receives a partial EFMI Accreditation for one year, until the program will reach this criterion. The application for accreditation will be dropped if more than two (2) criteria are judged as unsatisfactory. At least one year period of time is needed after a dropped application for a new judgement iteration.

3. Results

The AC2 by EFMI had successfully examined a wide variety of specializations including Biomedical Informatics, Health Informatics, Medical Informatics, Medical Technology, Bioinformatics, Biomedical Engineering, at undergraduate / postgraduate / doctoral / postdoctoral Studies. The information provided by an on-line catalogue[2], providing information about European programmes and courses of different Universities/Colleges/Institutions with 1000 BMHI programs and related specializations.

[2] http://efmi-ac2.bmhi-edu.org/

4. Conclusion

The first results by the EFMI AC2 procedure has already been initiated, and a number of programs have already been accredited. The first accreditation of Austrian UMIT received an award during MIE2021 in Athens. An accreditation process in BMHI needs to be based on updated educational recommendations by international organisations [9]-[12] in the field that will encompass all health care professions needs and requirements in clinical, nursing, and public health institutions, universities and health organisations.

References

[1] Mantas J. Accreditation and Certification in Health Informatics: Principles and Procedures. In ICIMTH 2017 Jan 1 (pp. 272-275).
[2] Kolokathi A, Hasman A, Chronaki C, Madsen I, Moen A, Randell R, Mantas J. Education in biomedical and health informatics: A European perspective. Studies in health technology and informatics. 2019;264:1951-2.
[3] Mantas J. Accreditation and Certification in Europe on Biomedical and Health Informatics as provided by EFMI. Bilten Hrvatskog društva za medicinsku informatiku (Online). 2021 Jul 11;27(2):1-8.
[4] Masaev SN, Dorrer GA, Vingert VV, Yakimova EA, Klochkov SV. Dublin Descriptors. In Journal of Physics: Conference Series 2020 Nov 1 (Vol. 1691, No. 1, p. 012021). IOP Publishing.
[5] Mantas J. The Accreditation Procedure Ensures Quality of Education in Biomedical and Health Informatics. Stud Health Technol Inform. 2020 Jun 26;272:484-486. doi: 10.3233/SHTI200601. PMID: 32604708.
[6] Mantas J. The European Educational Programs in Biomedical and Health Informatics. Stud Health Technol Inform. 2020 Jun 26;272:482-483. doi:10.3233/SHTI200600. PMID: 32604707.
[7] Mantas J. Accreditation and Certification in Health Informatics: Principles and Procedures. Stud Health Technol Inform. 2017;238:272-275. PMID: 28679942.
[8] Jaspers MW, Mantas J, Borycki E, Hasman A. IMIA Accreditation of Biomedical and Health Informatics Education: Current State and Future Directions. Yearb Med Inform. 2017 Aug;26(1):252-256. doi: 10.15265/IY-2017-011. Epub 2017 Sep 11. PMID: 28480478; PMCID: PMC6239229.
[9] Mantas J, Hasman A. IMIA Educational Recommendations and Nursing Informatics.Stud Health Technol Inform. 2017;232:20-30. PMID: 28106578.
[10] Hasman A, Mantas J. IMIA Accreditation of Health Informatics Programs. Healthc Inform Res. 2013 Sep;19(3):154-61. doi: 10.4258/hir.2013.19.3.154. Epub 2013 Sep 30. PMID: 24175114; PMCID: PMC3810522.
[11] Mantas J, Hasman A, Shortliffe EH. Assessment of the IMIA educational accreditation process. Stud Health Technol Inform. 2013;192:702-6. PMID:23920647.
[12] Mantas J, Ammenwerth E, Demiris G, Hasman A, Haux R, Hersh W, Hovenga E, Lun KC, Marin H, Martin-Sanchez F, Wright G; IMIA Recommendations on Education Task Force. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics. First Revision. Methods Inf Med. 2010 Jan 7;49(2):105-120. doi: 10.3414/ME5119. Epub ahead of print. PMID: 20054502.

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

Subject Index

This page intentionally left blank

Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

Author Index