HEALTHINFO 2019

The Fourth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing

ISBN: 978-1-61208-759-7

November 24 - 28, 2019

Valencia, Spain

HEALTHINFO 2019 Editors

Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, Germany

Shada Alsalamah, King Saud University, Saudi Arabia

HEALTHINFO 2019

Forward

The Fourth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing (HEALTHINFO 2019), held on November 24 - 28, 2019- Valencia, Spain, tackles with particular aspects belonging to health informatics systems, health information, health informatics data, health informatics technologies, clinical practice and training, and wellbeing informatics in terms of existing and needed solutions.

The progress in society and technology regarding the application of systems approaches information and data processing principles, modeling and information technology, computation and communications solutions led to a substantial improvement of problems in assistive healthcare, public health, and the everyday wellbeing. While achievements are tangible, open issues related to global acceptance, costs models, personalized services, record privacy, and real-time medical actions for citizens' wellbeing are still under scrutiny.

We take here the opportunity to warmly thank all the members of the HEALTHINFO 2019 technical program committee as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and efforts to contribute to the HEALTHINFO 2019. We truly believe that thanks to all these efforts, the final conference program consists of top quality contributions.

This event could also not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the HEALTHINFO 2019 organizing committee for their help in handling the logistics and for their work that is making this professional meeting a success.

We hope the HEALTHINFO 2019 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in health informatics research. We also hope Valencia provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

HEALTHINFO 2019 Steering Committee

Enrico Vicario, University of Florence, Italy
Wendy MacCaull, St Francis Xavier University, Canada
Chun-Hsi Huang, University of Connecticut, USA
Toshiyo Tamura, Waseda University | Future robotics organization, Japan
Saif Khairat, University of North Carolina - Chapel Hill, USA

HEALTHINFO 2019 Researsch/Academia Advisory Committee

Eleanor Horton, School of Nursing, Midwifery and Paramedicine | University of the Sunshine Coast, Australia
Doris Lindörfer, Ludwig-Maximilians-Universität München | IBE - Institut für medizinische

Informationsverarbeitung, Biometrie und Epidemiologie, Germany
David Greenhalgh, University of Strathclyde, UK

HEALTHINFO 2019 Publicity Chair
Jaime Lloret, Universitat Politecnica de Valencia, Spain

HEALTHINFO 2019

Committee

HEALTHINFO Steering Committee
Enrico Vicario, University of Florence, Italy
Wendy MacCaull, St Francis Xavier University, Canada
Chun-Hsi Huang, University of Connecticut, USA
Toshiyo Tamura, Waseda University | Future robotics organization, Japan
Saif Khairat, University of North Carolina - Chapel Hill, USA

HEALTHINFO Researsch/Academia Advisory Committee
Eleanor Horton, School of Nursing, Midwifery and Paramedicine | University of the Sunshine Coast, Australia
Doris Lindörfer, Ludwig-Maximilians-Universität München | IBE - Institut für medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Germany
David Greenhalgh, University of Strathclyde, UK

HEALTHINFO Publicity Chair
Jaime Lloret, Universitat Politecnica de Valencia, Spain

HEALTHINFO 2019 Technical Program Committee

Emmanuel Agu, Worcester Polytechnic Institute, USA
Jalel Akaichi, University of Tunis, Tunisia
Jesús B. Alonso Hernández, Institute for Technological Development and Innovation in Communications (IDeTIC) | University of Las Palmas de Gran Canaria (ULPGC), Spain
Jens Allmer, Hochschule Ruhr West, University of Applied Sciences, Medical Informatics and Bioinformatics, Germany
J. Antão B. Moura, Federal University of Campina Grande (UFCG), Brazil
Marjan Askari, Universiteit Utrecht, Netherlands
Panagiotis D. Bamidis, Aristotle University of Thessaloniki, Greece
Hugo Barbosa, Lusofona University of Porto / Faculty of Engineering of the University of Porto, Portugal
Rodolfo Barriviera, Instituto Federal do Paraná, Brazil
Kawtar Benghazi, University of Granada, Spain
Manuel Campos Martínez, University of Murcia, Spain
Chinmay Chakraborty, Birla Institute of Technology, Jharkhand, India
Po-Hsun Cheng, National Kaohsiung Normal University, Taiwan
Kevin Bretonnel Cohen, University of Colorado - School of Medicine, USA
Vassilis Cutsuridis, University of Lincoln, UK
Michael Czaplik, University Hospital RWTH Aachen, Germany
Arianna Dagliati, The University of Manchester, UK
Giuseppe De Pietro, Institute for High Performance Computing and Networking (ICAR) - Italian National Research Council (CNR), Italy / Temple University's College of Science and Technology, Philadelphia, USA
Steven A. Demurjian, The University of Connecticut, USA

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Argumentation Schemes for Clinical Interventions
Towards an Evidence-Aggregation System for Medical Recommendations

Olivia Sanchez-Graillet and Philipp Cimiano

Cognitive Interaction Technology Cluster of Excellence (CITEC)
Bielefeld University, Germany
Email: olivia.sanchez@uni-bielefeld.de

***Abstract*—The paradigm of evidence-based medicine requires that medical decisions are taken based on available, verified and high quality evidence. Such evidence has to be obtained from multiple relevant studies, considering their potential biases and shortcomings. Rationalizing and aggregating evidence from multiple studies is key to evidence-based decision making. Towards a system that is able to aggregate and summarize the evidence available in multiple studies, we have defined two argument schemes, which respectively provide reasons as to why a certain therapy may be regarded superior to another in terms of efficacy and safety. The argument schemes can be automatically instantiated via the semantic query language SPARQL from a knowledge base in which clinical studies have been formalized according to our own clinical trial ontology. The argument schemes are meant to be part of the framework of a configurable system that generates clinical recommendations by aggregating and summarizing the evidence from different clinical studies. We demonstrate the instantiation of the argument schemes in a study case on glaucoma and show that they are able to capture the reasoning behind determining such intervention superiority.**

Keywords–Argumentation in Medicine; Argumentation Schemes; Evidence-based Medicine; Summarization of Clinical Evidence

I. INTRODUCTION

Medical decisions are taken based on available, verified and high quality evidence. Randomized Clinical Trials (RCTs) are considered as the gold-standard for clinical research [1]. There are thousands of such controlled clinical studies and related publications available in open-access databases, such as PubMed [2]. Since clinicians are interested not only in obtaining effective intervention outcomes, but also in that the outcomes rely on high quality evidence, it is necessary to collect evidence from multiple relevant studies, whilst considering the possible presence of bias and the shortcomings of those studies [3]. The summarization and comparison of the aggregated information is normally done in the form of systematic reviews and meta-analyses.

Rationalizing and aggregating evidence from multiple clinical trials are crucial tasks for evidence-based decision making [4]–[7]. Criteria for grading the level of evidence have been already developed (e.g., The Grading of Recommendations Assessment, Development and Evaluation GRADE [8], [9]). However, applying such criteria to aggregate and summarize the evidence available in the vast number of relevant publications, requires an extensive manual effort. As a part of a system for generating medical recommendations based on clinical trial evidence, in this paper we present argument schemes that provide reasons as to why a certain therapy is regarded as superior to another – in terms of efficacy and safety – by aggregating the evidence found in multiple studies. The studies are formalized in a knowledge base structured according to our own Clinical Trial Ontology (C-TrO) [10]. The arguments are automatically instantiated from the knowledge base by extracting evidence via the SPARQL query language [11]. The argument schemes can be, in principle, used to rationalize the evidence for therapies of any health condition or disease. In this paper, we show the use of the schemes through a study case on glaucoma.

The remainder of this paper is structured as follows. Section II describes the proposed argumentation schemes and Section III the knowledge base defined on basis of C-TrO. The exemplary use of the schemes on glaucoma is presented in Section IV. Our conclusions and plans for future work are given in Section V.

II. ARGUMENT SCHEMES FOR AGGREGATING AND SUMMARIZING CLINICAL TRIAL EVIDENCE

In the definition of the argument schemes presented in this paper, we followed the model proposed by Walton et al. [12], in which argument schemes are seen as defeasible inference patterns that make explicit how a certain conclusion follows from a set of premises or assumptions. Thus, such arguments consist of a set of premises (assumptions), a conclusion, and critical questions that could invalidate the conclusion if satisfied. In our case, the conclusion of an argument corresponds to the claim that a certain therapy is superior to another given the evidence available in the form of multiple studies. Such type of arguments represent tools to systematically reason about the available evidence and thus support decision making. The arguments are defeasible reasoning patterns in the sense that the conclusion that one therapy is superior to another may be challenged and even invalidated by additional information (e.g., biases in the publications, lack of significance of size of effect, etc.).

Through the empirical analysis of published clinical trials of different types, and meta-analyses and systematic reviews on different health conditions, we identified the basic forms of argument schemes for inferring superiority of interventions in terms of efficacy and safety. The schemes can be applied in different configurations. For example, considering a given type of population, patient preconditions, country, etc. The different scenarios can be formed from the information contained in C-TrO. These schemes allow to structure available evidence as a basis to reason about the superiority of a certain treatment over another one. Each scheme states a major premise that is

 ISBN: 978-1-61208-759-7

Major premise: For people who suffer a given disease/health-disorder, it is desirable that a certain outcome indicator (or measurement) related to that disease/health-disorder changes, either increasing or decreasing.
Minor premise: It has been shown in a bigger number of clinical trials that $T1$ changes (either increasing or decreasing) a given disease/health-disorder indicator from the baseline in terms of an aggregation method in greater magnitude than $T2$.
Conclusion: $T1$ is a more effective medication treatment compared to $T2$ for changing the given disease/health-disorder indicator in the desired direction.

Critical Questions:
CQ1: Is the change (either increasing or decreasing) of the given disease/health-disorder indicator statistically significant (p-value)?
CQ2: Is the size of effect of $T1$ bigger than the one of $T2$?
CQ3: Are $T1$ and $T2$ applied to a similar number of patients across the different studies?

Figure 1. Scheme for superiority in terms of efficacy.

assumed to hold independently of the current level of evidence. Whilst the minor premise summarizes the current level of evidence as supporting the conclusion. The critical questions proposed pretend to challenge the validity of the conclusion based on the available information. The proposed argument schemes are described in what follows.

A. Argument Scheme for Superiority Based on Efficacy

In this argument scheme, the major premise expresses the general objective of the primary outcome of the intervention in question, and the minor premise considers the magnitude of the differences between the intervention results. The first critical question considers the statistical significance of the results; the second one refers to the size of the population that receives the intervention, which is important to consider since the p-value may vary according to this size; and the third one is about the absolute size of effect, i.e., the magnitude of the difference between groups. An intervention in which both size of effect and statistical significance are reported, tends to be more convincing than one in which only the size of effect is mentioned.

Figure 1 presents the corresponding argument scheme and its critical questions, where $T1$ and $T2$ are different drug treatments. In this scheme, it is implied that when there is a smaller number of clinical trials in which the outcome indicator in $T1$ changes in a bigger magnitude than in $T2$, the conclusion would be that $T1$ is less effective than $T2$. If the number of clinical trials is the same (or very similar), then $T1$ and $T2$ would be considered as being equally effective.

B. Argument for Superiority Based on Safety

In this argument scheme, the major premise expresses the general objective of the intervention outcome relative to safety. The minor premise considers the magnitude of the adverse effect that can be expressed in different ways, such as:

- Absolute magnitude, which refers to the number of people affected by a given adverse effect (e.g., "The most significant side effect of latanoprost was increased pigmentation of the iris which was observed in 15 patients").

Major premise: For people who suffer a given disease/health-disorder and who are under a medication treatment, it is desirable not to suffer any adverse effect.
Minor premise: It has been shown in a number of clinical trials that the administration of $T1$ leads to less incidence of adverse effects compared to the administration of $T2$.
Conclusion: Therefore, $T1$ is superior to $T2$ in terms of its safety profile, leading to less cases of the adverse effects.

Critical Questions:
CQ1: Are the adverse effects statistical significant?
CQ2: Are the size of effect of the adverse effects bigger for $T2$ than for $T1$?

Figure 2. Scheme for superiority in terms of safety.

- Relative magnitude, which refers to the percentage of people affected by a given adverse effect (e.g., "The most frequent drug adverse event was reported in 0.5% patients") .
- Uncertain magnitudes that denote uncertainty about the presence of an adverse effect (e.g., "The presence of an adverse effect was suspected").
- Modal words that indicate the degree of affection (e.g., "slightly affected") or expressions like "bigger degree (or amount)".

The corresponding argument scheme and its critical questions are presented in Figure 2. The critical questions are related to the statistical significance of the observed adverse effects.

C. Critical Questions Relative to the Quality of the Evidence

The following are the critical questions that apply to both schemes.

CQ3: How reliable is the evidence from these studies?

- ***CQ3.1*** Is there a risk of bias?
- ***CQ3.2*** Is the study randomized?
- ***CQ3.3*** Is the study blind?
- ***CQ3.4*** Is the study multi-center?
- ***CQ3.5*** Is the study intention-to-treat?

These critical questions are based on the following reasons:

- Intention-to-treat (ITT) studies are more realistic and unbiased than pre-protocol studies because they include all the patients in the results, while pre-protocol studies exclude patients who deviated from the protocol.
- Multi-center studies are more inclusive than single-center studies.
- Blind (or double-blind) studies are more objective than those of different type.
- There might be a risk of bias when a conflict of interest exists.

The context in which the argument schemes and critical questions are applied can be constrained by considering further

 ISBN: 978-1-61208-759-7

information. For example, the population's country, gender and age range. This evidence is available in the knowledge base, which is described in the next section.

III. The Knowledge Base and C-TrO

As part of the system for generating medical recommendations, we have developed the C-TrO ontology [10] that describes clinical studies with the adequate level of formalization and granularity for instantiating the proposed argument schemes and for defining different contexts of interest in which the argument schemes could be used.

The knowledge base follows the C-TrO structure and is described in Resource Description Framework (RDF) triples [13]. Below is an abbreviated example of the RDF triples corresponding to a clinical trial, one of its arms and one of the arm's interventions.

```
:CT_3 rdf:type :ClinicalTrial ;
   :hasObjectiveDescription "Latanoprost, a ..."
   :hasArm Arm_31, Arm_32 ;
   :hasPopulation CT3_Population ;
   :hasCTDesign DoubleBlind, Randomized .
:Arm_31 rdf:type ctro:Arm ;
   :hasNumberPatients 134 ;
   :hasIntervention :CT3_Intervention1 .
   :hasPrimaryOutcome :CT3_A1_OC1 ;
:CT3_Intervention1 rdf:type ctro:Intervention ;
   :hasFrequency "Once_at_evening";
   :hasInterval "Daily" ;
   :hasDuration "3 months";
   :hasMedication :CT3_I1_M1 .
```

IV. Study Case on Glaucoma

Glaucoma is a disease that damages the optic nerve and can lead to permanent visual loss. The damage of the optic nerve usually occurs when the Internal Ocular Pressure (IOP) increases. Therefore, the reduction of IOP is a desired outcome in an intervention for glaucoma.

For our study case, we formalized the clinical trials included in the meta-analysis carried out by Zhang et al. [14] which compares the efficacy and safety of the drugs timolol and latanoprost. The quality of the evidence was assessed by considering the design characteristics of the clinical trials such as masking, randomization, etc. The main outcome indicators studied were the percentage of IOP reduction for efficacy, and the relative risk for side effects (e.g., hyperaemia, conjunctivitis, etc.). The meta-analysis suggested that latanoprost was more effective than timolol in lowering IOP.

Figure 3 shows the instantiation of the argument scheme for the treatment superiority in terms of efficacy. This instantiation is carried out by executing a SPARQL query that represents the argument scheme with the information referent to glaucoma over the knowledge base. This query retrieves the appropriate evidence.

In this scheme, the major premise states the main objective of medical interventions for glaucoma, which refers to the greatest reduction of the diurnal IOP mean from baseline. The minor premise mentions that in the eleven studies compared, it was found that latanoprost interventions were more efficacious in reducing the diurnal IOP mean than timolol interventions. The table below this premise contains the evidence that supports this assertion (only five out of eleven studies are shown). Each row is a pairwise comparison of the interventions of each clinical trial. The first column contains the trial identifiers, the second column contains the publications that describe the clinical trials, and the remaining columns display the reduction of IOP mean (in $mmHG$ units) by the latanoprost and the timolol treatments, respectively.

Given this evidence, the conclusion in this argument scheme states that the latanoprost treatments are more effective than the timolol treatments under the specified conditions. The critical question is related to the statistical significance (p-value) of the IOP mean reduction from the baseline. Since the p-values of the interventions of some clinical trials were not reported, the conclusion of the argument can be weakened.

Major premise: For people who suffer glaucoma it is desirable that the *diurnal mean IOP* is reduced.
Minor premise: It has been shown in eleven clinical trials that *latanoprost* treatments reduced the *diurnal mean IOP* from baseline in a greater magnitude than the *timolol* treatments.

Evidence (Mean IOP Reduction (mmHg))			
CT_Id	**Reference**	**Latanoprost**	**Timolol**
CT_1	Alm A et al,1995	7.8	6.7
CT_2	Aquino MV et al.,1999	11.1	9.1
CT_3	Camras CB et al.,1996	6.7	4.9
CT_4	Diestelhorst M et al.,1998	4.9	2.1
CT_5	Mastropasqua L et al,1999	4.8	4.6

Conclusion: *latanoprost* treatment is a more effective medication treatment compared to *timolol* treatment for reducing the diurnal mean IOP.

CQ1: Is the reduction of the diurnal mean IOP statistically significant? Only some p-values were reported.

Figure 3. Instantiation of the scheme for efficacy.

The argument schema in terms of safety for glaucoma is instantiated as shown in Figure 4.

Major premise: For people who suffer glaucoma and who are under a medication treatment it is desirable not to suffer any adverse effect.
Minor premise: It has been shown in eleven clinical trials that the administration of the *timolol* treatment leads to less incidence of *Conjunctival_hyperemia* than the *latanoprost* treatment.

Latanoprost		Timolol	
Adverse Effect	**No.**	**Adverse Effect**	**No.**
ConjunctivalHyperemia	7	ConjunctivalHyperemia	2
IncreasedPigmentation	2	ReducedHeartRate	2
IrisPigmentationChange	1	ReducedBloodPreasure	2
		ChangeBloodVelocity	1
		Smarting	1
		IrisPigmentationChange	1

Conclusion: The *timolol* treatment is superior to the *latanoprost* treatment in terms of its safety profile, leading to less cases of the adverse effect *ConjunctivalHyperemia*.

Figure 4. Instantiation of the scheme for safety.

The major premise states that the ideal outcome of the medical interventions is that they do not cause any adverse effect. The minor premise states that across the eleven studies it was found that the presence of *conjunctival hyperemia* occurred more times in the latanoprost interventions than in

 ISBN: 978-1-61208-759-7

the timolol interventions. The table below this premise contains the evidence that supports that assertion (i.e. seven mentions in the case of latanoprost and only two mentions in the case of timolol).

Given this evidence, the conclusion in this scheme states that the timolol treatment is safer than the latanoprost treatment relative to conjunctival hyperemia. This schema can be applied to any other adverse effect of interest. In this study case, we have analyzed glaucoma and only the effect of two drugs. However, the scheme can consider any other disease and drugs that are contained in the knowledge base.

V. CONCLUSIONS

We have presented basic argument schemes for deciding the superiority of medical interventions in terms of efficacy and safety based on the aggregation of clinical trial evidence. From these schemes, more specific argument schemes could be derived. A study case on glaucoma was used as a proof-of-concept of our argument schemes, showing the feasibility of their use.

Other approaches that rely on argument schemes are those for identifying argument schemes in genetic research articles [15] and in letters for genetic counselling [16]. More recently, Mayer et al. [17] presented a method for recognizing evidence in the form of premises and conclusions in RCT abstracts, and Mayer et al. [18] developed a method for classifying the type of evidence found in clinical trials. Furthermore, several argumentation-based approaches for medical and health-care decision support have been proposed [19]–[23]. Hunter and Williams [20] developed an argument-based framework to aggregate clinical trial evidence, in which inductive arguments are generated from a set of evidence. Afterwards, the superiority of the interventions is determined according to given preference criteria across the generated arguments. In contrast to this approach, we have defined argumentation schemes supported by an ontology to rationalize the decision as to which intervention is superior, as well as to aggregate the corresponding level of evidence. Furthermore, our knowledge base is populated with evidence extracted from published clinical trials, while in the Hunter and William's method, the input evidence is taken from partially aggregated and synthesized information contained in medical guidelines.

As future work, we plan to validate our argument schemes via user studies. We want to extend our approach so that efficacy and safety are not considered as independent dimensions, but that they can be weighted against each other in the form of a meta-argument with an internal structure that resolves the trade-off depending on the relative weight given to safety over efficacy or vice versa. Finally, we are currently developing an information extraction system that automatically encodes the study results and clinical evidence in a knowledge base following the C-TrO ontology.

ACKNOWLEDGMENT

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG) within the project Rationalizing Recommendations (RecomRatio), as part of the Priority Program "Robust Argumentation Machines (RATIO)" (SPP-1999).

REFERENCES

[1] A. Akobeng, "Understanding randomised controlled trials," Archives of Disease in Childhood, vol. 90, 2005, pp. 840–844.

[2] "PubMed", URL: https://www.ncbi.nlm.nih.gov/pubmed [Last access: August, 2019].

[3] D. L. Sackett, W. M. C. Rosenberg, J. A. M. Gray, R. B. Haynes, and W. S. Richardson, "Evidence based medicine: what it is and what it isn't," BMJ, vol. 312, no. 7023, 1996, pp. 71–72.

[4] G. Tsafnat, A. Dunn, P. Glasziou, and E. Coiera, "The automation of systematic reviews," BMJ, vol. 346, 2013, p. f139.

[5] A. M. OConnor, G. Tsafnat, S. B. Gilbert, K. A. Thayer, and M. S. Wolfe, "Moving toward the automation of the systematic review process: A summary of discussions at the second meeting of international collaboration for the automation of systematic reviews (ICASR)," Systematic Reviews, vol. 7, no. 1, 2018, p. 3.

[6] M. H. Murad and V. M. Montori, "Synthesizing evidence: Shifting the focus from individual studies to the body of evidence," JAMA, vol. 309, no. 21, 2013, pp. 2217–2218.

[7] H. Bastian, P. Glasziou, and I. Chalmers, "Seventy-five trials and eleven systematic reviews a day: How will we ever keep up?" PLoS Medicine, vol. 7, no. 9, 2010, p. e1000326.

[8] H. Balshem et al., "GRADE Guidelines: 3. Rating the quality of evidence," J Clin Epidemiol, vol. 64, no. 4, 2011, pp. 401–406.

[9] G. H. Guyatt et al., "What is quality of evidence and why is it important to clinicians?" BMJ, vol. 336, no. 7651, 2008, pp. 995–998.

[10] O. Sanchez-Graillet, P. Cimiano, C. Witte, and B. Ell, "C-TrO: An ontology for summarization and aggregation of the level of evidence in clinical trials," in The Joint Ontology Workshop (JOWO) 2019: Ontologies and Data in the Life Sciences (To appear).

[11] "The SPARQL Query Language", URL: https://www.w3.org/TR/rdf-sparql-query [Last access: August, 2019].

[12] D. Walton, C. Reed, and F. Macagno, Argumentation Schemes. Cambridge University Press, 2008.

[13] "The RDF Working Group", URL: https://www.w3.org/RDF/ [Last access: August, 2019].

[14] W. Y. Zhang, A. L. W. Po, H. S. Dua, and A. Azuara-Blanco, "Meta-analysis of randomised controlled trials comparing latanoprost with timolol in the treatment of patients with open angle glaucoma or ocular hypertension," Br J Ophthalmol., vol. 85, no. 8, 2001, pp. 983–990.

[15] N. Green, "Towards creation of a corpus for argumentation mining the biomedical genetics research literature," in Proc. of the 1st Workshop on Argumentation Mining, 2014, pp. 11–18.

[16] N. L. Green, "Annotating evidence-based argumentation in biomedical text," in IEEE Intl. Conf. on Bioinformatics and Biomedicine (BIBM), 2015, pp. 922–929.

[17] T. Mayer, E. Cabrio, M. Lippi, P. Torroni, and S. Villata, "Argument mining on clinical trials," in Proc. of COMMA 2018, vol. 305. IOS Press, 2018, pp. 137–148.

[18] T. Mayer, E. Cabrio, and S. Villata, "Evidence type classification in randomized controlled trials," in Proc. of the 5th Workshop on Argumentation Mining, 2018.

[19] N. Gorogiannis, A. Hunter, and M. Williams, "An argument-based approach to reasoning with clinical knowledge," Intl. J. of Approx. Reasoning, vol. 51, no. 1, 2009, pp. 1–22.

[20] A. Hunter and M. Williams, "Aggregating evidence about the positive and negative effects of treatments," Art. Int. in Medicine, vol. 56, no. 3, 2012, pp. 173–190.

[21] R. Craven, F. Toni, C. Cadar, A. Hadad, and M. Williams, "Efficient argumentation for medical decision-making," in Proc. of the 13th Intl. Conf. on Principles of Knowledge Representation and Reasoning (KR), 2012, pp. 598–602.

[22] L. Longo and L. Hederman, "Argumentation theory for decision support in health-care: a comparison with machine learning," in Proc. of the Intl. Conf. on Brain and Health Informatics, 2013, pp. 168–180.

[23] M. A. Grando, L. Moss, D. Sleeman, and J. Kinsella, "Argumentation-logic for creating and explaining medical hypotheses," Art. Int. in Medicine, vol. 58, no. 1, 2013, pp. 1–13.

 ISBN: 978-1-61208-759-7

Predicting the Chances of Live Birth for Couples Undergoing *In Vitro* Treatments Using Decision Trees

Bruna Estácio da Veiga[1], Fernando Ferreira[2], José Manuel Fonseca[3]
UNINOVA
Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa
Caparica, Portugal
[1] email: bf.veiga@campus.fct.unl.pt
[2] email: flf@fct.unl.pt [3] email: jmf@uninova.pt

José Luis Metello
CIRMA – Centro de Infertilidade e Reprodução Medicamente Assistida
Hospital Garcia de Orta E.P.E.
Almada, Portugal
email: jmetello@gmail.com

***Abstract* —In developed countries, the prevalence of infertility ranges from 3.5% to 16.7%. There are several factors that affect the success rate of *in vitro* treatments and so every couple has a singular probability of success which can be predicted. As these treatments are complex and expensive with a variable probability of success, the most common question asked by *in vitro* fertilization patients is ''What are my chances of conceiving?". Classical statistics and artificial intelligence models have been published in the literature. So far, artificial intelligent prediction models are not aimed at live birth but rather at pregnancy and use undergoing treatment features. The main aim of this study is to develop a classification tree model that estimates the chance of a live birth before *In Vitro* Fertilization (IVF) treatments. This decision tree might result in a new clinical support system that helps physicians to deal with the couple's expectations.**

Keywords-artificial intelligence; decision tree; machine learning; in vitro; infertility.

I. INTRODUCTION

Artificial Intelligence (AI) is a promising tool for a wide range of applications in Medicine. AI is a "branch of computer science that attempts to both understand and build intelligent entities, often instantiated as software programs" [1]. Clinical Decision Support Systems (CDSS) have been developed to help enhance patient care and improve clinical outcome prediction. The first generation of AI systems relied on clinical knowledge and the computation of vigorous decision rules. Nowadays, AI in healthcare has reached machine learning techniques that can rely on complex interactions [2]. When the CDSS are trained with classified data and recognize patterns in those data, they constitute a type of IA named supervised learning. CDSS should be integrative and understandable, which are attributes of decision trees. Although CDSS are very helpful, these systems produce outputs that respond to the question "what?". However, the responsibility to know "why" belongs to the physicians [3].

One of the clinical areas where CDSS could be of great support to physicians is in predicting the output of in vitro treatments. According to the study performed by Boivin et al. [4], the prevalence of infertility ranged from 3.5% to 16.7% in developed countries. Based on these authors' estimates, 72.4 million women are currently infertile, and, of these, 40.5 million are currently seeking infertility medical care [4]. For Portugal, most recent data estimated that 9.8% of couples are infertile [5].

Infertility is defined as a clinical condition "characterized by the failure to establish a clinical pregnancy after 12 months of regular, unprotected sexual intercourse or due to an impairment of a person's capacity to reproduce either as an individual or with his/her partner" [6].

Nowadays, the two most recurrent Medically Assisted Reproduction (MAR) techniques are *In Vitro* Fertilization (IVF) and its subtype Intracytoplasmatic Sperm Injection (ICSI). According to the European Society of Human Reproduction and Embryology, the number of MAR cycles between 1997 and 2014 increased by 13%, reaching 776556 cycles in Europe in 2015 [7]. Although the probability of having a live birth is low with IVF, there are more and more people using this type of treatment. In fact, this probability may increase by up to 30% to 40% with MAR techniques [8]. In Portugal, the last report of MAR shows that the treatment success rate varies between 25-30% [9]. MAR treatments entail issues, such as stress, anxiety, or depression due to time and money involved [10]–[14]. Furthermore, these treatments may result in medical complications, such as ovarian hyperstimulation syndrome or premature births [15].

Pondering complications and benefits, the couple may decide if it is their will to proceed or not with treatment. In that way, infertile couples usually ask clinicians about their chances of conceiving.

There are several factors involved in predicting the output of *in vitro t*reatments. Female's age and infertility diagnosis are usually the main factors physicians take into account [16]. However, hormone doses, physiological factors, and sperm quality are couple' characteristics that interfere with the estimate of the success probability [17].

As the chances of in vitro treatment success depend on various factors and those factors are often correlated, there are many models in the literature that try to predict infertility treatment output for couples undergoing IVF-ICSI [8][16][26][18]–[25].

Traditional statistic models, such as binary logistic regression, are very common in *in vitro* treatment's outcome prediction. The first predictive model ever built in this

 ISBN: 978-1-61208-759-7

context is from Templeton et al. in 1996 using a logistic regression model to predict the probability of live birth for an individual woman using the woman's age, number of previous live birth or pregnancies not resulting in a live birth and whether these were a result of previous IVF treatments, female causes of infertility, duration of infertility and the number of previous unsuccessful IVF treatments [16]. Since then, many authors used logistic regression to predict the chances of live birth for couples undergoing IVF-ICSI[18]–[20][22][27]. The majority of the developments resided in adding new variables, such as serum Anti-Müllerian Hormone (AMH) value [19], Body Mass Index (BMI), or ethnicity [18]. Although logistic regression is the most recurrent approach to binary output problems and usually has high discriminatory ability, a major limitation is a need for the dataset features to be independent of each other [28].

Data mining techniques are another possible approach to this predicting problem. In 1998, Jurisica et al. proposed a model based on k-nearest neighbors classifiers [23]. They used a case-based reasoning system that exploits past experiences to suggest possible modifications to an IVF treatment plan in order to improve overall success rates. They built an interactive system for physicians that uses both pre-treatment features and also ongoing treatment features. This model's accuracy was 60.6%. In 2011, Guh et al. [24] developed a hybrid intelligence method that integrates genetic algorithm and decision learning techniques for knowledge mining. Their study counted on 70 different attributes (before and after treatment features) and had an accuracy of 72,3%. A complete study from Güvenir et al. [25] in 2015 compared the RIMARC (Ranking Instances by Maximizing the Area under ROC curve) algorithm with naive Bayes classifier and random forest and concluded that RIMARC has a potential to be used successfully to estimate the probability of success in medical treatment. RIMARC algorithm is based on pre-treatment features, and their output feature was clinical pregnancy. RIMARC's accuracy was 84.4%. Hafiz et al. [8] published in 2017 a comparative study between five classifiers based on before and after treatment features: support vector machines, adaptive boosting (Adaboost), recursive partitioning (RPART), random forests (RF) and one nearest neighbour (1NN) and concluded that RPART and RF had the highest values of Area Under ROC Curve (AUC) (0,82 and 0,84, respectively). In 2016 Milewska et al. [26] used classification trees to obtain a group of patients characterized most likely to get pregnant while using *in vitro f*ertilization. However, they used undergoing treatment features, such as quality of oocytes obtained by the stimulation. For the training group, the area under ROC (Receiver Operating Characteristic) curve (AUC) was 0.75–0.76, while for the validation group, it was from 0.66 to 0.68. Trimarchi et al. [29] findings obtained from more traditional statistical approaches seem to validate the results obtained by the data mining techniques both in terms of accuracy and number of variables considered. More recently, Tran et al. [30] created a deep learning model named IVY, which was an objective and fully automated system that predicts the probability of fetal heart pregnancy directly from raw time-lapse videos without the need for any manual morphokinetic annotation or blastocyst morphology assessment. They achieved an AUC of 0.93 [95% CI 0.92–0.94] in 5-fold stratified cross-validation.

