Mariella Särestöniemi ·
Pantea Keikhosrokiani · Daljeet Singh ·
Erkki Harjula · Aleksei Tiulpin ·
Miia Jansson · Minna Isomursu ·
Mark van Gils · Simo Saarakkala ·
Jarmo Reponen (Eds.)

Communications in Computer and Information Science 2084

Digital Health and Wireless Solutions

First Nordic Conference, NCDHWS 2024
Oulu, Finland, May 7–8, 2024
Proceedings, Part II

Part 2

Springer

OPEN ACCESS

Communications in Computer and Information Science 2084

Rationale
The CCIS series is devoted to the publication of proceedings of computer science conferences. Its aim is to efficiently disseminate original research results in informatics in printed and electronic form. While the focus is on publication of peer-reviewed full papers presenting mature work, inclusion of reviewed short papers reporting on work in progress is welcome, too. Besides globally relevant meetings with internationally representative program committees guaranteeing a strict peer-reviewing and paper selection process, conferences run by societies or of high regional or national relevance are also considered for publication.

Topics
The topical scope of CCIS spans the entire spectrum of informatics ranging from foundational topics in the theory of computing to information and communications science and technology and a broad variety of interdisciplinary application fields.

Information for Volume Editors and Authors
Publication in CCIS is free of charge. No royalties are paid, however, we offer registered conference participants temporary free access to the online version of the conference proceedings on SpringerLink (http://link.springer.com) by means of an http referrer from the conference website and/or a number of complimentary printed copies, as specified in the official acceptance email of the event.

CCIS proceedings can be published in time for distribution at conferences or as post-proceedings, and delivered in the form of printed books and/or electronically as USBs and/or e-content licenses for accessing proceedings at SpringerLink. Furthermore, CCIS proceedings are included in the CCIS electronic book series hosted in the SpringerLink digital library at http://link.springer.com/bookseries/7899. Conferences publishing in CCIS are allowed to use Online Conference Service (OCS) for managing the whole proceedings lifecycle (from submission and reviewing to preparing for publication) free of charge.

Publication process
The language of publication is exclusively English. Authors publishing in CCIS have to sign the Springer CCIS copyright transfer form, however, they are free to use their material published in CCIS for substantially changed, more elaborate subsequent publications elsewhere. For the preparation of the camera-ready papers/files, authors have to strictly adhere to the Springer CCIS Authors' Instructions and are strongly encouraged to use the CCIS LaTeX style files or templates.

Abstracting/Indexing
CCIS is abstracted/indexed in DBLP, Google Scholar, EI-Compendex, Mathematical Reviews, SCImago, Scopus. CCIS volumes are also submitted for the inclusion in ISI Proceedings.

How to start
To start the evaluation of your proposal for inclusion in the CCIS series, please send an e-mail to ccis@springer.com.

Mariella Särestöniemi · Pantea Keikhosrokiani · Daljeet Singh · Erkki Harjula · Aleksei Tiulpin · Miia Jansson · Minna Isomursu · Mark van Gils · Simo Saarakkala · Jarmo Reponen
Editors

Digital Health and Wireless Solutions

First Nordic Conference, NCDHWS 2024
Oulu, Finland, May 7–8, 2024
Proceedings, Part II

Springer

Editors
Mariella Särestöniemi
University of Oulu
Oulu, Finland

Daljeet Singh
University of Oulu
Oulu, Finland

Aleksei Tiulpin
University of Oulu
Oulu, Finland

Minna Isomursu
University of Oulu
Oulu, Finland

Simo Saarakkala
University of Oulu
Oulu, Finland

Pantea Keikhosrokiani
University of Oulu
Oulu, Finland

Erkki Harjula
University of Oulu
Oulu, Finland

Miia Jansson
University of Oulu
Oulu, Finland

Mark van Gils
University of Oulu
Tampere, Finland

Jarmo Reponen
University of Oulu
Oulu, Finland

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-031-59090-0 ISBN 978-3-031-59091-7 (eBook)
https://doi.org/10.1007/978-3-031-59091-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Foreword

Digital transformation is reshaping healthcare tools, organizations, and operating models. The concept of digital health encompasses previously described electronic healthcare services, such as telemedicine or eHealth, augmented with advanced data processing and computing methods like artificial intelligence (AI). Broadly, digital health methods are believed to bring modern healthcare services to areas where they were previously unavailable. Similarly, advanced methods are seen to enable more precise diagnostics and treatments using collected digital health data. Additionally, advancements in monitoring healthcare quality and effectiveness are anticipated, along with progress in research and education through better secondary use of health information. The development of wireless technologies has brought healthcare services to users' mobile devices, allowing citizens to participate in their healthcare in unprecedented ways.

Finland has been one of the leading countries in digitalization within the European Union's DESI measurements for several years. Patient information systems in public healthcare have been solely electronic since 2007. Finland has also implemented a comprehensive national health data repository and exchange platform, KANTA, allowing citizens to access their own health information, too. Currently, every resident in Finland has an account in this national service, which proved its value especially during the COVID-19 pandemic.

Digital health is a timely research topic, particularly as Finland undergoes significant regional changes in the organization of health and social services. As part of this transformation, services are increasingly delivered through digital channels, with more responsibility given to citizens. Similar changes are underway to varying degrees in other Nordic countries. At the European level, the adoption of the EU's AI Act and the preparation of the European Health Data Space (EHDS) are significant. Addressing these challenges and opportunities requires not only advanced basic research and the development of methods and innovations but also applied research on treatment methods and digital care pathways, as well as research evaluating operations and expertise.

This Nordic Conference on Digital Health and Wireless Solutions (NCDHWS 2024) was organized by the Digital Health (DigiHealth) and 6G Enabled Sustainable Solutions (6GESS) research programs at the University of Oulu. These programs bring together multidisciplinary research activities in medical and health sciences, health economics, information, and sensor technology, as well as wireless communications. Our ambitious aim was to organize for the first time an international multidisciplinary conference providing an excellent opportunity for professionals, researchers, and industry leaders from different fields to discuss and share insights on the latest developments in digital health and related technology.

The host city Oulu is known for being at the forefront of technology innovation. It is a hub for research and development in the fields of digital health and wireless communications. Electronic health records, telemedicine and mobile health services have been in clinical use in the city for more than 25 years. There is active research

aiming not only towards new territories like AI, novel sensor technology and edge computing, but also striving for scientific assessment of the impact of these innovations in real life.

The organization of this new type of multidisciplinary conference combining habits and customs of different fields would not have been possible without the hard work of many colleagues and experts who dedicated their time and expertise to make this conference successful. I would like to thank our multidisciplinary Organizing Committee, my co-chair Simo Saarakkala, and our coordinators Tuire Salonurmi and Sanna Tuomela for fruitful discussions and practical ideas shared in our regular meetings. Additionally, I would like to thank our Program Committee with International Reviewers and our Publication Chairs, both led by Mariella Särestöniemi for compiling excellent scientific content for the conference.

Furthermore, I would like to thank our excellent invited speakers for bringing their expertise in digital health and related technologies to the audience, and all our International Advisory Committee members for their kind assistance. I am grateful to Karoliina Paalimäki-Paakki for leading the Student Volunteer Committee and I would like to thank those student volunteers whose work with practical arrangements during the conference was invaluable. I would also like to thank our media team Katja Longhurst and Sallamaari Syrjä for maintaining our digital presence in various channels. My sincere thanks belong to Minna Komu for marketing the conference to companies and to Oulu University Hospital and to the Wellbeing Services County of North Ostrobothnia for supporting us in conference arrangements and providing a unique research environment. Finally, I would like to collectively thank all those who in various roles contributed to the success of NCDHWS 2024, either as an organizer or as a participant.

Oulu Jarmo Reponen
May 2024

Preface

The Nordic Conference on Digital Health and Wireless Solutions (NCDHWS) is a new international multidisciplinary conference which brings together experts and professionals from different fields. Organization of such a conference necessitates the collaboration of multidisciplinary committees, jointly devising strategies to harmonize the diverse conference cultures from various fields. Thus, all the NCDHWS 2024 conference committees, including program committee chairs, organization committee, program committee, and international advisory committee, have representatives from several different fields of engineering, medicine, and health sciences. Additionally, altogether 27 different countries are represented in our committees, and even more different nationalities.

As an example of combining habits and customs of different fields in our conference organization, we accepted three different submission types: full papers (10–19 pages), short papers (6–9 pages) and abstracts (1–3). Full and short papers appear in the main text of the book and are indexed, whereas abstracts appear in the backmatter of the volumes. This first NCDHWS conference turned out to be successful: we received 100 submissions including 57 full papers, 11 short papers, and 32 abstracts. The review process for this conference was double-blind and our large program committee (i.e., international reviewer committee) consisted of 122 reviewers. For each full and short paper, we selected 3–5 reviewers and for abstracts 2–3 reviewers with suitable scientific background. OpenReview was used as the submission platform since it automatically checks for potential conflicts of interests for each reviewer. OpenReview also automatically ensures confidentiality by shielding program committee chairs from accessing evaluations of the papers where they are authors, and by shielding the identities of the reviewers of their own papers.

Reviewers were advised to score papers based on submission type using OpenReview's scoring table and to provide detailed comments to authors to improve the paper quality. OpenReview has an evaluation scale from 1–10 in which score 6 stands for "marginally above acceptance threshold". Additionally, OpenReviewer lets reviewers score their level of confidence. After the review process, the program committee chairs calculated for each paper the average score, weighted with the reviewers' confidence level, and made final decisions on the acceptance of the papers and their presentation type (oral/poster). Program committee chairs decided to accept all the papers reaching weighted average score 6, i.e., reaching the acceptance threshold level suggested by OpenReview scoring. Program committee chairs naturally did not handle reviewer assignments nor did they make final decisions on their own papers. From the submitted papers, 50 full papers and 7 short papers achieved OpenReview's score 6 and hence were accepted for the proceedings. All authors were requested to make corrections and improvements on their papers based on reviewers' comments before final camera-ready submission. For final versions, we carried out plagiarism checks with Turnitin and asked authors to take actions if the plagiarism scores were high.

This Springer proceedings consists of two volumes. Full and short papers appear in the main parts of the volumes ordered similarly to session themes of the conference. Abstracts appear in the back matters of both volumes following the same thematic session order. We would like to express our gratitude to the Springer Nature team who helped us with all practicalities in this book edition process.

Additionally, we would like to thank all the members of our committees: organization committee, program committee as well as international advisory committee, who altogether made the organization of this exciting multidisciplinary conference possible. We would like to greatly acknowledge the keynote and invited speakers who took time out from their busy schedules and traveled up to Oulu to give us inspiring talks on their research. Finally, we would like to express our sincere thanks to all the authors for choosing NCDHWS 2024 to present their research results. With all their interesting presentations combined with excellent invited speeches, this multidisciplinary gathering was successful and fruitful for new collaborations.

May 2024

Mariella Särestöniemi
Pantea Keikhosrokiani
Daljeet Singh
Erkki Harjula
Miia Jansson
Aleksei Tiulpin
Minna Isomursu
Mark van Gils
Simo Saarakkala
Jarmo Reponen

Organization

General Chairs (Conference President and Vice-president)

Jarmo Reponen	University of Oulu, Finland
Simo Saarakkala	University of Oulu, Finland

Program Committee Chairs

Mariella Särestöniemi (Chair)	University of Oulu, Finland
Erkki Harjula	University of Oulu, Finland
Minna Isomursu	University of Oulu, Finland
Miia Jansson	University of Oulu, Finland
Jarmo Reponen	University of Oulu, Finland
Aleksei Tiulpin	University of Oulu, Finland
Mark van Gils	University of Tampere, Finland

Publication Chairs

Mariella Särestöniemi	University of Oulu, Finland
Pantea Keikhosrokiani	University of Oulu, Finland
Daljeet Singh	University of Oulu, Finland

Organization Committee

Jarmo Reponen (Chair)	University of Oulu, Finland
Simo Saarakkala	University of Oulu, Finland
Erkki Harjula	University of Oulu, Finland
Matti Hämäläinen	University of Oulu, Finland
Minna Isomursu	University of Oulu, Finland
Miia Jansson	University of Oulu, Finland
Karoliina Paalimäki-Paakki	OAMK, Finland
Minna Komu	OuluHealth, Finland
Ari Pouttu	University of Oulu, Finland
Tuire Salonurmi	University of Oulu, Finland
Mariella Särestöniemi	University of Oulu, Finland

Jani Tikkanen	Pohde, Finland
Aleksei Tiulpin	University of Oulu, Finland
Sanna Tuomela	University of Oulu, Finland
Mark van Gils	University of Tampere, Finland

Program Committee

Ijaz Ahmad	VTT Research Centre, Finland
Outi Ahonen	Laurea Applied Science, Finland
Daria Alekseeva	Tampere University, Finland
Slawomir Ambroziak	Gdansk University of Technology, Poland
Daisuke Anzai	Nagoya Institute of Technology, Japan
Atakan Aral	Umeå University, Sweden
Aslak Aslaksen	Bergen University Hospital, Norway
Tunc Asuroglu	VTT Technical Research Centre of Finland, Finland
Robin Augustine	Uppsala University, Sweden
Kirsti E. Berntsen	Norwegian University of Science and Technology, Norway
Roberto Blanco	University of Turku, Finland
Kerryn Butler-Henderson	RMIT University, Australia
Stefano Caputi	University of Florence, Italy
Henry Carvajal	Universidad de Las Américas, Equador
Ioanna Chouvarda	Aristotle University of Thessaloniki, Greece
Omedev Dahia	Lovely Professional University, India
Nils Dahlström	University of Linköping, Sweden
Jeppe Eriksen	University of Aalborg, Denmark
Hany Ferdinando	University of Oulu, Finland
Pål Anders Floor	Norwegian University of Science and Technology, Norway
Mark van Gils	University of Tampere, Finland
Heidi Gilstad	Norwegian University of Science and Technology, Norway
Guido Giunti	Trinity College Dublin, Ireland
Casandra Grundstrom	Norwegian University of Science and Technology, Norway
Erkki Harjula	University of Oulu, Finland
Ying He	University of Sydney, Australia
Sari Heikkinen	Laurea University of Applied Sciences, Finland
Juuso Heikkinen	Oulu University Hospital, Finland
Helinä Heino	University of Oulu, Finland

Minna Hokka	Diaconia University of Applied Sciences, Finland
Syed Sajid Hussain	Norwegian University of Science and Technology, Norway
Piia Hyvämäki	Oulu University of Applied Sciences, Finland
Matti Hämäläinen	University of Oulu, Finland
Iiris Hörhammer	Aalto University, Finland
Milla Immonen	Lapland University of Applied Sciences, Finland
Minna Isomursu	University of Oulu, Finland
Antti Isosalo	University of Oulu, Finland
Miia Jansson	University of Oulu, Finland
Vesa Jormanainen	University of Helsinki, Finland
HemDutt Joshi	Thapar Institute of Engineering and Technology, India
Pirjo Kaakinen	University of Oulu, Finland
Kimmo Kansanen	Norwegian University of Science and Technology, Norway
Outi Kanste	Laurea University of Applied Sciences, Finland
Pasi Karppinen	University of Oulu, Finland
Jani Katisko	Oulu University Hospital, Finland
Kaisa Leena Kauppinen	University of Oulu, Finland
Pantea Keikhosrokiani	University of Oulu, Finland
Ali Khaleghi	Norwegian University of Science and Technology, Norway
Jussi Koivunen	Oulu University Hospital, Finland
Jorma Komulainen	University of Eastern Finland
Elina Kontio	Turku University of Applied Sciences, Finland
Juha Korpelainen	Oulu University Hospital, Finland
Hilkka Korpi	Oulu University of Applied Sciences, Finland
Tuomas Koskela	University of Tampere, Finland
Pirkko Kouri	Finnish Society of Telemedicine and eHealth, Finland
Narasimharao Kowlagi	University of Oulu, Finland
Elizabeth Krupinski	Emory University School of Medicine, USA
Sumit Kumar	Lovely Professional University, India
Rajeev Kumar	Chitkara University, India
Tanesh Kumar	Aalto University, Finland
Maria Kääriäinen	University of Oulu, Finland
Janne Lehtiranta	University of Turku, Finland
Ove Lintvedt	eHealth Research Center of Norway, Norway
Karen Livesay	RMIT University, Australia
Tinja Lääveri	Aalto University, Finland
Madhusanka Liyanage	University College Dublin, Ireland

Ratko Magjarevic	University of Zagreb, Croatia
Terence McSweeney	University of Oulu, Finland
Alexander Meaney	University of Helsinki, Finland
Kristina Mikkonen	University of Oulu, Finland
Pooja Mohanty	Norwegian University of Science and Technology, Norway
Ayan Mondal	Indian Institute of Technology Indore, India
Lorenzo Mucchi	University of Florence, Italy
Teemu Myllylä	University of Oulu, Finland
Mikko Nenonen	University of Oulu, Finland
Khanh Nguyen	University of Oulu, Finland
Huy Hoang Nguyen	University of Oulu, Finland
Anne Oikarinen	University of Oulu, Finland
Aleksandr Ometov	Tampere University, Finland
Diana Moya Osario	Linköping University, Sweden
Karoliina Paalimäki-Paakki	OAMK, Finland
Egor Panfilov	University of Oulu, Finland
Anagha P. Parkar	Haraldsplass Deaconess Hospital Bergen, Norway
Juha Partala	University of Oulu, Finland
Mauricio Perez	Uppsala University, Sweden
Teodora Popordanoska	KU Leuven, Belgium
Pawani Porambage	VTT Technical Research Centre of Finland, Finland
Tarja Pölkki	University of Oulu, Finland
Prabhat Ram	University of Oulu, Finland
Jarmo Reponen	University of Oulu, Finland
Peeter Ross	Technical University of Tallinn, Estonia
Heidi Ruotsalainen	University of Oulu, Finland
Simo Saarakkala	University of Oulu, Finland
Juha Salmitaival	Aalto University, Finland
Tuire Salonurmi	University of Oulu, Finland
Pedro Moreno Sanchez	University of Tampere, Finland
Daniel Pinto dos Santos	University Hospital of Cologne, Germany
Kaija Saranto	University of Eastern Finland, Finland
Kamran Sayrafian	National Institute of Standards and Technology, USA
Aleksi Schrey	University of Turku, Finland
Kamal Kumar Sharma	Ambala College of Engineering and Applied Research, India
Shallu Sharma	Bennett University, India
Charenjeet Singh	Lovely Professional University, India
Daljeet Singh	University of Oulu, Finland

Simone Soderi	IMT School for Advanced Studies, Italy
Ursula Sokolaj	Norwegian University of Science and Technology, Norway
Putra Sumari	Universiti Sains Malaysia, Malaysia
Mariella Särjestöniemi	University of Oulu, Finland
Atthapongse Taparugsanagorn	Asian Institute of Technology, Thailand
Jani Tikkanen	OYS - Oulun Yliopistollinen Sairaala, Finland
Aleksei Tiulpin	University of Oulu, Finland
Paulus Torkki	University of Helsinki, Finland
Johanna Uusimaa	University of Oulu, Finland
Gillian Vesty	RMIT University, Australia
Morten Villumsen	University of Aalborg, Denmark
Sidsel Villumsen	University of Aalborg, Denmark
Alpo Värri	University of Tampere, Finland
Fan Wang	University of Oulu, Finland
Handy Wicaksono	Petra Christian University, Indonesia
Lotta Ylinen	Tampere University, Finland
Cheah Yu-N	Universiti Sains Malaysia, Malaysia
Nasriah Zakaria	Al Maarefa University, Saudi Arabia

International Advisory Committee

Najeeb Al-Shorbaji	Middle East and North Africa Association of Health Informatics, Jordan
Slawomir Ambroziak	Gdansk University of Technology, Poland
Robin Augustine	Uppsala University, Sweden
Rimma Axelsson	Karolinska Institutet, Sweden
Paolo Bifulco	University “Federico II” of Naples, Italy
Matthew Blaschko	KU Leuven, Belgium
Bernd Blobel	University of Regensburg, Germany
Akshay Chaudhari	Stanford University, USA
Luis M. Correia	University of Lisbon IST, Portugal
Flavio Esposito	Saint Louis University, USA
Vahid Farrahi	TU Dortmund University, Germany
Michael Fuchsjäger	Medical University Graz, Austria
Hassan Ghazal	Moroccan Society for Telemedicine & eHealth (MSfTeH), Morocco
Michele Y. Griffith	International Society for Telemedicine and eHealth, USA
Casandra Grundstrom	NTNU, Norway
Ingfrid S. Haldorsen	Haukeland University Hospital, Norway

Manami Hori	Tokai University Europe, Japan
Vesa Jormanainen	Ministry of Social Affairs and Health, Finland
Dipak Kalra	European Institute for Innovation through Health Data, Belgium
Hiroshi Kondoh	Tottori University, Japan
Ilkka Korhonen	GE HealthCare Finland, Finland
Elisabeth Krupinski	Emory University School of Medicine, USA
S. Yunkap Kwankam	International Society for Telemedicine and eHealth, Switzerland
Claudia Lindner	University of Manchester, UK
Anthony Maeder	Flinders University, Australia
Janne Martikainen	University of Eastern Finland, Finland
Lorenzo Mucchi	University of Florence, Italy
Vasiliki Mylonopoulou	University of Gothenburg, Sweden
Miika Nieminen	University of Oulu, Finland
Jérôme Noally	Universitat Pompeu Fabra, Spain
Kaija Saranto	University of Eastern Finland, Finland
Kamran Sayrafian	National Institute of Standards and Technology, USA
Päivi Sillanaukee	Ministry of Social Affairs and Health, Finland
Stein Olav Skrøvseth	Norwegian Centre for E-health Research, Norway
Anthony Smith	University of Queensland, Australia
Rosanna Tarricone	Bocconi University, Italy
Paulus Torkki	University of Helsinki, Finland
Yoshito Tsushima	Gunma University, Japan
Peter van Ooijen	University Medical Center Groningen, Netherlands
Gillian Vesty	RMIT University, Australia
Johanna Viitanen	Aalto University, Finland

Contents – Part II

Wireless Technologies and Medical Devices

Abstracts

Contents – Part I

Abstracts

Clinical Decision Support and Medical AI 1

Using Machine Learning Methods to Predict the Lactate Trend of Sepsis Patients in the ICU

Mustafa Kemal Arslantas[1], Tunc Asuroglu[2(✉)], Reyhan Arslantas[3], Emin Pashazade[4], Pelin Corman Dincer[5], Gulbin Tore Altun[4], and Alper Kararmaz[6]

[1] Department of Anesthesiology and Reanimation, Faculty of Medicine, Demiroglu Bilim University, Istanbul, Turkey
[2] VTT Technical Research Centre of Finland, Tampere, Finland
tunc.asuroglu@vtt.fi
[3] Clinic of Anesthesiology and Reanimation, Taksim Training and Research Hospital, Istanbul, Turkey
[4] Clinic of Anesthesiology and Reanimation, Kadıkoy Florence Nightingale Hospital, Istanbul, Turkey
[5] Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, Marmara University, Istanbul, Turkey
[6] Department of Anesthesiology and Reanimation, School of Medicine, Marmara University, Istanbul, Turkey

Abstract. Serum lactate levels are considered a biomarker of tissue hypoxia. In sepsis or septic shock patients, as suggested by The Surviving Sepsis Campaign, early lactate clearance-directed therapy is associated with decreased mortality; thus, serum lactate levels should be assessed. Monitoring a patient's vital parameters and repetitive blood analysis may have deleterious effects on the patient and also bring an economic burden. Machine learning and trend analysis are gaining importance to overcome these issues. In this context, we aimed to investigate if a machine learning approach can predict lactate trends from non-invasive parameters of patients with sepsis. This retrospective study analyzed adult sepsis patients in the Medical Information Mart for Intensive Care IV (MIMIC-IV) dataset. Inclusion criteria were two or more lactate tests within 6 h of diagnosis, an ICU stay of at least 24 h, and a change of $\geq$1 mmol/liter in lactate level. Naïve Bayes, J48 Decision Tree, Logistic Regression, Random Forest, and Logistic Model Tree (LMT) classifiers were evaluated for lactate trend prediction. LMT algorithm outperformed other classifiers (AUC = 0.803; AUPRC = 0.921). J48 decision tree performed worse than the other methods when predicting constant trend. LMT algorithm with four features (heart rate, oxygen saturation, initial lactate, and time interval variables) achieved 0.80 in terms of AUC (AUPRC = 0.921). We can say that machine learning models that employ logistic regression architectures, i.e., LMT algorithm achieved good results in lactate trend prediction tasks, and it can be effectively used to assess the state of the patient, whether it is stable or improving.

Keywords: Sepsis · Serum Lactate Value · Machine Learning · Intensive Care Unit

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 3–16, 2024.
https://doi.org/10.1007/978-3-031-59091-7_1

1 Introduction

Serum lactate level is traditionally considered a biomarker of tissue hypoxia and is often accompanied by sepsis [1]. Measuring and monitoring blood lactate concentration in sepsis and septic shock can reflect the severity of the illness and the response to therapeutic interventions [2–4]. It has been shown that the decrease in blood lactate values measured in the first hours of admission to the intensive care unit (ICU) over time is associated with better survival [5]. Persistently elevated or increasing lactate levels, indicating inadequate blood flow to organs and tissues (hypoperfusion), are associated with a higher risk of complications and death [6].

For adults with sepsis or septic shock, international guidelines suggest using serum lactate levels to guide resuscitation. This approach helps ensure patients with high initial lactate levels receive targeted treatment aimed at lowering lactate levels [7]. Recently, some randomized control trials demonstrated that early lactate clearance-directed therapy is associated with decreased mortality as compared to the usual care [8]. Because the lactate level measurement is based on time consuming laboratory analysis, technologies that can predict lactate trends quickly, accurately, and noninvasively can be of significant help to clinicians. Despite extensive efforts over the years, there are currently no commercially available intravenous (IV) chemical sensors (i.e., in the bloodstream) for continuous real-time monitoring of lactate levels in ICU patients [9]. Frequent blood draws for serum lactate testing expose patients to risks like infection from venipuncture or central line use, and potential anemia from repeated sampling [10, 11]. A non-invasive method could predict lactate trend of patients allowing clinicians to focus confirmatory testing on patients likely to experience deterioration. In addition, it may avoid unnecessary blood sampling and repetitive lactate measurements. Machine learning algorithms may be helpful to clinicians in this regard [12].

We performed this retrospective study with the hypothesis that a machine learning approach can predict lactate trends from non-invasive clinical variables of patients with sepsis.

2 Methods

2.1 Data Sources

MIMIC-IV is a database containing de-identified health data from over 60,000 ICU patients at Beth Israel Deaconess Medical Center (BIDMC). This database, maintained by MIT's Laboratory for Computational Physiology, is a valuable resource for medical research [13]. We obtained permission to use the anonymized MIMIC-IV dataset and followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [14] for reporting our findings. While STROBE focuses on observational studies, we additionally considered the recommendations offered by Stevens et al. [15] when preparing our manuscript, specifically for reporting machine learning analyses in clinical research. This research recommends statistical methods for machine learning analysis in clinical research, and machine learning analysis workflow is overviewed. Also, several key reporting elements according to the study designs are reported.

The study received ethical approval from both the institutions involved (MIT and BIDMC) and waived the need for individual patient consent because it utilized completely anonymized data already publicly available. Our research adhered to all relevant data privacy guidelines and regulations.

2.2 Study Design

Our retrospective study examined a subgroup of adult sepsis patients from the MIMIC-IV dataset. Sepsis was defined using the Sepsis-3 criteria: suspected infection and an acute increase in the SOFA score of at least 2 [16]. The SOFA score, reflecting organ dysfunction, was calculated using hourly clinical and laboratory data from the first day of each patient's ICU stay. The sepsis criteria were satisfied at the earliest time at which a patient had SOFA $\geq$ 2 and suspicion of infection (time of suspected infection: the culture time [if before antibiotic]; or the antibiotic time [if before culture]). According to the diagnostic criteria, we enrolled adult patients (age $\geq$ 18 years) with at least two serum lactate measurements recorded (within 12 h, starting 6 h before the initial sepsis diagnosis) and with an ICU stay $\geq$24 h.

2.3 Definition of Outcomes

We first needed to define trends to accomplish this above-mentioned lactate trend analysis. Therefore, three trend states were constructed according to value change in blood lactate. For the 12-h observation period, 1mmol per liter and above change was considered a trend indicator. We calculated the difference between two lactate values with a maximum interval of 6 h. According to this setup, all samples in the data cohort had been labeled as increase, decrease, or constant. Trend definition can be seen in Fig. 1.

Lactate value	Trend
• Last lactate $\geq$ 2 mmol/L and increase $\geq$ 1 mmol/L	Increase trend
• Initial lactate and last lactate values < 2 mmol/L • Initial lactate $\geq$ 2 mmol/L and change < 1 mmol/L	Constant trend
• Initial lactate $\geq$ 2 mmol/L and decrease $\geq$ 1 mmol/L	Decrease trend

Fig. 1. Trend definition of lactate values.

2.4 Variable Selection

According to the clinical literature, we identified nine variables that are most relevant in lactate trend analysis. These variables are age, initial lactate value, last lactate value,

and the time interval between two lactate measurements, the averages of hemodynamic and respiratory monitoring parameters measured in this time interval (heart rate, systolic blood pressure, diastolic blood pressure, mean blood pressure, oxygen saturation, and PaO_2/FiO_2 ratio) (Table 1).