In data mining of clinical data, it is wanted to provide models understandable to humans once it is imperative that physicians understand the conclusions, and that can explain it to their patients [3]. In other words, decision trees are graphical, allowing easy visualization and computationally translate the typical human reasoning in which the process is of eliminating hypotheses corresponding to the tests performed on each node. Therefore, the decision tree learning technique can be more useful in this context once it can create a model in terms of intuitively transparent *if-then* rules [31].

This study aims to develop, for a Portuguese hospital, a validated decision tree model that estimates the chances of live birth on couples before they start their IVF non-donor cycle on pre-treatment. The work presented in this short paper is only part of the whole study done. The ultimate goal is to use the most accurate model for the development of a clinical support interface. Other models based on logistic regressions and bayesian classifiers were developed. In addition, further data are pending collection for validation.

This paper is structured in introduction, methods, results, discussion, and conclusion. After presenting the problem in the introduction section, the chosen method to build the decision tree will be described, in the next section. The results section contains the tree model, which is analysed in discussion section. At the end of this paper, there is a conclusion section that summarizes the key points of this paper.

II. METHODS

This was a retrospective study of the data from 39 cycles. The cycles were performed between 2012 and 2016 in the Centro de Infertilidade e Reprodução Medicamente Assistida (CIRMA) at Hospital Garcia de Orta, E.P.E., Almada, Portugal.

After approval from the Hospital's Ethics Committee for Health, it was considered as a primary outcome the existence of live birth (at least one baby was born alive and survived for more than 1 month). Pregnancy was not considered as an output because the aim of this study was to predict the success of the treatment which is not complete measurable with pregnancy due to miscarriages.

In terms of the baseline characteristics used to develop the Portuguese model and taking into account the literature in this area [16][32][33], we used the following features: woman's and man's age (years), duration of infertility (months), cause of infertility (categorised as diagnosis of tubal, endometriosis, disovulation, male factor, both female and male factor, multiple female factors, unexplained or other), woman's and man's BMI (Kg/m2), serum anti-Müllerian hormone (AMH) (ng/mL), Antral Follicle Count (AFC) (number of follicles), woman's and man's ethnicity (Asian, Caucasian, Gipsy, Indian, Black or Mixed), woman's and man's smoking status (never, previous, present) and

 ISBN: 978-1-61208-759-7

woman's and man's previous live births (yes or no). This was the complete set of features available on the database.

Data was pre-processed aiming to find missing values and for that reason, two cases were not considered for that model. In other words, the model was constructed with data from 737 couples.

To build the decision tree, we used the Salford Predictive Modeler's CART® (Classification and Regression Trees) modeling software from Minitab Statistical Software. The CART methodology was developed in the '80s by Leo Breiman, Jerome Friedman, R.A. Olshen and Charles Stone and was first presented in their paper from 1984 [34]. The CART modeling engine, Salford Predictive Modeler's implementation of CART, is the only decision tree software embodying the original proprietary code. Their method allows the construction of binary decision trees and so CART only asks yes/no questions [35]. Binary trees can be specifically applied in this context because the output is binary (with or without live birth), adjusting better to the form of human reasoning.

CART analysis generates simple and practical clinical decision rules. Every value of each variable is considered as a potential split (parent nodes), and the CART method divides the selected range of variables to obtain an optimal binary split into two subgroups (child nodes)[34]. From this, CART analysis generated an optimal classification tree (minimal cost tree), and numerical rank for each input used to build the tree by relative importance. The software took into account that the database is unbalanced once option DATA was chosen in the priors of model setup. Gini impurity criterion was adopted as the node splitting rule. CART performed ten-fold cross-validation, and AUC was used to evaluate the accuracy of the model and compare it to the logistic regression model.

III. RESULTS

The CART analysis showed that the best discriminators for classification were AFC, AMH, female's age, infertility cause, and female's BMI. Figure 1 shows the optimal tree generated by this analysis, which means the tree with minimal cost generated by CART software. The tree has six splits and produces seven terminal nodes. Only two terminal nodes result in live birth class: node 3 and node 6. Node 3 includes the couples with values of AFC>10.50, female's age≤35.5 years and disovulation as a cause of infertility. Node 6 represents the couples with AFC between 10.50 and 26.50, female's age≤35.5 years, AMH>1.58 ng/mL, female's BMI ≤ 25.50 Kg/m^2 and with a cause of infertility which is not disovulation. The prevalence of live birth in these two nodes were 76.5% and 56.1%, respectively. Terminal nodes 1,2,4,5 and 7, obtained from the optimal tree, are determined as a group of couples without a live birth.

Table I shows the confusion matrix (also known as an error matrix) on the test set, allowing visualization of the performance of the model in terms of correct/incorrect classified cases. Table II reports the evaluation metrics on the test set. The AUC test for the discriminatory ability of the final prediction model is 0.68021 on train set and 0.59621 on the test set.

TABLE I. DECISION TREE EVALUATION METRICS ON TEST SET

Actual class	***Total***	***Correct Percentage***	***Predicted class***	
			No live birth N = 416	Live birth N = 321
No live Birth	506	61.46%	311	195
Live Birth	231	54.55%	105	126

TABLE II. DECISION TREE EVALUATION METRICS ON TEST SET

Metrics	***Value***
Specificity	61.46%
Sensibilility	54.55%
Precision	39.25%
F1 score	45.65%
Accuracy	59.29%

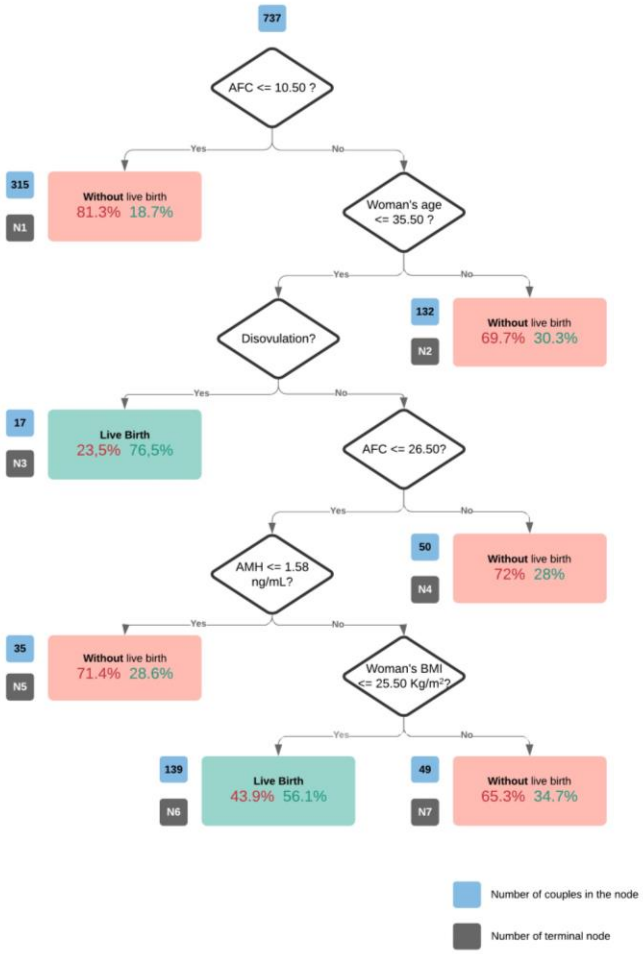

Figure 1. The CART live birth prediction model for couples undergoing in vitro fertilization.

In Figure 1, blue tags indicate the number of couples in the node; Grey tags indicates the number of the terminal node (N); The red percentage indicates the percentage of couples in the node without live birth; The green percentage indicates the percentage of couples in the node with live birth.

In the next section, the results obtained will be discussed.

IV. DISCUSSION

To our knowledge, this is the first decision tree reported in the literature for live birth IVF-ICSI prediction that accounts only on pre-treatment data from couples. The created tree provides the possibility of defining groups of couples for whom the probability of live birth is very small or very high. The overall rate of at least one live birth was 31.4% *a priori* and this tree returns a range of probabilities of success (green percentages represented in Figure 1) from 18.7% to 76.5%. This is the optimal tree computed by CART, and the features selected as better splitters were mainly woman's characteristics. The only man's characteristic that integrates this tree appears in "disovulation?" question which divides couples according to infertility cause.

The initial *if* question "AFC≤10.50?" splits the initial sample of 737 couples in two groups according to levels of AFC: AFC lower or equal to 10.50 classifies couples automatically with "No live birth" (Node 1) once only 18.7% of couples in that condition achieve a live birth. This result corroborates older literature findings in which women with low values of AFC had more difficulty to have success in IVF-ICSI [18][36].

Going down the tree, the group of women with a value of AFC higher than 10.50 are split according to their age: if the female's age is higher than 35.50 years then the tree classifies that group with "No live birth" (Node 2) because 69.7% of couples from our database did not achieve live birth. This result also agrees with literature because of the loss of fertility with aging in women, mainly due to the decrease of gamete's quantity and quality over the years [16][33][37].

Continuing down the tree, couples with younger women (female's age lower than 35.50 years) are parcelled out according to infertility cause. Couples in which women have disovulation are classified with "Live Birth" once 76.5% of couples in the Node 3 conditions accomplished a live birth. The tree continues with the remaining causes of infertility (tubal, endometriosis, male factor, both female and male factor, multiple female factors, unexplained or other). Once IVF/ICSI has higher success rates on male factor [38][39] and disovulation, we expected that the male factor was also a tree splitter.

Next, AFC values split the couples with all of the infertility causes mentioned before, which were not disovulation. Thus, AFC is used two times as a splitter in this tree, thus becoming the most important splitter according to CART software scoring variables output. Node 4 classifies women with AFC higher than 26.50 follicles with "No live birth". This was not expected according to literature [18][36]. However, this result shows that AFC is not sufficient as an ovarian reserve marker and that is why AMH appears as next splitter.

Node 5 classifies women with values of AMH serum lower than 1.58 ng/mL with "No live birth" since 71.4% of 35 women in that node did not reach the live birth. This result about AMH reinforces La Marca [19] results in which probabilities of live birth are higher on women with AMH 0.4ng/mL-<2.8 ng/mL and even higher on the group with AMH≥2.8 ng/mL. The value 1.58 ng/mL is in the middle category of La Marca study.

Continuing down the tree, women with higher values of AMH are split according to their BMI. Again, following the previous conditions of the tree, if the female's BMI is higher than 25.50 Kg/m2, then the prevalence of no live birth is 65.3%. In that way, it is better to have a value of female's BMI lower than 25.50 Kg/m^2 (taking into account all of the previous conditions that lead to Nodes 6 and 7). According to Adolphe Quetelet scale of BMI used worldwide [40], a BMI higher than 25 Kg/m^2 is related to pre-obesity and lower than that value is associated with normal weight. Various studies showed that overweight women have ovulatory problems and increased risks of abortion [41][42].

To our knowledge, since this is the first decision tree in this context with only pre-treatment features and with live birth as output, we do not have any direct possible comparison with other models especially in terms of variables included in the final model and AUC values. However, we can indirectly compare with models referred in introduction section, namely RIMARK, RF, RPART and 1NN and observe that they have much higher values of accuracy. Milewska *et al*'s model [26] with decision trees also has higher accuracy than our model because they used under treatment features and their output was "pregnancy" and not live birth.

An advantage of decision trees is that they are easily readable and understandable. Using such a proposed approach, it becomes easy for doctors to explain to couples their situation following the decision tree until the terminal node. On the other hand, classic logistic regressions result from computation coefficients that, most likely, will be far from intuitive to explain to all patients.

In future work, we plan to investigate the "without live birth" nodes in order to further analyze the causes of *in vitro* treatments unsuccess. Furthermore, we intend to collect more data to validate our model, possibly improve accuracy, and explore other artificial intelligence algorithms with deep learning approach.

V. CONCLUSION

In this study, clinical and lifestyle factors of 737 infertile couples were used to create a classification decision tree. This tree incorporates the five optimal features to provide a probability of live birth due to IVF-ICSI: AFC, female's age, AMH, infertility cause, and female's BMI. Decision trees building allows that the variables might be dependent and so this method gives successful results in terms of

evaluating these variables together and bringing up relations between variables. Furthermore, decision trees are intuitive and easier to explain to patients.

As we said before, the work presented in this short paper is only part of the whole study done. The ultimate goal is to use the most accurate model for the development of a clinical support interface. Also, further data are pending collection for validation. In that way, this decision tree might result in a new clinical support system that helps physicians to deal with the couple's expectations.

ACKNOWLEDGMENTS

We thank the Portuguese Foundation for Science and Technology FCT/ MCTES - UID/EEA/00066/2019 (JF).

Portions of the information contained in this publication are printed with permission of Minitab, LLC. All such material remains the exclusive property and copyright of Minitab, LLC. All rights reserved.

REFERENCES

[1] K.-H. Yu, A. L. Beam, and I. S. Kohane, "Artificial intelligence in healthcare," Nat. Biomed. Eng., vol. 2, no. 10, pp. 719–731, Oct. 2018.

[2] R. C. Deo, "Machine Learning in Medicine," PMC Author Manuscr., vol. 132, no. 20, pp. 1920–1930, 2018.

[3] Lobo, L, "Artificial Intelligence and Medicine," Artif. Intell. Med., vol. 69, no. 2, pp. 1–19, 1979.

[4] J. Boivin, L. Bunting, J. A.Collins, and K. G.Nygren, "International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care," Hum. Reprod., vol. 22, no. 10, pp. 1506–1512, 2007.

[5] J. L. Silva-Carvalho and A. Santos, "Estudo Afrodite: Caracterização da infertilidade em Portugal (Vol. 1. Estudo na Comunidade)," Porto Fac. Med. da Univ. do …, p. 74, 2009.

[6] F. Zegers-Hochschild et al., "The international glossary on infertility and fertility care, 2017," Hum. Reprod., vol. 32, no. 9, pp. 1786–1801, 2017.

[7] A. N. Andersen et al., "Assisted reproductive technology and intrauterine inseminations in Europe, 2005: Results generated from European registers by ESHRE," Hum. Reprod., vol. 24, no. 6, pp. 1267–1287, 2009.

[8] B. N. J. Pegah Hafiz, Mohtaram Nematollahi, and Reza Boostani, "Predicting implantation outcome of in vitro fertilization and intracytoplasmic sperm injection using data mining techniques," Int. J. Fertil. Steril., vol. 11, no. 3, pp. 184–190, 2017.

[9] Conselho Nacional de Procriação Medicamente Assistida, "Relatório da Atividade Desenvolvida pelos Centros de PMA em 2015," vol. 2015, p. 27, 2015.

[10] "Portuguese Fertility Association." [Online]. Available: http://www.apfertilidade.org/. [retrieved: September, 2019].

[11] N. Karaca et al., "Effect of IVF failure on quality of life and emotional status in infertile couples," Eur. J. Obstet. Gynecol. Reprod. Biol., vol. 206, pp. 158–163, 2016.

[12] S. Maroufizadeh, E. Karimi, S. Vesali, and R. Omani Samani, "Anxiety and depression after failure of assisted reproductive treatment among patients experiencing infertility," Int. J. Gynecol. Obstet., vol. 130, no. 3, 2015.

[13] V. Valoriani et al., "Differences in psychophysical well-being and signs of depression in couples undergoing their first consultation for assisted reproduction technology (ART): an Italian pilot study," Eur. J. Obstet. Gynecol. Reprod. Biol., vol. 197, pp. 179–185, Feb. 2016.

[14] M. Heredia, J. M. Tenías, R. Rocio, F. Amparo, M. A. Calleja, and J. C. Valenzuela, "Quality of life and predictive factors in patients undergoing assisted reproduction techniques," Eur. J. Obstet. Gynecol. Reprod. Biol., vol. 167, no. 2, pp. 176–180, 2013.

[15] R. Paulson, "In Vitro Fertilization," UpToDate, pp. 379–386, 2018.

[16] A. Templeton, J. K. Morris, and W. Parslow, "Factors that affect outcome of in-vitro fertilisation treatment," Lancet, vol. 348, no. 9039, pp. 1402–1406, 1996.

[17] M. M. A. Alireza Zarinara, Hojjat Zeraati, Koorosh Kamali, Kazem Mohammad, and Parisa Shahnazari, "Models predicting success of infertility treatment: A systematic review.," J. Reprod. Infertil., vol. 17, no. 2, pp. 68–81, 2016.

[18] R. K. Dhillon et al., "Predicting the chance of live birth for women undergoing IVF: A novel pretreatment counselling tool," Hum. Reprod., vol. 31, no. 1, pp. 84–92, 2016.

[19] A. La Marca et al., "Anti-Müllerian hormone-based prediction model for a live birth in assisted reproduction," Reprod. Biomed. Online, vol. 22, no. 4, pp. 341–349, 2011.

[20] S. M. Nelson and D. A. Lawlor, "Predicting live birth, preterm delivery, and low birth weight in infants born from in vitro fertilisation: A prospective study of 144,018 treatment cycles," PLoS Med., vol. 8, no. 1, 2011.

[21] E. R. Te Velde et al., "Comparison of two models predicting IVF success; The effect of time trends on model performance," Hum. Reprod., vol. 29, no. 1, pp. 57–64, 2014.

[22] J. A. Leijdekkers et al., "Predicting the cumulative chance of live birth over multiple complete cycles of in vitro fertilization: an external validation study," Hum. Reprod., vol. 62, no. August, pp. 1–12, 2018.

[23] I. Jurisica, J. Mylopoulos, J. Glasgow, H. Shapiro, and R. F. Casper, "Case-based reasoning in IVF: Prediction and knowledge mining," Artif. Intell. Med., vol. 12, no. 1, pp. 1–24, 1998.

[24] R. S. Guh, T. C. J. Wu, and S. P. Weng, "Integrating genetic algorithm and decision tree learning for assistance in predicting in vitro fertilization outcomes," Expert Syst. Appl., vol. 38, no. 4, pp. 4437–4449, 2011.

[25] H. A. Güvenir, G. Misirli, S. Dilbaz, O. Ozdegirmenci, B. Demir, and B. Dilbaz, "Estimating the chance of success in IVF treatment using a ranking algorithm," Med. Biol. Eng. Comput., vol. 53, no. 9, pp. 911–920, 2015.

[26] A. J. Milewska, D. Jankowska, and U. Cwalina, "Prediction of Infertility Treatment Outcomes Using Classification Trees," Stud. LOGIC, Gramm. Rhetor., vol. 47, no. 60, pp. 7–19, 2016.

[27] D. J. McLernon, E. W. Steyerberg, E. R. Te Velde, A. J. Lee, and S. Bhattacharya, "Predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation: Population based study of linked cycle data from 113 873 women," BMJ, vol. 355, 2016.

[28] E. Vittinghoff, D. Glidden, S. Shiboski, and C. McCulloch, Regression Methods in Biostatistics: Linear, Logistic, Survival and Repeated Measures Models. Springer, 2005.

[29] L. G. J. R. Trimarchi, J. Goodside, L. Passmore, T. Silberstein, L. Hamel, "Comparing data mining and logistic regression for predicting IVF outcome," Abstr. 59th Annu. Meet. Am. Soc. Reprod. Med., vol. 80, 2003.

[30] D. Tran, S. Cooke, P. J. Illingworth, and D. K. Gardner, "Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer," Hum. Reprod., vol. 34, no. 6, pp. 1011–1018, 2019.

[31] J. Han, M. Kamber, and J. Pei, Data mining: concepts and techniques, 3rd ed. Morgan Kaufmann Publishers is, 2012.

 ISBN: 978-1-61208-759-7

[32] A. Zarinara, H. Zeraati, K. Kamali, K. Mohammad, and M. M. A. Parisa Shahnazari, "Models predicting success of infertility treatment: A systematic review.," J. Reprod. Infertil., vol. 17, no. 2, pp. 68–81, 2016.

[33] M. Vander Borght and C. Wyns, "Fertility and infertility: Definition and epidemiology," Clin. Biochem., vol. 62, no. February, pp. 2–10, 2018.

[34] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, CART: Classification and Regression Trees. Wadsworth, Belmont.

[35] R. Timofeev, "Classification and Regression Trees (CART) Theory and Applications," Humboldt University, Berlin, 2004.

[36] K. Keane, V. F. Cruzat, S. Wagle, N. Chaudhary, P. Newsholme, and J. Yovich, "Specific ranges of anti-Mullerian hormone and antral follicle count correlate to provide a prognostic indicator for IVF outcome," Reprod. Biol., vol. 17, no. 1, pp. 51–59, 2017.

[37] A. Templeton, "Infertility and the establishment of pregnancy--overview," Br. Med. Bull., vol. 56, no. 3, pp. 577–587, 2000.

[38] A. C. Van Steirteghem et al., "Higher success rate by intracytoplasmic sperm injection than by subzonal insemination . Report of a second series of 300 consecutive treatment cycles," Hum. Reprod., vol. 8, no. 7, pp. 1055–1060, 1993.

[39] M. Alikani, M. Sc, A. Adler, B. Sc, and Z. Rosenwaks, "Intracytoplasmic sperm injection : a novel treatment for all forms of male factor infertility," Fertil. Steril., vol. 63, no. 6, pp. 1231–1240, 1995.

[40] World Health Organization, "Body mass index - BMI," 2019. [Online]. Available: http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi. [retrieved: September, 2019].

[41] D. Best and S. Bhattacharya, "Obesity and fertility," Horm. Mol. Biol. Clin. Investig., vol. 24, no. 1, pp. 1–6, 2015.

[42] R. J. Norman, "Impact of obesity on female fertility and fertility treatment," Women's Heal., vol. 4, no. Future Medicine, pp. 183–194, 2008.

 ISBN: 978-1-61208-759-7

A Blockchain Architecture for the Italian EHR System

Mario Ciampi, Angelo Esposito, Fabrizio Marangio, Giovanni Schmid, Mario Sicuranza
Institute for High Performance Computing and Networking
National Research Council of Italy
Via Pietro Castellino, 111 – 80131 Naples, Italy
e-mail: {mario.ciampi, angelo.esposito, fabrizio.marangio, giovanni.schmid, mario.sicuranza}@icar.cnr.it

***Abstract*—In the last years, significant changes of important socio-economic indicators, like population growth, life expectancy increase and patient mobility, have implied the need to provide new models for health provision. Thus, several efforts have been done to adequate and evolve the current e-health systems for enabling them to gather patient health data produced by health facilities in an interoperable way and according to shared business processes. However, even if such systems are now starting at collecting health data, it is still not possible to verify that all the tasks of a specific process are correctly executed. This work presents a permissioned blockchain architecture designed to manage the Electronic Health Records of the users, able to track the operations performed by the actors involved in a health process. The architecture proposed is compliant with both the Italian Regulation on Electronic Health Record and the recently introduced GDPR. A proof-of-concept of the architecture has been developed and validated against a relevant use case.**

Keywords-EHR; blockchain; patient-centric; architecture.

I. INTRODUCTION

An increasingly important problem for the well-being of modern societies is to have efficient, reliable and scalable health support systems. This is necessary to provide adequate healthcare – in the medium and long term – to populations whose lifetime expectation tends to increase constantly, but whose individuals often do not have a satisfactory health state, especially during their old age. Realizing these systems is an essential condition for containing public spending and the sustainability of national health systems. Indeed, they can be used to prevent health diseases, through the lifestyle monitoring of people and the use of innovative and non-invasive therapies based on precision medicine. In the attempt to achieve this goal, huge efforts are underway in EU countries to digitize health processes for increasing usability and reliability for patients and healthcare personnel, allowing for a reduction in time and costs. The areas in which improvements can and must be achieved are still many, and the margins of enhancement allowed by emerging technologies like permissioned blockchains for the secure and transparent processing of distributed workflows can be really substantial, such as to revolutionize prevention and treatment approaches. Indeed, current systems are rooted on data producers (e.g., hospitals and healthcare companies), while infrastructures and protocols designed to guarantee their adequate interoperability and a "patient-centric" approach are lacking, if not completely absent. This complicates and makes healthcare costlier for citizens, as well as favoring the incidence of accidental errors and frauds, often with serious consequences in terms of public health.

In this work, we propose a blockchain-based network for the decentralized management of Electronic Health Records (EHR), specifically designed according to the Italian EHR interoperability architectural model. We have developed a proof-of-concept prototype and performed a set of simulations for showing the effectiveness of our design and the advantages of deploying our system for the Italian National Health Service (NHS).

A. The Italian public health service

The Italian NHS is a system of facilities and services that have the purpose of guaranteeing all citizens, under conditions of equality, universal access to the equitable provision of health services. The Italian Constitution provides for legislative protection of the State and the Regions for the protection of health. The State determines the *essential levels of assistance* that must be guaranteed throughout the national territory, while the Regions plan and manage health care in their area in full autonomy [1].

In the last decade, many efforts have been made by national and regional institutional and technical organizations with the aim of improving the quality of health services and reducing costs by applying information and communication technologies in healthcare. The most relevant efforts concern the design and implementation of Health Information Systems (HISs) [2], with particular reference to the EHR, which has the aim of collecting all the health information related to a patient produced by the healthcare facilities and services on the national territory [3].

In order to overcome the problem of interoperability among the different regional EHR systems, the emanation of specific Italian norms since 2012 has allowed the competent Institutions (Agency for Digital Italy, Ministry of Health, Ministry of Economy and Finance, with the technical support of the National Research Council of Italy) defining the national EHR interoperability architectural model. This model is based on 21 regional IT platforms that interact each other by means of a national framework, namely National Interoperability Infrastructure (INI), as shown in Figure 1.

 ISBN: 978-1-61208-759-7

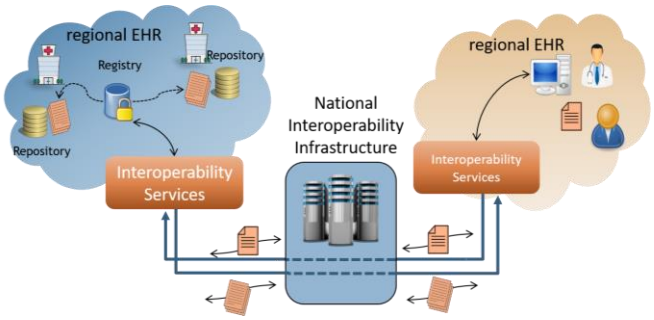

Figure 1. Italian EHR Interoperability Framework.

Each regional IT platform has the aim of indexing into a regional registry the metadata of the health digital clinical data related to its patients, whereas such data are stored into the data repositories managed by the health facilities [4].

The national interoperability infrastructure is conform to the registry/repository paradigm based on the IHE XDS Integration profile, which has the scope of facilitating the sharing of patient EHRs across health enterprises within an affinity domain (a group of healthcare facilities that intend to work together) [5]. With regards to the data structure, it is necessary to distinguish between clinical data and metadata. Clinical data are structured according to the HL7 CDA Rel. 2.0 standard format [6]. Such a format consists of two main sections: *header*, which contains context data (like patient name, author, etc.); *body*, which contains the clinical content. Each type of clinical document is structured according to Italian Implementation Guides, which are national localizations of the HL7 CDA Rel. 2 standard. It is worth noting that also clinical data represented in PDF are accepted by the national interoperability infrastructure. Metadata are a set of attributes related to the clinical data, which have the aim of facilitating their indexing and retrieval. Such metadata contain information like patient identifier, author, document reference and so on. They can easily be mapped to the data contained in the header of the HL7 CDA Rel. 2.0 documents. The structure and the types of the metadata comply with the IHE XDS profile. Moreover, a set of interoperability business processes have been formalized in order to specify all the activities performed by each actor involved. Such processes describe the steps to index, search for, and retrieving patient health data, wherever they are memorized on the national territory. All the platforms expose a regional node, which acts as an interface among the internal subsystems and the other regional nodes. The interactions among the regional nodes, based on consolidated international health informatics standards, are mediated by INI.

Along with the architectural model and the business processes, the functional and privacy requirements, as well as the technical specification for assuring interoperability have been defined [7]. Nowadays, almost all the regional EHR systems are able to i) collect patient health data, ii) permit their consultation to all the authorized actors (health professionals and operators, patients, caregivers, etc.), and iii) interact with other regional EHR systems to exchange information.

B. *Current issues and trends*

Despite the efforts made so far to develop a national federated architecture for the interoperability of EHR systems in Italy, significant actions are still to be taken in order to ensure an effective and correct implementation of the health business processes. In more detail, each business process is composed of a set of activities, part of which are performed by a regional EHR platform inside a Region and the other part of them is executed outside the Region by means of the interactions between a regional node and INI. Currently, the last one permits to control and track all the requests coming from the EHR platforms, whereas the interactions occurred within a Region are logged by the regional system. For these reasons, at the moment it is not possible to control that all the activities of a specific process are correctly executed, unless to analyze all the event logs generated by the distributed systems involved. Moreover, even considering a regional context, the operations performed are often tracked by different subsystems, not allowing this way the possibility to certify that the tasks executed are compliant to the desired workflows.

The definition of a security architecture, able to store in a reliable and effective way all the operations executed and easily integrable with the national architectural model, would allow ensuring patients, health professionals, and government organizations that the health data of interest are produced according to the specified and shared procedures. Such an architecture, proposed in this paper, would permit also the patients to: i) specify the policies for accessing their health data in a more flexible way, and ii) verify all the access requests performed by unauthorized users.

The rest of the paper is organized as follows. Section II describes relevant related works. Section III presents our contribution, giving the system requirements and its core architecture. Section IV provides details on the prototype developed, whereas Section V concludes the paper.

II. RELATED WORKS

In the last years, a massive amount of academic and industry work has been devoted to blockchain technologies and their applications in various sectors besides fintech. Healthcare, alongside with the supply-chain industry, has probably one of the highest prospects on opportunities from these technologies. A search for the term “blockchain” on PubMed returned 16 results in 2017, 77 results in 2018, and 88 results in the first eight months of 2019. Various companies have already implemented or are working on putting a blockchain system to the test for a healthcare use case (e.g., [8]-[12]), and as for July 2019 there are seven major healthcare blockchain consortia [13].

Below, for the sake of brevity, we will limit our discussion to three major projects, which have resulted in working implementations. Indeed, they exploit different and significant approaches to the management of EHRs that have influenced our work.

 ISBN: 978-1-61208-759-7

MedRec [14] is a project initiated in 2016 by MIT Media Lab and Beth Israel Deaconess Medical Center, with the aim to overcome four important issues in the healthcare context: fragmented data, slow access to medical data, systems interoperability, and patient agency. It provides a decentralized approach in which the permissions, data storage location, and audit logs are maintained in the blockchain, while all healthcare information remains in the already pre-existing EHR systems. The project has developed two blockchain platforms both built on Ethereum's technologies, but with major differences. Version 1.0 [15] was a small-scale, private network with specific APIs, whilst the current version 2.0 [16] is developed using Go-ethereum (Geth) and Solidity, but with changes to the amount of information stored on the blockchain for improving the scaling and privacy properties of transactions. Other major differences concern the consensus and governance protocols. MedRec 1.0 uses the Ethereum's proof-of-work protocol with appropriate parameters, where the mining process would be performed by medical researchers, who in turn would gain access to aggregated and anonymized data useful to further medical research. However, this approach poses concerns about the security and governance of patient data. In the current version, therefore, the EHR providers maintain the blockchain, resulting in a small and closed set of nodes that can reach consensus without the cost of mining. Providers use a proof-of-authority to append new blocks, and also to determine who is in their group.