These variable selections were used to reduce laboratory dependence on lactate trend analysis and therefore in a minimal-invasive manner. Preprocessing is a vital step to achieving robust machine learning models. These processes help reduce noise, remove redundant data, generate consistent data, and thus increase the performance of prediction models. We applied various preprocessing steps to the data cohort to improve data quality. Outliers in the dataset were removed to obtain consistency between data points. To make the range intervals more coherent unity-based normalization was applied. All ranges were transformed into 0 and 1. We received 18653 data samples after these preprocessing steps.

Feature selection strategies on clinical data provide the correct parameters to analyze a certain disease, treatment cost reduction and reduce computational burden [17]. To achieve these goals, we do a further investigation on variable space. We used the Correlation-based feature selection (CFS) algorithm as a feature selector. CFS algorithm acquires important and pertinent features using inner characteristics of data instead of machine learning approaches [18]. In many cases, some features have a high correlation with others. These features with high correlation characteristics produce redundant data and thus reduce the performance of prediction models. CFS algorithm evaluates the correlations between other features and discards features with high correlation [18]. According to the CFS algorithm, we identified four variables with less correlation than the others and can be used to predict lactate trends in sepsis patients. These are heart rate, oxygen saturation, initial lactate value, and time interval. The overall ranking of features can be seen in Table 2. In this table, feature ranks were identified according to their average merit value; a higher average merit value represents a lower correlation and a higher rank among feature sets [19].

2.5 Proposed Machine Learning Framework

Our proposed machine learning-based framework uses a clinical and demographical types of data and feeds these data to a classifier to oversee lactate trend in ICU settings. We utilized a traditional model for a supervised classification problem consisting of training and a test/evaluation phase.

First, training data consisting of annotated data samples are acquired from the MIMIC-IV dataset. Afterward, they go through a data preprocessing stage to increase data quality for the classification model. Every sample in training data has a lactate trend label (Increase/Constant/Decrease). These samples are trained with a classifier to construct a machine-learning model. A preprocessed test sample is fed to the classifier for the test stage, and the classifier predicts its lactate trend label. In conclusion, classification performance is reported in the evaluation phase (Fig. 2).

Table 1. Patient characteristics (N = 18653).

Variable	Lactate Trend			P-value[a]
	Decrease N = 3313 (18%)[a]	Increase N = 1328 (7.1%)[a]	Constant N = 14012 (75%)[a]	
Age (*y*)	65 (15)	65 (15)	65 (15)	0.015
Heart Rate (*bpm*)	89 (18)	95 (20)	89 (19)	<0.001
Systolic Blood Pressure (*mmHg*)	111 (14)	106 (15)	110 (15)	<0.001
Mean Blood Pressure (*mmHg*)	89 (18)	95 (20)	89 (19)	<0.001
Diastolic Blood Pressure (*mmHg*)	76 (10)	73 (11)	75 (10)	<0.001
Blood Oxygen Saturation Level (*%, SpO_2*)	98.35 (1.94)	97.75 (2.26)	97.89 (2.17)	<0.001
PaO_2/FiO_2 Ratio (*P/F Ratio*)	240 (118)	214 (119)	228 (114)	<0.001
Initial Lactate Value (*mmol/L*)	4.68 (1.66)	3.74 (1.97)	2.67 (1.57)	<0.001
Time Interval (*min*)	197 (81)	188 (82)	186 (84)	<0.001

[a]Mean (SD)
[b]Kruskal-Wallis rank sum test

Table 2. Ranking of features according to correlation analysis.

Rank	Average Merit Value	Feature Name
1	0.395	Initial Lactate Value
2	0.061	Oxygen Saturation
3	0.041	Time Interval
4	0.040	Heart Rate
5	0.030	Systolic Blood Pressure
6	0.025	PaO_2/FiO_2
7	0.024	Age
8	0.016	Mean Blood Pressure
9	0.011	Diastolic Blood Pressure

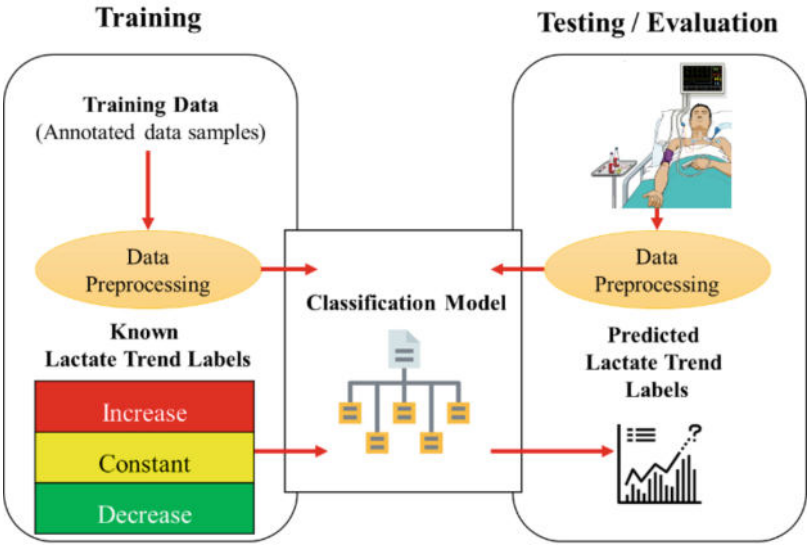


Fig. 2. Proposed machine learning framework for lactate trend prediction.

2.6 Selected Classifiers for Proposed Framework

We evaluated various classifiers on the MIMIC-IV dataset to predict lactate trends in sepsis patients. These classifiers are Naïve Bayes (NB), J48 Decision Tree, Logistic Regression (LR), Random Forest (RF), and Logistic Model Tree (LMT). Naïve Bayes is a traditional and simple machine learning approach that contemplates dataset attributes as an independent [17]. The outputs are considered class probabilities. Naïve Bayes acts on the Bayes theorem, which is the probability of any event occurring, given the probability of another event just occurring. The class with the highest probability is selected as the outcome. It became immensely popular in the machine learning area due to its advantages. These advantages are managing the overfitting problem very well and parallelizing the classification process [20].

J48 decision tree algorithm is an updated version of the popular decision tree algorithm ID3 [21]. It can be used in both numerical and categorical data. J48 aims to find a specific attribute that fully partitions the training data. This attribute has the highest in-formation gain value in the dataset [22]. By evaluating the probable values of this attribute, a branch pruning process starts, and J48 defines target values. In the meantime, J48 searches other high information gain attributes. This process continues until an explicit decision is made on the combination of attributes that gives a certain rule for determining the target value. At the end of the algorithm, all features are evaluated; therefore, all samples have a target value accordingly [22]. J48 became a popular machine learning tool in many areas due to its easy implementable and robust nature [21, 23, 24].

Random Forest (RF) belongs to the family of decision trees that employ a supervised ensemble learning strategy [25]. It gained popularity among classification and regression problem domains due to its robustness against overfitting and low computational load [25–27]. RF builds many decision trees that are based on the selection of a random subset

of variables that are called bootstrap samples. Other decision tree learners aim to find the best variable available, whereas RF uses random variables. The primary motivation for this approach is to reduce the correlation between these candidate random trees. This randomness approach is essential when making decisions because if highly correlated variables are available, it affects the prediction phase and leads to poor prediction performance. All predictions from random trees are combined to achieve the maximum result [26].

The logistic regression algorithm is mainly used for tackling classification problems and modeling class probabilities [28]. It aims to fit the data to a logistic curve to predict the occurrence probability of events [29]. It can handle nonlinear dataset effects.

LMT algorithm is a hybrid decision tree approach that utilizes logistic regression and decision tree learning [30]. Leaves of the tree have piecewise linear regression models constructed by logistic regression functions. To build these logistic regression functions LogitBoost algorithm is used [31]. Decision tree classifier algorithms do prune of the decision tree. Splitting of the decision tree is implemented via logistic variant information gain. The algorithm has many positive aspects; it can map linear relationships, overfitting can be easily avoided, and it is easy to implement. Because of its numerous advantages, in recent years, it has been used in many different research areas [30–32].

2.7 Evaluation Criteria

Experiments on predicting the lactate trend are evaluated with ten-fold cross-validation (CV). In this CV approach, the dataset is split into ten parts that have an equal number of samples. One part is selected for testing, and the rest are used for the training phase. The cross-validation process stops if all parts are used for the testing phase. Evaluation setting for three class classification is one versus all approach.

Area Under Curve (AUC) score and Area Under Precision-Recall Curve (AUPRC) metrics are used to assess the classification performance of machine learning algorithms. The AUC score is calculated by drawing a True Positive Rate (Sensitivity) and False Positive Rate (1-Specificity) curve. Then after drawing this curve area under the curve is calculated to assess the classification model. AUC score range is between 0 and 1. An AUC score of 1 means that the classification model can distinguish all samples. So, values that are close to 1 indicate better prediction performance. Compared to AUC, AUPRC prioritizes its ability to identify positive samples. In addition, AUPRC is preferred over the AUC as it is more sensitive and less prone to exaggerate model performance for unbalanced datasets.

3 Results

We evaluated RF, NB, J48, LR, and LMT classifiers on the lactate trend prediction task. We conducted our experiments based on three scenarios; the sepsis patient‘s lactate value has an increasing trend, sepsis patient has a steady lactate value trend, and sepsis patient‘s lactate value has a decreasing trend.

Table 3 shows classification results for the lactate trend increase scenario. As can be seen from Table 3 LMT and LR algorithms outperformed other classifiers and achieved

0.647 values in terms of AUC; but in terms of AUPRC, RF performed better. NB comes second. J48 decision tree performed worse when predicting lactate trend increase. LMT algorithm with four features (heart rate, oxygen saturation, lactate value before sepsis diagnosis, and time interval) achieved 0.630 in terms of AUC (AUPRC, 0.113).

Table 3. Classification results for the increasing trend of lactate.

Classification Model	AUC [95% CI]	AUPRC [95% CI]
RF	0.628 [0.614–0.642]	**0.143 [0.135–0.151]**
NB	0.637 [0.623–0.651]	0.127 [0.120–0.134]
J48 Decision Tree	0.555 [0.54–0.57]	0.102 [0.096–0.108]
LR	**0.647 [0.633–0.661]**	0.128 [0.121–0.135]
LMT	**0.647 [0.633–0.661]**	0.130 [0.123–0.137]
LMT (with 4 features)	0.630 [0.616–0.644]	0.113 [0.107–0.119]

Table 4 shows classification results for the constant lactate trend scenario. As can be seen from Table 4 LMT algorithm outperformed other classifiers (AUC, 0.803; AUPRC, 0.921). RF comes second, and LR comes third. J48 decision tree performed worse when predicting constant lactate trend. LMT algorithm with four features achieved 0.921 in terms of AUPRC.

Table 4. Classification results for the constant trend of lactate.

Classification Model	AUC [95% CI]	AUPRC [95% CI]
RF	0.794 [0.789–0.799]	0.914 [0.910–0.918]
NB	0.779 [0.774–0.784]	0.911 [0.907–0.915]
J48 Decision Tree	0.726 [0.721–0.731]	0.847 [0.841–0.853]
LR	0.792 [0.787–0.797]	0.915 [0.911–0.919]
LMT	**0.803 [0.798–0.808]**	**0.921 [0.917–0.925]**
LMT (with 4 features)	0.80 [0.795–0.805]	**0.921 [0.917–0.925]**

Table 5 shows classification results for the lactate trend decrease scenario. As can be seen from Table 5 LMT algorithm outperformed other classifiers (AUC, 0.847; AUPRC, 0.502). RF comes second, and LR comes third in terms of AUC. J48 decision tree performed worse when predicting constant lactate trend. LMT algorithm with four features achieved 0.844 in terms of AUC and 0.493 in terms of AUPRC. According to experimental results, we can say that machine learning models that employ logistic regression architectures overall achieved good results in lactate trend prediction tasks. Also, the LMT algorithm with just four variables achieved a noteworthy prediction performance compared with the LMT algorithm that uses all of the variables. Especially in constant

lactate and decreased lactate trends, LMT with four features achieved similar results to the LMT algorithm. We can say that the LMT algorithm with heart rate, oxygen saturation, lactate value before sepsis diagnosis, and time interval variables can be effectively used to assess the patient's state, whether it is stable or improving.

Table 5. Classification results for decreasing trend of lactate.

Classification Model	AUC [95% CI]	AUPRC [95% CI]
RF	0.842 [0.836–0.848]	0.491 [0.48–0.502]
NB	0.822 [0.816–0.828]	0.452 [0.441–0.463]
J48 Decision Tree	0.751 [0.743–0.759]	0.401 [0.391–0.411]
LR	0.826 [0.82–0.832]	0.473 [0.462–0.484]
LMT	**0.847 [0.841–0.853]**	**0.502 [0.491–0.513]**
LMT (with 4 features)	0.844 [0.838–0.85]	0.493 [0.482–0.504]

4 Discussion

The LMT models, one of the machine learning approaches, were the most accurate in predicting serum lactate trends from non-invasive clinical variables of patients with sepsis. In this method the AUC of increasing, constant, and decreasing lactate values were 0.647 [95% CI] [0.633–0.661], 0.803 [95% CI] [0.798–0.808], and 0.847 [95% CI] [0.841–0.853], respectively.

We observed different rankings of the importance of the variables for predicting lactate trends. For example, initial serum lactate measurement was a significant predictor of change in serum lactate values, followed by oxygen saturation, the time interval between lactate measurements, heart rate, SBP, P/F ratio, age, MBP, and DBP.

Multiple studies have been conducted on reducing the fatality rate associated with sepsis. Quickly identifying patients likely to experience severe sepsis or septic shock is essential for effective treatment. While lab tests (such as procalcitonin, C-reactive protein, and lactate) help predict sepsis, they can be time-consuming. This delay in diagnosis and treatment initiation highlights the need for faster prediction methods [33–37].

Signs of poor blood flow to tissues caused by sepsis can be both general and specific. General signs include low blood pressure, fast heart rate, decreased urine output, slow capillary refill, confusion, high blood lactate levels, and low blood oxygen saturation. Specific signs vary depending on the affected tissue. Notably, changes in vital signs, like heart rate, blood pressure, breathing rate, oxygen saturation, and body temperature, can appear several hours before serious complications or worsening of the patient's condition, providing valuable time for early intervention [38]. The Systemic Inflammatory Response Syndrome (SIRS) and qSOFA score (also known as the rapid Sepsis-Related Organ Failure Assessment) are primarily based on identifying changes

in vital signs. These criteria remain an essential clinical tool for the host's systemic response to inflammation, despite the discovery of several biomarkers [39].

Studies suggest that analyzing trends in intermittent vital signs could lead to earlier detection of clinical deterioration in patients, potentially improving outcomes in both general wards and emergency departments [40]. According to a study by Barfod et al. [41], abnormal vital signs (SpO_2, RR, BP, HR, GCS), especially abnormal RR, SpO_2, and GCS, are strong predictors for intensive care unit admission from the emergency department and in-hospital mortality. Today, some researchers are developing sepsis diagnosis and mortality prediction models by analyzing changes in vital signs using machine learning techniques. A machine learning-based sepsis prediction algorithm (InSight) developed by Mao Q et al. [42] provides high sensitivity and specificity for detecting and predicting sepsis, severe sepsis, and septic shock using only six common vital data acquired in the emergency department, general ward, and ICU.

In clinical conditions, the circulatory disorder may be characterized by abnormal hemodynamic parameters such as hypotension and tachycardia, abnormal tissue organ perfusion findings such as decreased urine output and changes in consciousness, and abnormal metabolic parameters such as increased lactate and metabolic acidosis [43].

Hyperlactatemia is common in patients with sepsis, which is a marker of disease se-verity and a strong predictor of mortality. Sepsis-associated hyperlactatemia may reflect the degree of activation of the stress response (and epinephrine release) [1]. In daily clinical practice, it is accepted that the increase in lactate levels over time primarily reflects a change due to increased production, decreased utilization, or both. As hyperlactatemia is often associated with poor circulation, we usually see a decrease in lactate levels in the improved circulation state, and we hypothesize (but cannot prove) decreased production [44]. However, since clearance is significantly reduced in stable septic patient shock states, continued hyperlactatemia may reflect decreased clearance rather than increased lactate production [5]. Lactate levels can help doctors predict a patient's risk of death, allowing them to determine the appropriate level of care. High lactate level indicates an increased risk of mortality, and it can help identify patients who need additional investigation and monitoring [45].

In our study, all models underperformed in predicting lactate increase, a more helpful indicator for disease severity. All the selected cases consisted of patients diagnosed with sepsis who already had high lactate levels. Therefore, we evaluated whether the upward lactate trend could predict further increase from the currently in-creasing state rather than the baseline level. This may have affected the predictive power of the model. In addition, the low performance can be attributed to the uneven distribution of the number of samples in each group in the cohort. The number of samples with an in-creasing trend is deficient compared to the others (Increase: 1328, Constant: 14012, Decrease: 3313 samples are available). Because of this imbalance, the model's predictive power may be affected. We can conduct future work to improve this situation. Increasing the number of patients with increased lactate in the dataset may be recommended, or methods such as synthetic data generation may be used at the training stage [46]. Though our current model offers valuable insights, we believe its performance can be significantly enhanced with access to a larger dataset. This would allow us to develop a model with improved discriminatory power, ultimately providing clinicians with more precise guidance on when to utilize

serum lactate testing. The reduction in the lactate trend and the prediction of stability can also reduce unnecessary testing. Also, additional parameters (such as the focus on infection and the medical treatments administered) will further improve the model's performance and facilitate the prediction of serum lactate trends.

Empirical results reveal that machine learning approaches that utilize logistic regression functions achieved higher AUC values than others. Due to its robust structure to overfitting, the LMT algorithm achieved high AUC values even with just four features. Experimental results also prove that the LMT algorithm can be combined with easily acquirable and routinely collected parameters to predict the lactate trend of sepsis patients. The LMT model has high computational complexity due to its hybrid logistic tree structure. If the model is trained with high dimensional data, it can lead to high CPU and memory consumption [47]. To overcome this issue, our approach only uses high importance parameters in the lactate trend prediction task. With this proposed approach, a quick and accurate solution based on easily acquirable wearable parameters can be implemented in the ICU setting to assess the trend of lactate value.

5 Conclusion

Lactate metabolism is affected by many factors; thus, predicting its level using ma-chine learning is not easy. Treatment can be tailored according to predicting the lactate trend rather than predicting one single value. Our study suggests that lactate change can be predicted with a suboptimal performance by using machine learning models that use patients' hemodynamic and respiratory parameters. Further clinical studies will help determine the full potential of this tool within a clinical context.

By adding more lactate-related parameters to the dataset, the performance of deep learning methods, a branch of machine learning, can be examined. Deep learning structures that have a reliable performance, such as LSTM (Long Short-Term Memory) and CNN (Convolutional neural network) can be combined with LMT to form a hybrid system and be used in predicting lactate trends. Lastly, synthetic samples can also be used in the models' training phase to increase machine learning models' prediction capability.

The need to find a better way to predict patients' survival is still ongoing. Machine learning is gaining more importance and attention as the clinical outcomes are well correlated with the systems' predictions. Clinicians prefer noninvasive and less costly approaches with accurate estimations of the patients. Predicting the lactate trend, in other words, the state of sepsis patient, whether it is stable or improving, in the ICU by LMT algorithm, which uses heart rate, oxygen saturation, lactate value before sepsis diagnosis, and time interval variables can be done effectively.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Garcia-Alvarez, M., Marik, P., Bellomo, R.: Sepsis-associated hyperlactatemia. Crit. Care **18**(5), 503 (2014). https://doi.org/10.1186/s13054-014-0503-3
2. Ryoo, S.M., et al.: Lactate level versus lactate clearance for predicting mortality in patients with septic shock defined by sepsis-3. Crit. Care Med. **46**(6), e489–e495 (2018). https://doi.org/10.1097/CCM.0000000000003030
3. Masyuk, M., et al.: Prognostic relevance of serum lactate kinetics in critically ill patients. Intensive Care Med. **45**(1), 55–61 (2019). https://doi.org/10.1007/s00134-018-5475-3
4. Liu, Z., et al.: Prognostic accuracy of the serum lactate level, the SOFA score and the qSOFA score for mortality among adults with Sepsis. Scand. J. Trauma Resusc. Emerg. Med. **27**(1), 51 (2019). https://doi.org/10.1186/s13049-019-0609-3
5. Nguyen, H.B., et al.: Early lactate clearance is associated with improved outcome in severe sepsis and septic shock. Crit. Care Med. **32**(8), 1637–1642 (2004). https://doi.org/10.1097/01.ccm.0000132904.35713.a7
6. Claridge, J.A., Crabtree, T.D., Pelletier, S.J., Butler, K., Sawyer, R.G., Young, J.S.: Persistent occult hypoperfusion is associated with a significant increase in infection rate and mortality in major trauma patients. J. Trauma Acute Care Surg. **48**(1), 8 (2000). https://doi.org/10.1097/00005373-200001000-00003
7. Evans, L., et al.: Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Crit. Care Med. **49**(11), e1063–e1143 (2021). https://doi.org/10.1097/CCM.0000000000005337
8. Pan, J., Peng, M., Liao, C., Hu, X., Wang, A., Li, X.: Relative efficacy and safety of early lactate clearance-guided therapy resuscitation in patients with sepsis: a meta-analysis. Medicine (Baltimore) **98**(8), e14453 (2019). https://doi.org/10.1097/MD.0000000000014453
9. Wolf, A., et al.: Evaluation of continuous lactate monitoring systems within a heparinized in vivo porcine model intravenously and subcutaneously. Biosensors (Basel) **8**(4), 122 (2018). https://doi.org/10.3390/bios8040122
10. Timsit, J.F., Ruppe, E., Barbier, F., Tabah, A., Bassetti, M.: Bloodstream infections in critically ill patients: an expert statement. Intensive Care Med. **46**(2), 266–284 (2020). https://doi.org/10.1007/s00134-020-05950-6
11. Jakacka, N., Snarski, E., Mekuria, S.: Prevention of iatrogenic anemia in critical and neonatal care. Adv. Clin. Exp. Med. **25**(1), 191–197 (2016). https://doi.org/10.17219/acem/32065
12. Mamandipoor, B., Yeung, W., Agha-Mir-Salim, L., Stone, D.J., Osmani, V., Celi, L.A.: Prediction of blood lactate values in critically ill patients: a retrospective multi-center cohort study. J. Clin. Monit. Comput. **36**, 1087–1097 (2021). https://doi.org/10.1007/s10877-021-00739-4
13. Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L.A., Mark, R.: MIMIC-IV (version 1.0). (2021). PhysioNet. https://doi.org/10.13026/s6n6-xd98
14. von Elm, E., et al.: The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. J. Clin. Epidemiol. **61**(4), 344–349 (2008). https://doi.org/10.1016/j.jclinepi.2007.11.008
15. Stevens, L.M., Mortazavi, B.J., Deo, R.C., Curtis, L., Kao, D.P.: Recommendations for reporting machine learning analyses in clinical research. Circ. Cardiovasc. Qual. Outcomes **13**(10), e006556 (2020). https://doi.org/10.1161/CIRCOUTCOMES.120.006556
16. Singer, M., et al.: The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA **315**(8), 801–810 (2016). https://doi.org/10.1001/jama.2016.0287
17. Chaudhuri, A.K., Sinha, D., Banerjee, D.K., Das, A.: A novel enhanced decision tree model for detecting chronic kidney disease. Netw. Model Anal. Health Inform. Bioinform. **10**(1), 29 (2021). https://doi.org/10.1007/s13721-021-00302-w

18. Sharma, A., Mishra, P.K.: Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis. Int. J. Inf. Technol. **14**(4), 1949–1960 (2021). https://doi.org/10.1007/s41870-021-00671-5
19. Ebrahimpour, M.K., Eftekhari, M.: Ensemble of feature selection methods: a hesitant fuzzy sets approach. Appl. Soft Comput. **50**, 300–312 (2017). https://doi.org/10.1016/j.asoc.2016.11.021
20. Bužić, D., Dobša, J.: Lyrics classification using Naive Bayes. In: 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1011–1015. IEEE, Opatija (2018). https://doi.org/10.23919/MIPRO.2018.8400185
21. Bhat, P., Malaganve, P.: Effect of J48 and LMT algorithms to classify movies in the web—a comparative approach. In: Saini, H.S., Sayal, R., Govardhan, A., Buyya, R. (eds.) Innovations in Computer Science and Engineering. Lecture Notes in Networks and Systems, vol. 171, pp. 1–13. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4543-0_58
22. Aşuroğlu, T., Açıcı, K., Berke Erdaş, Ç., Kılınç Toprak, M., Erdem, H., Oğul, H.: Parkinson's disease monitoring from gait analysis via foot-worn sensors. Biocybernetics Biomed. Eng. **38**(3), 760–772 (2018). https://doi.org/10.1016/j.bbe.2018.06.002
23. Ilyas, H., et al.: Chronic kidney disease diagnosis using decision tree algorithms. BMC Nephrol. **22**(1), 273 (2021). https://doi.org/10.1186/s12882-021-02474-z
24. Rahmayanti, N., Pradani, H., Pahlawan, M., Vinarti, R.: Comparison of machine learning algorithms to classify fetal health using cardiotocogram data. Procedia Comput. Sci. **197**, 162–171 (2022). https://doi.org/10.1016/j.procs.2021.12.130
25. Simsekler, M.C.E., Alhashmi, N.H., Azar, E., King, N., Luqman, R., Al Mulla, A.: Exploring drivers of patient satisfaction using a random forest algorithm. BMC Med. Inform. Decis. Mak. **21**(1), 157 (2021). https://doi.org/10.1186/s12911-021-01519-5
26. Açıcı, K., Erdaş, Ç.B., Aşuroğlu, T., Toprak, M.K., Erdem, H., Oğul, H.: A random forest method to detect Parkinson's disease via gait analysis. In: Boracchi, G., Iliadis, L., Jayne, C., Likas, A. (eds.) Engineering Applications of Neural Networks. EANN 2017. Communications in Computer and Information Science, vol. 744, pp. 1–13. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-65172-9_51
27. Aşuroğlu, T., Oğul, H.: A deep learning approach for sepsis monitoring via severity score estimation. Comput. Methods Prog. Biomed. **198**, 105816 (2021). https://doi.org/10.1016/j.cmpb.2020.105816
28. Choudhury, A., Gupta, D.: A survey on medical diagnosis of diabetes using machine learning techniques. In: Kalita, J., Balas, V., Borah, S., Pradhan, R. (eds.) Recent Developments in Machine Learning and Data Analytics. Advances in Intelligent Systems and Computing, vol. 740, pp. 1–13. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-1280-9_6
29. Kim, A., Song, Y., Kim, M., Lee, K., Cheon, J.H.: Logistic regression model training based on the approximate homomorphic encryption. BMC Med. Genom. **11**(4), 83 (2018). https://doi.org/10.1186/s12920-018-0401-7
30. Nazari, E., Aghemiri, M., Avan, A., Mehrabian, A., Tabesh, H.: Machine learning approaches for classification of colorectal cancer with and without feature selection method on microarray data. Gene Rep. **25**, 101419 (2021). https://doi.org/10.1016/j.genrep.2021.101419
31. Jha, S.K., Pan, Z., Elahi, E., Patel, N.: A comprehensive search for expert classification methods in disease diagnosis and prediction. Expert. Syst. **36**(1), e12343 (2019). https://doi.org/10.1111/exsy.12343
32. Tien Bui, D., et al.: Shallow landslide prediction using a novel hybrid functional machine learning algorithm. Remote Sens. **11**(8), 931 (2019). https://doi.org/10.3390/rs11080931
33. Opal, S.M., Wittebole, X.: Biomarkers of infection and sepsis. Crit. Care Clin. **36**(1), 11–22 (2020). https://doi.org/10.1016/j.ccc.2019.08.002
34. Herzum, I., Renz, H.: Inflammatory markers in SIRS, sepsis and septic shock. Curr. Med. Chem. **15**(6), 581–587 (2008). https://doi.org/10.2174/092986708783769704

35. Ljungström, L., Pernestig, A.K., Jacobsson, G., Andersson, R., Usener, B., Tilevik, D.: Diagnostic accuracy of procalcitonin, neutrophil-lymphocyte count ratio, C-reactive protein, and lactate in patients with suspected bacterial sepsis. PLoS ONE **12**(7), e0181704 (2017). https://doi.org/10.1371/journal.pone.0181704
36. Arslantas, M.K., Arslantas, R., Dincer, P.C., Altun, G.T., Kararmaz, A.: Prognostic value of the lactate-albumin difference in predicting 30-day mortality in critically ill patients. Shock **56**(1), 150–151 (2021). https://doi.org/10.1097/SHK.0000000000001613
37. Seymour, C.W., et al.: Time to treatment and mortality during mandated emergency care for sepsis. N. Engl. J. Med. **376**(23), 2235–2244 (2017). https://doi.org/10.1056/NEJMoa1703058
38. Zanotti Cavazzoni, S. L., Dellinger, R. P.: Hemodynamic optimization of sepsis-induced1 tissue hypoperfusion. Crit. Care **10**(Suppl 3), S2 (2006). https://doi.org/10.1186/cc4829
39. Gul, F., Arslantas, M.K., Cinel, I., Kumar, A.: Changing definitions of sepsis. Turk. J. Anaesthesiol. Reanim. **45**(3), 129–138 (2017). https://doi.org/10.5152/TJAR.2017.93753
40. Brekke, I.J., Puntervoll, L.H., Pedersen, P.B., Kellett, J., Brabrand, M.: The value of vital sign trends in predicting and monitoring clinical deterioration: a systematic review. PLoS ONE **14**(1), e0210875 (2019). https://doi.org/10.1371/journal.pone.0210875
41. Barfod, C., et al.: Abnormal vital signs are strong predictors for intensive care unit admission and in-hospital mortality in adults triaged in the emergency department - a prospective cohort study. Scand. J. Trauma Resusc. Emerg. Med. **20**(1), 28 (2012). https://doi.org/10.1186/1757-7241-20-28
42. Mao, Q., et al.: Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open **8**(1), e017833 (2018). https://doi.org/10.1136/bmjopen-2017-017833
43. Bakker, J., Postelnicu, R., Mukherjee, V.: Lactate: where are we now? Crit. Care Clin. **36**(1), 115–124 (2020). https://doi.org/10.1016/j.ccc.2019.08.009
44. Vincent, J.L., Bakker, J.: Blood lactate levels in sepsis: in 8 questions. Curr. Opin. Crit. Care **27**(3), 298–302 (2021). https://doi.org/10.1097/MCC.0000000000000824
45. Villar, J., Short, J.H., Lighthall, G.: Lactate predicts both short- and long-term mortality in patients with and without sepsis. Infect. Dis. (Auckl.) **12**, 1178633719862776 (2019). https://doi.org/10.1177/1178633719862776
46. Açıcı, K., Aşuroğlu, T., Erdaş, Ç., Oğul, H.: T4SS effector protein prediction with deep learning. Data **4**(1), 45 (2019). https://doi.org/10.3390/data4010045
47. Lee, S., Jun, C.-H.: Fast incremental learning of logistic model tree using least angle regression. Exp. Syst. Appl. **97**, 137–45 (2018). https://doi.org/10.1016/j.eswa.2017.12.014

Computed Tomography Artefact Detection Using Deep Learning—Towards Automated Quality Assurance

S. I. Inkinen[1], A. O. Kotiaho[2], M. Hanni[3,4], M. T. Nieminen[3,4], and M. A. K. Brix[3,4(✉)]

[1] HUS Diagnostic Center, Radiology, Helsinki University and Helsinki University Hospital, Haartmaninkatu 4, 00290 Helsinki, Finland
[2] Terveystalo Healthcare, Jaakonkatu 3B, 00100 Helsinki, Finland
[3] Department of Diagnostic Radiology, Oulu University Hospital, Kajaanintie 50, 90220 Oulu, Finland
[4] Research Unit of Health Sciences and Technology, University of Oulu, 90220 Oulu, Finland
mikael.brix@oulu.fi

Abstract. Image artefacts in computed tomography (CT) limit the diagnostic quality of the images. The objective of this proof-of-concept study was to apply deep learning (DL) for automated CT artefact classification. Openly available Head CT data from Johns Hopkins University was used. Three common artefacts (patient movement, beam hardening, and ring artefacts (RAs)) and artefact free images were simulated using 2D axial slices. Simulated data were split into a training set ($N_{train} = 1040 \times 4(4160)$), two validation sets ($N_{val1} = 130 \times 4(520)$ and $N_{val2} = 130 \times 4(520)$), and a separate test set ($N_{test} = 201 \times 4(804)$; two individual subjects). VGG-16 model architecture was used as a DL classifier, and the Grad-CAM approach was used to produce attention maps. Model performance was evaluated using accuracy, average precision, area under the receiver operating characteristics (ROC) curve, precision, recall, and F1-score. Sensitivity analysis was performed for two test set slice images in which different RA radiuses (4 pixels to 245) and movement artefacts, i.e., head tilt with rotation angles (0.2° to 3°), were generated. Artefact classification performance was excellent on the test set, as accuracy, average precision, and ROC area under curve over all classes were 0.91, 0.86, and 0.99, respectively. The precision, recall, and F1-scores were over 0.84, 0.71, and 0.85 for all class-wise cases. Sensitivity analysis revealed that the model detected movement at all rotation angles, yet it failed to detect the smallest RAs (4-pixel radius). DL can be used for effective detection of CT artefacts. In future, DL could be applied for automated quality assurance of clinical CT.