Patientory [8] is both the name of a digital health company established in 2015, and a no-profit association for developing and governing the PTOYNet [16] blockchain. PTOYNet is a fork of Quorum, which in turn is an enterprise-focused version of Ethereum mainly by developers of JPMorgan Chase. Quorum executes smart contracts within the Ethereum Virtual Machine, but uses alternatives to the mining-based consensus protocol of Ethereum; moreover, it has built-in the feature of transaction confidentiality thanks to end-to-end encryption. PTOYNet has been adapted from Quorum in order to store healthcare records and manage their transactions through the PTOY token, providing an ecosystem for healthcare organizations to collaborate and innovate in a completely decentralized fashion. In exchange for PTOY, patients and healthcare organizations are able to use the network to rent health information storage space and execute health-specific smart contract payments and transactions. Patientory Inc. gains its revenue from the Software as a Service (SaaS) annual contract, as well as population health management services from the aggregation of data on the platform: machine learning physician diagnoses support, patient-provider UIcare coordination, and patient engagement. In 2018 the company launched on the market a mobile distributed application (DApp) which leverages the services offered by the PTOYNet platform. At the time of writing, the approximate return on investment (ROI) in PTOY if purchased at the time of launch is -98.84% [17].

Medicalchain [10] is an infrastructure to securely store and share EHRs: any interactions with EHRs are recorded as transactions on the network, but the EHRs are encrypted and stored in data stores within appropriate regulatory jurisdictions. Its first implementation was released in February 2018 and is built on a double blockchain. The first blockchain is a permission-based Hyperledger Fabric architecture, which allows varying access levels to the EHRs: users can control who can view their records, how much they see and for what length of time. The second blockchain is Ethereum, which is used to run all the applications and services for the Medicalchain platform through the ERC20-compliant cryptocurrency token MedToken (MTN). MTNs have been offered through an initial coin offering (ICO) crowd selling process started on February 1st 2018. At the time of writing, Medicalchain has a current supply of 500,000,000 MTN with 308,656,962 MTN in circulation, with an approximate ROI of -98.49% [18].

The previous examples should point out the difficulties of realizing a blockchain-based EHR management system, both in terms of technical deployment and governance. These difficulties are exacerbated by the EU regulations in different ways. For example, the storage of EHRs in the ledger is not only inappropriate since blockchain systems do not have the requisites of massive databases, but it makes very difficult to enforce the right to data modification or erasure under particular circumstances, as stated by Articles 16 and 17 of the General Data Protection Regulation (GDPR) [19]. More generally, blockchains underline the challenges of adhering to the requirements of data minimization and purpose limitation in the current form of the data economy.

III. OUR CONTRIBUTION

We are working on a blockchain system for the EHR management compliant with both the recently introduced GDPR and the national EHR interoperability architectural model described in Section I.A. Indeed, our design centers around the functional requirements listed in Tables I, II and III. These requirements stem from the framework of fundamental rights of the GDPR, and the organizational constraints for the national EHR interoperability architectural model. They can be grouped into those deriving from needs related to patients and those arising from the needs of health organizations.

Patients' needs are related to their privacy and the rights to data access (Article 15 GDPR) and data portability (Article 20 GDPR), which provide patients with control over what others do with their personal data and what they can do with that personal data themselves.

 ISBN: 978-1-61208-759-7

TABLE I. REQUIREMENTS FOR PATIENTS

P1	Patients should have the right of control over their data on system. They must be able to specify who can do what on their own data
P2	Patients should have the ability to change at any time the access rights to their data
P3	Patients must be able to hide their data from specific healthcare practitioners
P4	Patients need to have the ability to know how and when their data are accessed and for which purpose. This will be possible through the *disclosure* property, as indicated in the EU directives. Patients should be able to provide access to healthcare practitioners that are not entitled to access their data
P5	Patients must be able to research and retrieve their health data in the system

TABLE II. REQUIREMENTS FOR HEALTHCARE ORGANIZATIONS

O1	The data holder must be the healthcare organization which generated data
O2	Healthcare organizations must provide protection to the data they hold. Every healthcare organization can manage security policies with a certain level of autonomy
O3	Every healthcare organization should be able to design its own security policy and to enforce it. The definition of the access policies must be implemented in total freedom and through a highly flexible mechanism
O4	Healthcare organizations should be able to change quickly and easily the access policies of a given document
O5	The access control should not add a significant administrative overhead
O6	Audit operations are required: it is necessary to track all the operations carried out by a healthcare organization

TABLE III. ADDICTIONAL REQUIREMENTS

A1	Identification and authorization of the actors' functions
A2	Document indexing functions: the Healthcare Assistance Region of the patient has the responsibility of maintaining index metadata related to all the documents related to its patients, even if such documents are produced and maintained by health facilities sited outside the Region
A3	Research and recovery of health data functions related to a specific patient
A4	Search and retrieval mechanisms and pseudo-anonymization data functions
A5	Backup and restore functions
A6	Functions for allowing a patient to send data produced by certified devices to organizations accredited to the blockchain for storage and management

A. System overview

Our system is a kind of permissioned network where, according to recent blockchain design principles [20], nodes are organized in *users*, *validating*, *endorsers*, and *ordering*.

- Users are just nodes which require services by submitting transactions, and in our context are patients, physicians and other personnel of the healthcare sector.
- Validating nodes have their own copy of the ledger: they are healthcare-related companies and institutions that check for transaction I/O versus the current status of the ledger.
- Endorsers are validating nodes which, on the basis of a consensus policy provided at the application layer, have got the additional task of checking transaction correctness both syntactically and by running them.
- Ordering nodes are nodes that – through a suitable consensus protocol for the ledger layer, implemented in a dedicated module – have to assemble transactions in blocks and select the next block of the chain for the relevant blockchain.

Ordering nodes do not need to store any blockchain, nor they are aware of transaction contents. They just assemble the endorsed transactions received in blocks and communicate the next block to the validating nodes for the relevant blockchain via a gossiping protocol.

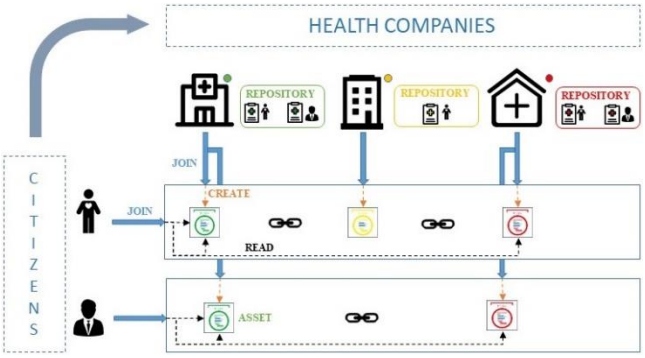

Figure 2. An high level view of our architecture.

Our system allows overcoming the current issues for the national interoperability of EHR systems in Italy. Indeed, the blockchain functionalities allow to have corroborate evidence that all the activities related of a specific process are correctly executed, provided that these activities are coded as appropriate transactions. The core architecture of our system is illustrated in Figure 2. It makes use of the building blocks described in the following sections.

B. Participants

We have identified the following four types of participants:

- *Patient*: any EU citizen, or any non-EU citizen with a valid permit to stay or a residence card.
- *Company*: any public health company, or any private health company authorized by the Ministry of Health.
- *Admin Officer*: an administrative official in charge of patient registration and accounting for a health company.
- *Company Doctor*: a physician working in a company registered in the network, who is in charge of carrying out diagnostic examinations or medical reports for patients, thus creating their health data.

Participants are created in a hierarchical way:

- Patients and Companies provide their info in order to be registered in the system.
- Companies create their own Admin Officers.
- Admin Officers of a given company create both the company's users (patients) and the physicians.

Patients get access to the system thanks to one of the two authentication methods prescribed in the National Interoperability Infrastructure, which are SPID or CNS. SPID [21] is the unique system of access with digital identity to the online services of the Italian Public Administration. CNS [22] is a device (i.e., a Smart Card or USB stick) that contains a "digital certificate" of personal authentication. They are identified by their *fiscal codes* (CF) (identify individuals and companies in Italy), and represented by

 ISBN: 978-1-61208-759-7

objects like that illustrated in Figure 3. Similar data structures are provided for admin officers and physicians.


```
{
  "$class": "org.electronic.health.record.Patient",
  "birthDate": "08-10-1987",
  "birthPlace": "Napoli",
  "age": 32,
  "companyList": [],
  "address": {
    "$class": "org.electronic.health.record.Address",
    "state": "Italia",
    "region": "Campania",
    "city": "Napoli",
    "houseNumber": "24",
    "street": "Via Roma",
    "zipCode": "80100"
  },
  "CF": "ALFDRS87D08F839L",
  "name": "Alfio",
  "surname": "Dorsi",
  "phone1": "3478123156",
  "email": "alfio.dorsi@gmail.com"
}
```


Figure 3. Example of Patient.

C. Assets

According to the national EHR interoperability architectural model described in Section I.A, patient health data are stored and accessed through the regional EHR systems. The patient's blockchain manages and stores just the transactions that produce or consume the patient's registration and authorization information and their health data. Specifically, each health document is represented in the patient's blockchain as an asset containing the link to the actual anonymized document, plus a set of metadata encoding the majority of fields described in the national technical specification for EHR interoperability [5]. Some of the fields specified in our assets are:

- *authorPerson:* defines the *CF* identifier of the author, in our case the physician that created the asset.
- *authorRole:* defines the role of the author (like physician of general medicine).
- *authorInstitution:* defines the *CF* identifier of the company in which the physician who created the asset works.
- *patientID:* the *CF* identifier of the participant for whom the document is created.
- *classCode*: defines the class of the document (prescription - PRS, medical report – REF, and so on).
- *confidentialityCode:* defines the level of confidentiality of the asset (unrestricted, low, moderate, normal, restricted, very restricted).
- *mimeType;* identifies the MIME type of the indexed document [5].

In Figure 4, we show an example of asset. As we can note, some fields are not filled. They are indeed optional fields that are automatically or manually filled only if necessary.


```
{
 "$class": "org.electronic.health.record.Doc",
 "docId": "d16e7",
 "creationDate": "2019-07-30T07:08:20.815Z",
 "authorPerson": "RSSDVD65D15F839N",
 "authorRole": "MMG",
 "authorInstitution": "Centro Diagnostico Radium",
 "XDSDocumentEntry_ClassCode": "WOR",
 "XDSDocumentEntry_Comments": "",
 "XDSDocumentEntry_ConfidentialityCode": "N",
 "XDSDocument_EntryFormatCode": "Prescrizione",
 "XDSDocumentEntry_eventCodeList": "P99",
 "XDSDocumentEntry_healthcareFacilityTypeCode": "Ospedale",
 "XDSDocumentEntry_mimeType": "text_x_cda_r2_xml",
 "XDSDocumentEntry_mimeTypePracticeSettingCode": "AD_PSC001",
 "XDSDocumentEntry_Title": "",
 "XDSDocumentEntry_TypeCode": "Prescrizione_farmaceutica",
 "patientCF": "DRSLSN87A13F839Z",
 "docType": "",
 "companyId": "CDRAD",
 "readAccess": [
   "PCCFRC00D03F205L"
 ],
 "hash": "",
 "dimension": "",
 "compDoctor":
"resource:org.electronic.health.record.CompanyDoctor#RSSDVD65D15F8
39N",
 "company": "resource:org.electronic.health.record.Company#CDRAD",
 "patient":
"resource:org.electronic.health.record.Patient#DRSLSN87A13F839Z"
}
```


Figure 4. Example of Asset.

D. Transactions

Transactions define the logic for the creation and updating of participants and assets. They are articulated in the following four sets, depending on their scope:

- *Creation and modification of participants*: various transactions permit to authorized parties to create and modify individual participants. Participants are univocally identified in the system by their fiscal code, which can be set and modified only by the creator of the participant, following the rules given in Section III.B. Some other types of data, like addresses or phone numbers, can be inserted or modified by the participants themselves after their creation. The whole process is managed through suitable *create* and *update* ACLs related to participants.
- *Creation and modification of assets*: consistently with the fact that assets represent patient's health data in the ledger, only agents (e.g., physicians, medical devices) previously authorized by a patient can create or update their assets. Only the creator of an asset can subsequently modify it, but in any case, this will be tracked in the ledger through a suitable transaction. By default, assets can be read by the patients to which they refer to and by their *general practitioners*, other than by their creators. If needed, the patient can give read access for the document to other participants in the network through a specific transaction, as detailed below. The creation, update

 ISBN: 978-1-61208-759-7

and access of assets are regulated by namesake transaction sets.

- *Access to documents*: this kind of transactions implements the P4 requirement of disclosure (see Section III.A) and are regulated by specific *read* ACLs.
- *Access to personal info*: patients must give their explicit consent to other participants (e.g., healthcare companies) for reading their personal information. This kind of transactions implements requirement P1 and are regulated by other *read* ACLs.

IV. IMPLEMENTATION AND RESULTS

We have implemented a prototype of a permissioned blockchain network, in order to assess – through a set of use case simulations – the proposed architecture against the functional requirements indicated in Section III. To implement our network, we used Hyperledger Composer v0.20.8 [23] and Hyperledger Fabric v1.2 runtime [24]. All our simulations have been performed on a Virtual Machine running Ubuntu 16.06.6 LTS. To test our architecture, we installed *composer-cli v.0.20*, *composer-rest-server v.0.20*, *generator-hyperledger-composer v.0.20*, *Yeoman* and *composer-playground v.0.20*.

In the following, we describe the use case where a general practitioner has prescribed an examination to a patient, as an illustrative example of a set of workflows that, thanks to the proposed architecture, are compliant to the requirements given in Section III. In the following images, the identifiers related to patients and health authorities are circled in red, so as to be able to identify the actors involved.

First, as shown in Figure 5, a company tries to access patient data. The system returns an error because the company has not the rights to read the patient's data. Therefore, as shown in Figure 6, the patient has to give the access to her/his personal information to the previous company; without this explicit consent, nobody can see her/his profile in the blockchain.

After the acceptance of the patient's request by a company admin officer, the company will be able to access to the patient's personal information (see Figure 7). This way, the request submitted (via a transaction not shown here) by the patient's general practitioner can be processed by the health company, which will reserve an examination date and will assign a specialist physician for the patient.

On the day of the examination, the physician will create the related asset by entering the required data (Figure 8), and such asset can be accessed read-only by the patient and her/his own general practitioner. No one can access the asset other than its creator, the patient and her/his general practitioner (Figure 9).

The patient can give read-only access to the document to other participants in the blockchain network (Figures 10 and 11); after that, they are allowed to read the asset. It is worth to stress here that only the patient has this capability, which correspond to the disclosure property P4.

The previous example and the other simulations we performed during our experimental tasks show that our blockchain network manages patient's data so to satisfy requirements P1-P5. Notice that our network manages assets that, as detailed in Section III.D, are composed of a set of metadata encoding the majority of fields described in the national technical specification for EHR interoperability, plus a link to get access to the actual anonymized health document provided for the patient. The patient's health document is not stored in the blockchain network, but in the data repository of the health company that produced it (see Figure 2). This way, our architecture satisfies requirements O1-O6 without sacrificing the audit requirement O6, often failed in current implementations. Last but not least, blockchain native functionalities allow to satisfy the requirements A1-A6.

```
To restart the REST server using the same options, issue the following command:
   composer-rest-server -c CDRAD@electronic-health-record -n never -u true -w tr
ue
Web server listening at: http://localhost:3000
Browse your REST API at http://localhost:3000/explorer
Unhandled error for request GET /api/Patient/DRSLSN87A13F839Z: Error: Unknown "P
atient" id "DRSLSN87A13F839Z".
```

Figure 5. A Company identified with ID CDRAD cannot see the patient identified by the ID DRSLSN87A13F839Z without her/his permission.

```
                -VirtualBox:~/fabric-dev-servers/electronic-health-record$ comp
oser transaction submit -c DRSLSN87A13F839Z@electronic-health-record -d '{
>   "$class": "org.electronic.health.record.GiveReadAccess",
>   "companyList": "CDRAD",
>   "pat": "resource:org.electronic.health.record.Patient#DRSLSN87A13F839Z"
> }'
Transaction Submitted.

Command succeeded
```

Figure 6. The patient DRSLSN87A13F839Z gives the read access to her/his information to the company.

```
                -VirtualBox:~/fabric-dev-servers/electronic-health-record$ curl
 -X GET --header 'Accept: application/json' 'http://localhost:3000/api/Patient/D
RSLSN87A13F839Z'
{"$class":"org.electronic.health.record.Patient","birthDate":"13-01-1987","birth
Place":"Napoli","age":32,"companyList":["NA204","CDRAD"],"address":{"$class":"or
g.electronic.health.record.Address","state":"Italia","region":"Campania","city":
"Napoli","street":"Via Toledo","houseNumber":"12","zipCode":"80132"},"CF":"DRSLS
N87A13F839Z","name":"Alessandro","surname":"De Rosa","phone1":"0815443121","emai
l":"alessandro.derosa@gmail.com"}                -VirtualBox:~/fabric-dev-serve
```

Figure 7. Making a curl operation on the REST server: the Company with ID CDRAD can now read patient's personal information.

```
                -VirtualBox:~/fabric-dev-servers/electronic-health-record$ curl
 -X GET --header 'Accept: application/json' 'http://localhost:3000/api/Doc/d16e7
'
{"$class":"org.electronic.health.record.Doc","docId":"d16e7","creationDate":"201
9-07-30T07:08:20.815Z","authorPerson":"RSSDVD65D15F839N","authorRole":"MMG","aut
horInstitution":"Centro Diagnostico Radium","XDSDocumentEntry_ClassCode":"WOR","
XDSDocumentEntry_Comments":"","XDSDocumentEntry_ConfidentialityCode":"N","XDSDoc
ument_EntryFormatCode":"Prescrizione","XDSDocumentEntry_eventCodeList":"P99","XD
SDocumentEntry_healthcareFacilityTypeCode":"Ospedale","XDSDocumentEntry_mimeType
":"text_x_cda_r2_xml","XDSDocumentEntry_mimeTypePracticeSettingCode":"AD_PSC001"
,"XDSDocumentEntry_Title":"","XDSDocumentEntry_TypeCode":"Prescrizione_farmaceut
ica","patientCF":"DRSLSN87A13F839Z","docType":"","companyId":"CDRAD","readAccess
":[],"hash":"","dimension":"","compDoctor":"resource:org.electronic.health.recor
d.CompanyDoctor#RSSDVD65D15F839N","company":"resource:org.electronic.health.reco
rd.Company#CDRAD","patient":"resource:org.electronic.health.record.Patient#DRSLS
N87A13F839Z"}                -VirtualBox:~/fabric-dev-servers/electronic-health
```

Figure 8. The physician which created an asset and can see the asset in the blockchain.

Figure 9. Only the physician which created the asset and the participant to which the asset refers to can access the data in blockchain. A participant identified with a different ID (PCCFRC00D03F205L) cannot see the asset.

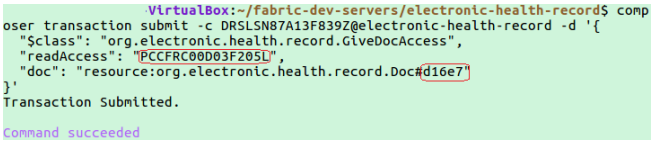

Figure 10. Patient with ID DRSLSN87A13F839Z can give read access permission to the participant with ID PCCFRC00D03F205L.

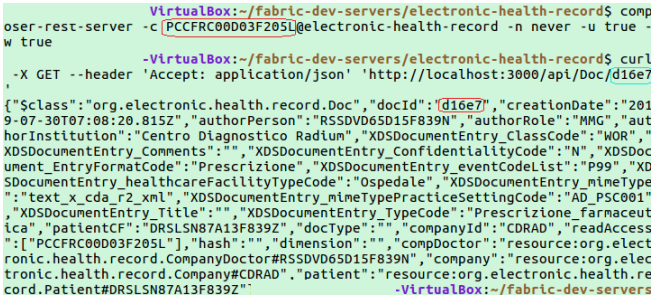

Figure 11. Now the participant with ID PCCFRC00D03F205L can read the document.

It is worth noting that some asset field values in Figure 4 (e.g., "Prescrizione") do not match the technical specification for EHR interoperability based on the IHE XDS framework. This is because of a current limitation of the Hyperledger Composer "Enumerated Types", which does not accept some special characters, e.g., the dot character. This issue can be overcome thanks to a lookup table processed at the application layer.

V. CONCLUSIONS AND FUTURE WORK

We have designed a blockchain-based architecture for the decentralized management of EHRs, which is compliant with the GDPR and allows to overcome some main issues concerning the current federated architecture for the national interoperability of EHR systems in Italy. Our proof-of-concept network represents just the core architecture of a federated EHR management system, and much more work is required to get a complete working system. First, our network has to be coupled with a suitable access control and security framework to protect patient's health data. This framework has to be designed according to the functional requirements illustrated in Section III, but it should compromise neither the usability of the system nor its scalability and management. Second, an accurate and full-fledged user interface has to be realized through the development of apps customized for the different kinds of network participants. Our next work will concern the design and implementation of the access control and security framework. Then we are going to realize a testbed for assessing the effectiveness of the EHR management system resulting by coupling it with the blockchain network illustrated in this work.

REFERENCES

[1] Italian Ministry of Health, http://www.salute.gov.it/portale/lea/dettaglioContenutiLea.jsp?lingua=italiano&id=5073&area=Lea&menu=vuoto, retrieved 09/2019.

[2] A.D. Black, J. Car, C. Pagliari, C. Anandan, K. Cresswell, T. Bokun, B. McKinstry, R. Procter, A. Majeed, and A. Sheikh, "The impact of ehealth on the quality and safety of health care: a systematic overview", PLOS Med. 8(1), 2011, pp. 1-16.

[3] F. Aminpour, F. Sadoughi, and M. Ahamdi, "Utilization of open source electronic health record around the world: a systematic review", Off. J. Isfahan Univ. Med. Sci. 19(1), 2014, pp. 57-64.

[4] M. Ciampi, M. Sicuranza, A. Esposito, R. Guarasci, and G. De Pietro, "A technological framework for EHR interoperability: experiences from Italy" Comm. in Computer and Information Science, Springer, 736, 2017, pp. 80-99.

[5] Integrating the Healthcare Enterprise, http://www.ihe.net/, retrieved 09/2019

[6] HL7 Version 3 Clinical Document Architecture (CDA) Release 2, https://www.hl7.org/implement/standards/product_brief.cfm?product_id=7, retrieved 09/2019.

[7] Agency for Digital Italy of the Presidency of the Council of Ministers, https://www.fascicolosanitario.gov.it/, retrieved 09/2019.

[8] https://patientory.com/technology/, retrieved 09/2019.

[9] https://enterprise.gem.co/, retrieved 09/2019.

[10] https://medicalchain.com/, retrieved 09/2019.

[11] https://encrypgen.com/the-dna-economy/, retrieved 09/2019.

[12] https://www.simplyvitalhealth.com/, retrieved 09/2019.

[13] https://hashedhealth.com/newsletter-sept-2019/, retrieved 09/2019.

[14] https://medrec.media.mit.edu/technical, retrieved 09/2019.

[15] A. Azaria, A. Ekblaw, T. Vieiraand, and A. Lippman, "MedRec: using blockchain for medical data access and permission management", in the proc. of the 2nd International Conference on Open and Big Data, IEEE, 2016, pp. 25-30.

[16] C. McFarlane, M. Beer, J. Brown, and N. Prendergast, "Patientory: a healthcare peer-to-peer EMR storage network" v1.19, https://patientory.com/, retrieved 09/2019.

[17] coinmarketcap/patientory, retrieved 09/2019.

[18] coinmarketcap/medicalchain, retrieved 09/2019.

[19] GDPR, https://eugdpr.org/, retrieved 09/2019.

[20] M. Vukolić, "Rethinking permissioned blockchains", in the proc. of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts, ACM, 2017.

[21] CNS, http://www.progettocns.it/, retrieved 09/2019.

[22] SPID, https://www.spid.gov.it/, retrieved 09/2019.

[23] https://hyperledger.github.io/composer/, retrieved 09/2019.

[24] https://hyperledger-fabric.readthedocs.io, retrieved 09/2019.

Building a Patient-Centered Blockchain Ecosystem for Caregivers: Diabetes Type II Case Study

Shada Alsalamah[1], Ghada Alsuwailem[1], Fatima Alrajeh[1], Samar Aharbi[1], Salmah AlQahtani[1], Razan AlArifi[1], Shaden AlShargi[1], and Shaikha AlOthman[2].

[1] Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh, KSA. e-mail: saalsalamah@ksu.edu.sa.

[2] MIT I-Corps, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA.

Abstract— **Diabetes is one of the fastest-growing health problems in the world, and is now reaching epidemic proportions in some countries. With the globally increasing numbers of diabetic patients and their need for professional care, it is difficult and time-consuming to share patient-care information among caregivers in a traditional way, which adds financial and psychological burden on home caregivers. Recent developments in advanced technologies and mobile health (mHealth) applications fail to equip them with the right ecosystem for patient-centered information sharing to allow for informed care decisions. Therefore, motivated by the need for home caregivers' empowerment to cope with the pressure, this paper studies the requirements for building an effective communication channel among caregivers and between them and their patients, and presents the potential benefits of mHealth applications and blockchain technology in achieving such requirements using a case study in Diabetes Type II patients in Saudi Arabia. Using interviews and questionnaires, this paper shows that there is the need for an ecosystem that supports diverse caregiving groups with multiple languages, distributes tasks between more than one caregiver to alleviate the burden on one caregiver, provides a treatment plan provision by a specialized care team to be viewed and followed by caregivers and patients, and alerts everyone in case of an emergency. Such requirements can be incorporated into an mHealth solution that builds a transparent blockchain-based patient-centered caregiving ecosystem.**

Keywords-Blockchain Technology; Diabetes Type II; mHealth; Patient-Centered care, Home caregiving.

I. INTRODUCTION

A caregiver, whether paid or unpaid, as a member of a person's social network helps patients with their daily-life activities [1]. Caregiving is most commonly used to address impairments related to people who are aging and/or with disabilities, diseases, conditions, injuries, or mental disorders. Supporting caregivers is one of the most important issues that has been neglected around the world while there is a growing demand for their help. According to recently published data, 25% of global population will be over 65 years old in the next 5 to 7 years [2]. Furthermore, there are no hospitals or nursing homes that can accommodate this huge and increasing number of patients. There are many difficulties related to communication between caregivers themselves and the caregivers within the care team regarding the accurate timestamped order in which to share informed decisions regarding the diabetic patient. Moreover, the care team faces many difficulties in order to manage and control the large number of patients. On the other hand, the patients themselves face a lot of difficulties, such as waiting times, and needing to see more than one doctor on the same day in order to receive the full treatment. Therefore, when they have disabilities, this process becomes very stressful for them. This project sheds light on diabetic patients of Type II, which occurs in patients over 40 years old and it is the most prevalent type of diabetes in Saudi Arabia [3]. The reason for selecting this type of category is the increasing number of diabetic patients in Saudi Arabia, which, according to previously published data, shows that 25% of Saudis are suffering from diabetes; this means that 4 million diabetic patients require 5.5 million consultations and follow-up visits each year [4]. Therefore, these numbers motivate us to do this project to help those people. Also, based on the interviews we conducted with diabetic patients [5] and diabetes caregivers [6], it is clear that there is a lack of Arabic speaking platforms that facilitate the caregiver's connectivity with diabetic patients.

A. Diabetes Statistics Globally Vs. Saudi Arabia

Diabetes is one of the fastest-growing health problems in the world and is now reaching epidemic proportions in some countries. It is mainly as a consequence of life-style behaviors, such as lack of exercise, unhealthy diet, obesity and being overweight. Over the past four decades, major changes have occurred in Saudi Arabia. Growth, prosperity and modern technology have brought pronounced changes in the lifestyle of people. In particular, eating habits are less healthy and the level of physical activity has declined. According to the International Diabetes Federation (IDS) 2019 report [7], there are currently 463 million people worldwide living with diabetes, and this is estimated to rise to 700 million by 2045. Additionally, the report reveals evidence that in 2019 the number of deaths resulting from diabetes and its complications is estimated to be 4.2 million, and health expenditure on diabetes has reached at least USD 750 billion, which represents 10% of total spending on adults [7].

Moreover, according to the IDS report, Saudi Arabia is ranked 9th in the top 10 countries with an estimated 2.5 thousand incidents (new cases) of Type I in children and adolescents in the age group 0-14 [7], while it is ranked 5th among the top 10 countries in terms of incidence rates (per 100,000 population per year) after Finland, Sweden, Kuwait, and Norway [7]. Therefore, Saudi Arabia is among the top 10 countries in the world with the highest prevalence of diabetes [7]. In Saudi Arabia, the national healthcare burden resulting

 ISBN: 978-1-61208-759-7

from diabetes is likely to exceed $0.87 billion [8], but this neglects the indirect costs associated with diabetes, such as absenteeism, loss of productivity from disease-related complications, unemployment due to disability and early mortality due to disease [8]. On the other hand, the social costs, such as pain and suffering and care provided by caregivers, as well as healthcare system administrative costs, cost of medications, clinician training programs, and research and infrastructure development are also omitted from this research study. Further studies are needed to confirm the present findings and to improve our understanding of the economic costs of diabetes and its related complications [8]. Diabetes should be treated under close collaboration between patients and caregivers in order to prevent long-term complications, and diabetics must have ongoing and accurate daily care to avoid early death.

B. Caregiving Ecosystem

In caregiving, some patients require modern homecare, which provides the patient with the correct and complete care. They are licensed, professional caregivers who devote all of their time attending to all of the patient's care needs [9]. mHealth is a general term for the use of mobile phones and other wireless technology in medical care [10]. It is also one of the most common applications for disease surveillance, treatment support, epidemic outbreak tracking and chronic disease management [10]. The scope of the patient-centered caregiving ecosystem for diabetic patients includes four aspects: diet monitored by a dietitian, physical training provided by a physical trainer, treatment plan implemented by a care team, and glucose level measured by the patient. Each and every aspect is managed by a specialist who collectively needs to collaborate for the implementation of patient-centric home caregiving. These aspects present the basic needs in diabetes caregiving ecosystems as illustrated in Figure 1 below.

Figure 1. Diabetes Caregiving Ecosystem

This paper aims to, first, study the requirements for building an effective communication channel among caregivers as well as between caregivers and their patients. Second, the paper presents the potential benefits of mHealth applications and blockchain technology in achieving such requirements using a case study in Diabetes Type II patients in Saudi Arabia. The remainder of this paper is organized as follows. Section II presents a review of the literature of all related work supporting home caregivers, while the methodology is explained in Section III. Section IV presents an analysis of the results, and Section V provides an overview of the proposed system design. The conclusions close the article in Section VI.

II. LITERATURE REVIEW

Although there are some applications available in the literature to support home caregivers, they have some limitations that mean that the key challenges cannot be addressed. *Thermo* [11] application helps take care of the family's health but it does not support diabetic patients. Furthermore, it is designed for management and maintains the record changes in temperature readings, displaying this in a timeline. However, it is only available in English, and does not provide services for Arabic-speaking caregivers. *Caring* [12] is designed for adding a person and making changes, adding notes and also sending reminders. This supports caregivers to connect, but there is a lack of necessary aspects relevant for a diabetic patient, such as setting an appropriate treatment plan for a diabetic patient and measuring blood sugar. *Medisafe* [13] solution could help in terms of remembering to take medication and it can select the shape and color of the pill for those who may not be able to read and it is also possible to invite a friend, but it is not permitted to add a caregiver. The mHealth application سكري[14] (i.e., Diabetes) is perfect to track blood sugar levels, meals and sports activities. It uses a medical doctor to monitor the patient's state remotely. However, it does not support home-caregivers and it is only available in the Arabic language. *Carely* [15] solution is a network connection between caregivers to share changes in caregiving and coordinate the responsibilities. However, it does not include important features, such as measuring blood sugar or sending an S.O.S alarm. Also, it is only available in the English language. *Saleem* [16] mHealth solution is mainly designed to connect patients with their medical doctors who can use it at their clinic to monitor vital signs remotely without the patient having to come to the clinic for each visit. It reduces the waiting time and therefore optimizes the hospital's resources. However, it does not connect home caregivers. *Fitbit* [17] solution is an American company that has products for activity tracking. It uses wireless and wearable devices such as watches, and it can measure data, such as the number of steps walked, heart rate, quality of sleep and steps climbed. This, however, does not measure sugar levels, support caregivers, or cater to Arabic language speakers. The above applications reviewed fall short of providing caregiving

 ISBN: 978-1-61208-759-7

requirements for diabetic patients in Saudi Arabia. For instance, when care is provided to the patient, the doctor must establish the treatment plans for the patient, while caregivers must confirm that the patient is following this treatment plan. According to the survey, the need for the measurement of blood sugar for the patient is the most important function.

III. METHODOLOGY

Using a pragmatic approach, this project implements a mixture of quantitative and qualitative methods. Initially, semi-structured interviews were conducted for primary data collection in order to identify challenges. Two interviews were conducted, the first with a consultant family physician [3] at the Ministry of Defense's Medical Services Directorate Medical Administration-MSD. The aim of this interview was to clarify the difference between the three different types of diabetes. The second interview was with a diabetes Type II patient [5] and his caregiver [6], who is also his wife. The aim of the interview was to clarify some points raised while analyzing the requirements. The question lists for each targeted interviewee are listed in Table 1 below.

TABLE I: INTERVIEWS WITH TARGETED INTERVIEWEE AND QUESTION LISTS.

Targeted Interviewees	Question Lists
Patients	• *Age:* • *Gender:* • *Type of diabetes:* • *Who are your caregivers?* • *Caregiver's age:* • *Caregiver's gender:* • *How difficult is maintaining your care?* • *Have you ever visited a doctor in an emergency situation?* • *If yes, can you explain your experience?* • *Have you ever forgotten to write down your blood glucose measurements and show them to your doctor?* • *"I often forget to take medicine on time." Do you agree with this sentence?* • *Do you agree to have more than one caregiver to obtain good care continuity?* • *Do you agree with the idea of gathering your caregivers via a system to provide good care continuity for you?* • *Do you prefer to have a daily blood glucose measurements recorder?* • *What are the diseases associated with diabetes in your opinion?* • *What is your favorite application?* • *May you talk about the beginning of your disease?* • *What was frustrating about that experience?*
Caregivers	• *Age:* • *Gender:* • *What is the type of patient diabetes?* • *Who is the patient who is being cared for?* • *Patient's age:* • *Patient's gender:* • *How difficult is maintaining your patient's care?* • *Have you ever had an emergency where you needed assistance on the spot?* • *If yes, can you describe that experience?* • *Have you ever thought about changing the way you help?* • *If yes, can you describe that experience?* • *Have you ever forgotten to write down the blood glucose measurements of your patient and show them to a doctor?* • *I often forget to give medicine to my patient on time." Do you agree with this sentence?* • *Do you agree to get help from other caregivers in order to provide good care continuity?* • *Do you agree that the idea of combining all caregivers in one system will save time and effort for communication?* • *What are the diseases associated with diabetes in your opinion?* • *What is your favorite application?* • *Can you talk about the beginning of your experience in caregiving?* • *What was frustrating about that experience?*

Furthermore, a questionnaire was designed to gather additional primary information regarding the proposed solution. The questionnaire targets caregivers and patients to reflect the opinion of a larger community of caregivers and patients for quantitative analysis. It was created using Google forms and classified into two categories, patient and caregiver. The patient's questionnaire has 15 questions and the first 6 questions provide an overview about the patient and his/her caregiver. The questions are diverse, consisting of multiple-choice questions, linear scale questions, true/false questions and open questions. Meanwhile, the caregiver category consists of 15 questions, and there were also Likert-type scale questions. Both questionnaires were distributed through social media and texting platforms, such as WhatsApp and Twitter.

IV. SYSTEM ANALYSIS RESULTS

This section presents all of the results and findings as follows.

A. Qualitative Results

Results from the interviews show that each of the three types of diabetes has different causes and care needs. Diabetes Type I patients are normally diagnosed from an early age due to different factors, such as hereditary and

endocrine diseases. Type II, which is the most prevalent, usually occurs after the age of 40 [3], while Type III is related to pregnancy diabetes. Furthermore, the care team in general face many difficulties in tracking the blood glucose test for patients, since most of the patients have more than one caregiver and there may be a lack of communication, which might cause difficulties when the doctor needs to know when the last time the blood test was high or low in a timely manner in order to decide on the appropriate treatment plan. Regarding the information the care team needs to check before deciding the treatment plan, the team normally needs to have the overall assessment for patients, such as job, residence, sports, name, weight and height. Moreover, they need to keep track of medical records for the patient as well as the history related to the patient's family, such as allergies, social needs and medication. There are some emergencies that require immediate medical intervention for diabetic patients, and it is centered around two important aspects: the patient's blood sugar drops, causing unconsciousness or blood sugar rises which causes the patient to be very stressed and might cause problems in vision. Therefore, there is a frequent need to change the treatment plan depending on the changes in current medical conditions and the patient's diet, injections, medicine and sport activities. Finally, results reveal the need to improve communication between care team members and between the patient and the care team in the future.

In regards to the patient's needs, results reveal the patient's difficulties in remembering treatment times and coordinating between medicines and food. Moreover, patients require continuous follow-up with a diet that is suitable with the health status. Also, the patient must visit the doctor periodically every three months in order for certain medical examinations to be conducted, such as "Dilated" eye exam, blood pressure and kidney function tests. In addition, diabetic patients require observation and care continuity by at least one caregiver. This includes measuring blood sugar, type of food and the provision of psychological comfort. Finally, findings show that the most important needs of both patients and caregivers, are mainly related to recording changes, confirming treatments and maintaining contact with the care team.

B. Quantitative Results

The total number of questionnaire responses was 68 patients and 97 caregivers. Results show that 60.3% of patients were Type II, while 38.2% of the cases were Type I. Therefore, this study decided to target Type II. Furthermore, 70.1% of patients want to have a daily blood glucose measurements recorder, while only 28.4% expressed they did not want that. This indicates that there is a need for a measurements' recorder. Regarding the caregiver's communication, 44.3 % strongly agree on the need for one platform to connect all caregivers, while 17.5% neither agree nor disagree, and only 9.3% disagree. This represents the need to address the aim of this study. Furthermore, results show that, firstly, there are interruptions in continuity of care. This is because of communication challenges within the caring ecosystem. This indicates the need to optimise patients' quality of life in delivering the required care. Secondly, some caregivers needed to provide emergency assistance on the spot, and some stated that they were unable to provide such help due to unpreparedness, for example the lack of knowledge of first aid instructions. This highlights the need for an S.O.S alarm and some guidance when an emergency occurs.

C. System Requirements

The results above identify the fundamental system design requirements for building an ecosystem that coordinates tasks between all caregivers through the mHealth solution, rather than keeping the pressure on one caregiver. Furthermore, the proposed solution is a patient-centered one that involves the patient as a caregiver to engage in the decision-making process, increasing their satisfaction and ultimately improving health outcomes. The proposed system has three users: *Care Team Member*, *Caregiver,* and *Patient*. The functional requirements for each user are further explained below.

1. *Caregiver*
    1. A Caregiver shall be able to register onto the system.
    2. A Caregiver shall be able to log in to the system.
    3. A Caregiver shall be able to view the patient's demographic information (caregiving history, work location and home location).
    4. A Caregiver shall be able to view a treatment plan.
    5. A Caregiver shall be able to create reminders to inform him/her when the time of treatment is.
    6. A Caregiver shall be able to record care changes about patient status.
    7. A Caregiver shall be able to add comments about the changes in the care timeline.
    8. A Caregiver shall receive an alarm to inform him/her of the time of treatment or patient's appointments.
    9. A Caregiver shall be able to confirm treatment.
    10. A Caregiver shall receive a notification to inform him/her that the treatment is confirmed from another caregiver/patient (or a Caregiver shall be able to view the confirmed treatment).
    11. A Caregiver shall be able to send an S.O.S alarm.
    12. A Caregiver shall receive an S.O.S alarm.
    13. A Caregiver shall be able to record measurements of blood sugar.
    14. A Caregiver shall be able to record care timeline.
    15. A Caregiver shall be able to view care timeline history.
2. *Patient*
    1. A Patient shall be able to register onto the system.
    2. A Patient shall be able to log in to the system.

 ISBN: 978-1-61208-759-7

3. A Patient shall be able to view the demographic information (personal information, caregiving history, work location and home location).
4. A Patient shall be able to add a new caregiver.
5. A Patient shall be able to view a treatment plan.
6. A Patient shall be able to create reminders to inform him/her of the time of treatment.
7. A Patient shall be able to record care changes about his/her status.
8. A Patient shall be able to add comments about changes in the care timeline.
9. A Patient shall receive an alarm to inform him/her of the time of treatment or appointments.
10. A Patient shall be able to confirm treatment.
11. A Patient shall be able to send an S.O.S alarm.
12. A Patient shall be able to record measurements of blood sugar.
13. A Caregiver shall be able to record care timeline.
14. A Patient shall be able to view care timeline history.

3. *Care Team Member*
   1. A Patient shall be able to register onto the system.
   2. A Patient shall be able to log in to the system.
   3. A Care Team member shall be able to view the patient's demographic information (caregiving history, work location and home location).
   4. A Care Team member shall be able to add a treatment plan.
   5. A Care Team member shall receive an alarm when an emergency occurs to the patient.
   6. A Care Team member should be able to view current location when an SOS alarm is pressed.
4. *System*
   1. The System shall be able to send an alarm to inform all caregivers of the time of treatment.
   2. The System shall be able to send notification after recording a change in care.

D. Blockchain-Based Distributed Medical Record

Interoperability of healthcare systems for data exchange has been a global issue for decades. This has been traditionally addressed using interoperability standardization [18]. The Fast Health Interoperability Resources (FHIR) has been one of the most recently and widely adopted interoperability standards in the healthcare sector, mainly due to its cost effectiveness [18]. However, with the first introduction of blockchain technology in the last decade, and after its recent developments beyond financial services and cryptocurrency applications, interoperability has been revolutionized with the applications of blockchain technology, which makes interoperability one of its mostly used use case applications of this new technology in most sectors even beyond healthcare [18][19]. Blockchain technology is considered a data structure that stores transactional records efficiently by supporting key features. Ultimately, this makes blockchain more capable in terms of solving existing healthcare interoperability problems more effectively, quickly and simply than traditional interoperability standardizations [18].

This project utilizes blockchain technology to provide effective support; this is because it is a digital ledger that allows caregivers to securely connect with each other, without the need for third-party middlemen [19]. Furthermore, blockchain technology creates an immutable, timestamped, trusted chain of components (changes in care continuity) that are distributed in a ledger among the network of caregivers and each caregiver has the same ledger [19]. There are four classes of accessibility and visibility in blockchain [20]: *Permissionless private* that provides access so that only specific people can visit, *permissionless public* that is accessible and visible to everyone, *permission private* is such that a specific person can access and visit and *permission public* provides visibility to anyone but only a specific person can access.

Private permission is suitable for this project so that a specific care team can write and read, and only caregivers that the patient knows can access and view information about him/her. Therefore, this study builds a private permissioned blockchain solution to address connectivity and patient homecare continuity issues. This is achieved by designing a blockchain-based social network for a patient-centered family caregivers' support system that can connect the patient's families, friends and neighbors to build an ecosystem of stakeholders to make shared, informed care decisions. Connected caregivers can exchange information about a patient, including, but not limited to, a patient's treatment, changes in vital signs readings, time of sleep, who is the last to care for the patient to ensure care continuity and they can also confirm taking treatments at specified times. In addition, the caregivers and the patient can add comments which increases the power of interaction and contact between the members of this system. Therefore, the blockchain ledger can be shared among all caregivers in this distributed network, and they will know who made the changes or added information in relation to patient care. Blockchain allows having multiple participants who need to view common information, or multiple participants' actions must be recorded or verified so other participants can trust the validity of the noted actions. Blockchain has strong potential in providing a solution [18][21]. Since these mentioned conditions meet our project aims, blockchain is the best platform solution.

V. SYSTEM DESIGN

In this section the system design is fully presented as follows.

A. Data Design

The system has the following users:

- ***Care Team Member*** includes medical doctors/healthcare practitioners who specialize in managing diabetes

 ISBN: 978-1-61208-759-7

patients. One of the specialist medical practitioners required by diabetic patients is an ophthalmologist who should check the retina annually, record when complications occur, and it is possible that the patient will require kidney and gangrene medical practitioners. Specialists, who could help the doctors in their job but are not certified as medical doctors, can provide the patient with the dietitian's plan at the beginning of the disease, and it may be the case that the disease is advanced and the patient was unaware. There is also a diabetes educator or health educator who follows-up with the patient's status at frequent intervals, measuring glucose, and providing instructions when there is low or high glucose.

- ***Caregiver*** includes family, friends or neighbors who help to take care of their patients.
- ***Patient*** is the person who needs to receive care.

The proposed system use case diagram is illustrated in Figure 2 below.

Figure 2. Proposed System Use Case Diagram.

The designed system builds a network that connects more than one caregiver and their patient. Patients may be able to care for themselves so when we refer to a caregiver, this might indicate the normal caregiver or the patient. In this system, there are features (illustrated in Figure 2) that help to provide good care, namely, the ability to view a treatment plan, create reminders, add comments about changes in the care timeline, confirm treatment, send an S.O.S alarm and record measurements of blood sugar; all these features are available to a caregiver. Also, the patient can add a new caregiver. Furthermore, a care team member shall be able to add a treatment plan that would be seen by caregivers. Three of the key proposed system use cases in Figure 2 including, *Add a New Caregiver/Care Team Member, Measuring Blood Sugar,* and *Performed Treatment* are further explained in use case activity diagrams below and illustrated in Figures 3-5, respectively.

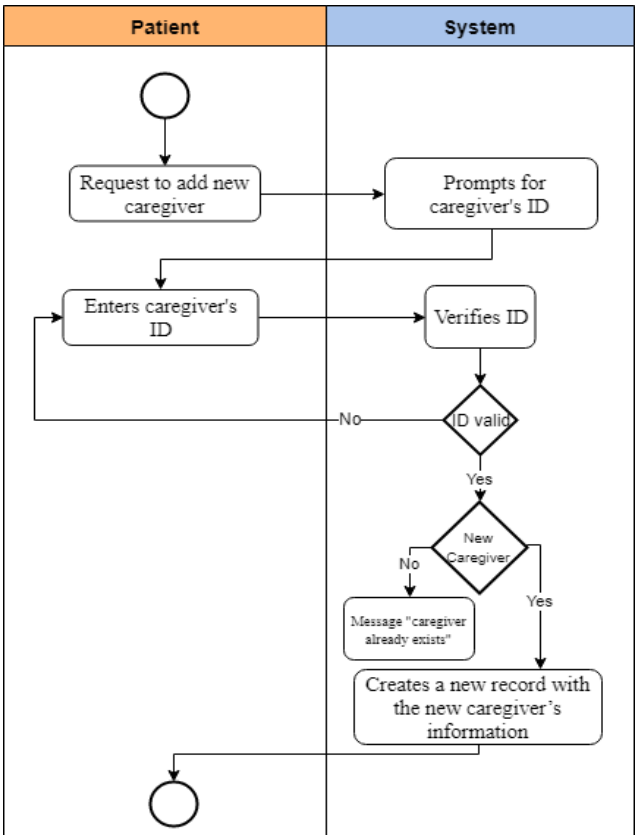

Figure 3. Add a New Caregive/ Care Team Member Activity Diagram.

Figure 3 shows the interaction between *Patient* and *System* (i.e., the proposed one) when a new *Caregiver* joins the patient's care team and so needs to be added to *System* to grant him/her all the needed rights. Therefore, only the *Patient* has the right to initiate this use case by requesting it from the *System,* which prompts for the new *Caregiver* details. Such information details can only be recorded by the *Patient*. Then *System* validates the details in terms of

 ISBN: 978-1-61208-759-7

completeness and accuracy for quality assurance purposes before it creates a new record for the new *Caregiver*.

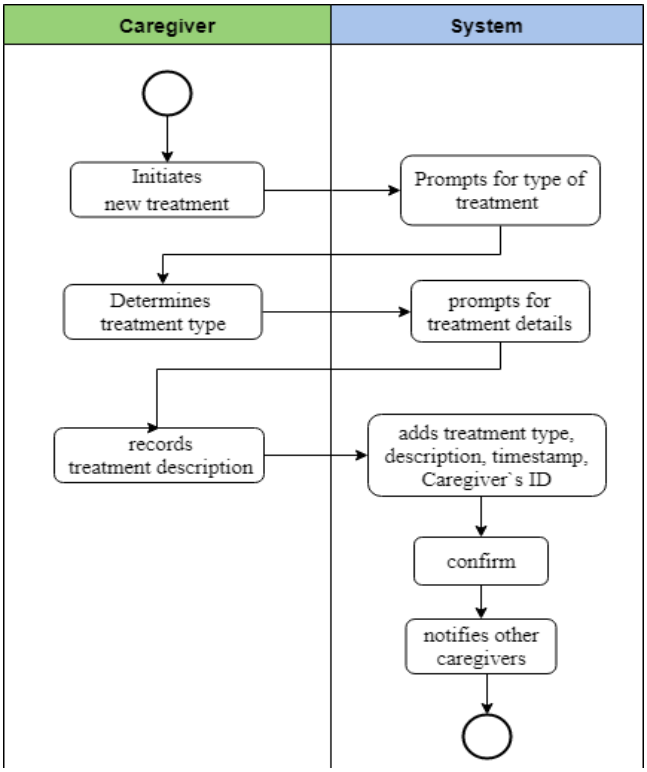

Figure 4. Performed Treatment Activity Diagram

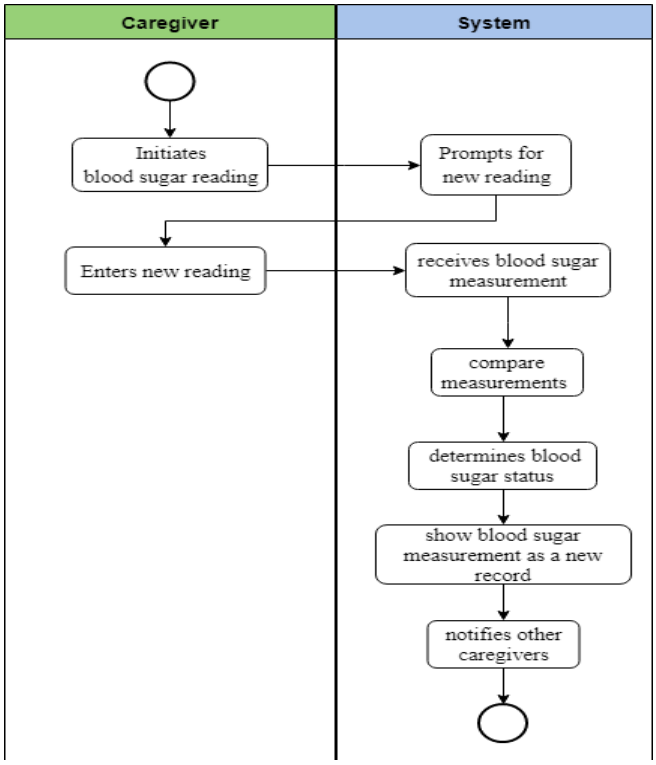

Figure 5. Measuring Blood Sugar Activity Diagram

Furthermore, once the *Patient* undergoes any kind of treatment or have a new blood sugar measurement reading, the *Caregiver* must record it to be shared with the rest of the care team as a change in care. Such change in care is shown in both *Performed Treatment* activity diagram (see Figure 4), and *Measuring Blood Sugar* activity diagram (see Figure 5), respectively. As for new treatments, the *System* prompts for details including, the treatment type, description, even the time it was performed, and the Caregiver's ID for authenticity. While in every new blood sugar measurement reading taken by a *Caregiver*, the *System* compares the latest reading with the previously recorded one to determine the blood sugar state. Finally, once the details are successfully recorded, for a new treatment or blood sugar measurement reading, the *System* notifies all registered *Caregivers* in the system earlier by the *Patient.* and notify the rest of the team about any changes in *Patient's* care.

VI. CONCLUSION

Caregiving is most commonly used to address impairments related to people who are aging and/or with disabilities, diseases, conditions, injuries, or mental disorders. Supporting caregivers is one of the most important issues that has been neglected around the world while there is a growing demand for their help. Moreover, diabetic patients, in particular, require home caregiving since there are no hospitals or nursing homes that can accommodate the huge and increasing numbers. Also, most of these diabetes patients are aged over 45 years according to the questionnaire, which means that these diabetic patients present with Type II and therefore require special care. Therefore, this study aims to address key challenges home caregivers face to support their loved ones. Moreover, the increasing number of diabetic patients motivates us to develop this application to serve or help those people. Besides, there are many difficulties related to communication between caregivers themselves as well as between the caregivers and the care team in terms of providing an accurate, timestamped order to share informed decisions regarding diabetic patients. On the other hand, the patients themselves face significant difficulties, such as waiting times and needing to see more than one doctor on the same day to receive the full treatment. Therefore, when they have disabilities, this process becomes very stressful for them. Also, there is a lack of Arabic speaking platforms that help with caregivers' connectivity for diabetic patients.

This paper sheds light on the challenges faced by patients and caregivers. It identifies the requirements for building a blockchain-based social network ecosystem to facilitate patient-centered family caregiving to empower and connect those patients' families, friends and neighbors to share patient-centered information with each other and make shared, informed care decisions. The proposed ecosystem design achieves this by allowing caregivers to track and record their patients' care management changes, health status and psychological status on a 24/7 basis. It also provides the caregivers with the treatment plan for diabetes and associated

 ISBN: 978-1-61208-759-7

diseases for comorbid diabetic patients which are decided by the care team member. This would facilitate the process of increasing the caregiver's connectivity in order to provide professional, timestamped care continuity for diabetic patients with Diabetes Type II in Saudi Arabia.

Scalability should be considered for future work related to the application to include more languages, involving communities from Arab countries and the rest of the world, as well as other types of diabetes, including Diabetes Type I and III and other chronic diseases. Also, consideration should be given to incorporating more IoT devices and wearables to connect with caregivers' and patients' smart devices. The proposed system design has great impact on the caregiving process for diabetic patients by allowing the patient's family to have continuous patient-centered care to help them provide the best care while reducing the burden of caregiving. Furthermore, it has a significant impact on the field of Computer Science by allowing analysists, developers, and engineers to interact with health care professionals in order to provide a solution that serves the diabetic patients of Type II along with their caregivers. This is to provide full treatment plan for diabetic patients suffering from different types of diseases and conditions associated with diabetes. Finally, this solution has an ethical and legal impact on our society as it provides transparency among all caregivers.

REFERENCES

[1] C. K. Hunt, "Concepts in caregiver Research", 2003, [online]. Available at: https://sigmapubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1547-5069.2003.00027.x [retrieved: April, 2019]

[2] United Nations Population Fund, "World Population Dashboard", 2019, [online] Available at: https://www.unfpa.org/data/world-population-dashboard [retrieved: January, 2019].

[3] A. AlGarzai, "Consultant family physician-Senior Trainer Quality management & patient safety specialist (GCI) Strategic Health management and leadership" (Interview), February 2019.

[4] International Medical Travel Journal, "Saudi Arabia Launches Healthcare Provitision Programme", 2018, [online] Available at: https://www.imtj.com/news/saudi-arabia-launches-healthcare-privatisation-programme/. [retrieved: March, 2019]

[5] F. Alharbi, "Diabetic Type II Patient. Personal Communication" (Interview), February 2019.

[6] A. Alkhalaf, "Caregiver Personal Communication" (Interview), February 2019.

[7] International Diabetes Federation, "The IDF Diabetes Atlas Ninth edition 2019", 2019, [online] Available at: https://diabetesatlas.org/en/. [retrieved: October, 2019]

[8] A. K. Alhowaish, "Economic costs of diabetes in Saudi Arabia," Journal of Family & Community Medicine. 2013 Jan-Apr; 20(1): 1–7. doi: 10.4103/2230-8229.108174.

[9] L. Sarafan, "What is Home Care?" 2018, [online] Available at: https://homecareassistance.com/pdf/ [retrieved: April, 2019].

[10] pHealth, "pHealth 2017", 2017, [online] Available at: http://www.phealth2017.eu/ [retrieved: April, 2019].

[11] Apple Store, "Thermo", 2019, [online] Available at : https://apps.apple.com/us/app/withings-thermo/id1108420798. [retrieved: April, 2019].

[12] Apple Store, "Caring", 2019, [online] Available at: https://apps.apple.com/us/app/caring/id1265929675?ls=1. [retrieved: April, 2019].

[13] Apple Store, 2019, "Medisafe", 2019, [online] Available at: https://www.medisafe.com/. [retrieved: April, 2019].

[14] Apple Store, 2019, "سكري", 2019, [online] Available at: https://apps.apple.com/sa/app/id1125752054. [retrieved: April, 2019].

[15] Apple Store, 2019, "Caring", 2019, [online] Available at: https://apps.apple.com/us/app/carely-family-caring-app/id703761701. [retrieved: April, 2019].

[16] Apple Store, 2019, "Saleem", 2019, [online] Available at: https://apps.apple.com/us/developer/saleem/id1400326537. [retrieved: April, 2019].

[17] Apple Store, 2019, "Fitbit", 2019, [online] Available at: https://apps.apple.com/us/app/fitbit/id462638897. [retrieved: March, 2019]

[18] I. Fish, and M. Barnard, "Blockchain vs FHIR: a showdown for healthcare integration?" IBM Healthcare & Life Sciences Industries Blog, 2018, [online] Available at: https://www.ibm.com/blogs/insights-on-business/healthcare/blockchain-vs-fhir-showdown-healthcare-integration/. [retrieved: October, 2019]

[19] ProQuest, "The Insurance Implications of Blockchain", 2019, [online] Available at: https://search.proquest.com/docview/1881388609?pq-origsite=gscholar.[retrieved: January, 2019]

[20] S. Al-megren, S. Alsalamah, L. Altoaimy, H. Alsalamah, and L. Soltanisehat, "Blockchain Use Cases in Digital Sectors: A Review of the Literature," 2018 IEEE Confs on Internet of Things, Green Computing and Communications, Cyber, Physical and Social Computing, Smart Data, Blockchain, Computer and Information Technology, Congress on Cybermatics; IEEE. Halifax, 2018, pp. 1417–1424. doi:10.1109

[21] M. Rouse, "mHealth (mobile health)", 2018, [online] Available at: https://searchhealthit.techtarget.com/definition/mHealth. [retrieved: April, 2019]

 ISBN: 978-1-61208-759-7

User Privacy in Mobile Health Applications

Analysis of e-Pulse application

Tolga Çakmak
Department of Information Management
Hacettepe University
Ankara, Turkey
e-mail: tcakmak@hacettepe.edu.tr

Şahika Eroğlu
Department of Information Management
Hacettepe University
Ankara, Turkey
e-mail: sahikaeroglu@hacettepe.edu.tr

***Abstract*— Mobile health (mhealth) applications that are widely used for many different purposes such as tracking chronical diseases and lifestyle management have many advantages as well as potential limitations and privacy concerns. In this regard, it is seen that the studies with different points of view on the personal data, data protection, and privacy and security features and architectures of these apps are increasing in the literature. This study aims to evaluate the privacy and data protection conditions of the e-Pulse app that is developed by the Ministry of Health in Turkey. In line with this aim, personal data security and privacy conditions of the e-pulse app that is the official mobile health app of Turkey are described with a checklist developed according to the literature review. Privacy and personal data protection issues of the app are highlighted and potential recommendations are discussed in the results of the study.**

Keywords-mobile applications; privacy policies; mobile health applications; data protection; data privacy.

I. INTRODUCTION

Users can do many things that are necessary for the routine life processes with mobile devices and mobile apps, which are being an integral part of daily life. Applications installed on smartphones enable users to benefit from a variety of mobile internet services including personalized services. In recent years, it is seen that mobile apps affect businesses, social life, and lifestyle. The rapidly increasing use of mhealth apps also helps not only to improve the quality of life of the citizens but also to improve health services. mHealth apps that are used by physicians and patients to manage and observe health information, with their features, such as getting blood test results, glucose reading and displaying medical images and medical information convert mobile devices (tablets, smartphones, etc.) into medical devices [1]. In this regard, potential advantages of mhealth apps are stated as offering a fast diagnosis, providing feedback for monitoring health status, promotion of healthy behavior, providing easy access to treatment and rehabilitation, receiving electronic prescriptions, faster access to consent and reducing waiting times [2]. In light of this information, it is possible to describe many mhealth apps as the information or data providers increasing awareness and literacy level about health.

mHealth apps those usage increased especially with wearable medical devices provide many advantages for medical tracking and consent. However, these apps potentially have some limitations and privacy concerns. In this regard, it is seen that many healthcare app providers are not certified by any authority institutions and they do not have training about health information privacy and confidentiality. This is stated as one of the reasons for the concerns [1]. For instance, app providers can request to access users' contacts stored in mobile devices without any permission. When the user allows the app to see or use the GPS data, apps often disclose that this information will be sent to advertising companies. Moreover, it appears that many app users do not know where the information goes or how the developer plans to use it [3]. mHealth apps require many personal data to provide their services. This situation brings the user's privacy and security concerns. These concerns can be listed as monitoring by the unauthorized agents and sharing location or health data to third parties. As a matter of fact, given the distributed and wireless nature of sensor networks, the difficulty of ensuring data security in mobile apps is mentioned in the studies [4]. Privacy violations that may occur in the use of mhealth apps may harm users. Users who are subject to loss of personal data can lose their reputation and social insurance or they can face with employment discrimination [5][6]. At this point, it is very important for the app providers to make efforts to produce beneficial apps and to take precautions for user privacy and confidentiality in the apps considering the potential threats. In addition, it is seen that privacy policies are insufficient level in terms of clarity of data collection, storage, and transmission processes. It would not be wrong to say that the difficulty of the readability and understandability of these policies is another factor for this insufficiency. It is expressed that the main reason for this situation is lack of a standardized format and terminology [1][7][8].

mHealth apps with their potentials have been the subject of many scientific studies in recent years. In this context, it is seen that different models are developed on the reduction of privacy concerns and the improvement of systems in mobile apps. In general, it is seen that different models are suggested in the studies, which emphasize the importance of the reliability of mhealth apps [9]-[11]. There may be privacy issues in the architecture of mhealth apps, which benefit

 ISBN: 978-1-61208-759-7

from significant information communication infrastructures. In this context, one of the most important areas for mhealth apps to reach wider use areas and to utilize their potential benefits more effectively is the security and privacy opportunities offered by these apps for personal data. The privacy and security of mhealth apps can be examined within the framework of three components (such as law, culture/custom, and technology). In addition to the technological safety and user analysis of apps, it is seen that governments and app producers are working on the resources that include guidelines and recommendations for the implementation of privacy and security in apps [12]-[15]. On the other hand, in spite of this guidance published in different sectors, the lack of formal and one-stop standards and guidelines for implementation is stated in the studies [5]. mHealth apps have access to user-identifiable information and personal health information, including diagnosis, symptoms, and use of therapeutic services. Legislation on privacy seeks to protect individuals by requesting consent for the collection, use, disclosure or storage of personal data, including sensitive health information. However, many mobile apps are known to access private data without users' permission.

The usability and effectiveness of mhealth apps are mostly based on user data. These apps can directly request data from users. Furthermore, they can gather user data through a social network connection or location data. It is possible to state that these apps can collect a wider range of data that the user may not be aware of, such as the user's lifestyle, location tracking, social network connections. This situation increases the exposure of users to privacy attacks. In this context, users need to be aware of the increasing privacy risks with ever-changing mobile technology. In mhealth applications, it is seen that there is no comprehensive protection policy for the privacy of personal data. Privacy protections are mostly limited to the protection required by the developer's privacy policy. In this regard, the study shows that only 30% of the most commonly used 600 mhealth apps have a privacy policy [16]. On the other hand, since there is no standard privacy policy content, different applications have different content policies and lack of users' awareness about the issue is an important factor.