Keywords: Computed tomography · Deep learning · Image artefacts · Quality assurance

1 Introduction

Image artefacts encountered in computed tomography (CT) limit the diagnostic quality of the images and may result in a re-scan of the patient. CT artefacts can be caused by the patient, they may be based on CT physics, or caused by malfunctioning hardware [1].

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 17–28, 2024.
https://doi.org/10.1007/978-3-031-59091-7_2

Patient-based artefacts are mostly due to movement during acquisition. These artefacts cause, for example, blurring and distortions in the reconstructed images. They can, however, be mitigated using different motion correction algorithms [2].

The most common physics-based artefact is the beam hardening artefact in which the polychromatic low energy photons of the x-ray beam attenuate most in the patient, increasing the mean energy of the x-ray spectrum. Another common artefact type is a metal artefact, resulting from beam hardening and photon starvation which cause streaking and cupping artefacts in the CT images. These artefacts can be alleviated by increasing x-ray tube peak kilovoltage, stronger beam filtration and by using beam hardening and metal artefact reduction algorithms [3].

Common hardware-based artefacts are ring artefact, tube arcing, and air bubble artefact in the oil coolant of the tube [1, 4]. Ring artefacts (RAs) can be caused, for example, by dead pixels in the x-ray detector or miscalibration. Artefacts arising from a gas bubble in the oil coolant of the tube are difficult to detect [5]. Hardware-based artefacts usually require maintenance service. However, RAs may be corrected using *e.g.* interpolation methods, filtering approaches, or flat-field recalibration [6].

Several automated detection approaches have been proposed for CT image quality assessment [7, 8]. However, these studies do not focus on artefact detection but on technical image quality assessment directly from clinical patient images. For example, Smith et al. 2017 developed a method to automatically estimate the detectability index, noise power spectrum, and modulation transfer function directly from patient CT images [7]. On the machine learning front, using deep learning (DL) convolutional neural networks in image analysis has gained great interest in image processing [9]. For CT image quality assessment in particular, a DL method was recently developed for deformable image registration quality assessment from lung CT images [10]. In another recent study focusing on MRI, a fast and automated DL method for assessing re-scan need in motion-corrupted brain series was developed [11].

Even though sophisticated artefact correction algorithms exist, modern CT scanners may produce image artefacts limiting the diagnostic quality which in the worst case may require a re-scan. Therefore, imaging technologists must carefully review the CT reconstructions after image acquisition. An automated artefact detection tool could optimize this process. On the other hand, artefact detection could be used for monitoring CT artefact prevalence as a performance management tool in hospitals. Despite the recent developments in CT image quality assessment [12–14], to the authors' knowledge, there are only a few studies focusing on CT artefact detection using DL [15, 16]. In the work by Madesta *et al.*, DL was applied for the detection of simulated respiration-related patient motion artefacts using a 3D patch-based approach in radiotherapy 4D CT imaging and for the subsequent correction of the artefacts using DL-based inpainting [15]. In the work by Prakash and Dutta, detector-related artifacts were simulated to projection data, and a deep learning approach was employed to detect streaks, rings, and bands [16]. In this study, a VGG-16 CNN architecture is applied for image artefact detection directly from clinical head CT patient images. Hypothesis is that DL can be utilized in the automated assessment of the clinical image artefacts.

2 Materials and Methods

Data: In this study, the openly available Head CT scan dataset available from Johns Hopkins University Data Archive was used [17]. The dataset consists of 35 subjects' non-contrast head CT scans. The dataset had to be curated as unwanted metal artefacts were present for some of the subjects' dental regions, and those slices were excluded manually from the image stacks. After metal artefact exclusion, the whole dataset consisted of 1501 slice images (1300/201 training phase/test phase).

Three different artefacts were simulated in the dataset using internally developed algorithms: ring artefact, beam hardening artefact, and movement artefacts (Fig. 1). The simulation was performed for each slice image in 2D, and for each slice, all three artefacts were simulated. First, the slice image was segmented into {air, adipose, water, brain, skull} material regions based on HU thresholding, in which the CT slice is segmented to different tissues based on the HU values of the individual pixels {air ≤ -100 HU, adipose ≤ -30 HU, water ≤ 20 HU, brain < 1000 HU, skull ≥ 1000 HU}. The energy-dependent linear attenuation coefficients (μ) for each segmented material (x) were extracted from attenuation table obtained from XCAT virtual phantom software package [18]. The attenuation process was modeled in discrete form as:

$$I_k = \sum_{E=0}^{E_{max}} I_{0,E} \exp\left(- \sum_{j=1}^{N} a_{k,j} \sum_{m=1}^{M} \mu_{E,m} x_{j,m} \right), \tag{1}$$

where, I_k is the transmitted x-ray intensity at sinogram index k, $I_{0,E}$ is the input spectrum, $a_{k,j}$ is the element of the forward projection matrix at row k and column j and $x_{j,m}$ is the thickness of material m^{th} material at pixel j. E_{max} is the maximum peak kilovoltage from the simulated x-ray tube spectrum. The 120 kVp and 70 kVp input spectra were simulated using the SPEKTR toolbox [19] in Matlab (2018b/v9.5.0, Mathworks Inc., Natick, MA, USA) (Fig. 1). The 120 kVp spectrum was applied for artefact free projection data simulation and for all artefacts except for beam hardening. Parallel beam projection geometry with 180 projections in 1-degree intervals was applied in the simulations and Poisson noise was simulated in the projection data. The projection data were flat-field corrected before filtered back projection (FBP) reconstruction with Ram-Lak filtering. Reconstructions were computed using the Astra toolbox (v.1.9.0.dev11) [20, 21]. The simulations and reconstructions were conducted in Python (v. 3.7).

The artefact-free projections and reconstructions were computed as explained above, and the artefacts were generated as follows:

Ring Artefact: RAs were inserted into the artefact-free FBP slice images as an image post-processing method. The number of RAs were chosen to contain either one single RA or several (1 to 20) RAs. The single RA corresponds to the situation where only one detector element is malfunctioning, and several RAs mimics a scenario of detector miscalibration or malfunctioning of multiple detector elements. When a single RA was simulated in the slice image, a circle mask with radius (4 to 245 range) was simulated and a value from uniform distribution (-1000 to 2000) was drawn, and the mask was added to the reconstructed image. For several RAs, diameters from 4 to 245 isocenter distances were simulated, and the value of the RAs in multiple RA case were drawn

from Gaussian distribution with a mean equal to image pixel values (excluding values < 0) and a standard deviation of 100. The RAs' thicknesses randomly varied from one to three pixels.

Beam Hardening Artefact: 70 kVp spectrum was used instead of 120 kVp in the beam hardening simulation, as it showed typical cupping artefact (Fig. 1c)).

Movement artefact: Movement artefact was simulated as head tilt rotation movement. The axial rotation axis was set at the center of the left edge of the slice image (Fig. 1d)). The amount of rotation (R) was drawn from a uniform distribution [−20, +20 degrees], and 180 evenly spaced intervals were generated from 0.1 to R degrees. Subsequently, the reconstructed image was rotated, and forward projection was generated to obtain the sinogram of the moved sample. One projection of this moved sample was stored in the sinogram containing movement. This was repeated 180 times to fill the sinogram of the moving target.

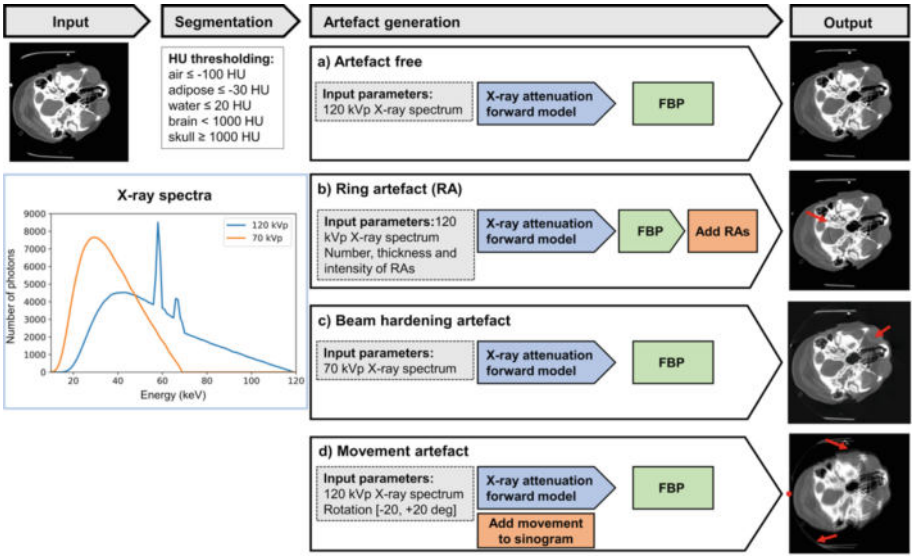


Fig. 1. Illustration of simulation workflow for different artefacts: a) Artefact free image, b) ringing artefact simulating dead pixels in detector (red arrow), c) beam hardening artefact image and d) movement artefact manifested as blurring of the image (red arrow). Also please note the "halo" in the movement artefact image in the field of view edge (red arrow). The rotation axis is annotated as a red dot.

Deep Learning Model for CT Artefact Classification: PyTorch (v1.8.1) framework was utilized, and an VGG-16 convolutional neural network architecture pre-trained with ImageNet dataset was applied for artifact classification [22]. The last fully connected layer was modified to produce four outputs. All model layers were trained during transfer learning for artefact detection. The reconstructed images were processed to a size of 1 × 512 × 512, and during training, they were augmented by randomly cropping to 1 × 450 × 450 size to obtain variability. Data augmentation was performed during training

phase before feeding input data to the network. Subsequently, they were resized to the required input size of 3 × 244 × 244 and fed as an input for the pre-trained VGG-16 network.

The total dataset ($N_{Total} = 1501 \times 4$ (6004) was randomly split into training set ($N_{train} = 1040 \times 4$(4160)) two validation sets ($N_{val1} = 130 \times 4$(520) and $N_{val2} = 130 \times 4$(520), and to a separate test set ($N_{test} = 201 \times 4$(804); two individual subjects). In the notation, four refers to {artefact-free, ring artefact, movement artefact, beam hardening} images simulated from one slice image. This yielded data splitting to 76.5% training, 13.2% validation and 10% for testing. The maximum number of epochs was set to 30. To avoid model overfitting, early stopping with patience of six epochs based on validation loss was applied. The model training was performed as follows: N_{train} was used in training, and N_{val1} was applied in early stopping evaluation. Then the model was evaluated on the N_{val2} set. The hyperparameter tuning was applied in this step.

After selecting the best performing hyperparameter combination, the final model was trained on $N_{train} + N_{val2}$ and N_{val1} was used for early stopping for that model. Final evaluations were performed to the separate test set. Hyperparameter tuning was applied to learning rates (LRs) {1e−4; 1e−5; 1e−6}, batch sizes {8,16}, and weight decays {1e−1, 1e−2, 1e−3} yielding 18 possible combinations. The final best performing combination was: LR = 1e−5; batch size = 8, and weight decay = 0.01, which was used in the final training using ADAM optimizer for 15 epochs. The cross-entropy loss function was applied as the loss function.

Validation of Artefact Detection Framework: For model performance validation, the area under receiver operating characteristics curve (ROC-AUC), average precision, accuracy, precision, recall, and F1-scores (2 × (precision × recall)/(precision + recall)) were determined. A Gradient-weighted Class Activation Mapping (Grad-CAM) approach was applied to provide visual explanations highlighting the regions important for the VGG-16 model decisions [23]. Finally, to assess the limits of reliability of the trained model (i.e. to test the detection limits for artefacts that were not so prominent), an additional sensitivity analysis was performed with two different slice images taken from the test dataset with different RA and movement artefact classes: 1. Bright and dark RAs with

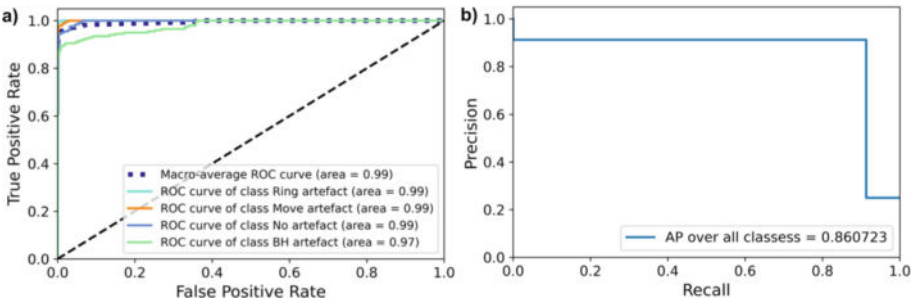


Fig. 2. a) Average receiving characteristics operating curve (ROC) for classification of simulated CT artefacts ($N_{test} = 804$) using VGG-16 classifier, and b) average precision-recall curve over all artefact classes. VGG-16 classifier had the best detection performance on simulated ring artefacts and the poorest performance was found on beam hardening artefacts.

increasing radius from 4 to 245 in 10 evenly spaced intervals were generated and fed to the trained model to assess how small a radius the RA model is capable of detecting. 2. Movement from 0.2° to 3° in 10 evenly spaced intervals was simulated in the image slices.

3 Results

The model performance was excellent for the test set (Table 1). The poorest precision was on the movement and beam hardening artefact classes. The overall ROC-AUC, average precision and accuracies were 0.99, 0.86 and 0.91, respectively (Fig. 2). When ROC-AUC was evaluated class-wise, the poorest performance was found on beam hardening artefact (area = 0.97) (Fig. 2).

Table 1. Precision, recall and F1-score values for simulated artefact-free and different image artefact types.

Image Type	Precision	Recall	F1-score
No artefact	0.99	0.71	0.85
Ring artefact	0.99	0.99	0.99
Beam hardening artefact	0.85	0.91	0.88
Motion artefact	0.84	0.99	0.92

The GradCAM visualizations showed that with all image classes, the model attention is focused on bony regions (Fig. 3). In the simulated artefact-free image, the model attention highlights the whole head region, whereas, for RA images, the model attention is intensified in the RA. For movement artefacts, the attention is focused on blurring artefact as well as in the field of view edge, which has a halo artefact (Fig. 3). For beam hardening artefact, the attention map focuses on the uniform brain region as well as in the bony regions (Fig. 3c).

The qualitative sensitivity analysis revealed that the model could classify all movement artefacts correctly. However, this was due to the visible edge in the field-of-view, which is also highlighted in the attention maps (Fig. 4). The model was incapable of detecting small RAs with a radius of 4 pixels (Fig. 5). Also, the bright RA near the bony regions was misclassified by the model even though the model attention highlights the RA region (Fig. 5e).

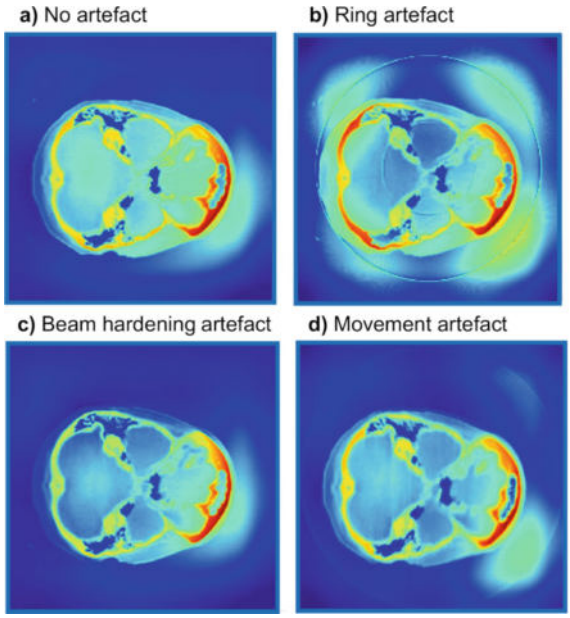


Fig. 3. Example illustrations of the artefacts and heat maps highlighting the regions where the model attention was the highest (pun location).

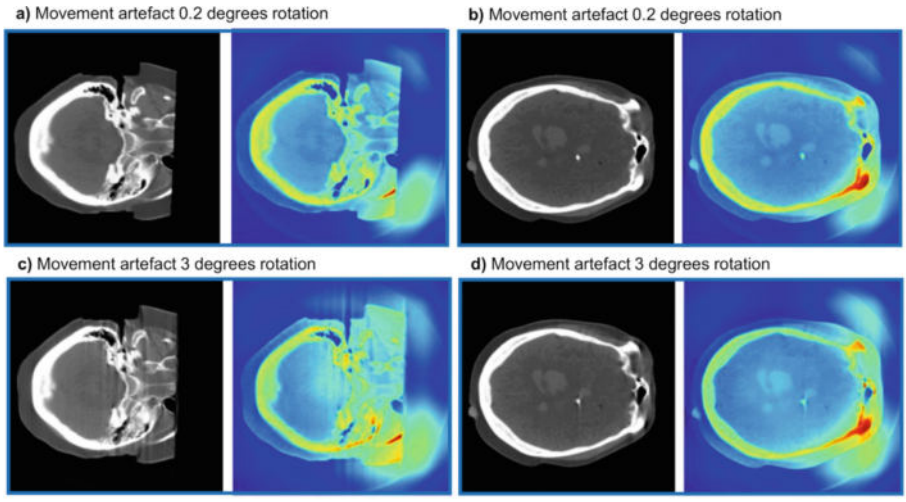


Fig. 4. Qualitative sensitivity assessment of the movement artefacts. The model could detect movement artefact in all images. The heatmap is produced using the Grad-CAM approach.

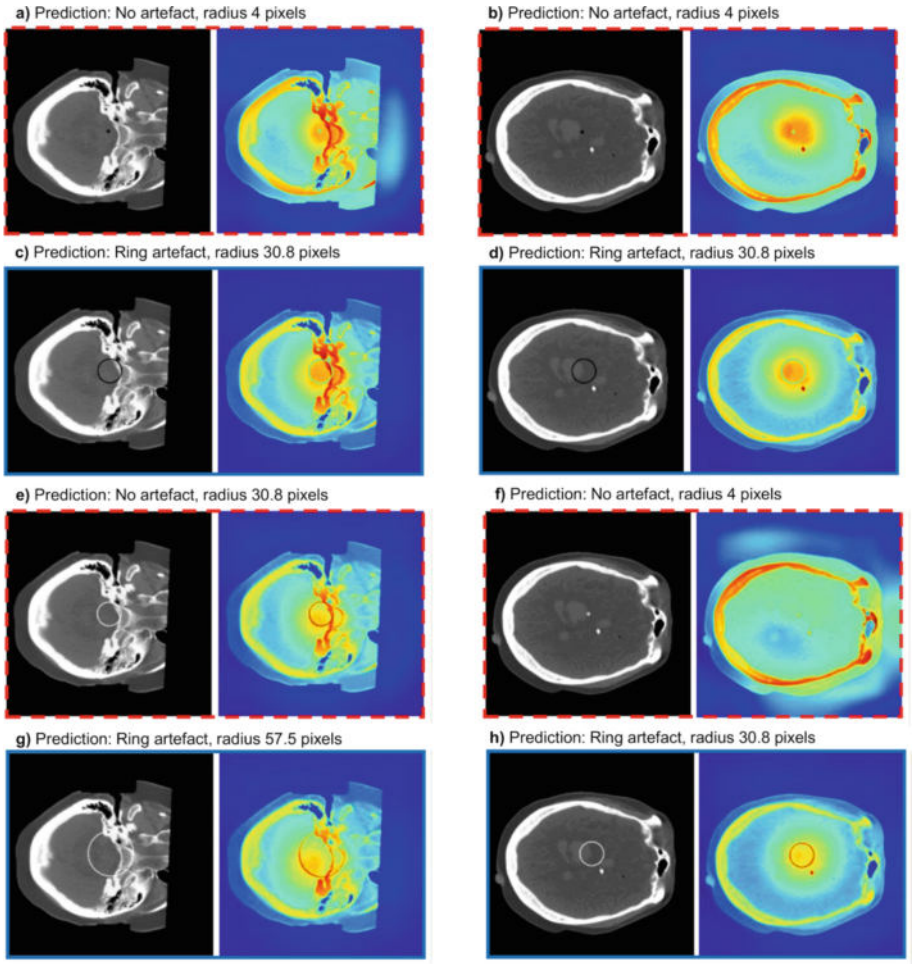


Fig. 5. Qualitative sensitivity assessment of the RAs. All images have a RA and the red dashed and blue solid borders denotes misclassification and correct classification of the model, respectively. The heatmap is produced using the Grad-CAM approach.

4 Discussion

Data-driven algorithms and especially convolutional neural networks have shown their applicability in various medical image processing and analysis tasks. This simulation study demonstrated that deep learning can be utilized as an effective classification and detection tool for CT image artefacts. Furthermore, the attention maps created using the Grad-CAM approach provided visual assessment for model performance evaluation.

In image quality assurance tasks, recent advances have been made in automating the analysis of technical image quality parameter directly from patient images [7, 8]. The approach developed in this study could be incorporated as a monitoring or artefact indicator tool to support imaging technologists who monitor the patient during the

clinical imaging process. DL could be applied to support this monitoring process as a post-analysis step. Although, majority of these CT image artifacts can be differentiated based on visual observation, the DL-based automated process may speed up the process of checking images from large datasets after acquisition, thus enabling faster maintenance action. Also, this approach could have applications not only for artefact detection but also in assessing the artefact prevalence of CT scanners. The developed method has the potential to improve quality assessment and decrease recall rates. Further, DL-based quality assurance could be used to identify protocols that require further optimization or scanners that require maintenance. To illustrate, if the classification tool frequently detects movement artefacts in routine use of a specific protocol, the rotation time could be decreased, the pitch factor increased to reduce scan times, or the imaging technologists could be trained to better advise/guide the patients not to move during scan process.

Although the overall model performance was excellent, there were misclassifications in the test data. Therefore, an additional sensitivity analysis was performed to assess the limits of the model. This analysis, combined with the Grad-CAM visualization, revealed that small RAs were left undetected as they were difficult to distinguish from anatomical structures and bright RAs near the skull were left undetected by the model. The model performance may be further improved by introducing more of these misclassified cases in the learning process. The model attention in movement artefact simulations focused on the edge of the field-of-view region, which is not ideal as the out of field artefacts may be pre-corrected for clinical images.

This study has the following limitations. Only simulated artefact dataset from the head region was utilized as a comprehensive collection of clinical images with various artefacts was not available. This was because some of the artefacts are not commonly encountered in clinical practice. For illustration, RAs are usually monitored by radiographers using quality assurance phantoms. In addition, only one openly available head CT dataset was utilized, and it contained images only from the head region. However, CT artefacts and artefact types differ depending on which body region is being imaged (e.g., in the thorax region, respiration-related motion artifacts are more common). Therefore, in future studies, the dataset should include other body regions. However, this was considered out of scope in this proof-of-concept study. One solution to produce a non-simulated dataset with various artefacts in the whole-body region would be to scan anthropomorphic phantoms with artefacts. Imaging phantoms enable a controlled and systematic process for artefact generation with different CT imaging protocols. For example, movement artefacts could be easily produced by introducing movement e.g. using a linear actuator. X-ray detector-related artefacts are more difficult to generate, but for example, contrast material contamination in the detector cover or mylar window produces RAs. In this proof-of-concept study, only a pre-trained VGG-16 neural network architecture was utilized as it performed very well initially. However, in future studies, other more sophisticated network architectures should be investigated in combination with real measured *i.e.,* non-simulated, artefact datasets.

Moreover, the grad-CAM heatmaps generated from the trained VGG-16 model highlighted well the artefact regions. However, other architectures should also be experimented with in more detail in future studies. In addition, the verification of model performance with clinical data needs to be addressed in the future. Furthermore, the

clinical data should have a variety of CT scanner vendors and protocols to increase the variability in the noise texture [14], image resolution, as well as in contrast. Finally, exposing the model for different X-ray energy spectra, and collimations may cause challenges in real patient data as the contrast and overall image quality appearance would differ.

5 Conclusion

In summary, a classification pipeline for CT image artefact detection was developed using VGG-16 model convolutional neural network architecture. The model performance was excellent on a simulated dataset, and the developed method shows promise for medical image quality assurance as an integrated part of the routine diagnostic workflow. However, before this, the results need to be verified with a clinical dataset of various artefacts and different CT scanners from different vendors.

Acknowledgments. This study was supported by Academy of Finland (Project no. 316899).

Disclosure of Interests. The authors have no relevant conflicts of interest to disclose.

References

1. Barrett, J.F., Keat, N.: Artifacts in CT: recognition and avoidance. Radiographics **24**(6), 1679–1691 (2004). https://doi.org/10.1148/rg.246045065
2. Sun, T., Kim, J.-H., Fulton, R., Nuyts, J.: An iterative projection-based motion estimation and compensation scheme for head x-ray CT. Med. Phys. **43**(10), 5705–5716 (2016). https://doi.org/10.1118/1.4963218
3. Kyriakou, Y., Meyer, E., Prell, D., Kachelrieß, M.: Empirical beam hardening correction (EBHC) for CT. Med. Phys. **37**(10), 5179–5187 (2010). https://doi.org/10.1118/1.3477088
4. Törmänen, J., Rautiainen, J., Tahvonen, P., Leinonen, K., Nieminen, M.T., Tervonen, O.: The 'Air in the CT X-ray tube oil' artifact—examples of the quality control images and the evaluation of four potential clinical patients' head computed tomography cases. J. Comput. Assist. Tomogr. **41**(3), 489–493 (2017). https://doi.org/10.1097/RCT.0000000000000532
5. Long, Z., et al.: Technical note: display window setting: an important factor for detecting subtle but clinically relevant artifacts in daily CT quality control. Med. Phys. **43**(12), 6413–6417 (2016). https://doi.org/10.1118/1.4966698
6. Lifton, J., Liu, T.: Ring artefact reduction via multi-point piecewise linear flat field correction for X-ray computed tomography. Opt. Express **27**(3), 3217 (2019). https://doi.org/10.1364/OE.27.003217
7. Smith, T.B., Solomon, J., Samei, E.: Estimating detectability index in vivo: development and validation of an automated methodology. J. Med. Imaging **5**(03), 1 (2017). https://doi.org/10.1117/1.JMI.5.3.031403
8. Christianson, O., Winslow, J., Frush, D.P., Samei, E.: Automated technique to measure noise in clinical CT examinations. Am. J. Roentgenol. **205**(1), W93–W99 (2015). https://doi.org/10.2214/AJR.14.13613

9. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**(2012), 60–88 (2017). https://doi.org/10.1016/j.media.2017.07.005
10. Galib, S.M., Lee, H.K., Guy, C.L., Riblett, M.J., Hugo, G.D.: A fast and scalable method for quality assurance of deformable image registration on lung CT scans using convolutional neural networks. Med. Phys. **47**(1), 99–109 (2020). https://doi.org/10.1002/mp.13890
11. Sreekumari, A., et al.: A deep learning-based approach to reduce rescan and recall rates in clinical MRI examinations. Am. J. Neuroradiol. **40**(2), 217–223 (2019). https://doi.org/10.3174/ajnr.A5926
12. Verdun, F.R., et al.: Image quality in CT: from physical measurements to model observers. Physica Med. **31**(8), 823–843 (2015). https://doi.org/10.1016/j.ejmp.2015.08.007
13. Christianson, O., et al.: An improved index of image quality for task-based performance of CT iterative reconstruction across three commercial implementations. Radiology **275**(3), 725–734 (2015). https://doi.org/10.1148/radiol.15132091
14. Juntunen, M.A.K., Rautiainen, J., Hänninen, N.E., Kotiaho, A.O.: Harmonization of technical image quality in computed tomography: comparison between different reconstruction algorithms and kernels from six scanners. Biomed. Phys. Eng. Exp. **8**(3), 037002 (2022). https://doi.org/10.1088/2057-1976/ac605b
15. Madesta, F., Sentker, T., Gauer. T., Werner, R.: Deep learning-based conditional inpainting for restoration of artifact-affected 4D CT images. Med. Phys. (2023). https://doi.org/10.1002/mp.16851
16. Prakash, P., Dutta, S.: Deep learning-based artifact detection for diagnostic CT images. In: Bosmans, H., Chen, G.-H., Gilat Schmidt, T. (eds.) Medical Imaging 2019: Physics of Medical Imaging, p. 158. SPIE (2019). https://doi.org/10.1117/12.2511766
17. Muschelli, J., Ullman, N.L., Mould, W.A., Vespa, P., Hanley, D.F., Crainiceanu, C.M.: Validated automatic braintraction of head CT images. Neuroimage **114**, 379–385 (2015). https://doi.org/10.1016/j.neuroimage.2015.03.074
18. Segars, W.P., Sturgeon, G., Mendonca, S., Grimes, J., Tsui, B.M.W.: 4D XCAT phantom for multimodality imaging research. Med. Phys. **37**(9), 4902–4915 (2010). https://doi.org/10.1118/1.3480985
19. Punnoose, J., Xu, J., Sisniega, A., Zbijewski, W., Siewerdsen, J.H.: Technical note: Spektr 3.0 - a computational tool for x-ray spectrum modeling and analysis. Med. Phys. **43**(8), 4711–4717 (2016). https://doi.org/10.1118/1.4955438
20. van Aarle, W., et al.: The ASTRA toolbox: a platform for advanced algorithm development in electron tomography. Ultramicroscopy **157**, 35–47 (2015). https://doi.org/10.1016/j.ultramic.2015.05.002
21. van Aarle, W., et al.: Fast and flexible X-ray tomography using the ASTRA toolbox. Opt. Express **24**(22), 25129 (2016). https://doi.org/10.1364/OE.24.025129
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations (ICLR 2015), Conference Track Proceedings, pp. 1–14 (2014)
23. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: visual explanations from deep networks via gradient-based localization. Int. J. Comput. Vis. **128**(2), 336–359 (2020). https://doi.org/10.1007/s11263-019-01228-7

Assessment of Parkinson's Disease Severity Using Gait Data: A Deep Learning-Based Multimodal Approach

Nabid Faiem[1](✉), Tunc Asuroglu[1], Koray Acici[2], Antti Kallonen[1], and Mark van Gils[1]

[1] Faculty of Medicine and Health Technology, Tampere University, 33720 Tampere, Finland
nabid.faiem@tuni.fi

[2] Department of Artificial Intelligence and Data Engineering, Ankara University, 06830 Ankara, Turkey

Abstract. The ability to regularly assess Parkinson's disease (PD) symptoms outside of complex laboratories supports remote monitoring and better treatment management. Multimodal sensors are beneficial for sensing different motor and non-motor symptoms, but simultaneous analysis is difficult due to complex dependencies between different modalities and their different format and data properties. Multimodal machine learning models can analyze such diverse modalities together, thereby enhancing holistic understanding of the data and overall patient state. The Unified Parkinson's Disease Rating Scale (UPDRS) is commonly used for PD symptoms severity assessment. This study proposes a Perceiver-based multimodal machine learning framework to predict UPDRS scores.

We selected a gait dataset of 93 PD patients and 73 control subjects from the PhysioNet repository. This dataset includes two-minute walks from each participant using 16 Ground Reaction Force (GRF) sensors, placing eight on each foot. This experiment used both raw gait timeseries signals and extracted features from these GRF sensors. The Perceiver architecture's hyperparameters were selected manually and through Genetic Algorithms (GA). The performance of the framework was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and linear Correlation Coefficient (CC).

Our multimodal approach achieved a MAE of 2.23 ± 1.31, a RMSE of 5.75 ± 4.16 and CC of 0.93 ± 0.08 in predicting UPDRS scores, outperforming previous studies in terms of MAE and CC.

This multimodal framework effectively integrates different data modalities, in this case illustrating by predicting UPDRS scores using sensor data. It can be applied to diverse decision support applications of similar natures where multimodal analysis is needed.

Keywords: Multimodal machine learning model · Perceiver · Gait analysis · GRF sensors

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 29–48, 2024.
https://doi.org/10.1007/978-3-031-59091-7_3

1 Introduction

1.1 Parkinson's Disease (PD)

PD is the fastest growing neurological disorder according to the Global Burden of Diseases, Injuries, and Risk Factors (GBD) studies [1–3]. The World Health Organization (WHO) estimated that about 8.5 million individuals were living with PD worldwide in 2019 [4]. In the last thirty years, there has been a significant increase in prevalence and mortality rates of PD. Several components contributed to this upward trend, such as a growing elderly population, environmental and social factors, and extended duration of the disease. If the current trend persists, it is projected that the number of individuals with PD could exceed 17 million by the year 2040 [5], which will pose enormous challenges for any healthcare system.

Diagnosis of PD is typically performed by neurologists specialized in movement disorders and involves different neurological tests and patient interviews. However, the diagnosis of PD remains a difficult task due to its overlapping characteristics with other neurodegenerative diseases and the subjectivity in short-term assessment. A global shortage of specialized neurologists increases the risk of misdiagnosis, potentially preventing targeted treatment and increasing disease severity. PD symptoms are typically evaluated using a rating scale called Unified Parkinson's Disease Rating Scale (UPDRS), which is a widely accepted score. This PD severity rating scale has four distinct components that include both non-motor and motor parts [6]. Early diagnosis and frequent assessment of PD symptoms is required for more targeted medical intervention and to support remote monitoring for better treatment management, thereby improving the quality of life of PD patients.

1.2 Gait Analysis

Gait analysis can be a useful tool for measuring gait abnormalities as gait worsens with the disease progression. The analysis is conducted in a specialized laboratory equipped with video systems, motion-capturing cameras, floor-based force sensors, and electromyography (EMG) systems [7]. Although this complex laboratory setup provides accurate results, the availability of these systems is limited due to expensive infrastructures and lack of skilled personnel, particularly in developing countries and remote places. Ground Reaction Force (GRF) non-invasive wearable sensors can offer a cost-effective and accessible tool for gait analysis to provide a comprehensive overview of gait pattern. These sensors are designed to capture joint movements and muscle activities effectively [8]. They are small and typically placed in the insole or underneath shoes to measure kinetic force, temporal, and spatial characteristics of gait variability.

Gait abnormalities in PD patients under cognitive load become more severe with the disease progression [9]. Dynamic changes in gait can be detected through regular monitoring of daily activities, medications, social interactions, or environmental conditions outside of the artificial laboratory in real-life settings. Therefore, there is a need for inexpensive monitoring methods that can be used not only during healthcare encounters but also to improve treatment intervention and management throughout the patient's lifetime [10].

The gait signals obtained from the controlled laboratory settings are more structured, as the participants follow a specific protocol in a strictly controlled environment [11]. Gait analysis has traditionally focused on temporal or frequency domain analysis with standard hypothesis testing. This approach was mainly used because the datasets were simpler and more structured [12]. However, real-world gait signals are unstructured and noisy, and may require multimodal sensing for more comprehensive analysis taking into account a wider context. Therefore, the GRF based gait signal alone may not be enough for estimating PD symptoms severity in remote monitoring applications. Analyzing other gait or tremor signals with non-motor symptoms combined could give us a better understanding about the progression of PD symptoms in real-life scenarios.

1.3 Multimodal in Decision Support

Analyzing multimodal data from diverse sources is a challenging task due to possible complex, non-linear relationships, and temporal dependencies between modalities [13]. If these modalities are analyzed separately with own methods (signal processing and other approaches), and their results are combined and processed thereafter, there is a risk of losing potential joint information between them. Integrating these modalities requires harmonization and standardization prior to analysis in a computational model [14]. Multimodal machine learning is an evolving field of machine learning where multiple modalities can be combined simultaneously to support or aid each other in enhancing the predictive performance of the model.

Recently, DeepMind's Perceiver architecture has shown promising results in processing different data modalities [15]. This architecture is designed on top of Transformer networks, it is capable of processing different modalities including time series, images, and other signals. The core concept behind the Perceiver architecture is the use of an iterative attention mechanism [15]. This mechanism allows the model to concentrate on distinct parts of signals to capture the underlying pattern. By integrating multiple modalities, the model may generate robust predictions as it can observe patterns from different modalities and identify the relationship between input modalities and output.

A Perceiver architecture-based multimodal machine learning framework could be an effective solution for simultaneously analyzing both raw gait timeseries signals and extracted hand-crafted features from GRF sensors, allowing them to complement each other to improve the predictive ability for PD diagnosis and PD symptoms severity estimation. The iterative nature of the Perceiver architecture, along with the weight sharing strategy can enhance the predictive performance by efficiently reusing the same input multiple times [15].

Optimizing hyperparameters is a challenging task in any deep learning model, particularly when parameters are selected manually. To overcome this issue, state-of-the-art optimization techniques like Random search (RS), Grid search (GS), Bayesian optimization (BO), and Genetic Algorithms (GA) techniques explored in earlier studies [16]. GA has the advantage of simulating different hyperparameter settings to find the optimal configuration to achieve better prediction performance. This algorithm has been used for hyperparameter optimization in the detection of femoral neck fracture [17] and the diagnosis of nutritional anemia [18].

1.4 Goal of the Study

The main advantage of using raw gait signals is that it eliminates the manual processing steps and simplifies analysis prior to computational model. However, to incorporate explicitly expert-based knowledge in the form of well-defined features, and to evaluate the capability of the computational model for analyzing multimodal data, we utilized the extracted features from gait timeseries signal as separate modality. This is because expert-derived features represent interpretable biomechanical metrics, while the gait timeseries signal provides sensor data covering variation in spatiotemporal domains. In addition, the effect of GA optimization performance is assessed in this approach.

This research proposes a multimodal machine learning framework based on the Perceiver architecture for predicting the severity of PD symptoms. The study also examines the framework's performance for the diagnosis of PD. The major contributions of this paper can be summarized as follows:

1. Study if UPDRS can be predicted with gait timeseries signals from GRF sensors and compare framework's performance with other studies,
2. Compare performance of multimodal vs single model approaches,
3. Study GA optimization performance.

The rest of the paper follows as outlined below: Sect. 2 discusses state-of-the-art data analysis methods related to PD diagnosis and severity assessment. Studies related to applying the Perceiver architecture in disease diagnosis for other diseases than PD are also presented in this section. Section 3 describes the materials and explanation of each component of the proposed framework. Section 4 presents the results of the performance evaluation of the proposed approach. Section 5 discusses the results comparing different modalities, as well as the limitations, challenges, and future scope of this research. Section 6 concludes the paper.

2 Related Work

Despite the need for regular assessment of PD symptoms to improve treatment management, most existing studies focused on diagnostic solutions for detecting PD. Fewer studies have addressed estimating the exact symptoms severity of PD as regression problem. This study includes references to PD diagnosis research to provide a comprehensive overview of how wearable sensors and data analysis can improve diagnosis. The majority of these studies explored single modality data whereas only a small number of studies have used a multimodal approach. These multimodal studies are mostly based on small populations and/or imbalanced datasets [19]. Prior to machine learning, standard hypothesis statistical tests like t-tests, Mann-Whiteney U test, and ANOVAs were employed for PD detection from gait data [12]. Recent literature shows successful implementation of machine learning and deep learning techniques for PD detection. Machine learning techniques like Random Forests (RF) [20] and Support Vector Machines (SVM) [21] have been explored for PD diagnosis from gait data. Convolutional Neural Network (CNN) and long short-term memory (LSTM) deep learning algorithms are mostly used in research for PD diagnosis [22]. Most studies use a single modality analysis for PD diagnosis and the PhysioNet Gait database is most used in these studies for gait analysis

[22]. This PhysioNet Gait dataset contains original UPDRS to reflect the severity of PD. The total UPDRS score ranges from 0 to 199, encompassing both motor and non-motor components, with 199 representing severe disability and 0 indicating the healthy state. The maximum score from the motor part of the scale is 108 [6, 23].

2.1 PD Severity Estimation

Aşuroğlu et al. [24] proposed a hybrid deep learning regression approach emphasizing on local pattern recognition for predicting PD symptoms severity from the gait signal. Their proposed framework is based on the combination of CNN and locally Weighted Random Forest (LWRF) that use multi-channel gait data to predict exact UPDRS scores. The convolutional part of their framework extracts local characteristics from the extracted time and frequency domain features and the LWRF part exploits the local relationships from these characteristics. Their proposed model achieved a state-of-the-art performance and outperformed the previous study. Aşuroğlu et al. [25] conducted a prior study focused on the same regression problem for estimating PD symptoms severity. In this study they used a decision tree-based supervised machine learning model. This study was the first one that utilizes multichannel GRF wearable sensors-based gait data in general. They utilized the same time and frequency domain features as [24] for the prediction. Their developed ML model exploited the local patterns from these features to better predict UPDRS scores.

2.2 PD Diagnosis

In this section, studies that used the PhysioNet gait dataset for PD diagnosis are discussed to maintain a consistent comparison. El Maachi et al. [26] used raw gait timeseries data for PD detection using a deep learning model based on the 1D CNN (1D-Convnet). This model was designed to simultaneously process 18 one-dimensional signals obtained from 16 GRF foot sensors and the total force from each foot, eliminating the need for manual feature extraction. The first part of their model used 18 parallel 1D-CNN signals for local spatial information extraction. Following this, a fully connected layer that integrates the relevant CNN spatial features for PD diagnosis. Their proposed algorithm achieved an accuracy of 98.7%, a sensitivity of 98.1% and a specificity of 100.0%. Alharthi et al. [27] used deep CNN architecture for PD diagnosis. They transformed the raw GRF sensor signal into a 3D matrix to provide input to the model. Their approach was designed to learn the spatiotemporal GRF signals without manual feature extraction. Their proposed model was robust against noise and effectively addressed the variability of human movement between individuals.

Pham et al. [28] focused on a single GRF sensor from the gait data for PD detection. From the timeseries signal, they extracted time-frequency and time-space features. With the extracted features they trained bi-LSTM (bi-LSTM). Their results showed better performance compared to conventional LSTM and other prior studies in terms of accuracy (100.0%), sensitivity (100.0%), specificity (100.0%) and F1 score (1.0). They also reported that their model was more efficient in terms of computational power and processing time as it used only one sensor. Balaji et al. [29] introduced a deep learning approach using LSTM for PD detection and severity classification, eliminating the need

for handcrafted features. They passed gait cycles from GRF sensors to train the LSTM network. Their network consists of four LSTM layers, four dropout layers followed by a fully connected layer and a SoftMax layer. Their approach achieved accuracy of 98.6% for PD detection.

Vidya et al. [30] used a hybrid CNN-LSTM model to explore the spatial and temporal dependence of GRF timeseries signals to differentiate between healthy subjects and different PD severity levels. They selected the optimal number of GRF sensors using a variability analysis. They applied the empirical mode decomposition (EMD) technique to extract the significant intrinsic mode functions (IMFs) through power spectral analysis to capture the non-linear and non-stationary characteristics of the timeseries signal. The dominant IMFs from the optimal GRF signals were then used to train the hybrid model for PD stage classification. They reported that their proposed hybrid model achieved better performance than other studies that used gait analysis to classify healthy subjects and PD severity stages.

Nguyen et al. [31] introduced a Transformer-based deep learning model that emphasized both temporal and spatial characteristics of gait signals to differentiate between healthy control subjects and PD patients. They applied one temporal Transformers for each gait sensor, and the dimensionally reduced outputs from these temporal Transformers were concatenated before being fed into a spatial Transformer. This spatially encoded feature set was then passed to two fully connected layers and an output layer for the final classification. Their model achieved accuracy of 95.2%, a sensitivity of 98.1% and a specificity of 86.8%.

2.3 Multimodal Data Analysis Using Perceiver

Although the Perceiver architecture is a recent development, it has already been implemented in other studies focusing on disease diagnosis other than PD.

Josef et al. [32] estimated the speed of human motion using IMU-based wearable sensors. They evaluated the performance of different deep learning methods, including the Perceiver architecture. In their experiment, they collected IMU data from a single foot, shin, and thigh. The Perceiver architecture, along with other deep learning techniques, outperformed conventional feature-based methods in estimating speed. Aadam et al. [33] evaluated the performance of the Perceiver architecture for classification of emotion from raw EEG signals. In this experiment, they used EEG signal from DEAP [34] dataset, and they used two modalities for the analysis. The first modality consisted of EEG signal from all channels as 1D vector and the spatial locations of electrodes as the second modality. They found that the Perceiver model performed better for multimodal configuration compared to single modality.

3 Materials and Methods

This section introduces a Perceiver architecture-based multimodal machine framework for PD symptoms severity estimation. The proposed framework can simultaneously process both raw gait timeseries signals and extracted features from GRF sensors as

multimodal input to predict UPDRS score. This framework can also process each modality input separately. Figure 1 depicts the workflow of the proposed framework which includes the Perceiver model and hyperparameter optimization using GA.

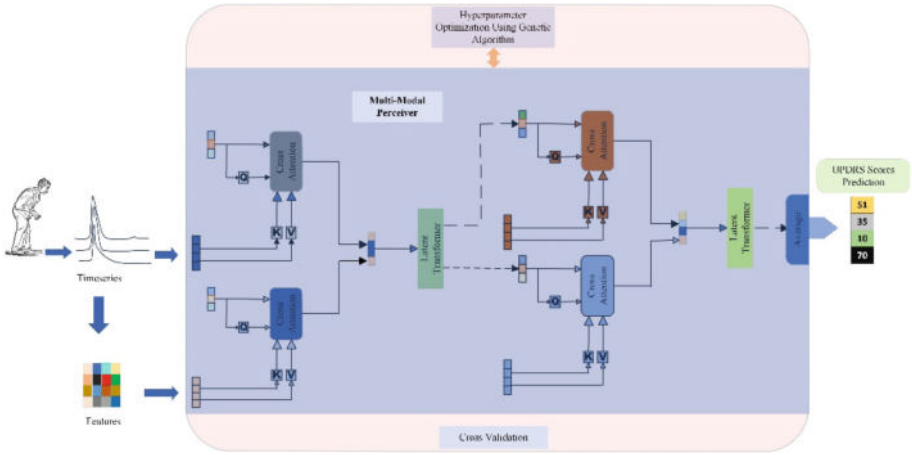


Fig. 1. Workflow of the proposed Perceiver architecture-based multimodal machine learning framework for UPDRS score prediction. The framework simultaneously processes raw gait time-series signals and extracted features from these signals to complement each other through the iterative process of Perceiver network. GA selects hyperparameters of the Perceiver architecture in the first cross validation (CV) and then applies them to subsequent CVs.

3.1 Dataset Description

The performance of the proposed architecture was evaluated using the dataset of PhysioNet [35] that includes walking sequences from 93 idiopathic PD patients and 73 healthy control subjects. The mean age of the PD patients was 66.3 years, and 63% of the patients were male. The average age of control subjects was 63.7 years among which 55% were men. This dataset was collected by three independent research groups (Yogev et al. [36], Hausdorff et al. [37], and Silvi Frenkel-Toledo et al. [38]) at the Laboratory for Gait & Neurodynamics, Movement Disorders Unit of the Tel Aviv Sourasky Medical Center.

Gait patterns were measured from each participant for two minutes by placing eight GRF sensors under each foot. Participants were asked to walk in two different scenarios: normal walking at a self-selected speed and dual-task walking. In the dual-task protocol, subjects were instructed to perform arithmetic tasks by serially subtracting seven from a pre-defined number.

These GRF sensors measure force (in Newton) with a sampling rate of 100 Hz and the force distribution of these sensors could be used to measure the gait impairment of subjects. The GRF sensors-based measurement system provides better foot distribution compared to force-sensitive resistors (FSR) due to their larger size and sensing area [39].

The dataset also includes demographic information about the participants, such as their gender, age, height, weight, and PD severity values as UPDRS scores. In this dataset, the mean UPDRS score of PD patients was 32, with a minimum score of 13 and a maximum of 70.

3.2 Pre-processing and Feature Extraction

The first 20 s and the last 10 s from each sensor data segment were excluded to minimize the start and end effects. After that, a median filter with a length of 3 was applied to remove outliers or large spikes. This filter reduces sudden walking fluctuations, and the small filter size preserves maximum force signal without distortion.

Gait cycles are repetitive in nature therefore sensor timeseries signals may contain redundant information. Hand-crafted features that reflect the spatiotemporal and frequency characteristics of timeseries signals may be useful for PD detection or PD severity estimation. Therefore, we extracted the frequency and time domain features from the median filtered signal of each GRF sensor. These seven-frequency domain and sixteen-time domain extracted features, presented in Table 1, effectively capture relevant information that has been demonstrated in previous studies [24, 25]. Extracted features were then standardized by subtracting the mean and scaling to unit variance, as defined in Eq. 1. In this equation, 'μ' represents a mean of a specific feature that is calculated from all subjects. 'σ' represents the standard deviation of that feature also calculated from all subjects, x represents feature value for a single subject and z is the standardized value of x.

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

Table 1. Time and Frequency based features extracted from each sensor of all participants.

Feature Domain	Features calculated from each GRF sensor
Time	mean, harmonic mean, median, range, interquartile range (IQR), mean absolute deviation, maximum amplitude and minimum/maximum spread, skewness, kurtosis, root mean square (RMS), energy, power, and entropy
Frequency	mean, minimum, maximum, normalized, energy, power, and phase

3.3 Perceiver Architecture

The Perceiver architecture has the potential of combining multiple data modalities to improve a model's predictive capability. The Perceiver leverages an iterative attention mechanism to scale high-dimensional multimodal data without making any domain specific assumptions. This architecture employs scalable Fourier features-based position encoding to preserve the temporal, spatial or spatiotemporal characteristics of the input data. These encoded features are then concatenated with input data before processing

in the main architecture. As illustrated in Fig. 2, the Perceiver architecture is composed of two main components: the cross-attention module and the latent Transformer. The cross-attention module reduces the dimensionality of the input data into a lower dimensional latent bottleneck. The latent Transformer then processes further to learn complex patterns in the data. This concept allows the construction of large networks of multiple cross-attention and the latent Transformer blocks for processing complex and high-dimensional multimodal data without changing the underlying architecture. The weight sharing strategy across each block can also improve predictive performance by efficiently reusing the same input multiple times. Finally, the model generates predictions by averaging the output of the final latent Transformer over the index dimension depending on the classification or regression task [15].

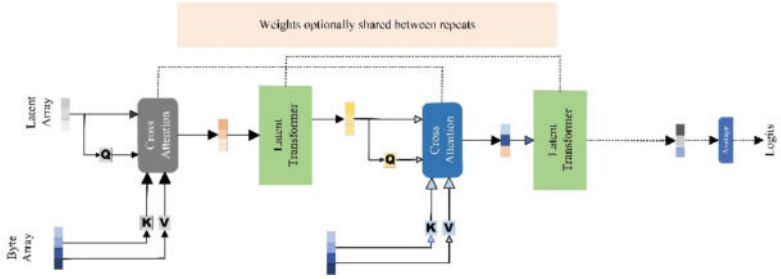


Fig. 2. The Perceiver architecture uses an iterative approach that alternates between cross-attention layers and several latent self-attention blocks. By alternating between these two types of blocks, the model can receptively attend to the input data and find patterns between the input and output layers more effectively. This mechanism allows the model to reduce the dimensionality of the input data while preserving the most critical information. The weight sharing strategy of the architecture can also improve predictive performance.

A potential disadvantage of the Perceiver architecture is that while the size of the latent array facilitates detail mapping of the input data, the bottleneck effect may limit the extent of detail. The use of multiple cross-attention layers can improve the precision of information extraction from the input data. However, this comes at the cost of increased computational resources that can lead to longer processing times [15].

3.4 Proposed Framework

The proposed framework uses the Perceiver architecture to identify the relationships between GRF signals with corresponding targets (either for classification or regression purposes). This work uses a multimodal version of the perceiver architecture that is capable of processing both unimodal and multimodal data in a single run [40]. This multimodal architecture leverages the attention mechanism to dynamically focus on distinct parts of these modalities to enhance the overall performance.

After the pre-processing phase, the timeseries signal of each participant is reshaped to 9025 samples $\times$ 16 sensors and the features set is reshaped to 368 $\times$ 1 from 23 features $\times$ 16 sensors. This conversion is necessary as the multimodal version [40] requires the

inputs in specific format for processing. These reshaped datasets are then converted to a one-dimensional tensor vector to feed into the Perceiver architecture.

To address the class imbalance during training, we use the weighted random sampler algorithm from PyTorch without oversampling. This means that each training batch includes samples from both classes (0 = healthy and 1 = PD) proportional to their class weights, until all samples from the minority class have been utilized. After that, training continues with samples from the majority class. The training batch size for these experiments is set to six. Validation is conducted in a single batch, where samples are randomly selected without considering class balance to simulate the real-world scenario.

3.5 Experimental Setup

The Perceiver model's hyperparameters are initially selected in a non-optimized manner using a trial-and-error method, where the optimal selection is determined based on the lowest prediction error. We first begin with predetermined hyperparameters and then adjust them to achieve the lowest possible prediction error. Through this process, we then select a network depth of 4, a cross-attention layer of 6 and latent-attention layer of 6 and the weighting sharing between the cross-attention and the latent self-attention layers. For optimization, we use Adam optimizer with a learning rate of 10^{-4}.

The proposed framework is implemented in Python with the PyTorch library on the JupyterLab development environment. This experiment is conducted on a computer with an AMD Ryzen 5 5600X 6-Core Processor, 16 GB RAM, and a 12 GB NVIDIA GeForce RTX 3060 graphics card. CUDA library of GeForce is used to utilize graphics card. Typical training sessions for processing either multimodal data or timeseries signals lasts approximately 80 h. The duration of training session extends when GA optimization is employed.

3.6 Evaluation

The accuracy of the proposed framework for predicting PD severity symptoms depends on the error between actual and predicted UPDRS scores. We use Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Correlation Coefficient (CC) to evaluate the performance of the proposed framework. MAE measures the mean difference between two continuous variables (mean difference between actual and predicted UPDRS scores).

$$MAE = \frac{|p_1 - a_1| + \cdots + |p_n - a_n|}{n} \tag{2}$$

$$RMSE = \sqrt{\frac{(p_1 - a_1)^2 + \cdots + (p_n - a_n)^2}{n}} \tag{3}$$

Here n is the size of the sample, p and a are the target and estimated value (output of the algorithm), respectively.