In mobile apps, the protection of patient privacy can prevent individuals from particular activities and lead them to remain passive. This may result in the loss of the right of the person to develop his / her material and spiritual existence. Respect for the private lives of persons benefiting from health services is important for the protection of their personal rights, dignity, and personalities. The protection of the patient's information is primarily the responsibility of healthcare providers. Although this responsibility of healthcare providers, users are required to improve their self-efficiency and awareness about privacy issues. Additionally, the state and authority institutions have also vital responsibilities regarding the privacy of the patients, and the implementation of related policies efficiently.

In this study, the privacy policies of the e-pulse application, which was developed by the Ministry of Health in Turkey are analyzed. Accordingly, the methodology of the research is discussed in the second part. The third section presents the findings. In the fourth and final section of the study, conclusion, and recommendations are given.

II. METHODOLOGY

This study aims to analyze privacy policies, data sharing and transmission processes, and privacy settings of the e-pulse mobile app, which is a mobile health application developed by the Republic of Turkey Ministry of Health within the scope of e-government approaches. The main feature of this app is to help citizens for tracking their daily activities such as steps and heart rates and their health records created as a result of the interactions with hospitals. In line with this information, the research questions of the study are as follows:

- What features does the e-pulse application have in relation to the protection of personal data?
- What kind of improvements does e-pulse need in terms of security of personal data and user privacy?

The qualitative research processes were followed in this study. In light of research questions, a checklist is developed according to the literature review. The checklist, which is used to analyze the e-pulse app consists of three sections titled privacy policy document, data sharing and handling, and privacy settings and security. Questions in each section have three answers (such as yes, no, partially) and a textbox for a description of current status. In addition, the checklist and the questions were reviewed through interviews with experts on personal data security. The data were gathered according to two techniques within the scope of content analysis. Firstly, the app was observed and its features used according to questions in the checklist. Secondly, the policy and help documents were analyzed to get deeper answers. Additionally, the gathered data were reported by the titles given in the checklist. Percentage values presented in each section were calculated by dividing the number of completed questions to the number of questions in the section.

III. FINDINGS

In this part of the study, the e-pulse platform, which is a personal health application developed within the scope of e-government applications in Turkey is evaluated in terms of the privacy policy document, data sharing and handling, and personal account information. The findings obtained from the evaluations are presented in this section.

A. Test-bed: e-Pulse App

mHealth apps, which are increasing in number and are widely used by individuals, are developed by governments as well as private developers. They are presented to citizens within the framework of health information systems and e-government applications. According to benefiting from management information systems in healthcare, the electronic pulse (e-pulse) app is developed as a mhealth app by the Ministry of Health in Turkey. The e-pulse app is a mobile and web-based information sharing and retrieval platform that enables citizens to reach all examination information, appointment, diagnosis, treatment, prescription, and medication details, allergy information, laboratory test

 ISBN: 978-1-61208-759-7

results and radiological images with their reports in all health facilities in Turkey. Containing many goals, the app aims to prevent patients from recurrent examinations, to allow both users and physicians to access recent and previous health data, and to make a contribution to the efficiency and effectiveness of health services.

The patients can evaluate the quality of health services they received and express their opinions and complaints by using the app. Patients also can record their data (such as blood pressure, sugar, and pulse data, steps) to their profiles and display this information comparatively. Patients who can access their health data from anywhere with the E-Pulse app will be able to use this data without having to contact the hospitals. The E-Pulse app can be used from a personal computer with its web-based structure. It is also accessible from any smart device with any of the operating systems (Windows Phone, IOS or Android).

B. Privacy Policy Document

At the beginning of the assessment, the privacy document is evaluated in terms of its structure and clarity. It is seen that the application has a privacy document, but this document also includes usage and copyright explanations of the app. Although the scope of the document is not limited to the privacy policy, the document is named as “privacy - data protection statement”. Additionally, this document does not contain a summary and a section of links that help readers to navigate through the document. Moreover, the document does not have the definitions of privacy terms and dates reflecting the latest revisions and updates. Plus, it is seen that the application has a feature to notify users about any updates in privacy conditions.

Secondly, the scope of the privacy policy document is analyzed according to expressions related to data collection, sharing and storing processes. In this context, findings indicate that the privacy document clearly informs users about the gathered data and how these data are used. In contrast to this finding, the methods being used for data collection processes are not explained in the document. It is also seen that the document states which information will be shared with third parties and the sharing purposes. Accordingly, findings reflect that the data protection measures are slightly presented in the document. On the other hand, the document does not cover how long the gathered data will be stored and its deletion procedures. Although the contact information is provided in help documents, the privacy document does not contain this information.

Evaluations about the privacy policy document of the e-pulse app show that the document is insufficient in terms of scope and structure. The document meets only five (38%) of 13 criteria that are expected to be given in a privacy policy document.

C. Data Sharing and Handling

In the second section of the survey, data sharing and handling functions of the e-pulse system analyzed with 12 questions. These questions are categorized under three titles. These categories are data sharing with third parties, policies, and accessibility.

Findings illustrate that the app is in a very sufficient level in terms of data sharing with third parties. Since the health organizations, mostly hospitals, feed the app with patient transactions in data creation processes, the data handled in third parties is not used in the app. Additionally, the app does not make a request for creating persistent cookies. The third-party tracking cookies blocked by web browsers are not used in the app. Lastly, the privacy policy document strongly emphasizes the conditions related to data sharing and usage conditions with third parties.

Analysis indicates that the e-pulse application provides a secure connection (https) in the web platform and provides access to whole data about the user. It is also seen that the app authenticates the user via e-government account information. Accordingly, users are allowed to modify or delete some of the data such as comments, weight, and size. In contrast, users can not make any changes in their health records. They can only view this data by using the app. It is a remarkable finding that the privacy document partially informs the users about these features. The app also does not provide an explanation of data transmission and storage processes in the privacy document. Findings reveal that the app successfully performs 10 of 12 questions. Findings also demonstrate that one question is partially carried out. In this context, the efficiency of the app in data sharing and handling issues was measured as 87.5%.

D. Privacy Settings and Security

As the third section of the research instrument, privacy settings and security features offered for personal accounts were assessed with 11 questions. In this regard, the authentication process of the app during the login was analyzed. Findings reflect that the app uses a two-factor authentication process and validate users by sending an SMS. Additionally, the app allows users to set which profile information or transaction will be shared with whom.

The app provides five options related to sharing patient transactions with third parties. These options are sharing health data with the family doctor, only the doctor who examines the user, all doctors in the hospital, all doctors employing for the Turkish Ministry of Health or sharing the health data with no one. When the user selects this option, the app asks for approval with SMS. Accordingly, although the user selects "do not share my data with doctors" option, the app allows whole doctors in the hospital where the user made an appointment will see the health data without approval. The related notification is also given under this option. It is a remarkable finding that the privacy policy document does not contain any information about these options, but help documents partially provide these details by giving screenshots.

The app also helps users to make their privacy settings by providing a default setting about sharing health data.

 ISBN: 978-1-61208-759-7

Moreover, users are expected to make their privacy settings in the first use of the app and they are directed to settings panel after the download.

The app allows users to snooze, close or re-open their accounts. In parallel with the findings of the data sharing settings, the privacy policy document does not cover these functions. Users can only learn their rights related to snoozing or closing their accounts via help documents linked on the official web page of the app. Lastly, the app has data validation features and asks users to provide feedback about the misinformation presented in their profile panel. In general, the assessments about privacy settings and security reveal that the app provides many options and features related to privacy and security. It is seen that the technical part of the app meets the basic criteria for privacy and security settings. Although given advantages, documentation of these settings is seen insufficient level especially regarding privacy policy documents. As a result of these insufficiencies, the app performs eight questions completely and it is seen that the requirements in one question are partially carried out. The performance of the app in this part is measured as 70.3%.

E. General Overview

In addition to specific results presented in previous sections of the study, the app was analyzed in terms of general scores obtained according to responses to questions in each section. In this context, the general overview of the app is illustrated in Figure 1 to display obtained percentages of each component, and to compare the levels of each section.

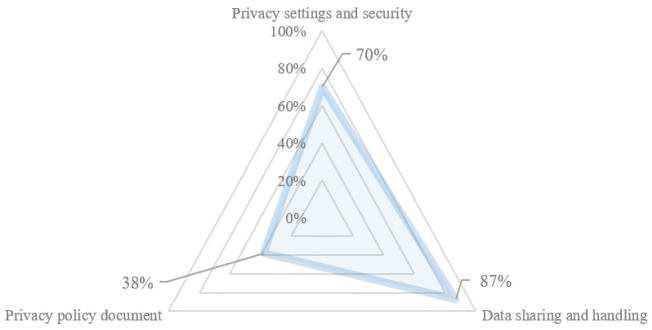

Figure. 1. A general overview of the analyzed components

According to the general situation illustrated in Figure 1, the level of data sharing and handling functions (87%) which are mostly based on the technical processes are higher than the other two sections analyzed in the study. In contrast, the privacy policy document is the weakest component (38%) of the app.

IV. CONCLUSION

The e-Pulse app is an app that is used in Turkey to meet goals expected from mhealth apps such as ensuring that patients have more information about their health information and their participation in treatment decisions. As is in other mhealth apps, processes like patient privacy, security, and privacy of personal data are important for the e-pulse app. Since the app is developed by the Turkish Ministry of Health, all data created in Turkish health institutions are accessible by the app. This situation makes the app's data processes more important. According to the analysis conducted with a checklist developed in this study, it is understood that the app has a privacy, use and copyright agreement. Assessments based on the structure and clarity of the agreement reveal that although the agreement has headings related to the security of personal data, limits of responsibility and copyrights, the provided information under these titles has not been made clear and detailed. It is seen that the processes about data sharing with third parties are expressed in the agreement. In contrast, information about data collection, processes and storage conditions are not expressed. Assessments reflect that the responsibilities during the loss or vulnerability of personal data are not accepted by the app developers. This situation is considered a significant deficiency in the security and confidentiality of the data of the users. Although the application provides users' confidentiality, use, and copyright agreement, the fact that this agreement does not contain detailed information on the confidentiality, security, and information about the protection of patient privacy. These insufficiencies constitute a weakness for the users who need to be clearly informed before using the app. It is clear that there are deficiencies in the documentation for the functions and features related to the privacy and data collection processes of the app. Additionally, evaluations of the application's privacy policy document point out that the document is insufficient in terms of scope and structure. In this context, it is recommended privacy policy document of the app should be improved in terms of data collection, processing and storage issues.

The results presented under the Privacy Settings and Security section of the study reflect that the app provides users five different security options. In this part, users are allowed to select one or more options given by the app. It is remarkable that there is a description given under one of these options. It is located under the "no physician needs to see my data (with SMS code or password required by this option is checked)" option. The description is "If you check this box and make an appointment with the Central Patient Appointment System, all physicians in the relevant hospital will be able to access your health records without obtaining further approval during that examination day." Although the user checks "the physician is not allowed to see the user's data" option, all physicians in the hospital could see the records on the appointment day. This situation conflicts with patient privacy that is also emphasized in the Patient Right Regulation in Turkey [17]. The 16th article of this regulation is "The patient may examine the file and records, which have information on his/her health status, directly or through his/her representative or legal representative and take a copy. These records can only be seen by those directly involved in the treatment of the patient". In addition, the app's privacy policy document does not provide detailed information about

 ISBN: 978-1-61208-759-7

these sharing options. On the other hand, two-factor authentication function and user feedback features of the app were evaluated as positive security and validation methods in the study.

It is understood from the results of the analysis conducted on the e-pulse that the app has deficiencies in the scope of the privacy, use and copyright agreement and this document needs to be improved regarding this issue. In this respect, the application developers are recommended to inform the users about the data collection, storage, and distribution in the privacy policy of the app. As another suggestion, the policy developers of the app are recommended to ensure the readability and intelligibility of the policy documents. Lastly, as stated in the literature, the lack of standardized format and terminological unity in these policies are considered as one of the weaknesses in this topic. In this regard, publications like standards and guidelines that can be published by authority institutions are seen as one of the major steps.

REFERENCES

[1] M. Rowan and J. Dehlinger, "A Privacy Policy Comparison of Health and Fitness Related Mobile Applications," Procedia Computer Science, vol. 37, pp. 348-355, 2014, doi:10.1016/j.procs.2014.08.051.

[2] G. Catan, R. Espanha, R. Veloso Mendes, O. Toren, and D. Chinitz, "The Impact of eHealth and mHealth on doctor behavior and patient involvement: an Israeli and Portuguese comparative approach," Stud Health Technol Inform. vol. 210, pp. 813-817, 2015.

[3] A.M. McDonald and L.F. Cranor, "The Cost of Reading Privacy Policies," vol. 4:3, pp. 543-568, 2008.

[4] D. Ding, M. Conti, and A. Solanas, "A smart health application and its related privacy issues," 2016 Smart City Security and Privacy Workshop (SCSP-W), pp. 1-5, 2016, doi:10.1109/SCSPW.2016.7509558.

[5] L. Parker et al. "A health app developer's guide to law and policy: a multi-sector policy analysis," BMC Medical Informatics and Decision Making. vol. 17, pp.1-13, 2017, doi:10.1186/s12911-017-0535-0.

[6] Privacy and Mobile Apps, Office of the Information Commissioner. [Online]. Available from: https://www.oic.qld.gov.au/guidelines/for-government/guidelines-privacy-principles/applying-the-privacy-principles/privacy-and-mobile-apps 2019.03.28.

[7] M. Al Ameen, J. Liu, and K. Kwak, "Security and Privacy Issues in Wireless Sensor Networks for Healthcare Applications," J Med Syst., vol. 36, pp. 93-101, 2012, doi:10.1007/s10916-010-9449-4.

[8] T. Giannetsos, T. Dimitriou, and N.R. Prasad, "People-centric sensing in assistive healthcare: Privacy challenges and directions," Security and Communication Networks, vol. 4, pp. 1295-1307, 2011, doi:10.1002/sec.313.

[9] P. Dadhich, K. Dutta, and M.C. Govil, "Trust Enhanced Authorization for Distributed Systems," International Journal of Scientific & Engineering Research, vol. 2, pp. 1-7, 2011.

[10] C. Lin and V. Varadharajan, MobileTrust: a trust enhanced security architecture for mobile agent systems, International Journal of Information Security. vol.9 (3), pp. 153–178, 2010. doi:10.1007/s10207-009-0098-x.

[11] Z. Yan, Y. Chen, Y. Shen, "PerContRep: A Practical Reputation System For Pervasive Content Services," J Supercomput, vol. 70, pp.1051-1074, 2014, doi:10.1007/s11227-014-1116-y.

[12] App Review - App Store - Apple Developer. [Online]. Available from https://developer.apple.com/app-store/review/ 2019.03.28.

[13] GSMA Public Policy | Privacy Design Guidelines for Mobile Application Development. [Online]. Available from https://www.gsma.com/publicpolicy/resources/privacy-design-guidelines-mobile-application-development 2019.03.28.

[14] Medical devices: software applications (apps). [Online]. Available from https://www.gov.uk/government/publications/medical-devices-software-applications-apps 2019.03.28.

[15] Office of the A.I. Commissioner, Mobile privacy: a better practice guide for mobile app developers - Office of the Australian Information Commissioner (OAIC). [Online]. Available from https://www.oaic.gov.au/agencies-and-organisations/guides/guide-for-mobile-app-developers 2019.03.28.

[16] A. Sunyaev, T. Dehling, P. Taylor, K. Mandl, "Availability and Quality of Mobile Health App Privacy Policies," Journal of the American Medical Informatics Association, vol. 22, pp. e22-e28, 2015, doi:10.1136/amiajnl-2013-002605.

[17] Patient Rights Guide, Legislation for Physicians and Managers of Medical Chamber. [Online]. Available from http://www.ttb.org.tr/mevzuat/index.php?option=com_content&view=article&id=984:hasta&catid=26:etik&Itemid=65 2019.05.06.

 ISBN: 978-1-61208-759-7

Motivation Enhancement in mHealth via Gamification

Felix Kettner, André Langer, Kim Schmidt, and Martin Gaedke

Distributed and Self-organizing Systems
Chemnitz University of Technology
Chemnitz, Germany
Email: felix.kettner@informatik.tu-chemnitz.de
andre.langer@informatik.tu-chemnitz.de
kim.schmidt@mail.de
martin.gaedke@informatik.tu-chemnitz.de

***Abstract*—The lack of therapy compliance (adherence) is an already well-studied phenomenon in the health sector. It is known that numerous factors can influence regular medication intake. A promising approach for increasing therapy compliance are patient-related factors. In order to increase motivation, Gamification-based approaches have already been selectively tested here. However, the number of mHealth applications using the potential of Gamification in this context in an effective way is limited so far. Our research therefore focuses on game-design-elements (GDE) that are particularly suitable to increase motivation in the mHealth sector. This includes a well-founded analysis of existing studies and applications with the identification of potential Gamification approaches. A novelty here is the use of comprehensive GDEs to increase motivation in the mHealth area. The result is a classification of GDEs according to their effectiveness and a conceptual design on how to apply them effectively, which is implemented prototypically in a progressive WebApp. Finally, the suitability of our approach for generating motivation was assessed through a user survey. The results show a significant influence on the use of GDEs on the development of Flow, and thus on motivation. Differences in the isolated effect of individual game design elements could not be proven.**

***Index Terms*—Adherence; Gamification; Motivation theory; Medication; mHealth**

I. Introduction

Digitization is a continuously advancing process affecting almost all domains in the everyday life [1]. The number of digital offers also increases in the health sector [2]. Already in 2015, the number of mobile available applications exceeded 103,000 in the *health and fitness* and *medicine* categories [3]. The spectrum ranges from simple, general reminder applications to apps that are highly specialized on certain diseases.

According to the world-health organization (WHO), only 50% of all patients comply with their agreed treatment plan in industrialized nations [4]. In case of a lack of adherence, those patients are often confronted with negative health consequences. In addition to individual effects, such as the worsening of the disease, relapses, drug resistance and a concomitant reduction in quality of life, there are also societal risks, such as an increased risk of spreading contagious diseases and economic costs [4].

According to Neckermann and Schinköthe [5], an improvement of adherence is especially in the domain of the patient- and therapy-related factors indicated. Verbrugghe et al. [6] divide the patient-related factors into conscious and unconscious reasons for non-adherence. The conscious factors are those that may have a negative impact on a patient's motivation to take medication, like the lack of motivation or information [7] and are a promising starting point for increasing adherence [5].

Other studies indicate that the majority of younger citizens already owns a smart-phone [8] and has a medium or high interest in e-health [9]. Therefore, the use of digital health-care applications can be one adequate way to get in touch with patients and help them to build adherent habits.

However, there are hardly any studies on the efficacy of these applications [10]. Although the promising possibilities of using digital applications for adherence enhancement are discussed [5], [11], valid long-term studies are still pending. So, while the effect of individual applications could hardly be validated so far, a few application providers already started to use a method for which studies are available: Gamification. It has already been shown that adherent behavior can be significantly increased with the help of game design elements (GDEs) [12]–[14]. However, GDEs so far seem to be used only isolated in applications for adherence enhancement, or their usage is limited to an application for certain diseases.

This work discusses the potential of Gamification in conjunction with mobile health applications and examines the research question, which principles can be used to design and use GDEs in increase the patient's motivation effectively in order to contribute to a better adherence. We focus on patient-related conscious factors in order to generate intrinsic motivation through the incentive of extrinsic motivation and to promote sustainable habitual education. Therefore, we present the following results:

- a categorization to classify GDE according to their effectiveness,
- a concept for the combined use of GDE in mHealth apps in the field of medication intake,

 ISBN: 978-1-61208-759-7

- evidence of the significant suitability of GDE to support patient compliance through a proof-of-concept based user study.

The rest of the work is structured as follows: Section II introduces the theoretical foundations of Gamification and the theory of motivation, self-determination and the concept of Flow. Section III presents existing solutions in the same problem area. Section IV defines requirements for GDE use in mHealth and shows our concept for the effective use of Gamification An example implementation and the evaluation are described in Section V. Section VI finally provides a summary and outlook on further research.

II. GAMIFICATION FOR THERAPY COMPLIANCE

We focus on the definition of Sailer [14] for *Gamification*:

> *"Gamification is a process of playfully designing activities in a non-game context through the use of game design elements."*.

Gamification can be used both to generate motivation and in the mHealth area to achieve a more regular medication intake. In general, the term motivation refers to those psychological processes that serve to initiate, direct, and sustain mental and physical activities [15].

A. Self-determination theory

The basis for this is the self-determination theory according to Deci and Ryan [16]. They postulate three basic psychological needs, which are innate, universal and pursuited by the human psyche. These are:

- competence
- relatedness
- autonomy

According to this, every person strives to interact effectively with their social environment and gain a sense of confidence and effectiveness in their own actions (competence), to feel connected to a social structure and other people (relatedness), as well as acting to their own interests and values (autonomy) [16]. Furthermore, Deci and Ryan [17] distinguish between intrinsic and extrinsic motivation. Intrinsic motivation refers to the performance of an activity on the basis of its own inherent satisfaction, while extrinsic motivation requires behavior directed towards a result that is separable from it.

B. Flow

Regarding the maintenance of motivation, supplements from the Flow theory according to Csikszentmihalyi et al. [18] are considered as an extension. They provide a psychological explanatory model for behavior that is repeatedly demonstrated and maintained despite the lack of extrinsic rewards, and thus corresponds to intrinsically motivated behavior according to Deci and Ryan [17]. Flow describes a subjective state in which an individual is engrossed in an activity and forgets about time, fatigue, and everything else apart from the activity itself [18]. The central elements of Flow are [19]: merging of action and awareness, centering of attention, loss of reflective self-consciousness, sense of control over action and environment, demands for action and clear feedback and autotelic nature of action.

In order to effectively use Gamification to increase motivation, it must meet the requirements of self-determination theory according to Deci and Ryan [16] and allow Flow [20]. Thus, through the proper use of GDEs, which generates Flow, the basic needs for autonomy, competence, and social inclusion can be satisfied and intrinsic motivation generated.

C. Review of game design elements

In the following, we examined various GDEs for their suitability for generating Flow. A rating was done in five different gradations. The criteria used for this are derived from the central elements of Flow in Section II-B. Elements, which contribute without restriction to the generation of Flow are marked as *"conducive"*. Once certain prerequisites have to be fulfilled for an element to be supportive, it is considered *"conditionally conducive"*. Elements that have no influence on the respective Flow aspect, are rated with *"no effect"*. Elements representing an obstacle under certain conditions are considered as *"conditionally obstructive"*, and elements that generally have a negative effect on Flow are marked as *"obstructive"*. Figure 1 shows a summary of the review.

Flow \ Game-design-elements	Avatar	Badges	Leaderboards	Progression	Performance graph	Levels	Missions	Narrative	Points	Team leaderboards
Centering of Attention	(✓)	(✓)	o	o	o	(✓)	(✓)	(✓)	(✓)	o
Loss of self-consciousness	(✓)	(✓)	o	o	o	(✓)	(✓)	(✓)	(✓)	o
Sense of control	(✓)	(✓)	(✗)	(✓)	(✓)	(✓)	(✓)	o	(✓)	o
Demands for action	(✓)	✓	✓	✓	✓	✓	✓	(✓)	✓	✓
Feedback	(✓)	✓	✓	✓	✓	✓	o	(✓)	✓	✓
Autotelic	✓	✓	(✓)	✓	✓	✓	(✓)	(✓)	✓	(✓)

✓ conducive (✓) conditionally conducive o no effect
✗ obstructive (✗) conditionally obstructive

Fig. 1. Assignment of GDEs to central Flow criteria.

D. Categorization and Summary

Figure 1 shows that GDEs are well suited to generate and maintain Flow, but not all equally and easily. The rating in the previous section allows to classify different GDEs into three categories, as shown in Table I.

GDEs, which can be used in principle for all areas of Flow and have no negative effects on motivation, even in an implementation not focusing on flow, are grouped into category I and are thus to be used preferentially.

If certain GDEs do not expect a positive effect on individual Flow aspects and, at the same time, have no negative effect,

 ISBN: 978-1-61208-759-7

TABLE I. CATEGORIES OF GAME DESIGN ELEMENTS

Category	Description	Game design elements
I	for all aspects of Flow (conditionally) conducive	avatar, badges, levels, points
II	for no aspect of Flow (conditionally) obstructive but also not for all aspects (conditionally) conducive	progression, performance graph, missions, team leaderboards
III	for a minimum of one aspect of Flow (conditionally) obstructive	leaderboards

they are assigned to category II. A use of these elements is nothing to oppose.

Category III includes those GDEs that can negatively affect at least one area of Flow. In motivational applications, the use of such elements should be avoided whenever possible.

In addition to the right balance between the skills of the user and the requirements of the application, a sensory overload has to be avoided and a clear set of rules has to be implemented. GDEs should not only be used for their own sake, but always with regard to the intended effect on the user and taking Flow and self-determination theory into account.

III. RELATED WORK

Various papers provide an overview of existing applications for increasing adherence and their effectiveness. A good first overview in this area offers [11]. But there is no focus on Gamification. Other work shows the use of Gamification for increasing motivation [14], [21] without considering the use for increasing adherence.

The most advanced approaches are the use of Gamification with single diseases, such as diabetes [12], [13] with the disadvantage of a very small target group. So, in lack of a broad scientific basis the following analysis refers to existing applications.

Prerequisite for the selection of an application for closer examination are the support with medical adherence and the use of GDEs. Thereby 11 applications where found within the categories reminder applications, applications for building habits, to convey information and for life-logging. Out of these only 4 where to be considered. This is because to the fact that the others where not longer available or at least not available in Europe. The remaining 4 applications are Asthma Action Hero, Habitica, Mango Health and MySugr.

Asthma Action Hero and MySugr are specialized to certain diseases and Habitica is not designed for medical purposes but highly gamified and customizable enough for the intended use. Figure 2 shows the results of the analysis using the requirements introduced in Section IV-A.

The advantage of the introduced concept in this work is the full consideration of all needed aspects to increase medical adherence by using Gamification with a wide target group. So the providing of disease related information, a scientifically founded approach for generating and maintaining motivation and also functional requirements are taken into account.

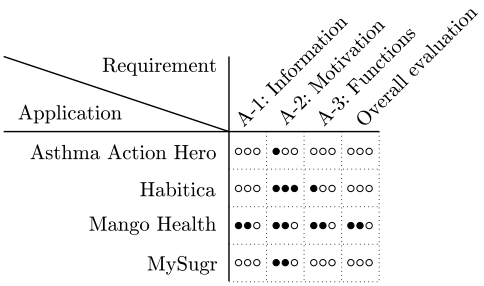

Fig. 2. Rating of Existing Applications.

IV. CONCEPT

Patients-related, conscious factors for non-adherence are a promising starting points for improving adherence to therapy. By adequately communicating facts, the lack of information can be reduced as a motivation obstacle. Furthermore, generated and already existing motivation must be sustainably maintained.

The requirements to be met derive from the foundations of self-determination theory according to Deci and Ryan [16], Flow according to Csikzentmihalyi et al. [18] and the study of GDE in Section II.

A. Requirements

Below is a list of requirements for an mHealth application that uses Gamification to generate motivation and Flow to improve adherence.

A-1: Mediation of disease-related information

A-2: Generating motivation
- **A-2.1:** Faciliating Flow
  - **A-2.1.1:** Centering of attention is made possible
  - **A-2.1.2:** Loss of reflective self-consciousness is made possible
  - **A-2.1.3:** Sense of control over action and environment is made possible
  - **A-2.1.4:** Demands for action are generated
  - **A-2.1.5:** Clear and direct feedback is given
  - **A-2.1.6:** Autotelic use is made possible
- **A-2.2:** Principles of self-determination theory are considered
  - **A-2.2.1:** Feeling of competence is made possible
  - **A-2.2.2:** Feeling of autonomy is made possible
  - **A-2.2.3:** Feeling of relatedness is made possible

A-3: Functional requirements
- **A-3.1:** Individual medication input possible
- **A-3.2:** Reminder function available

B. System components

This results in three essential services that must be taken into account when designing an application concept:

- Fulfillment of functional requirements,
- Generation of motivation and
- Education and information transfer.

 ISBN: 978-1-61208-759-7

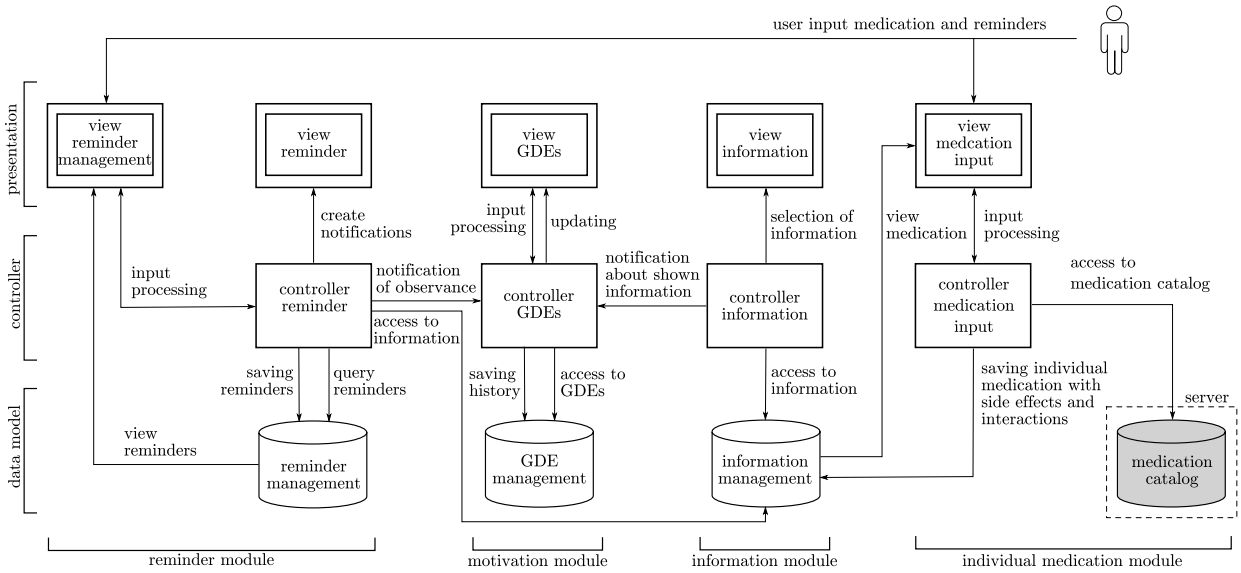

Fig. 3. Application Modules as a Complete System with Layers of the MVC Pattern.

This allows a modularization and thus separation of responsibilities based on the Model-View-Cotroller pattern (MVC). The range of functional requirements may still be further divided into the subcategories of the individual medication input and the reminder function. Figure 3 provides an overview of the application modules as a complete system. The following describes the functionality of each component.

1) Reminder module: The reminder module at the presentation layer level consists of the reminder administration view for entering the reminders and a reminder component for notifying the user of the due date of a scheduled medication and a demurring stock of medicine. The user inputs and reminders are processed by the control layer of the module. It also contains interfaces for storing and querying medication times in the data model layer, as well as for communicating with the controller of the GDEs.

2) Motivation module: The motivational module includes the presentation of the GDEs, their processing in the control layer, and the management of the GDEs in the data model layer. The controller of the GDEs has interfaces for communication with the controller parts of the reminder and information module in order to be able to process information about (non-) compliance with the deadlines and the information already presented. Application usage history data, such as obtained *points* or *levels*, is stored in the management part and retrieved from the control layer as needed.

3) Information module: The information module is responsible for the storage, selection and presentation of disease-related information. The drug-specific information on the interactions and side effects of the drugs entered are read from the control section of the module for individual drug input via the network from a corresponding database. Subsequently, this data is stored locally in the data memory of the information module. Furthermore, the module has an interface to the control layer of the motivation module at the level of the control layer. The content of the information module is divided into a learning part and a *quiz* part. In the learning part, the information is presented, while in the *quiz* section, this information is consolidated with the support of GDEs.

4) Individual medication module: In the presentation layer, a user interface is provided for individual drug input, alternatively, a barcode scan allows direct reading of the medication. The control layer of the module realizes the remote access to the drug catalog and the storage of the data at the level of the data model layer in the information module.

5) Cross-module content: To facilitate the use of the application, an optional tutorial is helpful. As an introduction to the functionalities and operation of the app, this must cover all modules in terms of content. The usage of tooltips is encouraged, which appear next to the corresponding buttons when first used. Overall, the design of the application should also take into account the guidelines of ISO 9142-10 *"Principles of Dialog Design"*.

C. Selection of Game Design Elements

Not all GDEs are equally suitable for all purposes. Furthermore, the combination of certain GDEs may result in a more immersive user experience as they complement and support each other. The use of *points* makes it easy to see a *level* progress, *badges* can easily be used hand in hand with a *progress bar* and for visualizing *missions*. There are multiple possibilities in the design and the combination of the individual GDEs.

Due to their particular suitability, all GDEs from *category I* are selected for use in the proof-of-concept (*points*, *badges*, *levels*, *avatars*). In addition, to improve and expand the interaction of the GDEs, from *category II progress bar* and *performance graph* are used.

The overall concept is based on the targeted use of the GDEs to generate Flow, taking account of the self-determination

 ISBN: 978-1-61208-759-7

theory. A novelty here is the use of comprehensive GDEs to increase motivation in the mHealth area.

V. Evaluation

To evaluate the fulfillment of the requirements from Section IV, an empirical analysis was carried out as a user survey with non-experts and an online questionnaire with explicit research questions, including a demonstrator.

The main component of the demonstrator is the implementation of the motivational module to test the interaction of the various GDEs to generate Flow. The implementation of the special functional requirements of reminders and medication input has been put into the background. Likewise, the module of information transfer is implemented only to the extent necessary to clarify the interaction with the motivation module.

The implementation is based upon React + Bootstrap and was done as a progressive web app and single-page application. A working version of the demonstrator can be found here (July 2019): https://www.purl.org/net/vsr/gamification.

A. Hypotheses

The following falsifiable hypotheses served as basis for the operationalization:

- Hypothesis I: *All GDEs used in the demonstrator are equally suitable for generating Flow.*
- Hypothesis II: *The development of Flow is not possible with the GDEs used in the demonstrator.*

Variables and indicators were defined and questions were formulated for the survey to prove these hypotheses. The most relevant questions from the survey are listed in Table II.

TABLE II. Selection of survey questions

No.	Question
5	How concentrated did you feel during application usage?
6	Did you think of anything else than the application during application usage?
7	How understandable did you find the context of the application?
8	How strong was your feeling in controlling the application?
9	How well did the application respond to your input?
10	How intuitive did you find the application control?
11	How motivating did you find the following items when using the application?
12	How necessary do you consider the following application elements to motivate you to take your medication?
13	How much did you feel motivated by the following elements to try the application further?

B. Results

In total, 91 people participated in the online survey over a 14-day period. Of these, 53 completed the questionnaire. The age range of the participants was between 18 and 75 years, focusing on the age group 26 to 35 years. The proportion of female participants was 69.81% (37 people), the male gender was 28.30% (15 people), and one person did not specify their sex. Of the 53 participants, 43 (81.13%) stated that they had already taken medication or did so regularly.

When testing for Flow, the mean of all answers to each question is above the expected mean ($p < 0.01$) as shown in Table III. This shows that the GDEs used in the demonstrator are able to produce Flow. Thus, Hypothesis II is refuted and the premise for Hypothesis I is met.

TABLE III. Pearson Chi-Square-Test of the Flow Related Questions

Question number	μ	σ	χ^2	α
5	3,68	1,09	47,14	0,001
6	1,89	0,32	31,72	0,001
7	3,47	1,01	17,24	0,01
8	3,53	1,01	23,49	0,001
9	3,85	1,05	75,93	0,001
10	3,72	1,03	47,93	0,001

TABLE IV. Friedman-Test for the Game Design Element Related Questions

No.	GDE	μ	σ	χ^2_F	α
11	Avatar	3,28	1,24		
	Points	3,23	1,15		
	Badges	3,38	1,20		
	Performance graph	3,32	1,12		
				3,86	$> 0,05$
12	Avatar	3,02	1,38		
	Points	3,16	1,23		
	Badges	3,19	1,48		
	Performance graph	3,57	1,34		
				6,48	$> 0,05$
13	Avatar	3,15	1,34		
	Points	3,00	1,27		
	Badges	3,28	1,34		
	Performance graph	3,13	1,23		
				0,46	$> 0,05$

Further questions served to examine differences in the effect size of the individual GDEs in the generation of motivation. However, the Friedman used showed no significant difference in the evaluation of the individual GDEs by the subjects ($p > 0.05$) as depicted in Table IV.

With the help of the user survey could be shown that through the use of the tested GDEs, as provided in the draft concept, Flow is enabled. From this, the usability of GDEs to increase motivation in medication adherence was derived. Hypothesis II could be falsified with the results of the survey, Hypothesis I on the other hand not.

However, the assumption of the different effects of the GDEs has not been confirmed. A possible explanation for this is that the effect on motivation is significantly dependent on the combination of the GDEs and their interaction in the field of dynamics, rather than on the individual elements. From this, it can be deduced that GDEs should not be used in isolation, but rather develop their effect in interaction with other GDEs. An affirmation of this thesis would also support the requirements created in Section IV.

One last question was aimed at the general acceptance of the demonstrator by the respondents. Of the 53 respondents, 34 (64.15%) stated, they would consider using a full version of the demonstrator when taking their own medication. The Pearson Chi-Square-Test shows a significant deviation ($p < 0.05$) of the actual distribution towards the readiness for use with an expected equal distribution.

 ISBN: 978-1-61208-759-7

VI. Conclusion and Future Work

The task and objective of this work was the analysis and testing of which GDEs are particularly suitable for increasing motivation in the mHealth domain. The focus was on increasing medical adherence.

First of all, a rating scheme for GDEs was designed, followed by categories for classifying GDEs and application concept requirements. Appropriate GDEs were considered to be *points*, *badges*, *levels*, *progress indicators*, *avatars*, and *performance graphs*. As a result, a WebApp was implemented as a proof-of-concept, including the conceptual motivational module. This demonstrator was tested by a user survey.

We found that using GDEs enables Flow. From this, the usability of the tested GDEs can be derived to increase motivation in the use of medication intake. However, the various GDEs studied are not significantly different in their individual effect on user motivation.

Further future research is indicated. In this work, we focused on the motivation module. For a more comprehensive evaluation and assessment of the concept, the implementation of a full-featured prototype, including an information module and other functional requirements, would be expedient. As part of the user survey, no impression of the duration of effect on the user motivation could be obtained. However, medications often have to be taken over a longer period of at least several days. To evaluate the long-term motivation, a study over a longer period would be necessary. To improve the long-term motivation, an extension with additional GDEs like *missions* or a *narrative* could be useful.

References

[1] S. Kemp, "Digital in 2018: Essential insights into internet, social media, mobile, and ecommerce use around the world," 2018, 2019.09.24. [Online]. Available: https://digitalreport.wearesocial.com/report_download

[2] Statista, "Volumen des eHealth-Marktes weltweit in den Jahren von 2013 bis 2020 (in Milliarden US-Dollar) (global ehealth market volume from 2013 to 2020 (in billions of dollars))," 2019, 2019.09.24. [Online]. Available: https://de.statista.com/statistik/daten/studie/712698/umfrage/volumen-des-ehealth-marktes-weltweit/

[3] "Chances and Risks of Mobile Health Apps (CHARISMHA)," 2016, 2019.09.24. [Online]. Available: https://publikationsserver.tu-braunschweig.de/receive/dbbs_mods_00060000

[4] E. Sabaté, *Adherence to long-term therapies: evidence for action*. World Health Organization, 2003.

[5] K. Neckermann and T. Schinköthe, "Adhärenz: Fehlende therapietreue und möglichkeiten der steigerung (adherence: Lack of adherence and potential for enhancement)," *Forum*, vol. 33, pp. 140 – 142, Apr 2018.

[6] M. Verbrugghe, S. T. L. Verhaeghe, K. Lauwaert, D. Beeckman, and A. Van Hecke, "Determinants and associated factors influencing medication adherence and persistence to oral anticancer drugs: A systematic review," *Cancer treatment reviews*, vol. 39, pp. 610 – 621, 2013.

[7] P. Behner, A. Klink, S. Visser, J. Böcken, and S. Etgeton, *Effekte einer gesteigerten Therapietreue: Bessere Gesundheit und höhere Arbeitsproduktivität durch nachhaltige Änderung des Patientenverhaltens (Effects of increased adherence: better health and higher productivity by sustained change of patients behavior)*. Bertelsmann Stiftung, 2012.

[8] M. Haas, "Smartphone-Markt: Konjunktur und Trends (smartphone market: business cycle and trends)," Bitkom - Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V., 2018, 2019.09.24. [Online]. Available: https://www.bitkom.org/Presse/Presseinformation/30-Milliarden-Markt-rund-um-das-Smartphone.html

[9] Statista, "Wie hoch ist Ihr generelles Interesse am Thema E-Health? (what is your general interest in e-health?)," In Statista - Das Statistik-Portal, 2017, 2019.09.24. [Online]. Available: https://de.statista.com/statistik/daten/studie/698066/umfrage/umfrage-zum-interesse-am-thema-e-health-in-deutschland-nach-geschlecht/

[10] I. Ahmed, N. S. Ahmad, S. Ali, S. Ali, A. George, H. Saleem Danish, E. Uppal, J. Soo, M. H. Mobasheri, D. King, B. Cox, and A. Darzi, "Medication adherence apps: Review and content analysis," *JMIR Mhealth Uhealth*, vol. 6, no. 3, p. e62, Mar 2018.

[11] A. Choi, A. W. Lovett, J. Kang, K. Lee, and L. Choi, "Mobile applications to improve medication adherence: Existing apps, quality of life and future directions," *Advances in Pharmacology and Pharmacy*, vol. 3, pp. 64 – 74, 2015.

[12] J. A. Cafazzo, M. Casselman, N. Hamming, D. K. Katzman, and P. M. R., "Design of an mHealth App for the Self-management of Adolescent Type 1 Diabetes: A Pilot Study," *JMIR*, vol. 14, p. e70, 2012.

[13] K. J. Rose, M. König, and F. Wiesbauer, "Evaluating success for behavioral change in diabetes via mhealth and gamification: Mysugr's keys to retention and patient engagement," *Diabetes Technology & Therapeutics*, vol. 15, p. A114, 2013.

[14] M. Sailer, *Die Wirkung von Gamification auf Motivation und Leistung; Empirische Studien im Kontext manueller Arbeitsprozesse (The effect of gamification on motivation and performance; Empirical studies in the context of manual work processes)*. Springer Fachmedien Wiesbaden, 2016.

[15] R. J. Gerrig, *Psychologie (Psychology)*. Pearson, 2015.

[16] E. L. Deci and R. M. Ryan, *Handbook of Self-Determination Research*. The University of Rochester Press, 2002.

[17] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," *Contemporary Educational Psychology*, vol. 25, no. 1, pp. 54–67, 2000.

[18] M. Csikszentmihalyi, S. Abuhamdeh, and J. Nakamura, *Flow*. Springer Netherlands, 2014, pp. 227–238.

[19] M. Csikszentmihalyi, "Play and intrinsic rewards," in *Flow and the foundations of positive psychology*. Springer, 2014, pp. 135–153.

[20] Z. Fitz-Walter, *Introduction to Gamification: 4 quick quests to get you started with gamification*. Gamification Geek, 2018.

[21] S. Nicholson, "A recipe for meaningful gamification," in *Gamification in education and business*. Springer, 2015, pp. 1–20.

 ISBN: 978-1-61208-759-7

Relationships between Quantitative and Subjective Evaluations of Assistive Effect on Standing Function of the Smart Suit

Hijiri Hiraishi

Graduate School of Information Science and Technology
Hokkaido University
Sapporo, Japan
Email: hiraishi@hce.ist.hokudai.ac.jp

Takayuki Tanaka

Graduate School of Information Science and Technology
Hokkaido University
Sapporo, Japan
Email: ttanaka@ssi.ist.hokudai.ac.jp

Mami Sakata

Graduate School of Engineering
Yokohama National University
Yokohama, Japan
Email: sakata-mami-dz@ynu.jp

Dai Kawahara

Graduate School of Engineering
Yokohama National University
Yokohama, Japan
Email: kawahara-dai-ph@ynu.jp

Keisuke Shima

Graduate School of Engineering
Yokohama National University
Yokohama, Japan
Email: shima@ynu.ac.jp

***Abstract*—The Smart Suit is a power assist suit that reduces the burden on the lumbar region by reducing the activity of the spinal column erector muscle and has a trunk stabilizing effect. The suit also has assistive effect on stabilizing human trunk motion by tightening the pelvis. By using this as prevention of work-related accidents, it leads to the elimination of labor shortages. However, for the smooth introduction of Smart Suit, it is important to match the objective auxiliary effect with the subjective auxiliary effect. In this study, we focus on the balance assistive effect of Smart Suit and clarify the objective support and subjective assistive effects.**

Keywords–Assist Tool; Standing Function; Quantitative and Subjective Evaluations.

I. INTRODUCTION

Japan has a rapidly aging population. According to the Annual Health, Labour and Welfare Report 2016 [1], the aging rate was 26.7% in 2015, whereas it was less than 5% in 1950. The aging rate is expected to increase, and one in 2.5 people in 2060 will be 65 or older. Thus, the working-age population supporting the elderly is decreasing and it is becoming difficult to maintain the labor force. To address this problem, one idea is to reduce the burden of labor, thereby reducing the risk of injury and illness and allowing people to work longer. KEIROKA technology reduces the physical burden without interfering with the movement of the worker. The Smart Suit [2][3] is an assistive technology used by various workers, including farmers, fishers, construction workers, industrial workers, and nursing care workers.

The Smart Suit is an assistive tool that uses a rubber belt to exert an assistive effect in a forward bending position. The suit reduces the activity of the spinal erector muscles by assisting the muscle force and stabilizes the trunk by increasing joint stiffness via the tightening force. However, there are individual differences in the exertion of the assistive effect and the agreement between the objective and subjective assistive effects. Therefore, the subjective assistive effect of the Smart Suit can indicate what the objective assistive effect of the Smart Suit feels like, allowing the efficiency of the suit to be improved. Understanding these relationships is expected to help the introduction and diffusion of Smart Suit. In this paper, we focus on the balance assistive effect of the Smart Suit and clarify the relationship between the objective and the subjective assistive effects.

II. RELATED WORKS

Although we measure the subjective assistive effect of the Smart Suit here, several studies have subjectively measured the work load. NASA-TLX (Task Load Index) developed by Hart et al. [5] is a widely used technique for measuring subjective mental workload. It uses a multidimensional construct to derive an overall workload score based on the weighted average of evaluations on six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration level. Subjective workload experience tasks, behaviors, and subject-related correlations are defined as a function of difficulty of manipulation within the experiment, various workload sources between experiments, and individual differences in workload definition.

Yamada et al. [6] proposed an assist system for subjective burden called Skill-Assist, which varies its mechanical impedance to give workers who have been working for many years a sense of achievement in being able to perform the skilled tasks they were capable of when younger again. A Skill-Assist control algorithm based on variable impedance control has been proposed.

The present study describes the relationship between subjective and objective assistive effects for the balance assistive effect of Smart Suit. However, we think that the results are relevant to other assist tools and workload support.

 ISBN: 978-1-61208-759-7

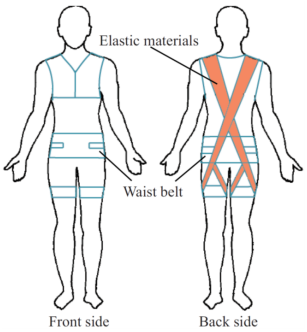

Figure 1. Schematic of elastic belts [2].

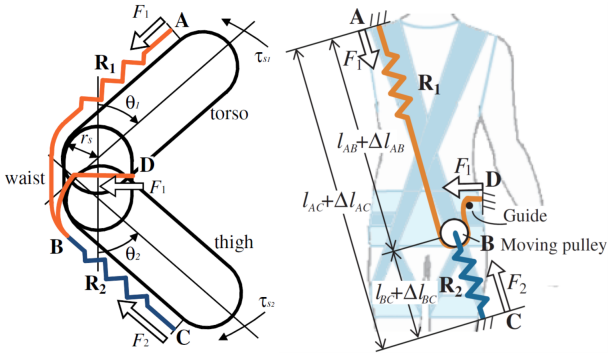

Figure 2. Assist mechanism of Smart Suit [2].

III. Assist Mechanism of Smart Suit

We describe the configuration of the Smart Suit and its effects. Figure 1 shows a schematic of the Smart Suit and Fig. 2 shows the assist mechanism. Elastic belts connect the thighs and shoulders to the back. The Smart Suit fits closely to the wearer's body and the assistive force is set according to the expected workload.

The elastic belt for the upper body, R_1, and the elastic belt for the thigh, R_2, are connected by a movable pulley at a point. The initial length of the path between A and C to the waist belt at point D after folding back at point B is the natural length of the elastic belt, and the change in path length when the wearer changes posture, Δl_{AC}, generates force F_1 at A and force F_2 at C. Assist torques τ_{s1} and τ_{s2} expand hip joint θ_1 and lumbar spine joint θ_2, respectively. The supporting torque reduces the wearer's joint torque and reduces the load on the muscles that move the joints. Simultaneously, force F_1 acts on the resilient waist belt at point D. F_1 increases the compression on the belt and lumbar spine at the tightening point. The combined torque τ_{s12} and force F_1 are given as

$$\tau_{s12} = \tau_{s1} + \tau_{s2} = \frac{6}{5} r_s k_s \Delta l_{AC} \quad (1)$$

$$F_1 = \frac{2}{5} k_s \Delta l_{AC} \quad (2)$$

where r_s is the moment arm of the elastic material and k_s is the coefficient of elasticity.

Equation (2) is derived from the balance of forces between elastic materials with a pulley configuration. The extended length of the entire line from shoulder to leg is divided by a ratio of 1:4 between the upper and lower elastic material.

According to Imamura, the rigidity of the trunk is increased and the posture is stabilized by force F generated during the forward bending posture [3]. In other words, wearing a Smart Suit exerts a sensory stabilizing effect and improves the stability of the entire body. And Imamura et al. presents an enhanced framework for evaluating an assistive effect of Smart Suit using a humanoid robot[4]. In this paper, we consider these assistive effects as balance assistive effects and perform subjective and objective evaluation.

IV. Evaluation of Standing Function Based on Virtual Light Touch Contact

We describe a Standing Function Evaluation System [7] used for quantitative measurement of the balance assistive effect of the Smart Suit. The Standing Function Evaluation System uses Virtual Light Touch Contact (VLTC) [8] devised by Sakata et al. Maintaining a standing position requires mitigating the risk of falling. A high standing function indicates good balance and low falling risk.

We describe the VLTC. Jeka et al. reported a phenomenon called light touch contact (LTC), in which touching a fixed point of a light force reduces postural fluctuation [9]. Because physical contact is required for LTC, VLTC provides the effect of LTC virtually with no touch.

Figure 3 shows a simple system configuration. See [7], [8], and [9] for details. The system consists of a Wii balance board as a force plate for measuring the Center Of Pressure (COP), a web camera for photographing the subject during measurements, a vibrator attached to the subject's finger for VLTC, and a computer for controlling the system. The measurement is performed with the subject standing on the Wii balance board. The subject touches a virtual wall configured around the body. In the measurement, the subject is switched between a state in which the virtual reaction force is presented when touching the virtual wall and a state in which no virtual reaction force is presented. Evaluate the change caused by . The measurement time is 40 s.

First, the COP is calculated for the sagittal plane (X direction) and the coronal plane (Y direction) from the Wii balance board. The following eight indices are used to evaluate the support standing function: d_1: total trajectory length of COP (L_{COP}); d_2: rectangular area of COP (S_{rect}); d_3: outer peripheral area of COP (S_{peri}); d_4: average velocity of COP (v_{COP}); d_5: average vector of COP (L); and d_{6-8}: index variations associated with virtual partition state changes.

The standardized indices (I_l) are determined using the standard deviation (σ^l) and the mean (μ^l) of the standard subject data for each index value.

$$I_l = \frac{d_l - \mu^l}{\sigma^l} \quad (3)$$

Considering I_l allows the inconsistencies between the measured and controlled values to be assessed. Then, the weighted sum of I_l ($S(I) = \Sigma_{l=1}^{N+8} w_l I_l$) is calculated for comprehensive evaluation of the stationary functions. It is assumed that the relationship between age and standing function can be expressed as a non-linear function,

 ISBN: 978-1-61208-759-7

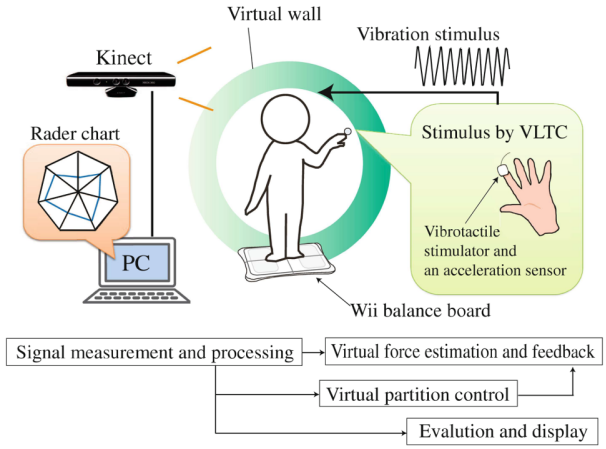

Figure 3. Overview of the Standing Function Evaluation System [7].

$$S_{age} = f_{age}(S(I)) \tag{4}$$

where $f_{age}(S(I))$ is a nonlinear function of the weighted sum, $S(I)$, for estimating the subject's age, and S_{age} is the subject's age estimated by the system, called "standing age" here. Similarly, it is assumed that the relationship between age and balance function can be expressed as a linear function in the current system as

$$B_{age} = g_{age}(I_1^{nc}) \tag{5}$$

where $g_{age}(I_1^{nc})$ is the total trajectory length of COP when VLTC is OFF, I_1^{nc} is used for estimating the subject's age , and B_{age} is the subject's other age estimated by the system, called "balance age" here.

Because standing age is calculated using several evaluation indices, it is an index of balance including the subject's sensory feedback. The smaller the standing age, the better the sense of balance. On the other hand, because the balance age is calculated from the index when VLTC is OFF, it indicates an individual's balance with no sensory factors. Therefore, the smaller the balance age, the better the potential balance.

V. Experiment for Quantitative and Subjective Evaluation

A. Experimental Setup

The introductory tests were conducted on nine men and women aged 20s to 50s to verify the effects of assistance of the Smart Suit according to the schedules shown in Table I. The subjects are working at a distribution center, and is inspecting and loading on a truck. They are in the middle and lower back posture when carrying luggage and working on lanes. Table II shows the results of physical fitness tests as the physical abilities of each subject, and their working style.

First, a 1-day advance measurement was performed to determine the subject's physical ability and to explain the introductory test, including the physical fitness test, the standing function evaluation, the significance of the Smart Suit, and the instructions for its use. For 5 days in the following week, the subjects performed their usual work while completing the subjective working environment survey [10] without wearing

TABLE I. SCHEDULE FOR INTRODUCTORY TESTS FOR EVALUATING THE ASSISTIVE EFFECT OF THE SMART SUIT.

	the first week	the second week	the third week
Days	Five days Labor environment investigation	Five days Wearing period	Five days Introduction test period
Conduct contents	awareness examination usual work	usual work	awareness examination usual work
Smart Suit	non-wear	wear	wear

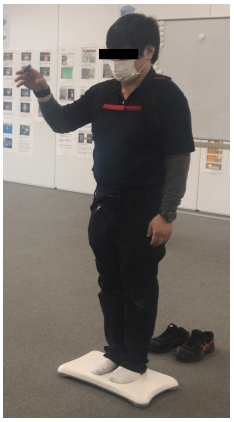

Figure 4. Measurements with the Standing Function Evaluation System wearing the Smart Suit.

a Smart Suit. In the next 5 days, they wore the Smart Suit, but they did not complete the subjective survey. In the subjective working environment survey, we conducted a questionnaire to ask about the languors of body and feelings at the start and end of work. The 5 days in the last week were the introductory test period, and the subjects did their usual work while wearing a Smart Suit and completed a subjective survey. After the introductory test period, a follow-up measurement similar to the pre-measurement was conducted and the subjects completed a feeling-of-use questionnaire for the Smart Suit. Figure 4 shows the advance measurement using the Standing Function Evaluation System with the Smart Suit.

The physical strength tests performed in the pre- and post-measurements were standing physical anteflexion, functional reach test, and grip strength measurement. These were measured twice each and the average value was used.

B. Experimental Results

1) Experimental Results of Subjective Evaluations: For the introductory test, we describe the results of the surveys of subjective experience and questionnaires for standing function evaluation and subjective evaluation. The results of the follow-up measurements were used. The results for the nine subjects are shown in Tables III and IV. The feeling of the assistive effect measured by the post hoc measurement questionnaire was evaluated by rating how much the assistive effect of the Smart Suit was felt on a 10-point scale. The response of subjects who could not determine whether there was an assistive effect was recorded as "none", that of the subjects who felt a subjective assistive effect from the Smart Suit was recoded as "yes", and that of the subjects who felt no assistive effect was recorded as "no". As shown in Table III, some subjects felt uncomfortable wearing at the time of walking or

 ISBN: 978-1-61208-759-7

TABLE II. PHYSICAL FITNESS TEST AND WORKING STYLE.

	Physical Fitness Test			Working Style	
NO.	Standing Position[cm]	FRT [cm]	Grasping Power[kgf]	Job Description	Working Road
a	7.0	41	46.0	Response to lane clogging	heigh
b	-7.5	37	50.5	Tractor docking	heigh
c	4.0	43	53.5	Loading and transporting luggage	heigh
d	11.0	46	25.0	Product inspection	low
e	-4.0	40	54.5	Refill items on the shelf	middle
f	1.5	50	48.0	Tractor docking	heigh
g	4.0	47	33.5	Store products on the shelf	low
h	-7.0	44	44.5	Sort the package	heigh
i	-3.5	41	34.5	Product inspection and sorting	middle

TABLE III. FEELING OF WEARING AND SUBJECTIVE EVALUATION.

	Profile		Feeling of wearing		Subjective evaluation	
NO.	Age [years]	Sex	At first	In the end	Feeling of Assistance Effect	Subjective Survey
a	46	male	Discomfort	Improved	8/10	none
b	33	male	Discomfort	Discomfort	6/10	No
c	53	male	Discomfort	Improved	8/10	none
d	38	female	Accustomed soon	No discomfort	2/10	Yes
e	42	male	Discomfort	Discomfort	7/10	Yes
f	55	male	Accustomed soon	No discomfort	8/10	none
g	28	female	Accustomed soon	No discomfort	1/10	Yes
h	56	male	Accustomed soon	No discomfort	8/10	Yes
i	52	male	Accustomed soon	No discomfort	7/10	none

at the beginning of use, but most subjects eventually became accustomed to using Smart Suit.

In the experience of the assistive effect in Table III, the two women may have felt the assistive effect less because their work did not involve much forward bending and their working roads are low as shown in Table II and III; thus, there was little opportunity to demonstrate the assistive effect of the suit. However, similar to subjective experiences of illness, the actual assistive effect does not necessarily agree with the feeling of the assistive effect. Thus, we identified subjects who did not feel the assistive effect, although there was an actual assistive effect .

Table IV shows the standing age and balance age measured using the Standing Function Evaluation System when the Smart Suit was not worn and when it was worn. The differences between the measurements show how much the standing function was improved by wearing the Smart Suit. In other words, the larger the difference between the measurements while not wearing the suit and wearing the suit, the larger the balance assistive effect. The effect was seen in six subjects; thus, several subjects benefitted from the Smart Suit.

2) Quantitative Evaluation Results.: The body sway measured in the Standing Position Function Evaluation System was analyzed. According to Yamamoto et al., body sway and low back pain risk are related [11]. People with lumbar lordosis had a large sway in the lateral direction (X direction), and those with a tendency to scoliosis or cervical tilt had a large sway in the anteroposterior direction (Y direction). Based on these observations, the risk of occurrence of lumbar pain could be reduced if the X direction and the Y direction of the center of gravity fluctuation are decreased when wearing the Smart Suit. Figures. 5 and 6 show the decrease in the effective value and speed of the body sway in each direction when the Smart Suit is worn. Six of the nine subjects showed a reduction in body sway in either the X or Y direction. We confirmed the

TABLE IV. STANDING AGE AND BALANCE AGE.

		Non-Wearing		Difference between wearing and non-wearing	
NO.	Age [years]	Standing Age	Balance Age	Standing Age	Balance Age
a	46	7	-15	-32	22
b	33	36	-30	7	5
c	53	55	36	10	9
d	38	33	6	12	13
e	42	38	-13	46	2
f	55	17	-15	16	5
g	28	32	-39	6	0
h	56	21	-21	19	-5
i	52	16	1	-6	-10

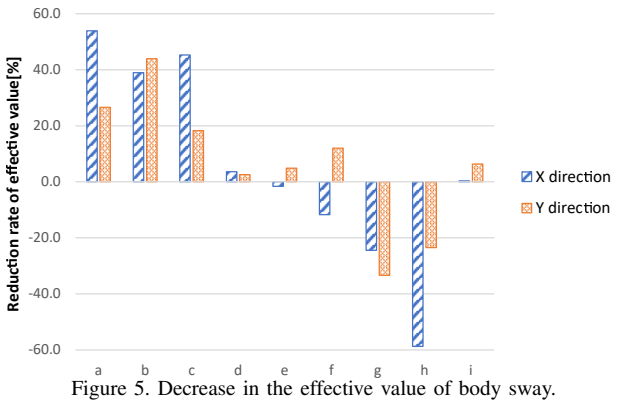

Figure 5. Decrease in the effective value of body sway.

correlation between the suppression of the body sway and the standing age and balance age that was used in the Standing Function Evaluation System (Table V). The difference in the balance age between wearing the suit and not wearing the suit and the suppression of body sway had coefficients of determination of 0.22 to 0.80. Therefore, by determining the difference in balance age, it is possible to determine whether the risk of lumbar pain is reduced.

The change in the sensory reweighting of the three sensory systems that contribute to standing in addition to the suppression of body sway is considered as the balance assistive effect of the Smart Suit. Sensory reweighting is a phenomenon in which a person adjusts the weighting of each sensory system when maintaining posture. Eikema et al. studied changes in posture and sensory reweighting when elderly people receive

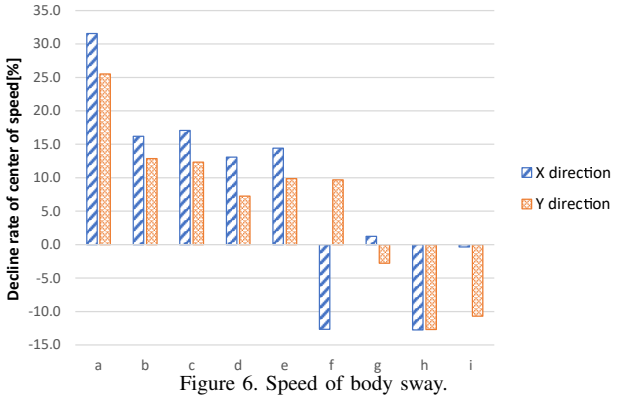

Figure 6. Speed of body sway.

 ISBN: 978-1-61208-759-7

TABLE V. COEFFICIENT OF DETERMINATION BETWEEN BODY SWAY AND THE STANDING FUNCTION EVALUATION SYSTEM.

	Difference of standing age		Difference of balance age	
	Effective value	Speed	Effective value	Speed
X direction	0.23	0.14	0.43	0.51
Y direction	0.08	0.07	0.22	0.8

sensory stimuli [12]. The balance assistive effect of the Smart Suit may affect the sensory system, similar to the VLTC of the Standing Function Evaluation System. We examined the effect of wearing a Smart Suit on the sensory system.