Initially gait data shorter than two minutes are discarded to maintain consistency across samples. This process resulted in a comprehensive dataset of 189 samples from

126 participants (68 PD patients and 58 Healthy control subjects). Gait timeseries data and the extracted features from each sample are structured according to the requirements of the framework to train either a single or a multimodal Perceiver model, individually or together. In the framework, we divide the structured dataset into ten randomly equal-sized subsamples using the ten-fold CV method. For each iteration, one subsample representing 10% of the dataset is reserved for validation and the remaining nine subsamples are reserved for training. This process is then repeated ten times. During training in each CV, samples are divided in proportion to preserve class balance as indicated in the Proposed Framework.

The final performance of each model is evaluated using the mean and standard deviation (SD) of MAE, RMSE and CC obtained from the ten-fold CV. The best performing model is identified by the lowest mean and SD in both MAE and RMSE, along with the highest mean and lowest SD in CC. The model is then benchmarked against the outcomes of previous similar studies. Similarly, we compare the mean and SD of these metrics between the multimodal and single model approaches to determine the best model. The performance of these models is visually illustrated using scatter plots of actual versus predicted UPDRS scores.

3.7 Hyperparameter Optimization Using GA

Hyperparameters define the complexity of the Perceiver architecture and its learning behavior [16]. Hyperparameters are difficult to optimize while improving model performance and reducing complexity of the architecture. This work attempts to optimize hyperparameters like the network depth, the number of cross-attention and self-attention blocks to minimize the prediction error of the framework. The manual tuning of these hyperparameters is time-consuming, so after an initial effort, we use GA for hyperparameter optimization. The main mechanism behind GA is given below [16]:

- First, initialize the population's equivalent of chromosomes and genes, randomly. These parameters represent the search space, hyperparameter and hyperparameter values, respectively.
- A fitness value of each member of the current generation is evaluated using a fitness function. The objective of the fitness function is to minimize the prediction error from the optimized hyperparameter settings. We use MAE as the fitness value for regression tasks and accuracy as the fitness value for classification tasks.
- The termination criterion for the regression task is set at a MAE of 2.5, while for the classification task it is set at an accuracy of 100%. These values are set high to achieve better performance by exploring more generations and combinations. If there is no change in the accuracy or the MAE score for two consecutive generations, then the evolution process will stop. Otherwise, if the termination criterion is not met, then proceed with the following steps:

  – Select parents from the mating pool.
  – Perform crossover and mutation operations on the chromosomes to produce the next generation population.
  – Evaluate the fitness of each child in the new generation.

To implement this algorithm, we use the TorchGA open-source library, which has implemented GA using PyTorch library [41]. The hyperparameters of this algorithm are presented in Table 2. In this study, we use the population size of 10 and repeat the algorithm until 10 generations have passed if the termination condition is not met. We configure three hyperparameters such as net depth, the number of cross-attention and self-attention layers and represent them as chromosomes. Their corresponding values, as shown in Table 2, are specified as genes. Parents are randomly selected from the mating pool, and a crossover operation is performed after the selection. We do not include mutation operations in this task.

We use ten-fold CV to evaluate PD diagnosis and symptoms severity estimation performance, applying both single and multimodal model with the same structured gait data and extracted features. The class balance is also maintained during the training. The optimal hyperparameters are identified through GA optimization using the fitness function from the first CV. The resultant hyperparameters from the first CV are then used in the remaining CVs.

Table 2. GA hyperparameters

Hyperparameters	Values
Population Size	10
Generation	10
Chromosomes (Genes)	Net depth (4,8), Cross-attention (1,2,4), Latent attention (2,4,8)
Parent Selection Type	Random
Crossover Type	Uniform

4 Results

4.1 Empirical Results

Table 3 presents ten-fold CV results of models with manually selected hyperparameters. The model that incorporated both Features and Timeseries as input modalities together is referred to as the Multimodal Data model. This model outperformed the other two models where Timeseries or Features were used as separate input modalities. From Table 3, it is observed that the Multimodal Data model achieved the highest performance among all models with an MAE of 2.23 $\pm$ 1.31, RMSE of 5.75 $\pm$ 4.16 and CC of 0.93 $\pm$ 0.08, indicating that it enhances prediction with a certain degree of variability. The model that utilized the Features modality demonstrated slightly lower performance with an MAE of 2.72 $\pm$ 1.57, RMSE of 6.79 $\pm$ 4.42 and CC of 0.91 $\pm$ 0.09. The model used the Timeseries modality has the highest errors, as indicated by an MAE of 3.18 $\pm$ 1.60, RMSE of 7.56 $\pm$ 3.89 and CC of 0.90 $\pm$ 0.08. The MAE scores of each model demonstrate relatively

low means and small SD. The RMSE scores of each model exhibits high means and larger SD, particularly due to the mispredictions of higher UPDRS scores.

Table 3. Performance comparison of different input modalities with manually selected hyperparameters (Network Depth: 4, Cross-Attention Layer: 6, Latent-Attention Layer: 6)

Input Modality	MAE Mean ± SD	RMSE Mean ± SD	CC Mean ± SD
Features	2.72 ± 1.57	6.79 ± 4.42	0.91 ± 0.09
Timeseries	3.18 ± 1.60	7.56 ± 3.89	0.90 ± 0.08
Multimodal Data	**2.23 ± 1.31**	**5.75 ± 4.16**	**0.93 ± 0.08**

Figure 3 indicates the relationship between predicted and actual UPDRS scores for three models with manually selected hyperparameters. The blue reference line in the plot represents the ideal scenario where predicted scores would perfectly align with actual scores. Figure 3 illustrates that the Multimodal Data model performed better than other models, as most samples were comparatively closer to the reference line than others.

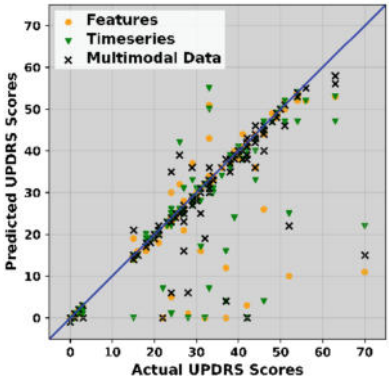


Fig. 3. Scatter plot showing the predicted against the actual UPDRS scores from Multimodal Data (black cross), Timeseries (green down-pointing triangle) and Features (yellow circles). Scores closely align with the y = x reference line (blue) indicate a strong correlation between predicted and actual UPDRS scores.

Table 4 presents the performance of three models with GA selected hyperparameters. It selects identical hyperparameters for the Timeseries and Multimodal Data modalities but selected a larger network for the Feature modality. Despite having a larger network depth, the Feature modality performed better than Timeseries modality. All models with GA selected hyperparameters showed poor performance. This suggests that suboptimal hyperparameters selected in the first CV could reduce GA optimization performance. However, the Multimodal Data outperformed the other two models in MAE, RMSE and CC.

Table 4. Performance comparison of different input modalities for the selected hyperparameters (D: Depth, C: Cross-attention layer and L: Latent-attention layer) using GA.

Input Modality	Hyperparameter (D, C, L)	MAE Mean ± SD	RMSE Mean ± SD	CC Mean ± SD
Features	8,1,2	3.64 ± 3.26	6.92 ± 5.54	0.92 ± 0.08
Timeseries	4,1,8	4.04 ± 1.82	7.79 ± 2.98	0.91 ± 0.07
Multimodal Data	4,1,8	2.58 ± 1.39	6.12 ± 3.46	0.93 ± 0.06

To compare the performance of models using these modalities with and without GA selected hyperparameters, we used scatter plot in Fig. 4. The plot visually compares the alignment between the predicted and actual UPDRS scores for each modality. In Fig. 4a, the model using the Features modality showed a decrease in performance when predicting UPDRS scores for healthy control subjects. Similarly in Fig. 4b, the model with the Timeseries modality showed a decline in performance for predicting UPDRS scores for both healthy control subjects and PD patients. The variation slightly improved when predicting using GA selected hyperparameters. However, the Multimodal Data model, as demonstrated in Fig. 4c, accurately predicted UPDRS scores for healthy control subjects in both cases. Similarly, the prediction UPDRS scores were relatively close to the reference line for PD patients in both scenarios.

In Table 5, we compared the performance of the Multimodal Data model with previous studies that predicted UPDRS scores from GRF signals. The Multimodal Data model, using manually selected hyperparameters, showed better performance in terms of MAE and CC compared to referenced studies. RMSE performance of the Multimodal Data model was slightly higher compared to one of the previous studies.

Table 6 demonstrates the classification evaluation metrics for three modalities with and without GA optimization. As can be seen from Table 6, the Multimodal Data model demonstrated the highest performance, with an accuracy of 97.3%, AUC of 0.98, sensitivity of 96%, and specificity of 100%. However, the performance slightly decreased with GA optimized hyperparameters. The models using the Features input modality also showed a relatively good performance, both with and without GA optimization, while the models that used Timeseries input modality had slightly lower performance in comparison. Overall, integrating multiple data modalities improved the predictive performance for PD diagnosis.

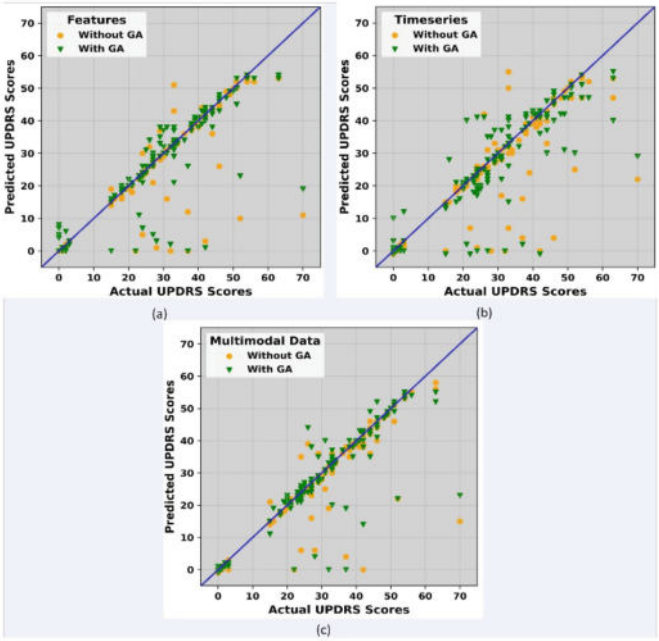


Fig. 4. Scatter plots comparing predicted and actual UPDRS scores for (a) Features Modality, (b) Timeseries Modality, and (c) Multimodal Data Modality with and without GA Optimization.

Table 5. Comparison with previous studies on PD severity estimation.

Authors	MAE	RMSE	CC
Aşuroğlu et al. [24]	3.01	4.56	0.90
Aşuroğlu et al. [25]	4.46	7.38	0.90
Present Method (Multimodal)	**2.23 ± 1.31**	5.75 ± 4.16	**0.93 ± 0.08**

5 Discussion

Currently used gait analysis methods are mostly limited to complex laboratories, but the advancement of wearable technologies demand increased support for remote monitoring. Regular assessment of PD symptoms severity can support and benefit all stakeholders, from patients to healthcare professionals, for better treatment management. Early diagnosis and monitoring may benefit individuals who are either developing motor symptoms or transitioning from non-motor to motor symptoms. This approach can also assist individuals who have manifested motor symptoms but have not received clinical diagnosis and clinical supervision.

Table 6. Performance Evaluation of PD Classifier for Different Input Modalities with and without GA Optimization.

Input Modality	Optimization	Hyperparameter (D, C, L)	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)
Features	Without GA	4,6,6	0.968	93.6	100	95.7
	With GA	4,1,2	0.972	94.4	100	96.3
Timeseries	Without GA	4,6,6	0.956	92.8	98.4	94.7
	With GA	4,1,8	0.949	94.4	95.3	94.7
Multimodal Data	**Without GA**	**4,6,6**	**0.980**	**96.0**	**100**	**97.3**
	With GA	4,1,8	0.956	92.8	98.4	94.7

The proposed framework successfully predicted UPDRS scores using the multimodal data, outperforming single model approaches in terms of MAE, RMSE and CC. Additionally, it outperformed previous studies in terms of MAE and CC. This multimodal approach might offer promising solution for combining different sensor modalities that capture the characteristics of motor and non-motor symptoms characteristics influenced by daily activities and treatment interventions. This approach would benefit from mutual support or co-learning from these modalities. However, multimodal machine learning is still a developing field, particularly in the biomedical sector. Special considerations are required for challenges like data linkage between modalities and dealing with noise and missing data. Training a Perceiver model on multimodal data requires extensive computational power due to large size of the data. Complexity and training time increase with larger batch sizes and large number of parameters of the Perceiver architecture. To effectively process multimodal data and reduce training time, a small batch size, optimized hyperparameters tailored to the data, and an advanced GPU are required.

The proposed framework predicted UPDRS scores using the gait timeseries signal, with promising results in terms of MAE, RMSE and CC. This approach could minimize the preprocessing steps required for automatic prediction in free living conditions. In this dataset, we observed higher gait variability in PD patients compared to healthy controls, particularly during dual task activities. Analyzing gait variability directly from the timeseries signal is challenging, as gait characteristics of PD patients and healthy subjects are influenced by several factors. Nonetheless, this framework effectively captures these variabilities to estimate PD symptoms severity.

In this study, we only used GA to optimize the hyperparameters of the first cross-fold due to computational limitations because training the Perceiver model is time-consuming task. These selected hyperparameters may not be sufficient for subsequent folds. In addition, although GA automates the hyperparameter tuning, this approach has limitations. This algorithm introduces additional hyperparameter configuration like population size, generation number, crossover, and mutation rate. In addition, the time complexity of this algorithm is considerably high [16].

In this study, we used a CV strategy but did not reserve any data for testing because of the relatively small sample size. As a result, performance of the model might be biased

towards this dataset. During training, class balancing was maintained until all samples from the minority class were used to ensure balanced learning. Without this approach, the model might have biased or overfitted towards the majority class.

This dataset has a limited population, particularly PD patients with high UPDRS scores. Patients with high severity typically exhibit motor disabilities even during normal walking. They are often excluded from experiments because of the risk of falling during dual-task activities. This exclusion can result in a misrepresentation of the actual PD population, leading to a high bias in the dataset. A complex model such as Perceiver, with its numerous parameters, could potentially reduce this bias and variation with extensive training at the risk of overfitting.

For future direction of this study, we aim to use a large and diverse dataset that would capture free-living gait characteristics from smart shoe and smartwatch, combined with non-motor symptoms collected via smartphone, could enable continuous assessment of PD symptoms.

6 Conclusion

Frequent assessment of PD symptoms may open new opportunities for personalized remote monitoring and effective treatment intervention in free-living environments. Free-living conditions introduce uncertain variables like different activity levels and environmental factors can complicate gait patterns. These complexities, combined with non-motor symptoms and medication effects, make it significantly challenging and impossible for any single sensor. A comprehensive monitoring system, combining gait sensors for motor symptoms and smartphone data for non-motor symptoms can capture the whole spectrum. However, analyzing multimodal data from these diverse sources in free-living conditions is a challenging task due to several factors such as data source integration and potential biases. The Perceiver architecture has shown promising results in combining data from different modalities effectively. Our study demonstrates the combination GRF based gait timeseries signal with extracted features predicted UPDRS scores accurately. This attention-based architecture outperforms previous studies for predicting PD symptoms severities. This architecture has potential to be used as multimodal machine learning framework for decision support solution in personalized treatment management, remote care, and digital twin-based applications.

References

1. Rocca, W.A.: The burden of Parkinson's disease: a worldwide perspective. The Lancet Neurol. **17**, 928–929 (2018). https://doi.org/10.1016/S1474-4422(18)30355-7
2. Feigin, V.L., et al.: GBD 2015 neurological disorders collaborator group: global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the global burden of disease study 2015. The Lancet Neurol. **16**, 877–897 (2017). https://doi.org/10.1016/S1474-4422(17)30299-5
3. Dorsey, E.R., et al.: GBD 2016 Parkinson's disease collaborators: global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the global burden of disease study 2016. The Lancet Neurol. **17**, 939–953 (2018). https://doi.org/10.1016/S1474-4422(18)30295-3

4. Launch of WHO's Parkinson disease technical brief. https://www.who.int/news/item/14-06-2022-launch-of-who-s-parkinson-disease-technical-brief. Accessed 01 Jun 2023
5. Dorsey, E.R., Sherer, T., Okun, M.S., Bloem, B.R.: The emerging evidence of the Parkinson pandemic. JPD. **8**, S3–S8 (2018). https://doi.org/10.3233/JPD-181474
6. Goetz, C.G., Poewe, W., Rascol, O.: Movement disorder society task force on rating scales for Parkinson's disease: the unified Parkinson's disease rating scale (UPDRS): status and recommendations. Mov. Disord. **18**, 738–750 (2003). https://doi.org/10.1002/mds.10473
7. Chen, P.-H., Wang, R.-L., Liou, D.-J., Shaw, J.-S.: Gait disorders in Parkinson's disease: assessment and management. Int. J. Gerontol. **7**, 189–193 (2013). https://doi.org/10.1016/j.ijge.2013.03.005
8. Tong, J., Zhang, J., Dong, E., Du, S.: Severity classification of Parkinson's disease based on permutation-variable importance and persistent entropy. Appl. Sci. **11**, 1834 (2021). https://doi.org/10.3390/app11041834
9. Mirelman, A., et al.: Gait impairments in Parkinson's disease. The Lancet Neurology. **18**, 697–708 (2019). https://doi.org/10.1016/S1474-4422(19)30044-4
10. Papapetropoulos, S., Mitsi, G., Espay, A.J.: Digital health revolution: is it time for affordable remote monitoring for Parkinson's Disease? Front. Neurol. **6**, 126864 (2015). https://doi.org/10.3389/fneur.2015.00034
11. Del Din, S., Godfrey, A., Mazzà, C., Lord, S., Rochester, L.: Free-living monitoring of Parkinson's disease: lessons from the field: wearable technology for Parkinson's disease. Mov. Disord. **31**, 1293–1313 (2016). https://doi.org/10.1002/mds.26718
12. Chandrabhatla, A.S., Jonathan Pomeraniec, I., Ksendzovsky, A.: Co-evolution of machine learning and digital technologies to improve monitoring of Parkinson's disease motor symptoms. NPJ Digit. Med. **5**(1), 32 (2022). https://doi.org/10.1038/s41746-022-00568-y
13. Chau, T.: A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractal methods. Gait Posture **13**, 49–66 (2001). https://doi.org/10.1016/S0966-6362(00)00094-1
14. Acosta, J.N., Falcone, G.J., Rajpurkar, P., Topol, E.J.: Multimodal biomedical AI. Nat. Med. **28**, 1773–1784 (2022). https://doi.org/10.1038/s41591-022-01981-2
15. Jaegle, A., et al.: Perceiver: general perception with iterative attention (2021). http://arxiv.org/abs/2103.03206
16. Yang, L., Shami, A.: On hyperparameter optimization of machine learning algorithms: theory and practice. Neurocomputing **415**, 295–316 (2020). https://doi.org/10.1016/j.neucom.2020.07.061
17. Beyaz, S.: Femoral neck fracture detection in X-ray images using deep learning and genetic algorithm approaches. Jt Dis Relat Surg. **31**, 175–183 (2020). https://doi.org/10.5606/ehc.2020.72163
18. Kilicarslan, S., Celik, M., Sahin, Ş: Hybrid models based on genetic algorithm and deep learning algorithms for nutritional Anemia disease classification. Biomed. Signal Process. Control **63**, 102231 (2021). https://doi.org/10.1016/j.bspc.2020.102231
19. Skaramagkas, V., Pentari, A., Kefalopoulou, Z., Tsiknakis, M.: Multi-modal deep learning diagnosis of Parkinson's disease—a systematic review. IEEE Trans. Neural Syst. Rehabil. Eng. **31**, 2399–2423 (2023). https://doi.org/10.1109/TNSRE.2023.3277749
20. Açıcı, K., Erdaş, Ç.B., Aşuroğlu, T., Toprak, M.K., Erdem, H., Oğul, H.: A random forest method to detect Parkinson's disease via gait analysis. In: Boracchi, G., Iliadis, L., Jayne, C., Likas, A. (eds.) Engineering Applications of Neural Networks, pp. 609–619. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-65172-9_51
21. Daliri, M.R.: Chi-square distance kernel of the gaits for the diagnosis of Parkinson's disease. Biomed. Signal Process. Control **8**, 66–70 (2013). https://doi.org/10.1016/j.bspc.2012.04.007
22. Sigcha, L., et al.: Deep learning and wearable sensors for the diagnosis and monitoring of Parkinson's disease: a systematic review. Expert Syst. Appl. **229**, 120541 (2023). https://doi.org/10.1016/j.eswa.2023.120541

23. Tsanas, A., Little, M., McSharry, P., Ramig, L.: Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. Nat Prec. (2009). https://doi.org/10.1038/npre.2009.3920.1
24. Aşuroğlu, T., Oğul, H.: A deep learning approach for Parkinson's disease severity assessment. Health Technol. **12**, 943–953 (2022). https://doi.org/10.1007/s12553-022-00698-z
25. Aşuroğlu, T., et al.: Parkinson's disease monitoring from gait analysis via foot-worn sensors. Biocybernetics Biomed. Eng. **38**, 760–772 (2018). https://doi.org/10.1016/j.bbe.2018.06.002
26. El Maachi, I., Bilodeau, G.-A., Bouachir, W.: Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait. Expert Syst. Appl. **143**, 113075 (2020). https://doi.org/10.1016/j.eswa.2019.113075
27. Alharthi, A.S., Casson, A.J., Ozanyan, K.B.: Gait spatiotemporal signal analysis for Parkinson's disease detection and severity rating. IEEE Sensors J. **21**, 1838–1848 (2021). https://doi.org/10.1109/JSEN.2020.3018262
28. Pham, T.D.: Time–frequency time–space LSTM for robust classification of physiological signals. Sci. Rep. **11**, 6936 (2021). https://doi.org/10.1038/s41598-021-86432-7
29. Balaji, E., Brindha, D., Elumalai, V.K., Vikrama, R.: Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network. Appl. Soft Comput. **108**, 107463 (2021). https://doi.org/10.1016/j.asoc.2021.107463
30. Vidya, B., Sasikumar, P.: Parkinson's disease diagnosis and stage prediction based on gait signal analysis using EMD and CNN–LSTM network. Eng. Appl. Artif. Intell. **114**, 105099 (2022). https://doi.org/10.1016/j.engappai.2022.105099
31. Nguyen, D.M.D., Miah, M., Bilodeau, G.-A., Bouachir, W.: Transformers for 1D signals in Parkinson's disease detection from gait (2022). http://arxiv.org/abs/2204.00423, https://doi.org/10.48550/arXiv.2204.00423
32. Justa, J., Šmídl, V., Hamáček, A.: Deep learning methods for speed estimation of bipedal motion from wearable IMU sensors. Sensors **22**, 3865 (2022). https://doi.org/10.3390/s22103865
33. Aadam, et al.: EmoPercept: EEG-based emotion classification through perceiver. Soft. Comput. **26**, 10563–10570 (2022). https://doi.org/10.1007/s00500-021-06578-4
34. Koelstra, S., et al.: DEAP: a database for emotion analysis; using physiological signals. IEEE Trans. Affective Comput. **3**, 18–31 (2012). https://doi.org/10.1109/T-AFFC.2011.15
35. Hausdorff, J.M.: Gait in Parkinson's disease (2008). https://physionet.org/content/gaitpdb/, https://doi.org/10.13026/C24H3N
36. Yogev, G., et al.: Dual tasking, gait rhythmicity, and Parkinson's disease: which aspects of gait are attention demanding? Eur. J. Neurosci. **22**, 1248–1256 (2005). https://doi.org/10.1111/j.1460-9568.2005.04298.x
37. Hausdorff, J.M., et al.: Rhythmic auditory stimulation modulates gait variability in Parkinson's disease: effects of RAS on gait variability in PD. Eur. J. Neurosci. **26**, 2369–2375 (2007). https://doi.org/10.1111/j.1460-9568.2007.05810.x
38. Frenkel-Toledo, S., et al.: Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease. Mov. Disord. **20**, 1109–1114 (2005). https://doi.org/10.1002/mds.20507
39. Bae, J., Tomizuka, M.: A tele-monitoring system for gait rehabilitation with an inertial measurement unit and a shoe-type ground reaction force sensor. Mechatronics **23**, 646–651 (2013). https://doi.org/10.1016/j.mechatronics.2013.06.007
40. Campagne, F.: Multi modality perceiver - pytorch (2021). https://github.com/fac2003/perceiver-multi-modality-pytorch
41. Gad, A.F.: PyGAD: An intuitive genetic algorithm python library (2021). http://arxiv.org/abs/2106.06158

Digital Care Pathways II

Dual-Perspective Modeling of Patient Pathways: A Case Study on Kidney Cancer

Anna Grøndahl Larsen[1](✉), Ragnhild Halvorsrud[1], Rolf Eigil Berg[2], and Märt Vesinurm[3]

[1] SINTEF Digital, Department of Sustainable Communication Technologies, Oslo, Norway
anna.g.larsen@sintef.no
[2] Department of Urology, Oslo University Hospital, Oslo, Norway
[3] Department of Industrial Engineering and Management, Aalto University School of Science, Espoo, Finland

Abstract. Patient pathway has become a key concept in the organization of healthcare. However, the materialization and operationalization of pathways often focus on work processes of health personnel, clinical decision-making, and deadlines, contradicting the strong patient-oriented perspective that is inherent in their definition. In this paper, we introduce a patient-centered perspective of kidney cancer pathways, reporting on a dual-perspective strategy to map and model patient pathways. Utilizing a multi-method approach, we map and model pathways from the perspectives of both healthcare personnel and patients and investigate the feasibility of the Customer Journey Modeling Language (CJML) for modeling patient pathways. To prevent confusion, the planned pathway as seen from the hospital perspective and the actual pathway experienced by the patient are referred to as 'pathway' and 'journey', respectively. In the paper, we describe methods to engage with healthcare professionals and patients to collect the necessary information to create precise models, and we show how precise modeling of patient pathways requires the integration of several information sources. Moreover, the study underlines the value of examining pathways from a dual perspective, as the two perspectives corroborate and supplement each other, illustrating the complexity of patient journeys. Finally, the findings provide insights into the feasibility of CJML, firstly underlining that the usefulness of visual models is context-dependent, and secondly, suggesting that the methods and subsequent visualizations may be useful as organizational, instructional, and communicative tools.

Keywords: Patient Pathway · Patient Journey · Customer Journey Modeling Language (CJML) · Feasibility Study

1 Introduction

Patient pathway has become a key concept in the organization of healthcare, intended to address issues pertaining to coherency, seamlessness, and accountability in a system with increasing complexity [1, 2]. Understanding and representing patient pathways is essential in healthcare, as it helps identify potential bottlenecks and informs targeted interventions to improve patient care [3]. Moreover, while patient pathways are mainly used

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 51–68, 2024.
https://doi.org/10.1007/978-3-031-59091-7_4

within hospitals, their use can also strengthen coordination between healthcare actors. A related concept is the patient journey, which takes the perspective of the patient throughout the pathway. Patient journey mapping is increasingly being adopted in healthcare settings to provide insights into the patient experience and to support communication with patients [4].

For simplicity, in the following, we use the term 'patient pathway' to describe pathways as designed by the healthcare system and the term 'patient journey' to describe actual, individual patient encounters throughout an illness period. We, however, maintain a patient-centered perspective for both. While many definitions of patient pathways suggest some level of patient-centered perspective [5], pathways tend to focus on clinical guidelines and other work processes behind the line of visibility of the patient [6]. As patient pathways involve numerous stakeholders and intricate decision-making processes, selecting a modeling approach that effectively captures both patient-provider interactions and the coordination of healthcare services is crucial to ensure the delivery of patient-centered care.

The complexity of healthcare processes, along with the need for effective communication and production planning, has led to a growing interest in using modeling languages to represent patient pathways. Several modeling languages have been used, including Unified Modeling Language [7], Business Process Modeling Language [8], extensions of such languages [9], Customer Journey Modeling Language [10], and a multitude of less formal patient journey maps [4], each with their strengths and weaknesses. The use of modeling languages for documenting patient pathways can reduce variability, help facilitate interdisciplinary collaboration, streamline decision-making processes, and ultimately improve patient outcomes [2, 5].

In this study we focus on the feasibility of the Customer Journey Modeling Language (CJML) for modeling of patient pathways, as planned by the healthcare system, and patient journeys, as experienced by individual patients. Offering a vocabulary, a metamodel, and purpose-specific diagrams, CJML is equipped with journey-specific constructs including touchpoints, actors, channels, phases, and user experiences [11]. CJML's patient-centric approach makes it particularly relevant for modeling patient pathways in healthcare. However, CJML may need adaptation to address the complexity and specific nuances of healthcare processes, as well as the interactions between healthcare institutions and personnel. Hence, there is a need for more in-depth research on CJML in healthcare contexts, including how it may be adapted and improved to suit the specificities of healthcare contexts.

Using kidney cancer as a case study, we explore how patient pathways and journeys can be identified from a dual perspective (both healthcare personnel and patients). We investigate how CJML can be utilized to visualize and compare these pathways and journeys. Additionally, we examine its application in healthcare settings, providing in-depth insights into the pathways and journeys from both perspectives, thereby enabling comparison. In specific, we address the following research questions:

1) How may we identify and make precise models of patient pathways and patient journeys using CJML?
2) What insights do we gain by a dual perspective exploring patient pathways and patient journeys?

3) How feasible are the CJML models in a healthcare setting?