The evaluation method is as follows. First, wavelet transformation was performed on body sway measured by the Standing Position Function Evaluation System. The wavelet-transformed time series gain was divided by the frequency band of each sensory system. The frequency band was 0.02–0.3 Hz for vision, 0.3–1.0 Hz for vestibular + tactile sense, and 1.0–3.0 Hz for position sense. Then, the time average and the variance of the gain of each frequency band were calculated to evaluate quantitatively how much each sensory system contributed to posture maintenance. A large gain in a sensory system indicates that the system is preferentially used for balance in posture maintenance.

The VLTC used in the Standing Function Evaluation System decreased the gain of the vestibular + tactile and position sensory systems, indicating that these systems were replaced by the effect of VLTC. Therefore, when the Smart Suit caused a similar decrease in the gains of these systems, it indicated that the Smart Suit exerted a balance assistive effect on the sensory systems.

Figure 7 shows a box and whisker plot of the sum of vestibular + tactile and position sensory system gains after wavelet transformation of typical subjects h and i for when the Smart Suit was worn or not worn and VLTC was OFF or ON. For subject i, when the VLTC was OFF the gain was significantly reduced when the Smart Suit was worn, whereas there was no significant difference (confidence of 95% or more) for subject h. In addition, for subject i, when the Smart Suit was not worn, the gain was significantly reduced when the VLTC was ON, whereas for subject h, the gain was similar to when the VLTC is OFF. These results show that wearing the Smart Suit assisted subject i with similar sensory systems to the VLTC. Although the assistance of the sensory system by the VLTC was observed in one subject, the assistive effect of wearing the Smart Suit was confirmed in subjects a–e, g, and i. In addition, subjects b, d, and i showed reduced gains compared with the other subjects when the VLTC was ON and the Smart Suit was worn. Thus, the assistance from the VLTC and from the Smart Suit do not need to be simultaneous.

VI. DISCUSSIONS

The balance assistive effect of the Smart Suit was evaluated by subjective evaluation, body sway, and sensory systems. There were some subjects who could not confirm whether they felt the assistive effect and some who thought that they did not feel the assistive effect. The presence or absence of the auxiliary effects did not necessarily coincide with these experiences.

Subjective evaluation showed that seven men felt the 10 levels of assistive effects. Among them, in the subjective

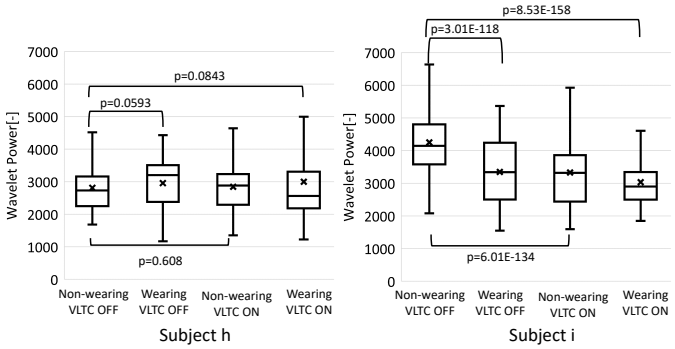

Figure 7. Sum of visual and vestibular + tactile gains for subjects h and i for when the Smart Suit is worn or not worn and VLTC is OFF or ON.

survey, two of these subjects indicated they felt subjective assistive effects. In addition, the subjective survey results for the two female subjects who did not feel the assistive effects indicated that they did experience assistive effects . Therefore, simply examining the feeling of the assistive effect is insufficient for indicating the subjective assistive effect, and it is necessary to use a survey method, such as a subjective survey. In addition, all of the nine subjects reported a subjective Smart Suit assistive effect either by feeling the assistive effect or through their answers in the subjective survey.

Considering the suppression of body sway in the left and right direction and the front and back direction by the Smart Suit as an objective balance assistive effect, the assistive effect was recognized in subjects a–f. However, although subject b reported no assistive effect in the subjective survey, their body sway was suppressed by the suit. In contrast, subjects h and i did not have their body sway suppressed, but reported a subjective assistive effect in the subjective survey, and in particular, subject h evaluated the experience of the assistive effect as high. Therefore, the suppression effect of body sway does not necessarily coincide with the subjective evaluation.

The evaluation of the assistive effect of the Smart Suit on the sensory system was confirmed in seven subjects (subjects a–e, g, and i). Although subject h experienced an auxiliary effect subjectively, no objective suppression of body sway or auxiliary effect on the sensory system was observed. Although subject f experienced an assistive effect, the objective suppression effect of the center of gravity sway was recognized in only the front and back direction and no assistive effect on the sensory system was observed. The other seven subjects showed either an objective suppression effect on body sway or an auxiliary effect on the sensory system, and these auxiliary effects are factors in determining an auxiliary subjective effect.

Although the suppression of body sway and the auxiliary effect on the sensory system are factors that determine the subjective auxiliary effect of the Smart Suit, they are not the only factors. This may be because the balance assistive effect is evaluated in the standing posture and that the muscle assistive effect of the Smart Suit is not considered. The Smart Suit was designed to exert an assistive effect via the expansion and contraction of the elastic material in the forward bending posture; thus, the subjective evaluation may be determined by the assistive effect for forward bending. Depending on the subject, the muscle strength assistive effect may also have a large effect on subjective evaluation.

 ISBN: 978-1-61208-759-7

We confirmed that a balance assistive effect was exhibited even when standing. The results suggested that the assistive effect might not only suppress the center of gravity sway, but also have an assistive effect similar to LTC on the sensory system. The subjective assistive effect, suppression of body sway, and assistive effect on the sensory system measured in this research are factors in judging whether the Smart Suit will have an assistive effect. Therefore, our results will allow quantitative evaluation of the balance assistive effect of the Smart Suit by suppressing body sway and assisting the sensory system.

VII. Conclusion and Future Works

In this paper, we evaluated the balance assistive effect of the Smart Suit subjectively and objectively. Subjective assistive effects were evaluated by the feeling of assistive effects and subjective experience. In the interviews, seven men out of nine subjects felt the assistive effect. On the other hand, according to the subjective survey, four subjects experienced the assistive effect, and these results were not necessarily consistent. Therefore, investigating subjective experience showed that there were subjects who needed the assistive effect even though they did not feel the assistive effect. The suppression of the body sway was evaluated by the effective value of the body sway in the lateral direction (X direction) and the longitudinal direction (Y direction) and the speed of the sway. Wearing the Smart Suit suppressed the body sway in six subjects, and the suppression was correlated with the difference between the balance ages measured by the Standing Function Evaluation System while not wearing and wearing the Smart Suit. Therefore, we showed that the body sway suppression effect can be measured quantitatively by using the Standing Function Evaluation System. In addition, by looking at the change in frequency gain of each sensory system in a Smart Suit based on the sensory reweighting model, we verified the auxiliary effect on the sensory systems using wavelet transform. A decrease in the frequency gain of the vision and vestibular + tactile systems was observed in seven subjects. We confirmed that the Smart Suit has an assistive effect on the sensory system similar to LTC. However, these evaluations of the relationship between the subjective and objective assistive effects did not necessarily agree. Although the suppression of body sway and the auxiliary effect on the sensory system were recognized as objective auxiliary effects, they could not explain the subjective auxiliary effect; however, they are probably important elements.

In future, we intend to examine not only the standing posture, but also the assistive effects in a forward bending posture that the Smart Suit was designed to exert. In addition to balance assistive effects, we also want to evaluate the muscle assistive effects and the physical abilities of the subjects. It is necessary to investigate how much these other effects affect the subjective auxiliary effect. Our goal is to establish a method to assess these effects comprehensively and conduct quantitative screening to see if a person can benefit from the Smart Suit.

Acknowledgement

This work was partially supported by JSPS KAKENHI Grant Number JP18H03276.

References

[1] Ministry of Health, Labour and Welfare, "Annual Health, Labour and Welfare Report 2016", Health, Labour Addministration Annual Report 2015, Oct. 2016.

[2] Y. Imamura, T. Tanaka, Y. Suzuki, K. Takizawa, and M. Yamanaka, "Motion-Based-Design of Elastic Material for Passive Assistive Device Using Musculoskeletal Model", J. of Robotics and Mechatronics, .23-6, pp. 978-990, Dec. 2011.

[3] Y. Imamura, "Analysis of Trunk Stabilization Effect by Passive Power-Assist Device", Journal of Robotics and Mechatronics, 26-6, pp. 791-798, Dec. 2014.

[4] Y. Imamura, Ko Ayusawa, Eiichi Yoshida and Takayuki Tanaka, "Evaluation Framework for Passive Assistive Device Based on Humanoid Experiments", International Journal of Humanoid Robotics, 15-03, 1750026, 1-25, 2018

[5] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research", Advances in Psychology, 52, pp.139-183, 1988.

[6] Y. Yamada, H. Konosu, T. Morizono and Y. Umetani, "Proposal of Skill-Assist: a system of assisting human workers by reflecting their skills in positioning tasks", IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028), 4, pp.11-16, 1999.

[7] M. Sakata, K. Shima, K. Shimatani and H. Izumi, "Simplified Standing Function Evaluation System for Fall Prevention", Proceedings of Intelligent Robotics and Applications- 9th International Conference(ICIRA), Part II, LNAI 9835, pp. 48-57, Aug. 2016.

[8] K. Shima, K. Shimatani, A. Sugie, Y. Kurita, R. Kohno and T. Tsuji, "Virtual Light Touch Contact:a Novel Concept for Mitigation of Body Sway", 2013 7th International Symposium on Medical Information and Communication Technology(ISMICT), 978-1-4673-5772-2, pp.108-111, Mar. 2013.

[9] John J. Jeka, "Light Touch Contact as a Balance Aid", Physical Therapy, 77-5, pp.476-487, May. 1997.

[10] Industrial Fatigue Research Group, "Subjective survey", ⟨http://square.umin.ac.jp/of/service.html⟩, Oct. 2017.

[11] S. Yamamoto, K. Sunami, M. Aoyagi, and H. Sakurai:"Evaluation of the reduction effect of low back pain occurrence risk by the body sway test", Industrial Hygiene Journal, 47, p.453, Apr. 2005.

[12] D.J. A. Eikema, V. Hatzitaki, V. Konstantakos and C. Papaxanthis, "Elderly adults delay proprioceptive reweighting during the anticipation of collision avoidance when standing", Neuroscience, 234, pp.22-30. Jan. 2013.

 ISBN: 978-1-61208-759-7

Spatiotemporal Activities in Human Brain during Recognizing Ambiguous Figure

Takahiro Yamanoi,
Faculty of Engineering,
Hokkai-Gakuen University,
Sapporo, Japan,
e-mail: yamanoi@hgu.jp,

Tomoko Yonemura,
Integrated Education Center,
Meikai University,
Chiba, Japan,
e-mail: yonemura@meikai.ac.jp,

Hisashi Toyoshima,
Sapporo Branch,
Japan Technical Software,
Sapporo, Japan,
e-mail: toyoshima@jtsnet.co.jp,

Yahachiro Tsukamoto,
Emeritus Professor,
Meijo University,
Nagoya Japan,
e-mail: yytsukamoto@ac.auone-net.jp,

Shin-ich Ohnishi
Faculty of Engineering
Hokkai-Gakuen University
Sapporo, Japan
e-mail: ohnishi@hgu.jp

***Abstract*—To treat higher order brain dysfunction, it is necessary to identify the location of each function. In Japan, a super-aging society is progressing, and many dysfunctions such as cerebral infarction have occurred. The fMRI analysis is not enough to identify the position of dysfunction in detail. The authors had tried to elucidate higher order brain functions. In the paper, the authors have measured electroencephalograms (EEGs) from subjects (MK and RE) who were observing four images of ambiguous monochrome line pictures. The equivalent current dipole source localization (ECDL) method has been applied to those Event Related Potentials (ERPs): averaged EEGs by each figures and trials. The paper reports the comparison results of "Saxophone player and Girl's face." In the case of the Saxophone player, the process was done over a latency from 400ms to 1000ms, however for the Girl's face image, the corresponding process was completed relatively quickly and ended the latency around 800ms. Especially in the case of Girl's face, ECDs were localized to the right and the left angular gyrus (AnG) around 370ms and to the right post central gyrus (PstCG) around 415ms, then by way of language areas, ECDs were localized again to the right and the left AnG around 520ms. It has been clarified in our previous study that activities on the angular gyrus (AnG) are important to discriminate the unusual shape of presented images. This fact is confirmed also in this work.**

Keywords-Electro-encephalogram; Electro-encephalogram; Equivalent Current Dipole Localization Method; Ambiguous Pictures; Brain Activity.

I. INTRODUCTION

To elucidate brain function, many instruments have been used nowadays, however, to detect activities in milli second order was nothing sufficient but using EEGs. The authors compare the results of an attempt to analyze EEG from two subjects M K and R E during recognizing an ambiguous image of a saxophone player and/or a girl's face among four types of images. First, the authors compared each latency of EEGs between in cases of the saxophone player and the girl's face. EEGs were averaged and summed by each 19channel for each image and the event related potentials (ERPs) were obtained. The authors analyzed spatiotemporal activities and pathways in the brain by using the equivalent current dipole localization (ECDL) method [1], as our previous EEGs experiments. To elucidate brain function, many instruments have been using nowadays, however, to detect activities in milli second order was not sufficient but using EEGs.

Figure 1. Presented Ambiguous Images (Grey dots are introducing points to one of them.)

Ambiguous figures are often used to elucidate the mechanism of perceptual alternation and attention (cf. [2]), which are concerned with the multi stable perceptual phenomena. Some of the present authors had tried an experiment by using the eye mark camera EMR-9 (NAC Co.), to see if imposing the introducing point in the figure is helpful to recognize each image. The insert of introducing point proved to be effective, because obtained results had differences in locus of each eye movement, however, that could not elucidate activities in the brain. The previous researches using fMRI to examine the brain information processing suggested that the activities on the frontal lobe [3] and the right parietal lobe [4] are largely involved in processing images. However, the results of eye mark camera were not able to point at the area. In this study, it is examined that the activity of the AnG can be seen where the previous research elucidated. In section II, procedure of the experiment and the analysis of EEG data are explained. In section III, the results of ECDL method are summarized. Section IV presents a discussion of the results and the future work.

II. EEG MEARSUREMENT EXPERIMENT AND ANALYSIS

Two subjects M. K. and R. E. are 22-year-old females and have normal visual acuity. They are right-handed. The subjects put on 19 active electrodes and watched the 21inch PC monitor screen 30 cm in front of them. Their heads were fixed on the table on a chin rest.

Stimuli are simple monochrome figures of ambiguous picture (Figure 1). First, a fixation point was presented, then a stimulus was presented, both of them were during 3000ms (Figure 2). EEGs were measured on the multi-purpose portable bio-amplifier recording device (Polymate AP1524, TEAC) by way of the electrodes and the frequency band is between 1.0Hz and 2000Hz. Outputs were transmitted to a recording PC. Each position of electrode was measured in the three dimensional coordinate on each experiment by the subjects. These electrode positions were used in applying ECDL method with MRI of the subject.

By use of the equipment, the authors have measured

 ISBN: 978-1-61208-759-7

EEGs on each visual stimulus. So as to effectively execute the ECDL method, both data were summed and averaged according to each channel of EEGs, each type of figure, and each subject to get ERPs. Then, the ECDL method was applied to each ERPs. Because of the number of the recording electrodes was 19, theoretically, at most three ECDs could be estimated by use of the PC-based ECDL analysis software "SynaCenterPro [2]" (NEC Corpo- ration). Selected results of the goodness of fit (GOF) of ECDL were more than 99 %. Estimated EEGs were super im- posed on MRIs from the subject.

The authors presented four types of monochrome images (Figure 1) to the subjects and measured EEGs during recognizing presented images. Among four images, the authors will report compared results between the saxophone player and the girl's face by estimated latency. One extra channel

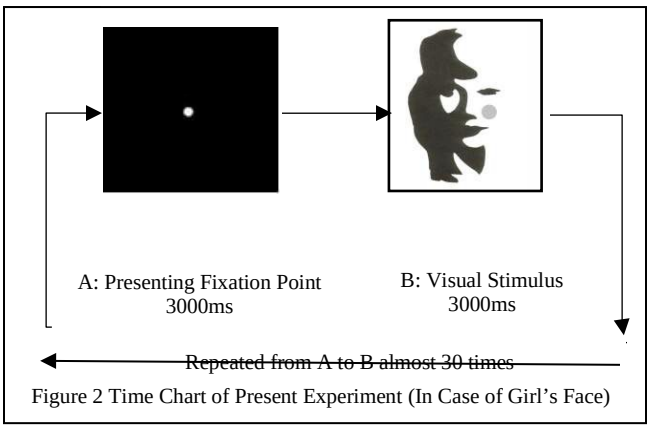

Figure 2 Time Chart of Present Experiment (In Case of Girl's Face)

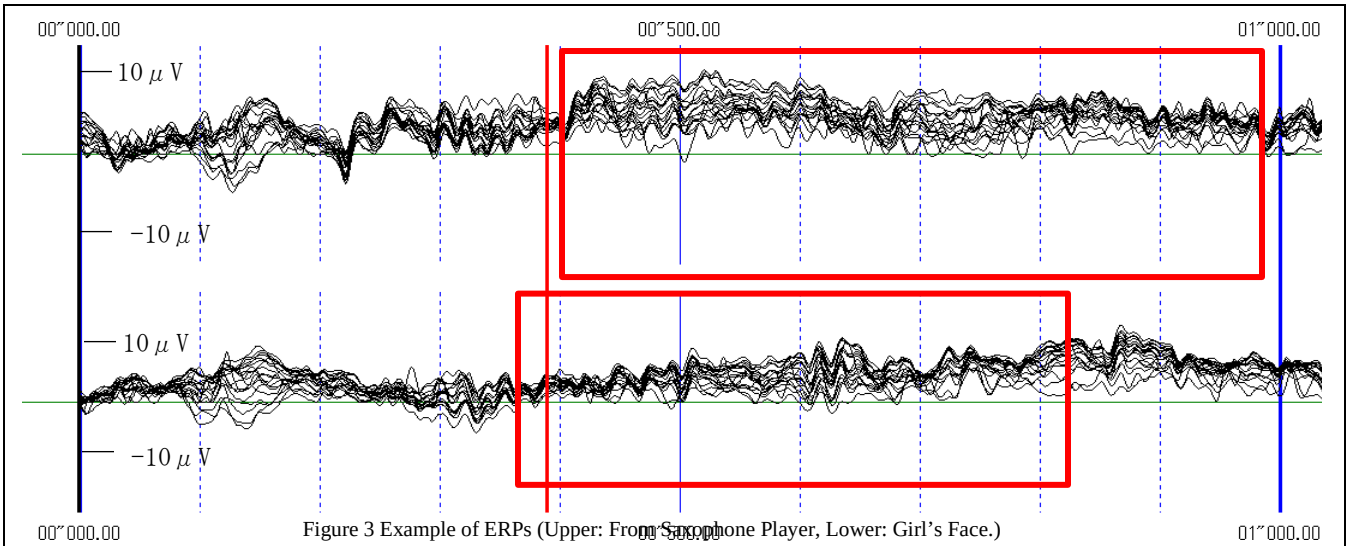

Figure 3 Example of ERPs (Upper: From Saxophone Player, Lower: Girl's Face.)

TABLE 1. RELATIONSHIP BETWEEN ESTIMATED ECD AND LATENCY IN EVENT RELATED POTENTIALS FOR SAXIOPHONE PLAYER (SUBJECT: M. K.)

Brain Part	L MFG	L PrCG	L ParaHip
Latency	336	339	406
R FuG	R PrCG	Broca	Wernicke
446	501	586	646
R ITG	R ParaHip	R ITG	Broca
653	672	676	687
R ParaHip	L AnG	Broca	
771	814	890	

TABLE II RELATIONSHIP BETWEEN ESTIMATED ECD AND LATENCY IN EVENT RELATED POTENTIALS FOR GIRL'S FACE (SUBJECT: M. K.)

Brain Part	L AnG	R AnG	L AnG
Latency	387	388	397
R PstCG	R SFG	Broca	L AnG
416	427	433	434
L AnG	R FuG	R FuG	R FuG
439	440	455	483
L AnG	R AnG	R Hip	
514	579	587	

TABLE III. RELATIONSHIP BETWEEN ESTIMATED ECD AND LATENCY IN EVENT RELATED POTENTIALS FOR SAXIOPHONE PLAYER (SUBJECT: R.E.)

Brain Part	L PrCG	Wernicke	R ParaHip
Latency	408	417	468
Broca	L AnG	R PrCG	L ParaHip
473	513	519	535
Broca	Wernicke	R FuG	L AnG
560	621	690	733
Wernicke	Broca	L AnG	
737	744	762	

TABLE IV. RELATIONSHIP BETWEEN ESTIMATED ECD AND LATENCY IN EVENT RELATED POTENTIALS FOR GIRL'S FACE (SUBJECT: R. E.)

Brain Part	L AnG	R AnG	R PstCG
Latency	362	369	415
L SFG	Wernicke	R ParaHip	R MFG
424	430	457	475
R Hip	R MFG	R MFG	L PrCG
482	489	495	517
L AnG	L AnG	R Hip	
525	527	574	

 ISBN: 978-1-61208-759-7

was used as trigger of the beginning of presentation of figure, and was recoded as a pulse. This pulse timing was used as a marking index, and summed by figure type and by recognizing result individually, to obtain ERPs.

In the experiment, a gazing point was presented for 4 seconds at the center of the screen to the subjects, and this was used as the screen masking, after that, a visual stimulus was presented for 3 seconds at the center of the screen. The above cycle was repeated 40 times, and these were put together as one course, and the measurement of two courses were made into one set. The visual stimuli were presented at random.

The authors applied the equivalent current dipole estimation (ECDL) method [1] as same as the previous researches. In general, ECDL method places ECD in the head model and calculates the theoretical value of electric potential distribution on the scalp "forward problem" and optimizes the ECD parameter so that the error between theoretical value and measured value is minimized. To solve so called "inverse problem" analytically result in solving the defect optimization problem, and is solved using a numerical analysis method in which initial values are set at grid points. The head model was modeled as three layers of concentric spheres of different conductivity: scalp, skull and cortex.

MRI of brain from each subject was used to set up the concentric sphere model of the head. In addition, the accuracy and reliability of the estimation results were evaluated by the values of Goodness of fit (GOF) and statistical confidence limits, respectively. The authors used PC version dipole estimation software (SynaCenterPro: NEC) for these analyses. Among the estimation results, the results with a GOF value of 99% or more and a 95% confidence limit of 1mm or less were adopted. In the dipole estimation software SynCenterPro used by the authors, the dipole of the estimated result is superimposed and displayed on the subject's MRI.

III. RESULTS OF ECDL ANALYSIS

After the latency around 400 ms, the ECDs were localized to the right ParaHip (R ParaHip) (Figure 4), the right fusiform gyrus (FuG), the Broca's area (Figure 5), the R ParaHip or Hip, the Broca's area, and the right fusiform gyrus (FuG) (Tables). Above mentioned spatiotemporal pathway accords with so called the ventral pathway which is related with the primitive process of visual recognition. These areas are also related to the integrated process of visual recognition of picture and the recalling of word. Especially, the angular gyrus is said to integrate information of some modalities, so there might have recalled a word already at this stage. On the right inferior temporal white matter, a process progresses from recognition of a picture to recalling of a word.

These ventral and dorsal processes are done in series or in parallel. The relationship is resumed in Table II and III. Moreover, there is a possibility that these areas are also the language areas because these subject's dominant language areas were considered to be located in the right hemisphere from the precedent research.

According to Table II, the authors found the spatiotemporal pathway of the human brain activities as follows. The authors call a pathway A until the first activity on the Broca's area and after the first activity on the Brocas's area then again activity.

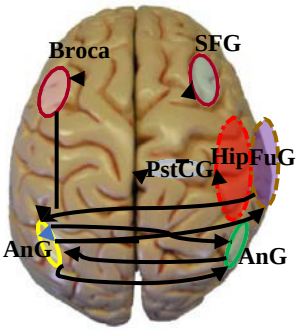

Figure 4 Spatiotemporal brain activity before R ParaHip.

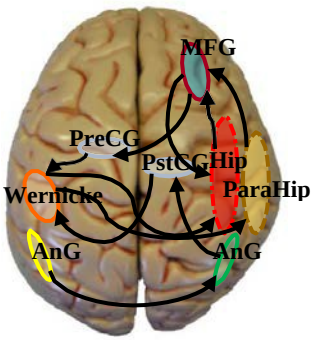

Figure 5 Spatiotemporal brain activity after R ParaHip.

Presented figures are shown in Figure 1. An example of gray introducing point is placed on each image. Figure 3 shows thee ERP of the subject M.K. In this result, the potentials so called as P300 and N400 could also be seen slightly. N400 is said to appear as a result of processing which involves semantic conflict. Although the presented images and their experimental situations are different, the authors analyzed the ECDs and compared the estimated latencies with the results from the precedent experiment on recalling of the fruit name [8]. It can be seen from Figure 2 that the response latency of the saxophone player's is slightly delayed from that of the girl's face image, and the potential of saxophone shows a relatively high negative potential.

As shown in Figure 3, P300 was faintly observed with similar latency. The same is true for N400. The para hippocampus (ParaHip) was estimated as the processing site at this latency. As for the whole process, in the case of the saxophone player, the process can be seen over a latency of 400ms to 1000ms, but for female face images, it can be confirmed that the corresponding process is completed relatively quickly and the latency is around 800ms.

As shown at the beginning, regarding the recognition of language, the path in the first half and the second half of processing is assumed. From this, in the subject MK, in the estimation from the ERP regarding to the saxophone player, a pathway obtained is:

 ISBN: 978-1-61208-759-7

L MFG → L PrCG → L ParaHip → R FuG →R PrCG → Broca → Wernicke →R ITG →R ParaHip → RITG → Broca →R Hip → L AnG → Broca,
the processing route in the second half after L MFG was obtained and shown in Figure 5.

It was relatively difficult to estimate the position due to the early response of the initial visual processing, however, result pathway,

L AnG → R AnG →L AnG → R Pst CG →R SFG →Broca → L AnG → R FuG → L AnG ⟶R AnG →R Hip
was obtained. Furthermore, within a short latency, several loop responses have been observed. These results were shown in Figure 4 of the first half pathway and Figure 5 of the second half. Although this result was roughly confirmed in the case of subject RE, it is considered that in addition to the comparison between the presented images, a detailed comparison and examination between the subjects is required

IV. Conclusion and Futurework

From the results from previous studies on fruits and animals [8], according to the difference between normal shapes (round ones) and non-normal shapes (long ones or angled ones), in the former case, no ECD was estimated on the right AnG. The presented figures in this study are ambiguous figures, on considering these as well as the previous study, it is apparently not ordinary shape, so it is thought that ECD is estimated to R AnG. If processing in 700ms just before the high potential in Figure 2 was reaction from the FuG, the AnG, and the TE in the right hemisphere, the shape pro- cessing is clearly performed in that period. In particular, the right AnG is supposed to be the shape processing area, that has been clarified in the previous studies by Yamanoi et al. It is interesting that the ECD on this area was also estimated for the ambiguous figure. Furthermore, the authors would further investigate and clarify in more detail what role the N700 component plays in shape processing, including the results of previous studies. The authors had recorded EEGs for the other ambiguous figures, so it will be hoped to continue analysis on the other cases and will compare the results each other. Although, the EEG experiment needs much time for preparation and for measure, it will be necessary to increase the number of subjects.

Acknowledgments

This research was partially supported by a grant from the Ministry of Education, Culture, Sports, Science and Technology for the national project of the High-tech Research Center of Hokkai-Gakuen University ended in March 2013. And also, this work was partly supported by JSPS KAKENHI Grant Number 16H02852.

The experiment was approved by the ethical review board of Hokkaido University.

References

[1] T. Yamazaki, K. Kamijo, and A. Kenmochi: “Accuracy of Multiple Equivalent Current Dipole Source Localization from EEG in Terms of Confidence Limits for Radial Distances”, Bio Medical Engineering, vol. 37, no. 4, pp. 336-341, 1999.

[2] J. Brascamp, P. Sterzer, R. Blake, and T. Knapen: “Multistable Perception and the Role of the Frontoparietal Cortex in Perceptual Inference”, Annual Review of Psychology, vol.69, pp.77-103, 2018.

[3] P. Sterzer and A. Kleinschmidt: “A neural basis for inference in perceptual ambiguity”, Proceedings of the National Academy of Sciences, vol.104, pp.323-328, 2007.

[4] R. Kanai, B. Bahrami, and G. Rees: “Human Parietal Cortex Structure Predicts Individual Differences in Perceptual Rivalry”, Current Biology, vol.20, pp.1626-1630, 2010.

[5] T. Yamanoi, H. Toyoshima, J-L. Vercher, E. Sanchez, and M. Sugeno: “Dominance of recognition of words presented on right or left eye -Comparison of Kanji and Hiragana”, Modern Information Processing from Theory to Applications, pp. 407-416, Elsevier Science B.V., Oxford, 2006.

[6] Y. Tanaka, T. Yamanoi, M. Otsuki, H. Toyoshima, S. Ohnishi, and T. Yamazaki: “Spatiotemporal Brain Activities on Recalling Body Names” (in Japanese), 22th Soft Science Workshop, pp.27-28, 2012.

[7] H. Toyoshima, T. Yamanoi, T. Yamazaki, and S. Ohnishi: “Spatiotemporal Brain Activity during Hiragana Word Recognition Task”, Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 15, no. 3, pp. 357-361, 2011.

[8] K. Sugimoto, T. Yamanoi, H. Toyoshima, M. Otsuki, S. Ohnishi, and T. Yamazaki: “Spatiotemporal Human Brain Activities on Recalling Tetrapod and Fruit Names”, Engineering Research, Bullutin of Graduate School of Engineering, Hokkai-Gakuen University, pp. 45-50, 2017.

 ISBN: 978-1-61208-759-7

Development of a System for Detection and Monitoring of Heart Diseases based on ECG and Activity Recognition

Pinar Bisgin, M.Sc.

HealthCare, Fraunhofer Institute for Software and System Engineering, Dortmund, Germany
Email: pinar.bisgin@isst.fraunhofer.de

***Abstract*—Patients with suspected chronic heart failure or symptomatic arrhythmias receive long-term Electrocardiogram ECG. The ECG device records the electrical activity of the heart in a variety of activities in everyday life. The patient must record a protocol during the measurement to demonstrate the various activities of daily life and at different times of the day. The protocol to be made is usually inaccurate because the patient has not usually logged every single activity. Through the fusion of activity detection and ECG, this information is passed on to the physician to aid in the assessment of the diagnosis. Furthermore, the method can be used for therapeutic purposes to consider cardiac activity in specific activities. Human daily activity recognition has gained much attention since it has a wide range of applications. In the preliminary work was a Body Sensor Network (BSN) developed consisting of accelerometer, gyroscope, and barometer. For that purpose 20 subjects are participated. This collected sensor data was evaluated using Machine-Learning algorithms. Finally, the averaged classification results could be obtained: achieving F1 score of (89.22 ± 3.18) % with the SVM algorithm. In upcoming work, the proposed activity classification module will be combined with a mobile ECG sensor. This combination can be provided with a possible and correct protocol, but also relieve the patient. In further steps, the data can be identified using the combined sensor data which have pathological patterns. For this purpose, classification algorithms will be used.**

Keywords–Human-activity recognition; activity recognition; wearable sensors; mobile healthcare system; ECG; fusion.