To address these research questions, we present a case study of kidney cancer care conducted by health service researchers and medical doctors at a major university hospital in Norway. Kidney cancer typically develops slowly and is often detected incidentally during investigations for other conditions [12]. The involved hospital treats approximately 100 kidney cancer patients annually. The vast majority of these patients undergo either partial or radical nephrectomy. This study aims to contribute to the ongoing discourse on the application of modeling languages in healthcare and emphasize the importance of patient-centered approaches in representing and understanding patient pathways.

2 Methods

The study adopted a multi-method approach dominated by qualitative research methodology. The first phase of the work was conducted in two parallel streams: 1. Detailed insights into patient pathways (healthcare's perspective) and 2. A longitudinal study of patient journeys (patient's perspective), see Fig. 1. Based on this, we modelled the kidney cancer pathway and patient journey using CJML. In the second phase the feasibility of CJML in healthcare settings was evaluated.

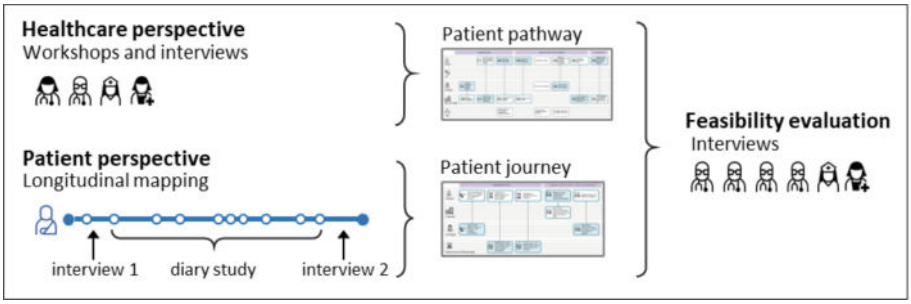


Fig. 1. Principal sketch of the study.

2.1 Recruitment

Healthcare personnel were recruited by one of the authors, while patients were recruited through medical doctor(s) involved in their treatment. A doctor in charge of kidney cancer treatment presented the study to selected patients, inquiring about their interest in participating and obtaining their consent to provide their name and contact information to the research team. If patients consented, the doctor called the research team and provided name and contact information to patients. A researcher then contacted the patient regarding participation in the study. After patients had formally given their informed consent, they were invited to a start interview. Patients were recruited purposefully shortly after they had been diagnosed with kidney cancer, and prior to surgery. Digital literacy was a requirement for participation.

The study meets the requirements of data protection legislation and research ethics and was approved by regional and institutional ethics committees. Participation in the study was voluntary and based on informed consent. To guarantee anonymity, we have used fictitious patient names and altered certain details of the patient journey.

2.2 Mapping and Modeling of Patient Pathways with Healthcare Personnel

The mapping of the patient pathway was conducted through a combination of the following activities: information search, observation, and workshops involving relevant healthcare personnel. Firstly, we carried out a desktop study reviewing all available material on national care pathways for kidney cancer. Second, we had several meetings with healthcare personnel at the Department of Urology, Oslo University Hospital. First, we had two virtual meetings with two doctors in the department, where we described the project. After these two meetings we were invited to observe a multi-disciplinary team meeting (MDT) where 10–15 doctors discuss the patients and decide on treatment. Then we arranged two workshops with two urologists to pinpoint the detailed kidney cancer pathway and had two shorter meetings with medical secretaries at the Department of Urology.

The goal of the workshops with the urologists was to get a detailed description of the patient pathway for kidney cancer, from diagnosis to treatment, from a hospital/department-level perspective. In the first workshop, we asked the urologists to describe the process from diagnosis to treatment. The workshop was taped and later transcribed, and we also took notes during the workshop. Based on these insights, we created an initial sketch of the kidney cancer pathways, highlighting areas that were unclear or required further investigation. This initial sketch served as the foundation for a second workshop with the urologists, aimed at verifying and specifying the description of the kidney cancer patient pathway. Again, we taped the discussion in addition to taking notes. Following this, we revised the pathway description, and subsequently sent it to the two urologists for feedback. In the next step, we used these data and pathway descriptions to model the kidney cancer patient pathway, using CJML. Hence, the analysis has been carried out as a continuous and iterative process, where we have collected data and, based on this, sketched out initial overviews of the pathway, before we have shown these initial overviews to practitioners, collected more data and revised the sketch of the planned pathway.

2.3 Mapping and Modeling of Patient Journeys

Patients were recruited for a longitudinal study that included an initial interview, a diary designed to map the patient journey in detail, and a debrief interview (cf. [13]).

Start Interview: The primary aim of the start interview was to get detailed information of the participant's patient journeys up to that point, including all contacts with healthcare services. The interviews were recorded and later transcribed, and all touchpoints were logged in a spreadsheet.

Diary: The primary goal of the diary was to collect the patient's touchpoints in real-time as they occurred. There are several advantages of using diaries: they enable continuous updates and facilitate longitudinal studies without too much intrusion in

patients' daily lives, they reduce the time between an occurrence and the account given of it, and they capture a level of detail that is difficult to achieve through interviews alone (cf. [14, 15]). Patients were presented to the diary and instructed in how to fil it in at the end of the start interview. They were asked to make an entry in the diary every time i) they had contact with healthcare services, not only consultations but also confirmations and reminders of appointments, and ii) they undertook an activity relevant to their illness, such as for example information retrieval. Moreover, for each diary entry they were requested to fill in date and time, who they had been in contact with and the communication channel used (e.g. SMS or face-to-face), provide a detailed account of the event and their experience of it, and rate their overall experience of each entry on a scale from 1–5. In practice, the diary was a Word document that included instructions and a table with four columns—date/time, actor/channel, description of touchpoint, and rating—and unlimited rows, where each row represented one touchpoint. To secure as detailed diaries as possible, patients received reminders to fill in their diary approximately every two weeks and were asked to regularly email us their diary with new entries. The responsible researcher continuously logged these diary entries into a spreadsheet and noted any questions, such as on missing information that needed clarification during the debrief interview. Thus, the analysis started during data collection. After patients had logged their activities in the diary for 2–4 months, we scheduled a debrief interview.

Debrief Interview: The primary goal of the debrief interview was to review the diary entries to ensure that all relevant information was included. During the debrief interview, the interviewer went through the touchpoints collected in the start interview and the diary and requested additional information where necessary. After finalizing and transcribing the debrief interview, the spreadsheet with touchpoints was updated. This updated spreadsheet was subsequently used as the starting point for modeling the patient journey.

Four kidney cancer patients were part of the study. However, in modeling the patient journey further below, we focus on the patient journey of one patient, from initial symptoms through surgery to scheduling of semi-annual checks, covering a period of five months. The patient was recruited to the study about two months after the initial symptoms and participated in the study for app. three months. Hence, the first phase of the patient journey, from initial symptoms until surgery, is documented in retrospect, through interviews and retrospective diary entries.

2.4 Feasibility Evaluation

In exploring the application of CJML to model patient pathways and their utility in healthcare settings, we interviewed four urologists (P1-P4), a cancer coordinator (P5), and a patient coordinator at the hospital admissions office (P6). We asked them to reflect on and evaluate the feasibility of our diagrams of the patient pathway (Fig. 2) and patient journey (Fig. 3). Participants had the opportunity to review the diagrams both before and during the interview. They were asked about their initial impressions of the diagrams, the extent to which they found the diagrams useful and comprehensible, and to reflect on how such visualizations may be used. Moreover, participants were asked whether important information was missing in the visualization of the kidney cancer patient pathway. Finally, at the end of the interview participants were asked to rate the

ease-of-use and the usefulness of the diagrams on a scale from 1 to 5, with the numbers signifying the following: 1 to a very small extent, 2 to a small extent, 3 neither/nor, 4 to a large extent, 5 to a very large extent.

The interviews were carried out on Microsoft Teams in September and October 2023, recorded, and automatically transcribed. The interviewer also took notes during the interviews. Immediately following each interview, key information, including quotations from the audio file, was entered into a spreadsheet. The spreadsheet included responses to all interview questions. This provided an overview and facilitated comparison between participants. Based on this, we began drafting initial findings, which were later discussed and refined by the authors through an iterative process, that included revisiting audio recordings, and re-examining the spreadsheet and transcripts.

3 Results

In this section, the results of our study are presented. First, the kidney cancer patient pathway, as seen from the hospital perspective, is described and visualized. Then, an actual patient journey, as seen from the patient perspective, is presented, before the two perspectives are compared. Finally, findings from the feasibility interviews are presented, exploring the use of CJML-based visualizations in healthcare settings.

3.1 Modeling of the Kidney Cancer Pathway (Healthcare Perspective)

Kidney cancer pathways vary between individuals and involve a large number of actors and touchpoints. In modeling the patient pathway, we have therefore made the following assumptions: We model the patient pathway of patients that seek medical care due to hematuria and exclude patients that are diagnosed amid investigations for other conditions. Moreover, we focus on the pathway for patients that undergo surgical treatment without post-surgery complications. All contacts and communication the patient has with the hospital is documented in the Electronic Health Record (EHR). To increase the readability of the diagram, we however only include EHR at selected points. Also, reminders of appointments are excluded. A model of the kidney cancer pathway using CJML is presented in Fig. 2.

Kidney Cancer Pathways. Standardized cancer patient pathways (CPPs) have been introduced in some countries with a guaranteed timeframe for timely diagnosis and treatment. From the hospital perspective, the planned pathway typically begins when they receive a referral from a general practitioner, although sometimes the patient is referred from another hospital. The patient has usually undergone a CT scan that revealed a renal tumor. A urologist at the Department of Urology then assesses the referral, and formally confirms the start of the kidney cancer patient pathway. Cancer pathway coordinators are responsible for arranging the appointments and acting as the patient's contact person. The patient is contacted via electronic mail (through the patient health portal), letter, or telephone.

A weekly multi-disciplinary team meeting (MDT) is held among healthcare professionals to discuss patient cases, aiming to provide the most comprehensive care possible, at the right place and time for each patient. A MDT Coordinator ensures that all patients

with a new cancer diagnosis are discussed, and their scans and biopsies are reviewed by the team. Leveraging the combined expertise of each team member and considering the specific needs of each patient, the MDT recommends a treatment plan. This plan is documented and discussed with the patient via a telephone call or a follow-up appointment. The outcomes of the MDT meeting can include surgery, biopsy, and active monitoring. In the following, we will focus on the planned pathway for patients undergoing surgical treatment.

Approximately one week prior to the scheduled surgery, the patient attends a pre-surgical assessment appointment with a surgical intern who evaluates the patient's health. The intern conducts a focused physical exam to ensure there are no medical risks that could predispose the patient to a medical emergency during the planned procedure. During the pre-surgical assessment, the patient also meets with a nurse. Finally, an anesthesiologist conducts a health assessment before surgery to gather information about any medical conditions the patient may have, their medications, and any previous experience with anesthesia.

The patient is admitted to the hospital either on the day of surgery or the day before. Upon arrival, they are greeted by a nurse who explains the processes and provides the patient with an identity bracelet to wear throughout their hospital stay. Additionally, the patient will meet the operating surgeon. After the surgery is completed, the patient is transferred to the recovery unit where they are closely monitored by a perioperative nurse. The patient meets with the surgeon to be informed of the outcome of the operation. Subsequently, the patient is moved to an inpatient room for postoperative care.

Patients are usually dismissed from the hospital 2–7 days after surgery. Upon discharge, they receive post-procedure instructions. Approximately 4 weeks after surgery, patients attend a scheduled appointment at the post-treatment clinic, where they are informed about the surgical pathology report and plans for further follow-up.

3.2 Modeling the Kidney Cancer Patient Journey (The Patient Perspective)

Here, we describe and model one of the actual patient journeys that was captured in the longitudinal study. This patient was diagnosed with kidney cancer following acute illness while travelling, which was unrelated to the cancer. In total, the patient journey included more than sixty touchpoints, necessitating several simplifications. Firstly, the diagram begins with the patient's initial contact with the hospital responsible for assessing and treating the cancer, thus omitting the early part of the patient journey as described below. Additionally, the following touchpoints have been omitted for simplicity: For each new hospital appointment, the patient was notified via SMS and through the patient health portal. Furthermore, 1 to 2 days before an appointment, the patient received SMS reminders. The resulting model of the patient journey is shown in Fig. 3.

Actual Kidney Cancer Patient Journey. Holger is in his fifties, works full-time, and lives a busy life. His patient journey began when he experienced severe abdominal pain while travelling. He visited a local emergency care unit and was subsequently referred to the nearest hospital, where he was diagnosed with an incarcerated hernia and underwent surgery. Simultaneously, a CT scan revealed a tumor in his kidney. Consequently, the

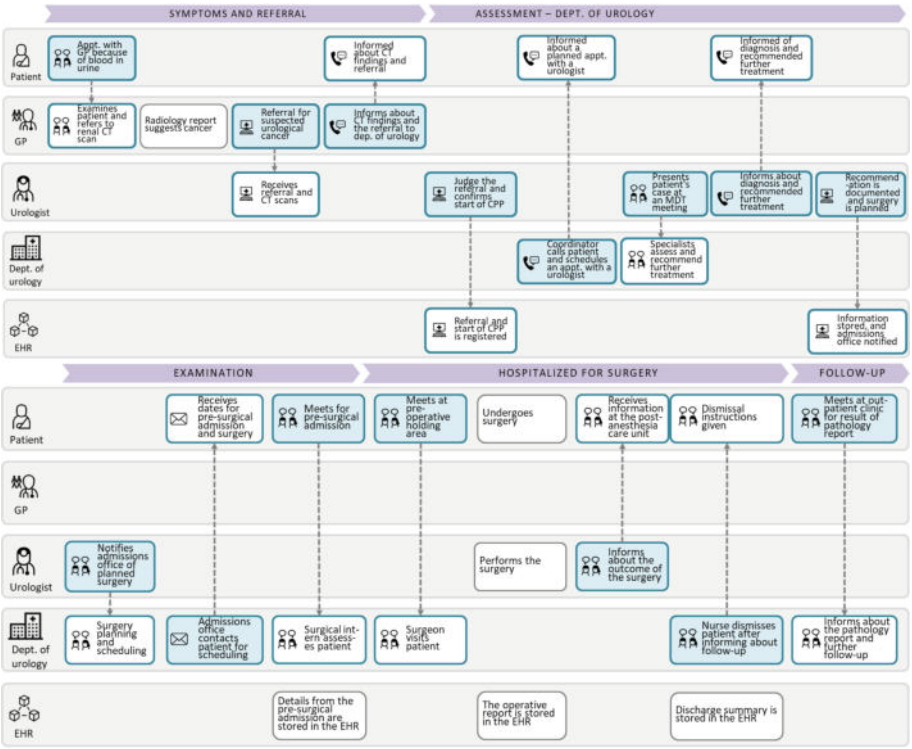


Fig. 2. Patient pathway—kidney cancer.

hospital referred the patient to his hometown hospital for follow-up and further evaluation. Approximately three weeks later, he was notified of a follow-up appointment at the Department of Urology at his hometown hospital. However, one day before the scheduled appointment, Holger received a phone call from the hospital stating that they had not yet received the CT images, necessitating a postponement of the appointment. After some back and forth, three days later, Holger received another phone call from the hospital, this time to schedule him for new CT images the following day. Consequently, Holger visited the hospital the next day for the CT scans and blood tests. Three days after this visit, a doctor from the hospital called him to discuss the need for a renal biopsy. Two days later, he returned to the hospital to undergo the biopsy.

Shortly after the biopsy, Holger was notified about a scheduled phone appointment with a urologist for the following day. However, the urologist called later than scheduled, and since Holger was unavailable, they postponed the call until to next day. During the rescheduled call, the urologist informed Holger that the tumor was malignant and recommended its removal through endoscopic surgery. Due to work obligations, Holger requested that the surgery be scheduled no sooner than four weeks later, which the urologist agreed to. A few days after the phone consultation, Holger received notifications about preoperative assessment and surgery appointments. Three weeks later, he attended

the preoperative assessment, where he met with a nurse and an anesthesiologist and had the necessary blood tests done.

A week and a half after the preoperative assessment, Holger was admitted to the hospital by a nurse and then had a consultation with the operating surgeon, who briefed him on the procedure. Shortly thereafter, he was escorted to the operating theatre by a nurse, administered anesthesia, and underwent the surgery. Hours later, he awoke and was moved to the ward, where he received continuous follow-up care from nurses. The day following the surgery, Holger had a consultation with a doctor (previously unknown to him). Throughout the day, he was under continuous care by nurses, including pain relief. The next day, Holger had a follow-up meeting with the operating surgeon, who briefed him on the surgery's outcome, further follow-up plans, and his dismissal from the hospital. Later that day, Holger was dismissed from the hospital and did not experience any further complications.

The day after he was dismissed, Holger was notified about a follow-up appointment at the hospital scheduled for one month later. Approximately ten days post-surgery he received an unexpected phone call from the operating surgeon, who informed him that he was now considered cancer-free. This was confirmed during the scheduled hospital appointment a few weeks later, after which Holger was put on a schedule for half-yearly CT-scans as a routine post-surgery check-up.

In total, Holger logged $n = 38$ touchpoints in the diary. For a majority of the touchpoints ($n = 27$), he rated his experience as 'very satisfied'. In an additional seven cases, he rated 'satisfied' or 'neutral', all of which related to the issue of transferring CT images between hospitals.

3.3 The Dual Perspective

The data and analysis presented above show that the two approaches corroborate and supplement each other. As illustrated above and summarized in Table 1 below, the patient pathway and the patient journey overlap in important respects, hence substantiating each other. For example, the actual patient journey includes the key steps presented in the visualization of the patient pathway (Fig. 2), such as being informed about diagnosis and recommended further treatment, receiving dates for pre-surgical admission and surgery, meeting for pre-surgical admission, being informed about the outcome of the surgery and meeting at outpatient clinic for results of the pathology report. Simultaneously, the planned pathway and the actual journey also differ. For example, in the actual patient journey, kidney cancer is found by coincidence and the patient is referred from a specialist rather from a GP, illustrating the diversity in patient journeys. Moreover, the patient journey includes touchpoints that are not presented in the patient pathway, such as CT images being re-taken, biopsy appointment, and the patient receiving an unscheduled phone call by the surgeon post-surgery, illustrating the complexity in actual patient journeys. This complexity is illustrated by Fig. 3 and further underlined when considering that the figure is a simplified version of the patient journey.

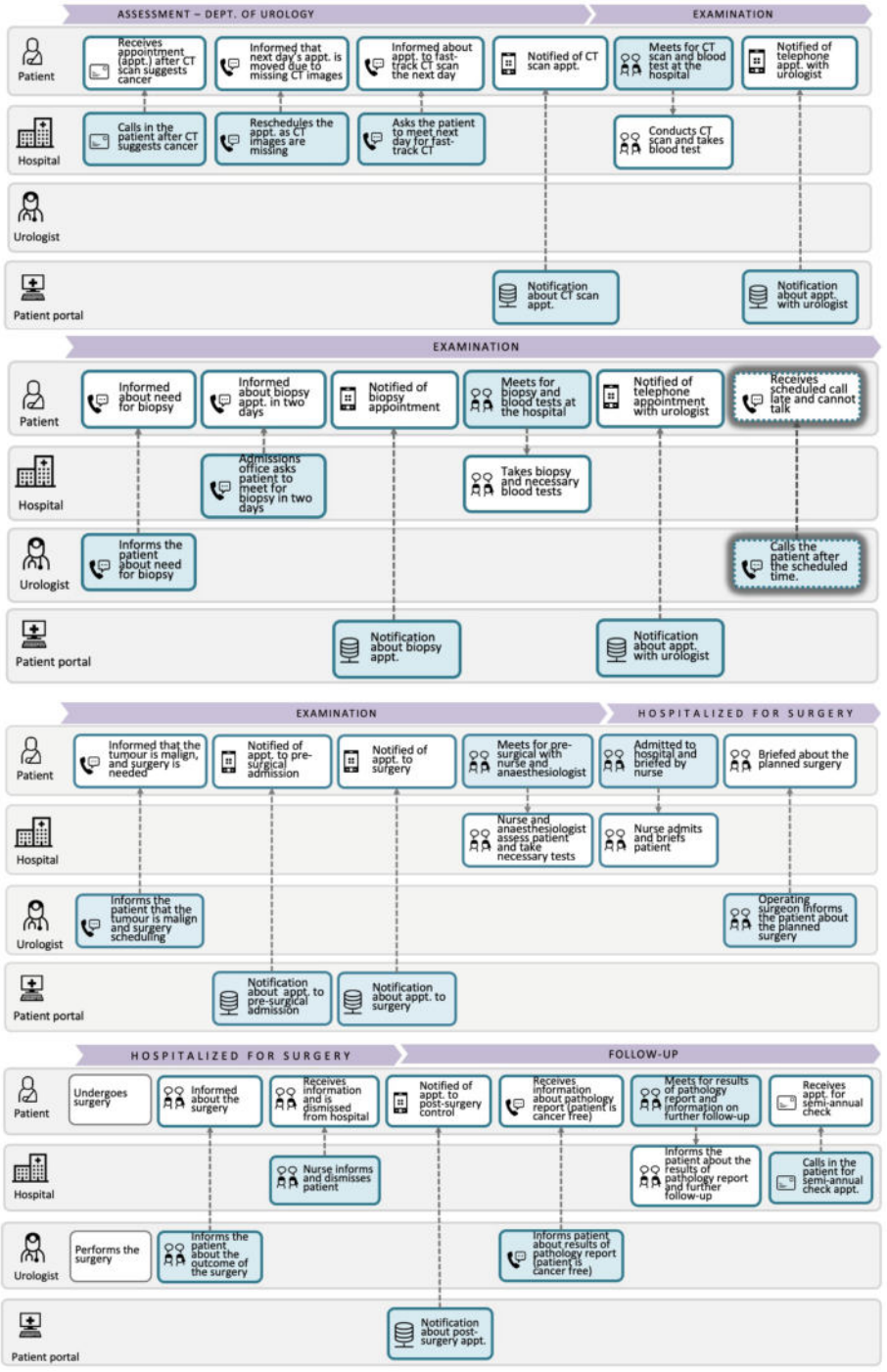


Fig. 3. Patient journey of a kidney cancer patient.

Table 1. Comparison of patient pathway and actual patient journey

Touchpoints that overlap	Additional touchpoints in patient journey	Additional touchpoints in patient pathway
- Notification of appointment with urologist - Phone appointment with urologist - Notification of appointment to pre-surgical admission and surgery - Pre-surgical admission - Surgery admission - Surgery - Information about outcome of surgery - Dismissal instructions - Meeting for result of pathology report	- CT images re-taken - Biopsy appointment - Receives scheduled phone call late, rescheduled - Hospitalization includes several touchpoints - Unscheduled phone call - Touchpoints including healthcare professionals beyond hospital dept	- Referral - Start of CPP - MDT meeting - Documentation and surgery planning

3.4 Feasibility Evaluation

We asked four urologists (P1–P4) and two coordinators (P5–P6) to assess the feasibility of the visualizations of the patient pathway (Fig. 2) and patient journey (Fig. 3). The findings show that the participants generally do not use diagrams or other visualizations of patient pathways as part of their work, although some have prior experience with similar ways of representing pathways. Hence, the findings indicate that visual representations of patient pathways are currently not much used in hospital settings.

Despite having little previous experience with visualizations of pathways, participants generally found the diagrams to be intuitive and easy to understand. Being asked to rate the ease-of-use of the diagrams on a scale from 1 to 5, the majority gave the rating 4 (two participants) and 5 (three participants), while one participant gave the rating 2 noting that he found the diagrams to contain too many boxes and too much information, making them complicated to read.

In terms of usefulness, participants did not consider the visual representations to be useful in their day-to-day work, noting that as experienced doctors and coordinators they are already familiar with the key points in the kidney cancer patient pathways, and in their daily professional lives they focus on the steps that they themselves oversee, such as surgery. A coordinator for example said that *"To me there is nothing new in this diagram. […] The diagram gives a good overview, but I cannot see that it would be useful for me in my work"* (P5). However, the participants pointed to how such diagrams may indeed be useful as organizational, instructional, managerial, and communicative tools. The participants in particular noted potential usefulness related to the following: to map current practices across roles and departments; to arrive at a common understanding across roles; to identify bottlenecks and increase efficiency; to communicate efficiency potentials to for example healthcare professionals, hospital managers, or health authorities;

to standardize and secure equal treatment to patients, and in training for new employees and in educational settings (see Table 2, below, for an overview). A urologist for example noted that:

> It may contribute to improve efficiency, coordination. You see that there are steps that could be slimmed down […] We complain all the time that there are large queues, but I also believe it is inefficient operations (P4).

Another urologist who occasionally had been part of a similar activity where everyone involved in a specific treatment wrote down steps and responsibilities on sticky notes, noted that:

> I think it is very useful if the right people sit together and together figure out how to do something about the pathway. [...] It is useful if all involved are meant to see the same problems. That is also why we do [the sticky note activity], so that everyone involved can agree on an idea, how something is to be done, and that all perspectives are taken into consideration (P3).

Hence, the findings point to how usefulness would depend on the context, the purpose, and the role and seniority of those using it. Granted this context dependency, participants generally found it challenging to rate the usefulness of the visualizations. For example, some gave the rating 1, relating usefulness to their day-to-day work, while others rated 4 or 5, relating usefulness to broader organizational, instructional, or communicative issues. One of the urologists pointed this out in noting that usefulness would depend both on the context and the motives for using it:

> If, for example, I was to present something for The Ministry of Health to point out that we should do so and so, this would be the way to do it. So, I would rate it anywhere between 1 to 5 (P3).

This also relates to the level of detail in visualizations, where the participants noted how the appropriate level of detail would depend on the purpose and the recipient. Some for example noted that if the visualizations were intended to quickly communicate the key steps in a kidney cancer patient pathway, for example to patients or other groups with limited previous knowledge, the current visualizations may be too detailed and complicated to read. Similarly, others noted that if visualizations were intended to communicate the complexity of patient journeys in educational settings, or to map out current practices, a high level of detail might be beneficial.

Being asked about whether such diagrams could be useful to show patients, the participants were generally skeptical, noting that individual patient journeys vary, diagrams of the patient pathway may feel overwhelming, leave patients with more questions than answers, and create confusion if the actual patient journey deviate from what is visualized in the diagram. One of the urologists said that "*No, I think it would confuse more than being beneficial*" (P4). Another one noted that: *"I am not so sure about that, because today patients are bombarded with information, they drown in information, and they are not able to deal with all the information"* (P1). Some of the participants, however said that very simplified diagrams with key touchpoints might be useful for patients.

Table 2. Areas of use for visualizations of patient pathways.

Organizational/managerial	Instructional	Communicative
- Map practices across roles, departments - Arrive at common understanding - Identify bottlenecks and increase efficiency - Standardize and secure equal treatment to patients	- In training for new employees - In educational settings	- To communicate efficiency potentials, e.g., to hospital managers, health authorities, healthcare professionals

4 Discussion

In this study, we utilized CJML to model a kidney cancer patient pathway from observations, workshops, and interviews with healthcare personnel, as well as a detailed patient journey of one patient from interviews and a longitudinal diary study. Moreover, we collected feedback from urologists and coordinators on the feasibility of using CJML modeled pathways in practical healthcare settings. Below, we discuss the key findings related to our three research questions. First, we discuss the effectiveness of our methods for precisely identifying and modeling patient pathways and journeys. Second, we discuss the insights provided by applying a dual perspective. Third, we discuss the feasibility of CJML models in healthcare settings.

Mapping the Patient Perspective. The longitudinal study offered a detailed view of the patient journey, capturing many of the touchpoints that patients experience. We find that the diary method, where patients log each new touchpoint, is well-suited as a basis for modeling patient journeys. However, the diaries alone are not sufficient to provide a full picture of the patient journey, as patients tend to forget to log some touchpoints, such as appointment notifications or phone calls with healthcare personnel. Therefore, interviews are essential to supplement and clarify diary entries, gather patients' assessment of touchpoints, uncover missed entries, and hence, to provide a fuller picture of the patient journey. For the patient journey detailed in this paper, about eighty percent of the touchpoints were logged in the diary, with the remaining twenty percent identified during interviews. This is consistent with previous findings, which suggest that informants typically report between 50–70% of touchpoints [16]. The diary method enriches the toolkit for collecting patient journey data, complementing focus groups (i.e., [17, 18]), surveys [19, 20], document analysis [21], and interviews (i.e., [22, 23]), by facilitating more detailed patient journey maps. Of the data collection methods available, the diary method ranks among the most detailed. However, the challenge with the method is that it is resource intensive, and there is a risk that participants drop out during data collection. Despite these challenges, our study reaffirms the method's suitability for accurately modeling patient journeys with CJML, thus providing healthcare personnel with an overview of what a patient journey may look like from the patient perspective, which they today largely seem to lack.

Relatedly, the method provides insights into patients' experiences with each touchpoint, which can be instrumental in identifying bottlenecks and improving healthcare services overall. An example of the metrics and presentation of patient experience can be found in [13]. By gaining knowledge on patient experiences throughout the patient journey and in relation to specific touchpoints, we obtain a richer, more detailed, and more dynamic understanding of patient experience as it evolves throughout a patient journey, compared to the snapshot provided by measuring patient experience at a single point during or after an illness period. Therefore, our method may be a valuable supplement to patient-reported experience measures (PREMs) [24]. The specificity of the measures entails that they capture particular challenges that patients face, making them well-suited as basis for improving healthcare services. Moreover, by offering insights into experiences over time, researchers can more accurately capture how the overall patient experience is shaped, and which factors that are most important to patients.

Mapping the Patient Pathway. Our study highlights the necessity of a iterative process to accurately map the patient pathway. This includes recurring workshops with healthcare personnel for data gathering and drafting initial pathway sketches, followed by meetings to review initial visualizations, where healthcare personnel provide feedback on the pathway's accuracy and completeness. Patient pathways are complex, and carrying out data collection in steps ensures that the data and subsequent visualizations are as comprehensive and accurate as possible. This approach aligns with current literature on the development of pathways [20].

The Dual Perspective. Our findings show that the dual perspectives corroborate and supplement each other. On one hand, the patient pathway and journey overlap in important respects, hence the two perspectives substantiate each other. On the other, their differences highlight the diversity and complexity of patient pathways, illustrating the importance of visualizations to provide an overview. The dual perspective yields insights unattainable through studying the pathway or journey in isolation. Studying patient journeys reveals a detailed picture of all the touchpoints that a patient indeed encounters. Moreover, we get insights into touchpoints across and beyond healthcare institutions and personnel throughout an illness period, and insights on patients' experiences with healthcare personnel, communication, and coordination. Conversely, studying patient pathways from the healthcare side provides visibility into aspects that are hidden from patients, such as electronic health records, internal procedures, meetings, organizational issues, and decision-making.

Feasibility. Our findings underline how the usefulness of CJML visualizations depends on the context. The results from interviews suggest that while healthcare personnel may not find such diagrams particularly useful in their everyday work, such models may be useful as organizational, instructional, managerial, and communicative tools, such as to map practices across roles, arrive at a common understanding, identify efficiency potentials, and communicate these to stakeholders, as well as in training and educational settings. Relatedly, the findings illustrate that the appropriate *level of detail* in visualizing patient pathways depends on the purpose and the intended recipient. For example, the appropriate level of detail is likely higher when representations are meant to provide an overview and identify bottlenecks within or across departments than when

visualizations, for example, are meant to communicate such bottlenecks and improvement potentials to hospital managers or health authorities. Moreover, if used in training or for educational purposes, very detailed visualizations may help healthcare personnel or students understand the complexity of patient pathways. Less detailed visualizations may be better suited if the goal is to provide an overview of the key steps in a planned patient pathway at a specific department or hospital. Hence, representations of pathways and journeys should have a clearly stated aim, with visualizations being adapted to the aim and the recipient. However, to get an overview of and be able to make relevant representations in different contexts, a detailed mapping of patient pathways would nevertheless be a prerequisite. In other words, all touchpoints may be important, but their importance in visual representations will depend on the context and purpose. This indicate a need to further formalize the modeling language so that, for example, touchpoints representing reminders may easily be omitted from visualizations.

Theoretical and Practical Contributions. Our results contribute to theory by highlighting the need for duality in healthcare systems modeling. Modern healthcare production systems are often designed largely from the production perspective [25], and although the importance of the patient perspective is often underlined, its role in the design of the service production system is often still small. Our results highlight the critical role of the patient in the design of healthcare systems, especially in long-term care processes. Considering a person with kidney cancer lives with the disease 24/7, how can we design a functional care system purely from the perspective of the actor who meets the person for a few hours every year? Practically, we contribute a concrete tool and example of how such a dual pathway model can be built and used to inform decision-making in patient processes. CJML is an openly available modeling tool with support resources available for healthcare professionals to make use of in their process improvement efforts.

5 Conclusion

In this paper we have examined the identification and modeling of patient pathways and journeys, the insights gained from applying a dual perspective, and the feasibility of CJML-based models in healthcare. The methodological approach, and the combined exploration of the healthcare- and patient perspective highlight the complexity of patient pathways and patient journeys, illustrating that visualizations can provide a valuable overview of this complexity. Although healthcare professionals may not find these visualizations crucial for daily tasks, our findings suggest that the methods and subsequent visualizations may be useful as organizational, instructional, managerial, and communicative tools.

The focus on one illness, a few patients, and a single hospital department allowed for an in-depth analysis of patient pathways and journeys. This approach has allowed us to explore how pathways and journeys can be precisely captured and modelled, how they compare, and the feasibility of CJML models in healthcare settings. However, the focus on a single illness from the perspective of a relatively limited number of healthcare professionals and patients is also a limitation. The insights provided in this paper should be explored further and tested in other contexts, with other illnesses, and

on larger samples. We maintain that our findings related to identifying and modeling patient pathways/journeys, and the feasibility of using these models in healthcare settings have value beyond kidney cancer care. Nevertheless, kidney cancer patient pathways may be relatively short and standardized compared to other illnesses, such as chronic diseases or other types of cancer, indicating a need for further studies focusing on other illnesses. In terms of feasibility, additional studies are needed to better understand the specific contexts in which visualizations may be useful and the appropriate level of detail for presenting patient pathways/journeys to different actors, such as patients, hospital administration, nurses, doctors, or even students. Relatedly, since the findings indicate a need for filtering touchpoints and providing visualizations at different levels of abstraction, future studies should further explore how the formalism of CJML can be extended to more effectively support the healthcare context.

Acknowledgments. We would like to thank the informants for their valuable feedback. Additionally, we are grateful to Per Magnus Mæhle from the Division of Cancer Medicine at the Oslo University Hospital for initially connecting the project with The Department of Urology. This work was funded by the Research Council of Norway under the Pathway project (no. 316342).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Vanhaecht, K.: The impact of clinical pathways on the organisation of care processes. Ph.D. Dissertation. University of Leuven (2007)
2. Campbell, H., Hotchkiss, R., Bradshaw, N., Porteous, M.: Integrated care pathways. BMJ **316**(7125), 133–137 (1998). https://doi.org/10.1136/bmj.316.7125.133
3. Gartner, J.B., Abasse, K.S., Bergeron, F., Landa, P., Lemaire, C., Côté, A.: Definition and conceptualization of the patient-centered care pathway, a proposed integrative framework for consensus: a concept analysis and systematic review. BMC Health Serv. Res. **22**(1), 1–24 (2022)
4. Davies, E.L., et al.: Reporting and conducting patient journey mapping research in healthcare: a scoping review. J. Adv. Nurs. **79**(1), 83–100 (2023)
5. Aspland, E., Gartner, D., Harper, P.: Clinical pathway modelling: a literature review. Health Syst. **10**(1), 1–23 (2021). https://doi.org/10.1080/20476965.2019.1652547
6. De Bleser, L., Depreitere, R., Waele, K.D., Vanhaecht, K., Vlayen, J., Sermeus, W.: Defining pathways. J. Nurs. Manag. **14**(7), 553–563 (2006). https://doi.org/10.1111/j.1365-2934.2006.00702.x
7. White, S. A.: Introduction to BPMN. IBM Cooperation (2004)
8. Dijkman, R., Hofstetter, J., Koehler, J. (eds.): Business Process Model and Notation, vol. 89. Springer, Heidelberg (2011)
9. Braun, R., Schlieter, H., Burwit, M., Esswein, W.: BPMN4CP revised – extending BPMN for multi-perspective modeling of clinical pathways. In: 49th Hawaii International Conference on System Sciences (HICSS), Koloa, pp. 3249–3258 (2016). https://doi.org/10.1109/HICSS.2016.407

10. Halvorsrud, R., Lillegard, AL., Røhne, M., Jensen, AM.: Managing complex patient journeys in health care. In: Pfannstein, M.A.., Rasche, C. (eds.). Service Design and Service Thinking in Healthcare and Hospital Management, pp. 329–346. Springer (2018). ISBN 978-3-030-00748-5
11. Halvorsrud, R., Sanchez, O.R., Boletsis, C.: Involving users in the development of a modeling langiage for customer journeys. Softw. Syst. Model. **22**, 1589–1618 (2023). https://doi.org/10.1007/s10270-023-01081-w)
12. Capitano, U., Montorsi, F.: Renal cancer. The Lancet **37**(10023), 894–906 (2016). https://doi.org/10.1016/S0140-6736(15)00046-X
13. Halvorsrud, R., Røhne, M., Celius, Moen, S.E.G., Strisland, F.: Application of patient journey methodology to explore needs for digital support. a multiple sclerosis case study. In: Proceedings of the 17th Scandinavian Conference on Health Informatics, Oslo (2019)
14. Mendoza, J., et al.: Strengths and weaknesses of digital diaries as a means to study patient pathways: experiences with a study of hypertension in the Philippines. Int. J. Qualit. Methods **21**, 1–11 (2021). https://doi.org/10.1177/16094069211002746
15. Thomas, J.A.: Using unstructured diaries for primary data collection. Nurse Res. **22**(5), 25–29 (2015). https://doi.org/10.7748/nr.22.5.25.e1322
16. Halvorsrud, R., Kvale, K., Følstad, A.: Improving service quality through customer journey analysis. J. Serv. Theory Pract. **26**(6), 840–867 (2016). https://doi.org/10.1108/JSTP-05-2015-0111
17. Schalkwij, V., et al.: ShareHeart: a patient journey map of patients with ischemia and non-obstructive coronary artery disease based on qualitative research. J. Clin. Nurs. **32**(13–14), 3434–3444 (2023). https://doi.org/10.1111/jocn.16409
18. Benson, M., et al.: Development of a patient journey map for people living with cervical dystonia. Orphanet J. Rare Dis. **17**(130), 1–9 (2022). https://doi.org/10.1186/s13023-022-02270-4
19. Philpot, L., et al.: Creation of a patient-centered journey map to improve the patient experience: a mixed methods approach. Mayo Clinic Proceedings: Innovations, Quality & Outcomes **3**(4), 466–475 (2019). https://doi.org/10.1016/j.mayocpiqo.2019.07.004
20. He, Q., Du, F., Simonse, L.W.: A patient journey map to improve the home isolation experience of persons with mild COVID-19: design research for service touchpoints of artificial intelligence in eHealth. JMIR Med. Inform. **9**(4), e23238 (2021). https://doi.org/10.2196/23238
21. McCarthy, S., O'Raghallaigh, P., Woodworth, S., Lim, Y.Y., Kenny, L.C., Adam, F.: Embedding the pillars of quality in health information technology solutions using "integrated patient journey mapping" (IPJM): case study. JMIR Hum. Factors **7**(3), e17416 (2020). https://doi.org/10.2196/17416
22. Ciria-Suarez, L., et al.: Breast cancer patient experiences through a journey map: a qualitative study. PloS One **16**(9), e0257680 (2021)
23. Røsstad, T., Garåsen, H., Steinsbekk, A., Sletvold, O., Grimsmo, A.: Development of a patient-centred care pathway across healthcare providers: a qualitative study. BMC Health Serv. Res. **13**(1), 1–9 (2013)
24. Bull, C., Byrnes, J., Hettiarachchi, R., Downes, M.: A systematic review of the validity and reliability of patient-reported experience measures. Health Serv. Res. **54**(5), 1023–1035 (2019)
25. Vissers, J., Elkhuizen, S., Proudlove, N.: Operations Management for Healthcare. Routledge, London (2022)

Digital Services in the Welfare, Social and Health Sector Organizations of the South Ostrobothnia Region

Merja Hoffrén-Mikkola(✉)

Seinäjoki University of Applied Sciences, Seinäjoki, Finland
merja.hoffren-mikkola@seamk.fi

Abstract. Digital services in healthcare and social services have increased due to national promotion and Covid19 pandemic. However, the regional differences may exist. Successful implementation and sustainability of digital services requires that attention is paid to addressing barriers and supporting facilitators at all levels in health care provision.

The purpose of this study was to investigate the effects of employee status, form of organization and organizational size on the views related to current state and the role of digital services, development barriers, development plans and the support needed for development in welfare, social and health service organizations operating in the South Ostrobothnia region. The study was carried out in the era of exceptional circumstances created by the Covid19 pandemic in the summer of 2021. The study was a quantitative cross-sectional study using an electronic survey. Respondents (n = 121) were managers, entrepreneurs and employees of welfare, social and health service organizations operating in the South Ostrobothnia region.

The results suggested that in more than four out of five welfare, social and health service organizations operating in the region of South Ostrobothnia, part of the services were already digital in the summer of 2021. These services had been extensively developed during the previous year, which was lived in exceptional circumstances caused by the Covid19 pandemic. Digital services were seen to function especially as enablers of customers in exceptional circumstances. However, managers or entrepreneurs also saw digital services as reaching new customers more important than employees. The acquisition of technology and human resources were felt to be the most significant barriers in the development of digital services, regardless of the employee status, form of organization and organization size. Regarding the use and development of digital services, information was felt to be necessary, especially about the characteristics of digital services, and financial support was also felt to be necessary for the development. However, the support needs were significant in many aspects related to digital service development. In particular, large organizations needed information on the cost-effectiveness of digital services.

The results can be used to support welfare, social and health service organizations in digital service development.

Keywords: digital services · health care · social services

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 69–82, 2024.
https://doi.org/10.1007/978-3-031-59091-7_5

1 Introduction

The digitalization of the social and health sector has been promoted in a targeted manner for a long time. The reform of the organization of health, social and rescue services carried out during Sanna Marin's term as prime minister (2019–2023) included the Future Health and Social Services Center program, where measures related to digitalization were a significant part of the realization of the goals [1]. These goals were aimed at improving the equal availability, timeliness and continuity of services, shifting the emphasis of operations to preventive and proactive work, ensuring the quality and effectiveness of services, and strengthening the multidisciplinary and interoperability of services.

Covid19 pandemic increased rapidly the use of digital services in social and health care sector [2, 3]. Already half a year after the start of the restrictions caused by the pandemic, electronic contact withdrawals had increased in services that were already in use before the pandemic [2]. In addition, already then also new digital services were developed. For example, customer and patient meetings were arranged more online and group activities were implemented remotely. However, the report by Jormanainen et al. (2020) [2] targeted to describe the development of digital services in the area of a large hospital district (HUS), which does not give full details about the changes in the service structure in other parts of the country, such as in South Ostrobothnia. Region specific differences may appear in Finland in the development of digital services in social services and health care sectors [3].

In the current government program, digital services and information management in the social and health sector are one of the contents of a functioning and sustainable welfare society [4]. The goal is to draw up long-term strategic goals to guide the development work of health and safety information management, digitalization, and research as well as development and innovation activities, so that the use of technology in the social and health sector would produce the desired benefits. The goal is to increase the share of electronic transactions and make digital transactions a priority for those customers for whom it is possible. Health care professionals have felt that electronic transaction services will be the main transaction channel in the future, while still securing in-person transactions [5]. The impact of digital services on patient satisfaction has been shown to be positive when digital services are accessible, easy to use, improve patient-provider communication and include the option to usual care [6].

Digital services in health care and social services sectors include for example digital peer support groups and similar communication services between customers or relatives or remote transactions and monitoring between customer and the service provider [7]. These can also include customer self-monitoring for example with symptom diaries or collecting the customer's health and well-being information with electronic forms. These services can be either asynchronous, meaning that two people aren't expected to be present and available at the exact same time, or synchronous, where there is simultaneous communication with two people [6]. Some of the digital services replace traditional reception visits and phone calls [7].

According to recent report [7], health care sector in Finland has more digital services as compared to social care sector. The most used digital services are those produced both publicly and privately in outpatient care of health services, both in special care and in primary health care. In social care, the solutions focus on remote home care solutions

and point-based solutions to electronic forms and applications. The most widely used are digital services, which support electronic communication between the client and the professional. These include, for example, separate remote receptions and electronic messages, which can be found in almost all service categories in both healthcare and social services sectors. However, there seems to be regional differences in the amount and type of digital services in Finland. These differences are partly explained by differences in naming accuracy of the services. In addition, the regional digital services are a combination of national solutions and solutions that can be purchased from the market or that can be developed in-house. Regional variation of digital services is thus formed depending on how much the regions have implemented self-developed or commercial, tendered digital services.

The health care and social services workforce in Finland is divided into the private and public sectors [8]. There are differences between different wellbeing services counties in how the workforce is divided into the public and private sector in the field of social security services. Ten years ago, less than half of the public organizations and an even smaller proportion of private organizations offered digital services for specific target groups [9].

Digital services are seen in the current government program as one of the ways to improve the functionality of the social and health service system and to curb expenditure growth [4]. Long-term cost benefits are often sought with digital services, although uncertainty is recognized in this regard, as broad estimates of the cost benefit potential of the digitalization of the social and health sector are still scarce [6, 10]. Doctors have seen the benefits of digital services for themselves as reduced telephone traffic, increased efficiency, freeing up time for medical evaluations, less crowded waiting rooms, and more accurate communication [11]. Healthcare professionals have seen increased flexibility, autonomy [5, 11] and time and money savings [11] as benefits of digital services for customers. According to healthcare professionals the benefits of digital services for themselves have been improved flow of work, enrichment of the professional's own job description, increased efficiency of information transmission and improved service availability [12]. Digital services have had a positive impact on healthcare professionals' satisfaction in endocrinology, palliative care, dermatology, and surgery [6]. Easy use and perceived usefulness of the digital services have been related to healthcare personnel satisfaction. However, the effects of digital services on the work of professionals are not only positive [12, 13]. As disadvantages of digital services, healthcare professionals have highlighted the decrease in face-to-face contact between the client and the professional and problems with the use of technology [12]. However, healthcare professionals' satisfaction in digital services has been shown to be understudied in literature [6].

In the development and evaluation of digital health services, the service's effectiveness, safety, costs, information security and data protection, as well as usability and accessibility must be taken into account [14]. The inclusion of digital services in healthcare changes treatment processes, which needs to be further developed from the point of view of healthcare professionals, so that the services work in the best possible way for both the professional and the customer [5]. Successful implementation and sustainability of digital services requires that attention is paid to addressing barriers and supporting facilitators at all levels in health care provision [6].

Due to the lack of region-specific data, the purpose of this study was to investigate digital services in welfare, social and health service organizations operating in the South Ostrobothnia region in Finland. The aim was to investigate the effects of employee status, form of organization and organizational size on the views related to different aspects of digital services and the development of these services. The study was carried out in the era of exceptional circumstances created by the Covid19 pandemic in the summer of 2021.

The research questions were:

1. What is the current state and the role of digital services in organizations and how do the views differ between respondents with different employee status, form of organization or organizational size?
2. What are the perceived development barriers related to digital services and how do the views differ between respondents with different employee status, form of organization or organizational size?
3. What are the development plans of the digital services in organizations and how do the views differ between respondents with different employee status, form of organization or organizational size?
4. What are the perceived support needs for digital services development and how do the views differ between respondents with different employee status, form of organization or organizational size?

2 Methods

2.1 Study Design

The study was a quantitative cross-sectional study using an electronic Webropol survey (version Webropol 3.0). Research population included welfare, social and health service organizations operating in the South Ostrobothnia region. These were private companies (micro, SME, large), public actors and service providers maintained by foundations or societies from the following business and service sectors: 1. Health services (primary health care, medical clinics), 2. Social services (housing services, family caregivers, home care), 3. Sports, youth and cultural services of municipalities, 4. Physiotherapy services, 5. Rehabilitation services (speech and occupational therapy), 6. Interpreter services (hearing and speech impaired services), 7. Psychologist services, 8. Exercise services, gyms, group exercise, 9. Services for the older adults, 10. Child protection services, 11. Substance abuse services.

Specialized medical care in its main features (e.g., specialized medical care at a Central hospital) and dental services on the municipalities were excluded from the study, although the survey was sent to organizations that offered some of these services.

2.2 Sample

The sampling method was mainly convenience sampling but with features from cluster sampling. An email list was collected with search of welfare, social and health service organizations operating in the South Ostrobothnia region. Organizations' contact information (e-mail addresses) was searched on the Internet using search engines and on the

websites of known organizations. E-mail addresses had to be findable and available on the organizations' websites or otherwise found on the internet. When contact information was found, the survey was sent to managers, entrepreneurs, as well as employees of organizations.

The survey was sent to a total of 1266 email addresses and 223 different organizations. The survey got through to 1252 e-mail addresses.

2.3 Data Collection and Ethical Considerations

The request to answer the survey was sent for the first time on 31 May 2021. Two reminder messages were sent to answer the survey: the first on 9 June 2021 and the second on 16 August 2021. The survey link was closed on 20 August 2021. This deadline for answering the survey was known and presented in the last reminder message.

The study followed the principles of the Helsinki Declaration (2013) [15] and the General Data Protection Regulation GDPR (2016) [16]. The survey was carried out using an anonymous questionnaire and the name of the organization that the respondent represented was not asked. Thus, the individual respondents could not be identified. The respondents were sent a cover letter explaining the purpose of the study, possible benefits to science and society, and an explanation of the voluntary nature of their participation.

2.4 Quantitative Cross-Sectional Survey

This is a sub-study of a wider study, in which entrepreneurs, managers and employees of welfare, social and health service organizations were asked about their current digital services and the use of welfare technologies as part of the services, as well as the effects of the Covid19 on the services. In addition, the survey inquired about the development ideas of the aforementioned sub-areas and the support needed for development. The survey was purpose-designed for the current study but partly used questions from previous studies by Seinäjoki University of Applied Sciences [17, 18] (permissions retrieved) targeted to welfare, healthcare and social services organizations. The questionnaire and descriptive results from the entire research group have been published in Hoffrén-Mikkola et al. (2021) [19]. In this study, the results and differences from the survey are reported on the current state and the role of digital services, perceived barriers to development, plans for the development of digital services and support needs for development according to employee status, form of organization and organization size. The survey question about the perceived barriers for development of digital services was from Kettunen et al. (2020) [17] and the question regarding support needs for development of digital services was from Toivonen & Vainionpää (2020) [18]. Respondents were introduced to digital services in the survey with the following description: *Digital services can be e.g., electronic appointment booking, chat service on the website, remote reception, consultation, training or other remote service, video-mediated service, electronic message delivery to the customer or the use of sensor-based data to support the service.*

2.5 Analyses

The data was analyzed with the IBM SPSS Statistics Version 29.0.0.0 (241) program. The results are presented as frequencies, relative frequencies (%) and/or means and standard deviations (SD). Statistical significances between different employee groups (manager or entrepreneur, employee), organizational forms (public operator, private company) and organizations of different sizes (1–9, 10–49, 50–249, $\geq$ 250 people) were tested for categorized and ordinal variables using Pearson Chi-Square test (2-sided). Opinion scale (4-point Likert scale) variables were treated as continuous variables. Statistical significances between different employee groups and organizational forms in continuous variables were tested with the nonparametric Mann-Whitney U test and for organizations of different sizes first with the Kruskal Wallis test and then pairwise comparisons with post hoc Bonferroni. The limit of statistical significance was set at 5% ($p < 0.05$). In the case of opinion scale variables, when the respondents had the opportunity to choose the "I can't say" option, these answers were classified as missing information before the test.

3 Results

The survey was answered by 121 people. Therefore, the response rate was 9.7%. Of the respondents, 53.7% (n = 65) were employees and 46.3% (n = 56) were managers or entrepreneurs. The majority (75.2%, n = 91) were from the public sector and 19.0% (n = 23) were from private companies. Six people (5.0%) were from societies and one person (0.8%) from a foundation-type organization. Since the majority of respondents were from the public sector and private companies, comparisons of organizational form were made only between these groups. The survey received responses from organizations of all sizes: 13.2% (n = 16) were from 1–4 people, 11.6% (n = 14) from 5–9 people, 31.4% (n = 38) from 10–49 people, 19.8% (n = 24) from 50–249 people and 24.0% (n = 29) more than 250 people organizations. In the organization size comparisons, the two smallest organizational groups (1–4 people and 5–9 people) were reclassified into one group (1–9 people organizations, n = 30, 24.7% respondents).

3.1 Current State and the Role of Digital Services

83.5% of the respondents answered that some of the services of the organization they represented were digital. Correspondingly, 16.5% reported that the organization they represented did not have any digital services. There were no statistically significant differences between employee status, form of organization or organization size in these views. Of those who had digital services, clearly more than half (61.4%) perceived some of these services had been put into use during the previous year. Employee status, form of organization or organization size had no statistically significant differences in this.

In the entire group of respondents, those respondents who had digital services, 47.5% felt that specifically the organization's digital services had either significantly or very significantly enabled customers during the exceptional circumstances created by the Covid19 pandemic. Employee status, form of organization or organization size did not

have statistically significant differences in these views. However, digital services were not seen so much as significant for reaching new customers, as 29.7% of the entire respondent group felt that digital services had either significantly or very significantly provided the organization with the opportunity to reach new customers. Form of organization or organization size did not have statistically significant differences in these views. However, managers or entrepreneurs saw digital services as reaching new customers more significant than employees ($p < 0.05$). Of employees 20.4% perceived that digital services had either significantly or very significantly provided the organization with the opportunity to reach new customers whereas of managers or entrepreneurs 40.4% perceived so.

3.2 Perceived Barriers for Development of Digital Services

The most significant barriers in the development of digital services in the organization that the respondent represented were perceived to be especially the acquisition of technology (software, tools), and human resources, which were perceived by 66.1% and 66.9% of the respondents as either significant or very significant barriers, respectively (Fig. 1). There were no statistically significant differences between employee status, form of organization or organization size in these views, although there were indications that small organizations perceived the barriers to be slightly greater than large organizations ($p = 0.075$, ns) and public actors greater than private companies ($p = 0.068$, ns).

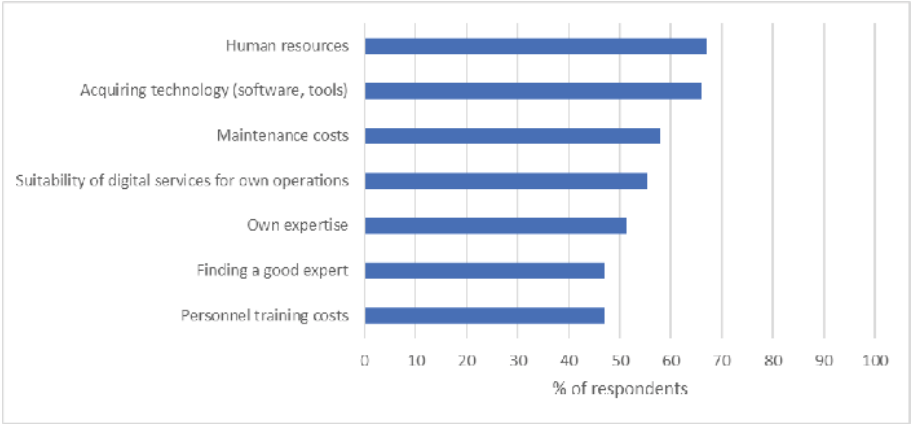


Fig. 1. Either significant or very significant barriers for development of digital services perceived by respondents. Numbers are percentages of respondents in the entire respondent group ($n = 121$).

3.3 Development Plans

In the entire group of respondents, slightly more than half (52%) estimated that the organization they represented had the goal of increasing the number of digital services

within a year, and 58% that it was intended to do so within five years. However, increasing the number of digital services also included a lot of uncertainty among the respondents (Tables 1 and 2).

Table 1. Respondents' views on organizations' goals to increase the number of digital services within one year relative to employee status, form of organization and organization size. P-values describe the statistically significant difference between respondent categories within independent variable. ** = statistically significant difference ($p < 0.01$).

Independent variable	Respondent category	Yes (%)	No (%)	I cannot say (%)	Diff (p-value)
Employee status	Manager or entrepreneur (n = 56)	57.1	17.9	25.0	p = 0.007 **
	Employee (n = 65)	30.8	18.5	50.8	
Form of organization	Public actor (n = 91)	45.1	12.1	42.9	p = 0.009 **
	Private company (n = 23)	34.8	39.1	26.1	
Organization size	1–9 people (n = 30)	26.7	40.0	33.3	p = 0.003 **
	10–49 people (n = 38)	34.2	18.4	47.4	
	50–249 people (n = 24)	62.5	8.3	29.2	
	≥ 250 people (n = 29)	55.2	3.4	41.4	

In particular, managers or entrepreneurs, public actors and large organizations perceived that the organization they represented was about to develop digital services within one year (Table 1) and within five years (Table 2). However, the differences in views between different groups of respondents were smaller in longer time scale (five years) as compared to shorter time scale (one year).

3.4 Support Needs

In the entire group of respondents, the most support related to the use and development of digital services was perceived for obtaining the necessary information about the characteristics of digital services (e.g., usability, reliability, level of development), where the mean ± standard deviation in the entire group of respondents was 3.08 ± 0.8 on a scale of 1–4, where 4 meant very significant need. Next, financial support was perceived to be necessary for the use and development of digital services (3.06 ± 0.90). However, support needs were perceived to be high for all categories since 57–71% of respondents reported either significant or very significant support needs for different categories (Fig. 2). Employee status or form of organization did not have a statistically significant effect on the perceived need for support. The size of the organization did not have a statistically significant effect on support needs, except for one area: larger organizations needed statistically significantly ($p < 0.01$) more information about the

Table 2. Respondents' views on organizations' goals to increase the number of digital services within five years relative to employee status, form of organization and organization size. P-values describe the statistically significant difference between respondent categories within independent variable. ns = no significant difference, * = statistically significant difference ($p < 0.05$).