I. INTRODUCTION

Patients with chronic heart failure or symptomatic arrhythmias, a long-term Electrocardiogram (ECG) is usually performed. Long-term ECG is an important tool for diagnosing cardiovascular diseases and assessing the involvement of the cardiovascular system in primary extracardiac disorders. The long-term ECG and the accompanying activity measurement can be used to monitor the state of health. The vital parameters of the patient can be recorded at a defined load. The resting pulse, pulse during exercise and the respiratory rate can be recorded - especially in the larger efforts, such as climbing stairs. The recording of activities with a longer period of time is very important. With the pacemaker already the activity can be measured. Thus, the health condition of the patient can be observed and deteriorations can be detected. Furthermore, it is suitable for controlling antiarrhythmic pharmacotherapies [1], i.e. Congestive Heart Failure (CHF). To diagnose these arrhythmias, patients are medically monitored in their everyday life. The monitoring can take several hours to several days. The measurement takes place at the home of the patients. The ECG device records the electrical activity of the heart in a variety of activities in everyday life. The patient must protocol the various activities they are performing during the measurement using a long-term ECG [2], because the doctor wants to evaluate the heart activity of the patient. The patient protocols his activities so that the doctor can look at each recorded heart activity, what the patient was doing, and whether the heart activity is in the normal range. After the data has been recorded, it is evaluated by a physician. However, the protocol to be made is usually inaccurate because the patient usually has not logged every single activity. If irregularities are identified by the doctor, it cannot be understood exactly what the patient was doing at a given time. This combination can be provided with a possible and correct protocol, but also relieve the patient. Furthermore, pathological patterns can be identified from the data by using classification algorithms.

The paper addresses past work, and aims to unveil future work as part of a PhD. The paper is organized as follows: Section 2 describes the preliminary work for activity recognition. Finally, section 3 describes some thoughts of the dissertation.

II. PRELIMINARY WORK

A. Body Sensor Network

In order to achieve therapeutic success, an attempt is made to recognize the activity by means of a BSN. A BSN includes a number of implanted, portable (near-the-body) or near-remote sensors and low power actuators. The purpose of BSNs is not only to monitor the patient's illness, but also to make predictions about the course of the disease. For long-term patient monitoring, daily activities must be detected with high accuracy. The focus is on activities in daily life such as rest, walking, jogging, and climbing stairs.

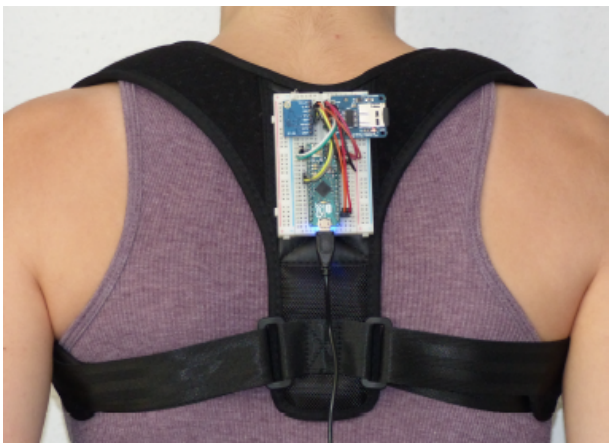

Figure 1. Wearable sensor platform fixed at a test subject's upper body.

 ISBN: 978-1-61208-759-7

The BSN is based on an Arduino Micro, which contains a ATmega32U4 microcontroller. The data is provided via I2C by the 10DOF sensor module GY86, which comprises the barometric pressure sensor MS5611, as well as the MPU6050 3D accelerometer and 3D gyroscope. A MicroSD breakout board+ was used for data storage. The entire platform is fixed at the test subject's upper body by use of a chest band (see Figure 1) [3]. The study was conducted with 20 healthy volunteers ($45/55$ % male/female) with age (24 ± 4) years (range 21–38 years). The study protocol was divided into two parts, which provided the training and test data and had an approximate duration of 20 min and 30 min, respectively. Both protocol stages comprised several phases of the activities standing, walking, jogging, upstairs and downstairs.

B. Activity Classification

Activity classification can aid in the evaluation of the medical data. The classification task is usually used to provide medical sensor systems with context information such as the ECG. For this purpose, positioned a sensor platform on the back of the subjects. Subsequently, a study was carried out with the sensor platform. The platform recorded sensor data stored on an integrated SD card. For evaluating the sensor platform with efficacy the conducted study was divided into two parts. In the two parts, the algorithm's training and evaluation data were collected on the classes walking, standing, running, ascending stairs and descending stairs. The goal was to evaluate daily activities with the algorithm.

This sensor data were evaluated using Machine-Learning algorithms like k-Nearest-Neighbors (KNN), Support Vector Machines (SVM), Neural Networks, Random Forest and AdaBoost. The process caused the calculation of various features and several feature selection techniques used to reduce dimensionality. It helped to improve the classification quality. Finally, parameter optimization was carried out for each classifier. The hyperparameter was optimized with respect to the F1 score [4]

$$\mathrm{F1} = \frac{2 \cdot \mathrm{P} \cdot \mathrm{S}}{\mathrm{P} + \mathrm{S}}, \tag{1}$$

which combines the two common measures of classification accuracy, sensitivity *S* and precision *P*

$$\mathrm{S} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad \mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{2}$$

where TP, FP, and FN denote true positives, false positives, false negatives, respectively. As a result, an optimal combination was obtained of parameters for the respective classifiers. After completion of optimization, the second part of the study was evaluated using a grid search. Then, averaging of the optimized parameters over all the subjects was done.

TABLE I. SVM CLASSIFICATION RESULTS.

Activity	*Sensitivity [%]*	*Precision[%]*	*F1-Score [%]*
idle	97.69 ± 1.76	99.02 ± 1.71	98.33 ± 1.04
walk	85.80 ± 7.11	94.88 ± 3.74	89.88 ± 3.79
run	95.11 ± 6.29	99.79 ± 0.94	(7.28 ± 3.60)
up	91.51 ± 11.81	76.85 ± 11.55	82.35 ± 7.55
down	85.91 ± 8.81	73.69 ± 13.38	78.26 ± 8.57
total	91.21 ± 3.39	88.85 ± 2.51	89.22 ± 3.18

With these values, finally, the averaged classification results could be obtained: F1-score of the SVM $(89.22 \pm 3.18)\%$, which achieved the best results. While KNN, Random Forest and AdaBoost had promising results, too, Neural Networks did not work on the data. The results of the classification using SVM of the activities *idle*, *walk*, *run*, *up* and *down* are summarized in TABLE I.

III. UPCOMING WORK

With the accelerometer, gyroscope and the barometer, it is possible to recognize the activities standing, walking, jogging, upstairs and downstairs. There were no motion artifacts detected by positioning the BSN on the back. The portable device activity classification module is developed on a continuous basis because of the following requirements: classification accuracy, energy consumption and computation time.

In my dissertation, I would like to expand the activity recognition by adapting the sensor system to real-time measurement. The research question is: the combination of heart data and activity helps to better differentiate physiological states. Furthermore, the proposed scheme has great potential in real-time applications due to its inability to dimensionality reduction, simple classifier structure, and good recognition performance.

In addition to the activity, a continuous ECG measurement should take place in order to detect direct changes. This component should also be integrated into the sensor network. With the combination of the ECG and activity detection, on the one hand, the physician can receive the assignment of heart activities during daily activities. On the other hand, an evaluation can be made by the algorithm, so that a home monitoring system can detect pathological patterns from the data without going to a doctor. For this reason, I have to involve clinical partners to consider patients with CHF. The CHF is a chronic progressive disease that affects the pumping capacity of the heart muscle. The heart is unable, despite sufficient blood supply, to promote sufficient cardiac output. Cardiac output refers to the volume that the heart spills into the circulation per minute [5]. Patients and clinical partners will be involved during these projects from the University Hospital Aachen and Bonn, Germany.

REFERENCES

[1] G. e. a. Sauer, "Position paper for performing quality checks on resting, exercise and long-term ecg," Journal of Cardiology, vol. 94, no. 12, 2005, pp. 844–857.

[2] R. Klinge, The Electrocardiogram: Guide to Education and Practice. Georg Thieme publishing company, 2015.

[3] J. Kirchner, S. Faghih-Naini, P. Bisgin, and G. Fischer, "Sensor selection for classification of physical activity in long-term wearable devices," in 2018 IEEE SENSORS, pp. 1709–1712.

[4] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," ACM Computing Surveys (CSUR), vol. 46, no. 3, 2014, p. 33.

[5] C. Prinz, Basic Knowledge of Internal Medicine. Springer publishing company, 2012.

 ISBN: 978-1-61208-759-7

3D Upper-Body Pose Estimation and Classification for Detecting Unhealthy Sitting Postures at the Workplace

Christopher Pramerdorfer
cogvis & TU Wien
Wiedner Hauptstraße 17/3a
1040 Vienna, Austria
Email: pramerdorfer@cogvis.at

Martin Kampel
TU Wien
Favoritenstraße 9-11
1040 Vienna, Austria
Email: kampel@cvl.tuwien.ac.at

Johannes Heering
Fitbase
Brauhausstieg 21
22041 Hamburg, Germany
Email: heering@fitbase.de

***Abstract*—Prolonged sitting in an unhealthy posture is a common cause of back pain and other health problems in office workers. People are often not aware that they are sitting in an unhealthy way, the problems this can cause in the long term, and how they should improve their posture. We present a system that is able to provide this information by analyzing people's postures over several days. The system is fully automatic and requires no worn devices. Instead, data from a depth sensor is used for periodic 3D upper-body pose estimation. This pose estimation is carried out by a convolutional neural network that was trained on synthetic depth data to overcome the lack of available real-world datasets. On this basis, each pose is assigned to one of several common classes of healthy and unhealthy sitting poses. This results in a large collection of body poses and classification results, which are used to generate a personalized posture report that includes suggestions for improving the sitting posture. We show experimentally that the system is able to estimate 3D poses and perform pose classification with high accuracy.**

Keywords–Workplace health promotion; Sitting posture estimation; Deep learning; Depth data analysis.

I. INTRODUCTION

Approximately 75% of all employees in industrial countries have jobs that require working in a seated position [1]. In the DACH-region (Germany, Austria, and Switzerland) alone, this is the case for around 15 million people. Prolonged sitting is a common cause of pain and health problems in office workers [2][3]. People are often not aware that they are sitting in an unhealthy way, which problems this can cause in the long term, and how to improve their posture [4].

In this paper, we present *ergoscan*, a system we developed to address this issue. Ergoscan periodically measures the head and upper-body pose of people sitting in front of their computer at the workplace. The system requires no worn sensors or any form of user participation, which users might consider intrusive, and does not utilize image or video data to protect the privacy of monitored persons. Instead, ergoscan processes depth data, enabling 3D pose estimation and classification with high accuracy. This information is collected periodically over several days and sent to a server in an encrypted and anonymized form. All processing is carried out locally such that no other data have to leave the system.

An ergoscan system consists of a 3D sensor and an ARM-based single-board computer, which are integrated in a single casing that is mounted at the top of the user's computer monitor as illustrated in Figure 1. The system does not require accurate placement or alignment to facilitate installation; it obtains this information automatically via calibration based on visible planar surfaces such as walls. Installation takes less than a minute and requires no tools or adhesives.

Figure 1. A ergoscan device mounted on top of a monitor.

Once installed, ergoscan periodically performs face detection using an efficient cascade detector [5] to determine whether a person is sitting in front of the screen. If a person is detected, a Convolutional Neural Network (CNN) with a novel architecture estimates their head and upper-body pose. CNNs achieve state-of-the-art performance in related tasks, such as pose estimation in color images, but require large datasets for training, which are not available in our specific problem domain [6][7]. As obtaining a suitable dataset would be a significant effort in this domain, we utilize synthetic depth data for training and show that this is an effective alternative.

The six keypoints located during pose estimation are nasion (intersection of the frontal bone and the two nasal bones of the human skull), chin center, front of the throat, manubrium, as well as the left and right shoulders. These keypoints were selected based on feedback by physiotherapists but the method is generic and can be adapted to any number of keypoints. Angles derived from the 3D coordinates of these keypoints are input to a random forest classifier [8], which assigns one of 15 classes of common healthy and unhealthy sitting postures that were defined together with experts. Two of these poses are visualized in Figure 2.

Systems remain at a particular workplace for up to one week. During this time, thousands of pose measurements and classifications are collected. Physiotherapists analyze this information in an aggregated form to identify unhealthy sitting

 ISBN: 978-1-61208-759-7

poses that are commonly assumed. On this basis, the monitored person receives a personal posture report with descriptions and visualizations of these poses, as well as suggestions and links to video tutorials for improving their sitting posture in order to prevent long-term health problems. Figure 2 shows visualizations from an example report.

Figure 2. Example visualizations from a sitting posture report.

We assessed the performance of ergoscan on a dataset of 1500 samples, with each depicting one of 31 people assuming the 15 prototype postures. On this dataset, ergoscan estimates 3D keypoint coordinates with an average error of 2 to 5 cm depending on the keypoint, and is able to perform pose classification at an accuracy above 99%.

This paper is structured as follows. Section II summarizes related work on body pose estimation and synthetic dataset generation. Our pose estimation and classification methods are described in Sections III and IV, respectively. Section V describes the experiments and discusses the results, and Section VI concludes the paper.

II. Related Work

Human pose estimation in color images via CNNs is a popular research topic. Two seminal works in this field are [9] and [10]. Both use networks with fully-connected layers for regressing keypoint coordinates. The former work is the first to demonstrate that 3D poses can be recovered from color images, although this is possible only up to scale. More recent works such as [6][7][11] instead perform dense keypoint prediction with fully convolutional networks, which improves accuracy but is slower and requires more memory due to the the additional upsampling path. As these resources are limited on ergoscan devices, we opt for keypoint regression.

In contrast, there is lack of recent works that utilize depth data. A reason for this is that these sensors are not as widespread as cameras (and camera phones), and consequently a lack of large datasets. Pose estimation in depth data was a popular research topic following the release of the Kinect depth sensor in 2010 [12][13][14]. In contrast to these methods, which perform pose estimation via regression forests, we utilize CNNs for this task due to their higher performance. This was shown in [15], which presented a patch-based method for 3D pose estimation in depth data using a combination of a CNN and a recurrent neural network. Two more recent methods are [16] and [17]. The former utilizes a CNN for estimating the coefficients of a linear combination of prototype poses that result in the pose depicted in the input depth map. The latter both processes and predicts 3D volumes, arguing that regressing 3D poses directly from 2D depth maps hinders optimization during training. Our method processes 2D depth maps, which is computationally more efficient, and avoids such problems during training via two-stage keypoint regression.

Synthetic datasets are a promising means for enabling data-driven solutions in problem domains for which no comprehensive datasets are available. To our knowledge [12] was the first work to demonstrate the potential of this approach for 3D pose estimation in depth maps. We adopt this approach and train the CNN on synthetic depth maps, however we create both body poses and 3D models in software rather than employing actors and motion capturing, which is less labor-intensive and requires no special equipment. The most comprehensive public dataset that includes depth maps of people is SURREAL [18]. However, this dataset does not reflect our problem domain in terms of body poses and keypoints.

III. Pose Estimation

Our pose estimation method takes a depth map and a face bounding box as the input and outputs $K = 6$ 3D keypoints.

A. Preprocessing

Preprocessing entails converting the input to be compatible with the CNN. First, the face bounding box is extended by a factor of four in order to capture the head and upper body of the monitored person. The resulting depth map patch is extracted and resampled to a fixed size of 96×96 pixels. This follows normalization of the pixel values, which are given in mm, based on the operating conditions of the ergoscan system, namely assuming a maximum person distance of $d_{max} = 1500$ mm as well as considering that the sensor is unable to measure distances closer to $d_{min} = 400$ mm. To do so, pixels greater than d_{max} are set to 0, which effectively removes most background objects, and then all values are scaled linearly from $[d_{min}, d_{max}]$ to $[0, 1]$ to facilitate transfer learning. While the resulting *normalized distances* are no longer metric, relative distances are preserved and the original distances can be recovered via the inverse mapping.

B. Pose Estimation Network

Pose estimation is carried out by a CNN that first predicts image coordinates of all keypoints and, on this basis, the corresponding normalized distances. We found this approach to be more stable during training, as suggested in [17].

The network architecture is illustrated in Figure 3 and consists of three stages. The first stage performs feature extraction, from which the second stage regresses image coordinates. This follows a novel stage for distance prediction that integrates information from the previous stages and the input image. The outputs are the image coordinates predicted by the second stage and the corresponding distances from the third stage.

The feature extraction stage is a ResNet-18 [19] that was pre-trained for classification on ImageNet [20] and then fine-tuned for pose estimation in depth data. We chose this architecture due to its high performance and ability to run on the target hardware at the required speed. After pre-training, we replaced the classifier with a keypoint regressor that forms the second stage of the network, and performed fine-tuning.

This second stage regresses keypoint image coordinates using the features extracted in the previous stage. The layer

 ISBN: 978-1-61208-759-7

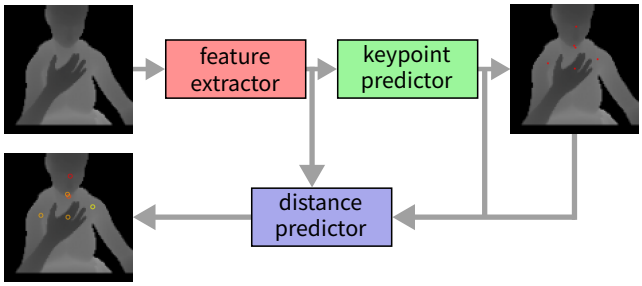

Figure 3. Overview of the stages and data flow through the network.

architecture is detailed in Table I and starts with a global average pooling layer to convert the feature tensors to vectors. This follows a linear layer with a ReLU activation and final linear layer with $2K$ neurons (K keypoints with two coordinates). Batch normalization and dropout layers are utilized to facilitate network optimization and for regularization, respectively.

TABLE I. IMAGE COORDINATE REGRESSION ARCHITECTURE.

Layer type
Global average pooling
Batch normalization
Dropout ($p = 0.25$)
Linear (512 neurons)
ReLU
Batch normalization
Dropout ($p = 0.25$)
Linear (12 neurons)

The first and second stages were trained together as follows. First, we trained only the second stage while utilizing the first (pre-trained) one as a static feature extractor. Once the validation loss had saturated, we fine-tuned both stages to allow the network to adapt to the depth data. This approach prevented the first stage from adapting in a detrimental way due to errors made initially by the untrained second stage. We minimized the Huber loss [21] between predicted and ground-truth image coordinates using stochastic gradient descent with a cyclic learning rate and momentum [22].

The distance regression stage predicts normalized distances for all keypoints. Our goal when designing this stage was to make all relevant information available to it, namely the keypoint coordinates predicted in the previous stage but also the corresponding normalized distances in the input depth map as well as the features extracted by the first stage. This is realized using a *distance lookup layer* that converts the predicted image coordinate vectors to integral coordinates and uses this information to access the corresponding normalized distances in the input depth map. If an image coordinate is out of bounds or if there is no distance information available, the layer assigns a normalized distance of -1 to signal the later stages that there are missing data. The layer returns the original keypoint coordinates as well as the corresponding distances, i.e. a $B \times 3K$ tensor with B being the minibatch size.

Table II summarizes the layer composition of the stage. Initially, there are two parallel branches. The branch that processes features starts identically to the keypoint prediction stage and outputs a $B \times F$ tensor. The other branch consists of the distance lookup layer. Dropout is omitted in this branch to preserve information. The outputs of both branches are then concatenated to a single $B \times F + 3K$ tensor. The remaining layers are consistent with the keypoint regression stage.

TABLE II. DISTANCE REGRESSION ARCHITECTURE.

Input-Features	-Keypoints	-Images
Global average pooling Dropout ($p = 0.25$)	Distance lookup	
Concatenation		
Batch normalization		
Linear (512 neurons)		
ReLU		
Batch normalization		
Dropout ($p = 0.25$)		
Linear (6 neurons)		

We added and trained this stage after the previous stages, which were not modified in this process. We again minimized the Huber loss but did not penalize errors for keypoints whose ground-truth image coordinates were outside the image.

C. Synthetic Training Set

The pose estimation network was trained solely on synthetic data. The key considerations during dataset design were realism and comprehensiveness in order to ensure that trained models would be able to generalize to actual sensor data. To this end, the goal was to capture a wide variety of realistic body types, poses, and office environments.

Pose animations were carried out using the *Blender* 3D modeling software with the *ManuelBastioniLAB* addon for person models and animations. These tools enable realistic and anatomically correct person animations in 3D. The amount and types of modeled poses were chosen based on studies on sitting postures in office environments as well as own analyses of office recordings. This was to ensure that the resulting set of poses would be both realistic and comprehensive. 5000 different poses were generated.

On this basis, 15000 different 3D person models were created. This highlights the potential of synthetic data – recruiting this many people for recording is infeasible for most companies and research institutes. Care was taken to ensure that the models capture a wide variety of realistic body shapes. For this purpose, character properties such as gender, age, height, weight, and body tone were varied, with each affecting body and face shapes in a realistic way. Each 3D model has hair and clothing for increased realism, and includes accurate ground-truth pose information. Figure 4 shows two examples.

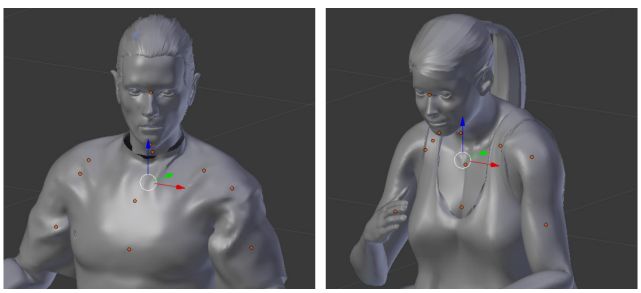

Figure 4. 3D person models for training purposes.

 ISBN: 978-1-61208-759-7

Depth maps were rendered by a custom renderer that randomly selects a person, a desk and chair for the person, and a piece of background furniture from a pool of available 3D models, arranges these objects in the scene in a realistic way, places a virtual camera at a varying location and orientation, and then renders a depth map. The renderer also exports metadata such as camera parameters and image coordinates of keypoints. All 3D models of furniture were manually selected from the ShapeNet dataset [23]. The pool of background furniture comprised 1663 models of shelves, cupboards, and couches. The virtual camera had VGA resolution and a field of view of 60 degrees, similar to current off-the-shelf depth sensors. Figure 5 shows a rendered depth map.

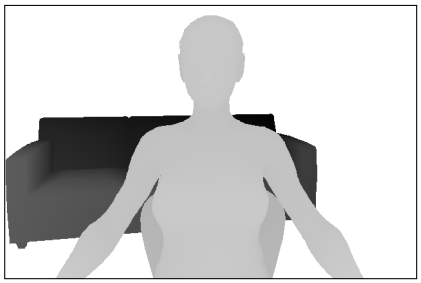

Figure 5. Visualisation of a rendered depth map (further objects appear darker). The desk is outside the field of view in this example.

For increased realism, sensor noise similar to that of the Kinect was simulated. Our method for simulating unsuccessful measurements (zero-pixels) was based on [24]. For each depth map, we computed a smoothed normal map and set depth map pixels whose normals were close to perpendicular to 0. We then applied additive Gaussian random noise with a standard deviation based on the measured distance, in approximation of the random noise of the sensor [25].

D. Postprocessing

To obtain a 3D pose from the CNN output, predicted normalized distances are first mapped to distances from the sensor. The predicted image coordinates are then converted to camera coordinates using the known sensor intrinsics and mapped distances, which in turn are mapped to world coordinates using the extrinsics estimated during system calibration.

IV. Pose Classification

Pose classification is based on angles rather than absolute world coordinates. This angle representation has the advantage of being invariant to the offset between monitored people and the sensor, which generally varies over time. It also increases robustness with respect to variations in person height as angle representations are invariant to uniform scaling.

Our angle representation of a given 3D pose is a collection of 20 informative 2D angles. We favor this 2D approach as we consider such angles – and derived classification rules – more intuitive than 3D angles. Each angle is calculated by computing the 3D vector between two specific keypoint coordinates, discarding a particular coordinate to obtain a 2D vector, and computing the angle between this vector and $(1, 0)$.

The rules for obtaining these angles, i.e. which 3D vectors to compute and which coordinates to discard, were found via feature selection on a training set consisting of 33000 3D poses estimated by several ergoscan devices. Each of these poses was assigned a ground-truth class label by experts. Feature selection was carried out by computing all 90 possible 2D angles for each of these poses, training a random forest on the resulting dataset, and determining the 20 most important angles according to the feature importances learned by the forest [8].

V. Experiments

We assessed the pose estimation and classification performance of ergoscan, and studied the performance impact of training on synthetic data.

A. Dataset

As there were no public datasets available that reflect our problem domain, we created such a dataset ourselves. For this purpose, 31 people were recruited, which were assuming the 15 prototype sitting poses under supervision. During this time, an ergoscan system computed 1500 pose estimates and classifications (100 per pose). Ground-truth keypoints and class labels for these samples were obtained using a professional motion capture system and manual labeling, respectively.

B. Pose Estimation Performance

We calculated the estimation error for a given pose estimate and keypoint as the Euclidean distance between the 3D keypoint coordinate measured by ergoscan and the corresponding ground-truth coordinate. We did so for each keypoint and sample, and report averages and standard deviations.

Figure 6 summarizes the results. The estimation errors are under 30 mm on average, with the exception of the shoulder keypoints. For the shoulders, the CNN often predicted keypoints that were too low. This was caused by missing data for the upper parts of the shoulders due to sensor limitations. The results are promising and confirm that it is feasible to train a CNN for pose estimation in depth data on synthetic data.

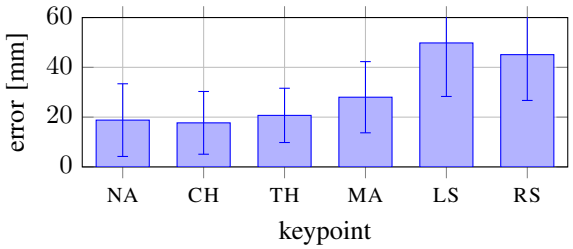

Figure 6. 3D keypoint estimation errors on the test dataset. NA: nasion, CH: chin, TH: throat, MA: manubrium, LS: left shoulder, RS: right shoulder.

C. Generalization Performance

We studied the decrease in performance incurred by training on synthetic data, which despite our efforts do not (and arguably cannot) perfectly match real sensor data. For this purpose, we split the synthetic dataset into a training set (52000 samples) and a test set (8000 samples), and retrained the network using the same hyperparameters. We then computed the per-keypoint estimation errors on the test set like before.

The results are shown in Figure 7. As expected, the estimation errors are significantly lower on synthetic data than on real data (Figure 6). As care was taken to render the synthetic data as realistic as possible, this indicates that a decrease in performance must be accepted in general when training on synthetic depth data.

 ISBN: 978-1-61208-759-7

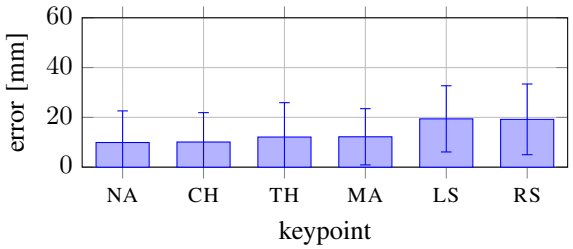

Figure 7. 3D keypoint estimation errors on the synthetic test set. NA: nasion, CH: chin, TH: throat, MA: manubrium, LS: left shoulder, RS: right shoulder.

D. Pose Classification Performance

Ergoscan misclassified only one of the 1500 samples in the test dataset. This confirms that ergoscan is able to classify poses with high accuracy and consequently that ergoscan can detect unhealthy sitting poses reliably.

VI. Conclusion and Future Work

We have presented ergoscan, a system for promoting a healthy posture in office workers by raising awareness. Ergoscan requires no user participation and monitors people's postures over several days to identify unhealthy postures that are frequently assumed. Posture monitoring is realized using a CNN for upper-body pose estimation with an architecture optimized for depth data analysis. On this basis, ergoscan automatically assigns each pose estimate to one of 15 common sitting poses. The results confirm that training CNNs on synthetic data can be a suitable approach if no comprehensive real datasets are available, and that ergoscan is able to perform pose estimation and classification reliably. We plan to investigate performance penalties due to synthetic training data in a more detailed and general way, and on this basis to develop improved sensor noise simulation methods. Another task planned for the future is providing realtime feedback to users via a website or smartphone app in addition to the reports.

Acknowledgment

This work was supported by the Austrian Research Promotion Agency (FFG-855696).

References

[1] S. Reinecke, R. Hazard, K. Coleman, and M. Pope, "A continuous passive lumbar motion device to relieve back pain in prolonged sitting," Advances in industrial ergonomics and safety IV, 2002, pp. 971–976.

[2] A. M. Lis, K. M. Black, H. Korn, and M. Nordin, "Association between sitting and occupational LBP," European Spine Journal, vol. 16, no. 2, 2007, pp. 283–298.

[3] B. Cagnie, L. Danneels, D. Van Tiggelen, V. De Loose, and D. Cambier, "Individual and work related risk factors for neck pain among office workers: a cross sectional study," European Spine Journal, vol. 16, no. 5, 2007, pp. 679–686.

[4] N. J. Delleman, C. M. Haslegrave, and D. B. Chaffin, Working Postures and Movements. CRC Press, 2004.

[5] S. Wu, M. Kan, Z. He, S. Shan, and X. Chen, "Funnel-Structured Cascade for Multi-View Face Detection with Alignment-Awareness," Neurocomputing, 2016.

[6] A. Newell, K. Yang, and J. Deng, "Stacked Hourglass Networks for Human Pose Estimation," in European Conference on Computer Vision. Springer, 2016, pp. 483–499.

[7] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," CoRR, vol. abs/1703.06870, 2017.

[8] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, 2001, pp. 5–32.

[9] S. Li and A. B. Chan, "3D Human Pose Estimation from Monocular Images with Deep Convolutional Neural Network," in Asian Conference on Computer Vision, 2014, pp. 332–347.

[10] A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," in Conference on Computer Vision and Pattern Recognition, 2014, pp. 1653–1660.

[11] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," in Conference on Computer Vision and Pattern Recognition, 2019, pp. 5693–5703.

[12] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-Time Human Pose Recognition in Parts from Single Depth Images," in Conference on Computer Vision and Pattern Recognition, 2011, pp. 1297–1304.

[13] R. Girshick, J. Shotton, P. Kohli, A. Criminisi, and A. Fitzgibbon, "Efficient regression of general-activity human poses from depth images," in International Conference on Computer Vision, 2011, pp. 415–422.

[14] H. Yub Jung, S. Lee, Y. Seok Heo, and I. Dong Yun, "Random Tree Walk toward Instantaneous 3D Human Pose Estimation," in Conference on Computer Vision and Pattern Recognition, 2015, pp. 2467–2474.

[15] A. Haque, B. Peng, Z. Luo, A. Alahi, S. Yeung, and L. Fei-Fei, "Towards Viewpoint Invariant 3D Human Pose Estimation," in European Conference on Computer Vision, 2016, pp. 160–177.

[16] M. J. Marín-Jiménez, F. J. R. Ramírez, R. Muñoz-Salinas, and R. Medina Carnicer, "3D human pose estimation from depth maps using a deep combination of poses," CoRR, vol. abs/1807.05389, 2018.

[17] G. Moon, J. Yong Chang, and K. Mu Lee, "V2V-PoseNet: Voxel-to-Voxel Prediction Network for Accurate 3D Hand and Human Pose Estimation from a Single Depth Map," in Conference on Computer Vision and Pattern Recognition, 2018, pp. 5079–5088.

[18] G. Varol, J. Romero, X. Martin, N. Mahmood, M. J. Black, I. Laptev, and C. Schmid, "Learning from Synthetic Humans," in Conference on Computer Vision and Pattern Recognition, 2017.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, no. 3, 2015, pp. 211–252.

[21] P. J. Huber, "Robust estimation of a location parameter," in Breakthroughs in statistics. Springer, 1992, pp. 492–518.

[22] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1," CoRR, vol. abs/1803.09820, 2018.

[23] A. X. Chang, T. A. Funkhouser, L. J. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, "ShapeNet: An Information-Rich 3D Model Repository," CoRR, vol. abs/1512.03012, 2015.

[24] C. Xu and L. Cheng, "Efficient Hand Pose Estimation from a Single Depth Image," in International Conference on Computer Vision, 2013, pp. 3456–3462.

[25] K. Khoshelham and S. O. Elberink, "Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications," Sensors, vol. 12, no. 2, 2012, pp. 1437–1454.

 ISBN: 978-1-61208-759-7