Independent variable	Respondent category	Yes (%)	No (%)	I cannot say (%)	Diff (p-value)
Employee status	Manager or entrepreneur (n = 56)	58.9	14.3	26.8	p = 0.017 *
	Employee (n = 65)	38.5	9.2	52.3	
Form of organization	Public actor (n = 91)	47.3	7.7	45.1	p = 0.139 ns
	Private company (n = 23)	43.5	21.7	34.8	
Organization size	1–9 people (n = 30)	40.0	23.3	36.7	p = 0.014 *
	10–49 people (n = 38)	31.6	13.2	55.3	
	50–249 people (n = 24)	62.5	8.3	29.2	
	≥ 250 people (n = 29)	65.5	0.0	34.5	

cost-effectiveness of digital services than small organizations. In bilateral tests, there was only a significant difference ($p < 0.01$) between organizations with 10–49 people (2.56 ± 0.96) and more than 250 people (3.31 ± 0.806).

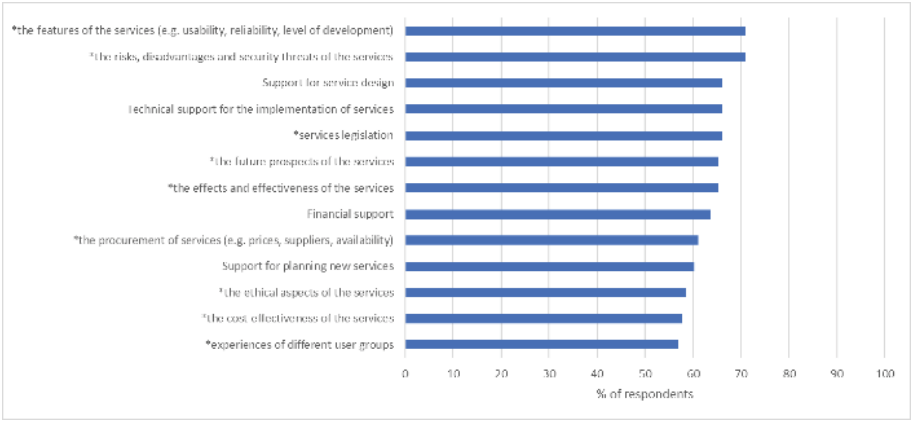


Fig. 2. Either significant or very significant support needs for development of digital services perceived by respondents. Numbers are percentages of respondents in the entire respondent group (n = 121). *Information about…

4 Discussion

This study supports the previous literature [2, 3, 20] that Covid19 pandemic increased the development of digital services in social and health care sector. Laukka et al. (2021) [3] reported that, of the psychiatric outpatient visits, 9% were conducted using phone or remote access in January 2020 and 48% in May 2020 which means that by remote transactions had increased 39 percentage units in five months. In the current study 61% of organizations that had digital services had developed some of these during the previous year with Covid19. This study adds to previous literature with the findings that the amount of digital services and the development of services during Covid19 has been similar in public actors and private organizations as well as in organizations of different sizes in South Ostrobothnia.

The results of the current study suggest that digital services in welfare, social and health service organizations were especially seen as enablers of existing customers in exceptional circumstances than reaching new customers. This is in line with previous literature in psychiatric outpatient services [3] and in specialized youth psychiatric care [20]. Laukka et al. (2021) [3] reported that digital and remote services enabled psychiatric outpatient services during Covid19 pandemic to those patients for whom they were suitable and who could use them and found out that one of the supporters of remote service usage was familiarity of the customers. The specialized youth psychiatric care employees in the study by Sirnelä-Rif et al. (2020) [20] felt that taking care of a young person they had known for a long time was successful also remotely during Covid19 pandemic. They perceived that this was because there was already a trusting relationship with the patient created. The familiarity of customers could explain the results of the current study that the digital services were more seen as enablers of already existing customers in exceptional circumstances than reaching new customers. Especially employees had this opinion in the current study. This can be because employees normally work closer with customers as compared to managers or entrepreneurs. Koivisto et al. (2023) [12] showed that different actors (nurse, manager, technology developer) perceive the role of a technology for care work in different ways because they have different goals for the technology. It has been reported in previous literature that healthcare professionals have emphasized the importance of trustful relationships with clients in telehomecare of older adults [21] as well as in telemonitoring of chronically ill patients [22]. The current study supports these results.

The human resources and acquisition of technology (software, tools) were felt to be the most significant barriers in the development of digital services according to managers, entrepreneurs and employees of the welfare, social and health service organizations of the current study. These may be explained first of all by the shortage of labour in healthcare and social services sector [8] and thereafter by the timing of the survey during the exceptional circumstances caused by Covid19 pandemic. Healthcare professionals have described that one of the negative factors of digital services in their work is insufficient resourcing to maintain the professional's technology skills, which include insufficient training and support, insufficient time to learn new things and number of tools and constant updates [12]. Human resources as one of the most significant barriers for development of digital services in the current study may be because the development of digital services takes time and resources from the organization. Because of shortage

of labour it may be impossible to find time to include both managers and employees for digital service development. Digital competence of occupational health doctors and occupational health nurses have been reported to include competence in developing the use of technology in one's own organization [12]. Especially during the Covid19 pandemic this may have felt impossible which may explain the results of the current study. In addition, the challenges for using remote services in psychiatric outpatient services shortly after the start of Covid19 pandemic were related for example to the lack of IT equipment and inoperative programs, which is supported by the current study where acquisition of technology was one of the most significant barriers for development of digital services. The organizations were forced to tailor services to digital format without proper time for planning and therefore were not prepared for this change in services with proper technology acquisition.

Regarding the use and development of digital services, support and information was felt to be necessary especially about the characteristics of digital services, and financial support was also needed for the development. However, the support needs were significant in many aspects related to digital service development. Professionals in occupational healthcare have needed most of all the instructions on the use of video conferencing and chat services [12]. The current study was not that specific since it included organizations from different kinds of service sectors and target groups and therefore different variety of digital services. In particular, large organizations that were about to develop digital services more in one year and five years time scale than smaller organizations needed information on the cost-effectiveness of digital services. The effectiveness and cost-effectiveness of technologies and digital services are indeed factors about which little is still known, as no systematic information has been collected [22] and the results are unclear [6, 10].

This study has a few methodological considerations and limitations that need to be addressed. The number of survey respondents (n = 121) and the response rate (9.7%) were small. This may be partly because the sample was large. The survey was sent not only to managers and entrepreneurs, but also to all employees of organizations whose contact information could be found on the internet. Almost the same number of managers or entrepreneurs and employees answered, so the response rate for managers or entrepreneurs is estimated to be higher than for employees. However, this was not investigated, and it was not possible to find out which group the respondents represented in the case of managers and entrepreneurs. The strength of the research can be considered that with this research setup, it was possible to examine employees' views widely and compare employees' perspectives about digital services with the views of managers or entrepreneurs. One reason for low response rate could be the timing of the survey during the holiday season in Finland. However, since the reminders to respond were sent both in June and in August, it can be argued that all had possibility to answer even during this season. It should also be noted that the study was not able to document differences between the professionals who responded to the survey and those who did not. Thus, it is not known how well the respondents represented the entire population. The questionnaire was extensive (8 pages, 26 questions, response time 15–20 min), so it is possible that the least busy managers, entrepreneurs and employees answered the survey, and this can possibly affect the results. The survey was also mainly purpose-designed and was not

validated. Last, it must be remembered that this survey was carried out already few years ago before the social and health care reform and the establishment of wellbeing services counties. As suggested also by Pennanen et al. (2023) [7], the use and prevalence of digital services should be studied more when wellbeing services counties are settled and have further developed their digital services. This applies also to South Ostrobothnia. The results of the current study may then be used for comparison.

Conclusions and Practical Implications. The results of the current study can be used to support welfare, social and health service organizations in digital service development. First, the study supports previous research that digital services work best when the customer relationship already exists, and the customer is known. This should be taken into account in service path development. It must be noted, as supported by the current study, managers or entrepreneurs and employees perceive the roles of a digital services in different ways. Therefore, all these groups should be included in digital service development in the organization. Second, organizations need extra resources, human resources as well as financial support, in development of digital services which is a challenge during the shortage of labour in healthcare and social services sector as well as with cost challenges faced by the newly developed wellbeing services counties. Finally, according to this study, the large organizations and public organizations in South Ostrobothnia were more certain in their views to develop more digital services especially in short time scale. Organizations should be supported in their plans with providing them support with the forms of more information about the characteristics of digital services as well as about the effectiveness and cost-effectiveness of these services since this information is scarce and partly unclear.

Acknowledgments. The author would like to acknowledge principal lecturer Katja Valkama, development manager Sami Perälä, RDI expert Jouko Lakaniemi and designer Tuula Salo from Seinäjoki University of Applied Sciences for their cooperation in collecting the survey data. The survey was carried out as part of the AKKE-funded E-P Hyte 2021 project, which was financed by the Regional Council of South Ostrobothnia.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Finnish Institute for Health and Welfare. Tulevaisuuden sosiaali- ja terveyskeskus -ohjelma. Updated 14.8.2023. Cited 15.8.2023. https://thl.fi/fi/tutkimus-ja-kehittaminen/tutkimukset-ja-hankkeet/tulevaisuuden-sosiaali-ja-terveyskeskus-ohjelma
2. Jormanainen, V., et al.: Covid19-epidemian vaikutukset hyvinvointiin, palvelujärjestelmään ja kansantalouteen: Asiantuntija-arvio syksy, vol. 2020, pp. 135–143 (2020). Finnish Institute for Health and Welfare. http://urn.fi/URN:ISBN:978-952-343-578-0
3. Laukka, E., et al.: Muutos psykiatristen avohoitokäyntien toteutuksessa ja johtajien kokemukset etäasiointia estävistä ja edistävistä tekijöistä COVID-19 aikakaudella. Finnish J. eHealth and eWelfare **13**(1), 49–62 (2021). https://doi.org/10.23996/fjhw.98228

4. Finnish Government. Vahva ja välittävä Suomi: Pääministeri Petteri Orpon hallituksen ohjelma. (20 June 2023). https://julkaisut.valtioneuvosto.fi/handle/10024/165042
5. Koivuluoma, M., Haverinen, J., Reponen, J.: Hoitoprosessien uudistamattomuus haastaa terveydenhuollon digitalisaatiota. Finnish J. EHealth and EWelfare **14**(3), 284–298 (2022) https://doi.org/10.23996/fjhw.111788
6. Lakoma, S., et al.:Impact of Digital Services on Healthcare and Social Welfare – An Umbrella Review. International Journal of Nusing Studies (2024). https://doi.org/10.1016/j.ijnurstu.2024.104692
7. Pennanen, P., et al.:Digitaalisten palvelujen vaikutukset sosiaali- ja terveydenhuollossa. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja, vol. 52 (2023). http://urn.fi/URN:ISBN:978-952-383-059-2
8. Tevameri, T.: Sote-palveluala: Työvoima, yksityisen sektorin rooli ja kansainvälistyminen. TEM toimialaraportit. vol. 2 (2023). http://urn.fi/URN:ISBN:978-952-327-605-5
9. Hyppönen, H., Hämäläinen, P., Reponen, J. (eds.). E-health and e-welfare of Finland: Check point 2015. National Institute for Health and Welfare (THL), Report 18/2015 (2015). https://urn.fi/URN:ISBN:978-952-302-563-9
10. Saranto, K., Kinnunen, U.-M., & Jylhä, V., Digitalisaatio ja sähköiset palvelut uudistuvassa sosiaali- ja terveydenhuollossa. Tampere University Press (2020). https://urn.fi/URN:NBN:fi:tuni-202011278265
11. Fagerlund, A.J., Holm, I.M., Zanaboni, P.: General practitioners' perceptions towards the use of digital health services for citizens in primary care: a qualitative interview study. BMJ Open **9**, e028251 (2019). https://doi.org/10.1136/bmjopen-2018-028251
12. Koivisto, T.: Digitoimijuus terveydenhuollon ammattilaisen työssä. Tampereen yliopiston väitöskirjat, vol. 881 (2023). https://urn.fi/URN:ISBN:978-952-03-3102-3
13. Vehko, T., Hyppönen, H., Ryhänen, M., Tuukkanen, J., Ketola, E., Heponiemi, T.: Tietojärjestelmät ja työhyvinvointi – terveydenhuollon ammattilaisten näkemyksiä. Finnish J. EHealth EWelfare **10**(1), 143–163 (2018). https://doi.org/10.23996/fjhw.65387
14. Haverinen, J., Keränen, N., Falkenbach, P., Maijala, A., Kolehmainen, T., Reponen, J.: Digi-HTA: Health technology assessment framework for digital healthcare services. Finnish J. EHealth EWelfare, **11**(4), 326–341 (2019). https://doi.org/10.23996/fjhw.82538
15. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA **310**(20), 2191 (2013)
16. European Parliament and Council of the European Union. General data protection regulation GDPR. Official J. European Union. L119, **59**, 1–88 (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TO
17. Kettunen, S., Joensuu-Salo, S., Mäntysaari, P., Aalto, A., Katajavirta, M.: Digitaalisuus muuttaa sosiaali- ja terveysalaa: osaamisen taso eteläpohjalaisissa pk-yrityksissä sekä esimerkkejä uudesta liiketoiminnasta. Seinäjoen ammattikorkeakoulun julkaisusarja B. Raportteja ja selvityksiä, vol. 150 (2020). https://urn.fi/URN:NBN:fi-fe202002125327
18. Toivonen, M., & Vainionpää, J.: Knowledge and attitudes of welfare entrepreneurs about welfare technology – a survey for South Ostrobothnian entrepreneurs. Finnish J. EHealth EWelfare **12**(2), 139–157 (2020). https://doi.org/10.23996/fjhw.87711
19. Hoffrén-Mikkola, M., Valkama, K., Perälä, S.: Digitaaliset palvelut ja hyvinvointiteknologiat Etelä-Pohjanmaalla: kysely hyvinvointi-, sosiaali- ja terveyspalveluorganisaatioille koronaaikana. Seinäjoen ammattikorkeakoulun julkaisusarja B. Raportteja ja selvityksiä, vol. 166 (2021). https://urn.fi/URN:NBN:fi-fe2021121560682
20. Sirnelä-Rif, E., Holttinen, T., Haapasalo-Pesu, K.M.: Nuorten hoitosuunnitelmaan tehtävä kirjaus etähoidon osuudesta. Lääkärilehti **36**, 1787 (2020). https://www.utupub.fi/handle/10024/172439?show=full

21. Heinonen, T., Lindfors, P., Nygård, C-H.: Etäkotihoitotyön sisältö ja kuormittavuus sekä mahdollisuudet työurien pidentäjänä. Gerontologia **36**(2), 128–142 (2022). https://doi.org/10.23989/gerontologia.109921
22. Carlqvist, C., Hagerman, H., Fellesson, M., Ekstedt, M., Hellström, A.: Health care professionals' experiences of how and eHealth application can function as a value-creating resource – a qualitative interview study. BMC Health Serv. Res. **21**, 1203 (2021)
23. Uusitalo, A.: Hyvinvointiteknologia, työhyvinvointi ja tuottavuus sosiaali- ja terveysalalla. Diplomityö. Lappeenrannan-Lahden teknillinen yliopisto LUT. Tuotantotalouden koulutusohjelma (2021). https://urn.fi/URN:NBN:fi-fe202104079594

A Persuasive mHealth Application for Postoperative Cardiac Procedures: Prototype Design and Usability Study

Renata Savian Colvero de Oliveira[1(✉)], Grace T Marcon Dal Sasso[2], Sriram Iyengar[3], and Harri Oinas-Kukkonen[1]

[1] Oulu Advanced Research On Service and Information Systems, University of Oulu, Oulu, Finland
renata.deoliveira@oulu.fi
[2] Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil
[3] College of Medicine, University of Arizona, Phoenix, USA

Abstract. The global burden of cardiovascular diseases (CVD) is a worldwide public health problem. In 2019, 18.6 million people died from CVD, representing a 17.1% increase compared to 2010. Also, some individuals who experience a cardiovascular event will require some form of cardiovascular procedure, such as a pacemaker or implantable cardioverter-defibrillator insertion, aneurysm repair, or heart valve replacement. Mobile health (mHealth) is a valuable tool for supporting individuals with CVD in self-management, providing medical recommendations, virtual consultations, reminders, and disease monitoring notifications. The main objective of this research was to enhance postoperative care for cardiac procedures. To achieve this, the research involved the development of a new mHealth application and the subsequent evaluation of its usability. The study constituted technological and usability research by using Design Science Research Methodology (DSRM). The design of the mobile application followed the principles of Persuasive Systems Design (PSD) model, which encompass a clear definition of the main task, user interaction through dialogue, system credibility, and social support, aiming to help change user behavior. The sample was non-probabilistic for convenience, and System Usability Scale (SUS) was applied to physicians and nurses as well as individuals in the information technology field. The sample comprised 18 participants, of whom 55.6% were female. The participants rated the application positively, with a median final SUS score of 95 (IQR 90–97.5). Finally, the mobile application presented high usability and user acceptance.

Keywords: Cardiovascular Diseases · Persuasive Systems Design · Usability Study · Mobile Health Applications

1 Introduction

Cardiovascular disease (CVD) is a major cause of death worldwide, making its burden a global public health concern. According to estimates, 18.6 million people died from CVD in 2019, a 17.1% increase from 2010 [1].

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 83–100, 2024.
https://doi.org/10.1007/978-3-031-59091-7_6

Low- and middle-income countries bear the greatest brunt of the substantial disease burden attributed to cardiovascular diseases (CVD). In the Americas region, CVD is accountable for 36.4 million years of life lost due to premature deaths, 40.8 million disability-adjusted life years annually, and 4.5 million years lived with disability [2]. Brazil mirrors these global statistics, with 30% of deaths attributed to CVD [3].

Some patients who have undergone a cardiovascular event may require cardiovascular procedures due to the disease's progression. These procedures may include the implantation of a pacemaker or implanted cardioverter defibrillator and the correction of an aneurysm [4]. It is essential to prevent post-operative complications, such as atrial fibrillation, kidney failure, reoperation due to bleeding, stroke, and pneumonia [5].

In the current landscape, a comprehensive meta-analysis of randomized controlled trials (RCT) has provided robust evidence supporting the beneficial impact of digital technology on CVD management. This analysis highlights significant improvements in several key areas, including total cholesterol levels, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, physical activity, dietary habits, and adherence to medication regimens [6]. The integration of mHealth solutions plays a pivotal role in this advancement by facilitating patient self-management of CVD. mHealth achieves this through the provision of direct access to medical consultation and advice, as well as the delivery of personalized reminders and notifications that aid in continuous disease monitoring and management [7].

Making healthier lifestyle choices and changing one's behavior are the most crucial ways to halt the progression of CVD. To modify unhealthy lifestyle choices including smoking cigarettes, eating poorly, and not exercising, younger patients should receive regular counseling [8]. Behavior modification is necessary and helpful even in cases where the patient has previously suffered a cardiovascular event to reduce the risk of further events. Taking prescription drugs as prescribed, scheduling frequent doctor's appointments, and controlling risk factors are a few instances of this [9].

In this scenario, the application of persuasive technology can yield significant benefits. Persuasive technology encompasses a broad range of digital tools and platforms specifically designed to influence attitudes or behaviors. This includes, but is not limited to, computers, websites, smartphones and their applications, tablets, wearable devices, and computer games [10]. Each of these technologies has the potential to engage users and effectively guide their behaviors or attitudes in desired directions. According to Fogg's [11] initial description, computing items are persuasive social agents. Stated differently, these technological products possess the ability to impact and trigger social reactions from their users, through various means such as rewarding individuals who provide positive feedback, modeling a particular behavior or attitude, or providing social support. Oinas-Kukkonen and Harjumaa [12] suggested a method for applying design principles to software requirements and their subsequent implementation as characteristics of a system. To make this more practical, they proposed the Persuasive Systems Design (PSD) model, which has four main categories of software features: primary task support (PRIM), dialogue support (DIAL), system credibility support (CRED) and social support (SOCI).

In a recent RCT, the PSD model was utilized to investigate the efficacy of a mobile health behavior change support system as an obesity control intervention. The intervention group lost weight more successfully than the control group, according to the trial (95%CI –3.8 to –1.6, p <0.001). Thus, the effectiveness of the PSD's use in mHealth in modifying behavior can be observed [13].

This study aimed to improve postoperative care for cardiac procedures. This entailed developing a mHealth application prototype and then evaluating its usability, firstly focusing in professionals on the technology and health domains.

2 Background

2.1 Cardiovascular Diseases

According to the World Health Organization (WHO), CVD is a group of disorders of the heart and blood vessels that include coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic heart disease, congenital heart disease and deep vein thrombosis and pulmonary embolism [14]. The main diseases are known as ischemic heart disease, or acute myocardial infarction and cerebrovascular accidents. Common reasons for the precipitation of these diseases are fatty deposits on the inner walls of blood vessels and bleeding from blood vessels in the brain, for example [14].

2.2 mHealth Applications for Cardiovascular Procedures

To identify the state of the art regarding this topic, a scoping review was conducted in November of 2022 and search keys were created for all the electronic bibliographic databases consulted, including National Institutes of Health US National Library of Medicine (PubMed) and Medical Literature Analysis and Retrieval System Online (MEDLINE), Web of Science, a Scopus, EMBASE and IEEE Xplore. The main descriptors were "Mobile Applications" AND "Postoperative Care" AND "Cardiac Surgical Procedures". After excluding 231 duplicates, 385 articles remained for inclusion and exclusion by title and abstract. After this first stage, 19 articles were selected for full reading. Of these, four were unrelated to the research topic, five did not have results (reports or protocols), one was due to unavailability of full text and three could not be retrieved even after contacting the corresponding author. Finally, six articles were included.

The studies have explored the detection of atrial fibrillation with the help of an electrocardiogram [15, 16], medication adherence [17] and collection of photoplethysmography data to detect patients' heart rhythms allowing them to report symptoms and send messages to their doctors [18]. In a similar direction regarding communication with the healthcare team, Atilgan et al. [19] used a telemedicine system that allowed patients to monitor their vital signs and report symptoms to their doctors via instant messaging and video conferencing. Aydin et al. [20] reported using two different applications, one for control patients' breathing, including breathing exercises, and training to increase lung capacity. The second application was related to medication adherence, and researchers selected it from those available on the Google Play Store and App Store.

This work focused on designing an application that optimized the patient's clinical experience by eliminating the need for additional portable devices to make it more accessible and cost-effective. By unifying the different functionalities, including reinforcing medication adherence, facilitating contact with the medical team, monitoring signs and symptoms after cardiac procedures, providing relevant educational content, and scheduling teleconsultations, all within a single application, we focused on not only improving the patient's post-operative care but also increase their active participation in their recovery process. This approach reflects not only practical efficiency but also a concern for accessibility and convenience for users, thus aligning with the principles of persuasive technology.

2.3 Persuasive Technology and Persuasive Systems Design

To be successful in changing behavior, Fogg's model stipulates that three elements must converge at the same time for a behavior to occur: motivation, ability, and triggers (e.g., an alarm that sounds, a text message, an advertisement). To achieve a target behavior, a person must have sufficient motivation, sufficient skill, and an effective trigger. Furthermore, these three factors must be present simultaneously [21].

The software features that comprise the PSD are split up into the four main categories. Carrying out the user's primary task, such as enabling users to independently track their progress, is referred to as PRIM. By reminding users of tasks related to the primary task or making the user interface visually appealing, computer-human dialogue, DIAL, makes sure that users receive assistance in maintaining their target behavior. Credibility, CRED, explains how to create a system that is more trustworthy by offering accurate, impartial, and fair information and directing users through reliable sources. Lastly, social support, SOCI, explains how to use social influence to motivate users of the system. Some examples of this include bringing together individuals who share a goal, helping them feel included, and offering opportunities for cooperation [12].

3 Methods

3.1 Study Design

This paper addresses prototyping and usability study. The research followed Design Science Research Methodology (DSRM) [22] and the design of the mobile application prototype followed the principles of PSD [12].

The six phases of the DSRM [22] were applied as follows:

(1) Problem identification: The lack of mobile applications for cardiac procedures utilizing persuasive technology.
(2) Definition of objectives: The main objective was to develop a mobile health application for the post-operative period of cardiac procedures and evaluate its usability.
(3) Solution development: Creation of a mobile application through the Adalo® platform.

(4) Evaluation of the solution: The evaluation was carried out using the usability scale (System Usability Scale, SUS) [23].
(5) Communication of Results: The results of the research process were communicated to third parties through this publication.
(6) Critical evaluation: A critical evaluation of the developed solution and the process used to develop it were made in the "discussion".

For the theoretical construction of the application prototype, the four persuasive software main categories of the PSD [12] model were used (Table 1).

Table 1. PSD categories and the presentation on the app

PSD category	Presentation on the application prototype
PRIM	The application was responsible for making the task of post-operative care for cardiac procedures easier for the patient, through various content and properties; Users had access to content specific to their surgical procedure, as well as podcasts and interviews with experts in the field; Users had access to their statistics and progress regarding application usage; The app allowed users to complete online questionnaires to assess their mood, pain, track depression and quality of life post-heart surgery
DIAL	The application had reminders and warnings regarding the use of medication, date, and time of appointments. It also asked about signs and symptoms, requested photos of the surgical wound, requested test results (e.g., prothrombin activity time). Also, the more the user used the application's functions, it was released extra content such as podcasts with experts. The application was made colorful and pleasant
CRED	The application's content was based on international guidelines and organizations, such as: World Health Organization, American Heart Organization, Enhanced Recovery After Cardiac Surgery Society, Society of Thoracic Surgeons, and the Guidelines of the Brazilian Society of Cardiology. There was a "References" topic, in which users had access to the websites and protocols used to build the application. The application had a section about the research team and their affiliations. Furthermore, the logo of the Federal University of Santa Catarina was inserted on the app home page
SOCI	It had a "Message Board", in which they can exchange ideas with peers and report how they are feeling after the surgery (which can be anonymous, with a fictitious name, if desired); The physicians and nurses could also participate

3.2 User Interface Design

To structure the interface, wireframes, originally in Portuguese, were created with the software Balsamiq®. They represent a visual scheme, blueprint, or model of a screen or web page design in an interaction design. It is the basis for an iterative prototype. Besides that, wireframes focus on the screen content and not the graphic details. Its

purpose is to illustrate high-level concepts [24]. Figure 1 shows the idea of building a message board in the first illustration, where the patient can choose a fictitious name, and the healthcare professional can also use it. The second shows the options and types of button ideas for the quality-of-life section, including explanatory videos, podcasts, and depression scales.

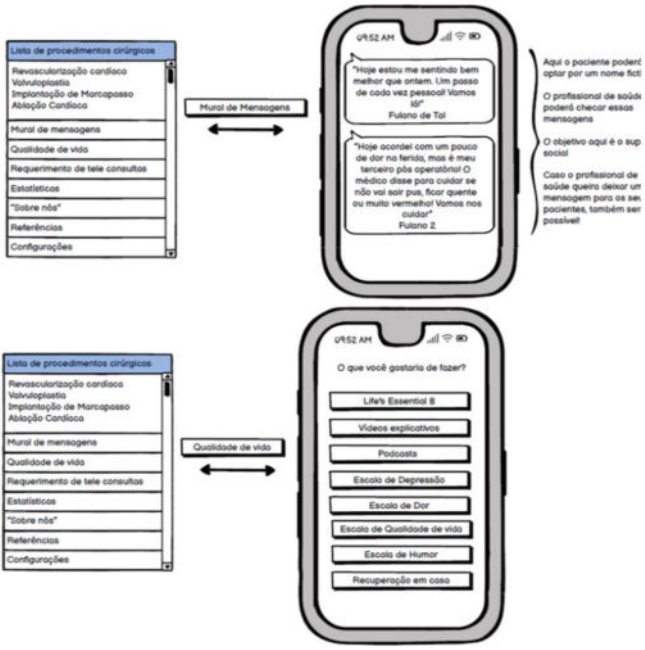


Fig. 1. Application wireframe (message board and quality of life section)

3.3 Population and Sample

The sample was non-probabilistic for convenience, it was chosen due to its practical benefits. This sampling method was accessible, cost-effective, and time-efficient, making it ideal for rapidly gathering feedback in the early stages of development. The survey invitation was emailed to 58 people residing in Southern Brazil, including healthcare professionals (doctors and nurses) and people in the technical field (systems analysis and development, software development, computer science, data science, computer engineering, and experience design of users). These health non-professionals were included to also assess the functional and design aspects of the technology. Of these, 18 responded to the participation email, answering the Informed Consent Form.

As inclusion criteria for health professionals, medical doctors and nurses were included, from any area or specialty, as at this time only the usability and not the content of the application were evaluated. Those who did not belong to the professional

category of doctor or nurse were excluded. As inclusion criteria for technology professionals, those from any area or specialty who agreed to participate in the research and who had experience in using and developing applications were included.

3.4 Data Collection and Organization

The prototype of the application was provided to professionals on June 9, 2023, and they were asked to fill out the usability scale through an online form, via GoogleForms®, on June 15, 2023. The collected data was organized in electronic spreadsheets and analyzed on June 20, 2023, to June 24, 2023.

3.5 Data Collection Instruments

Usability was assessed using the System Usability Scale (SUS) [23] after detailed description and guidance provided on GoogleForms®.

The SUS questionnaire [23] has been validated in Portuguese and in Brazil [25, 26], and comprises 10 items, classified by respondents on a 5-point LIKERT scale ranging from completely disagree to completely agree. The SUS items alternate between positive and negative items to avoid response bias, with the aim that participants agree or disagree after reading reflection, and not simply on impulse. To obtain the final score, which varies from 0 to 100, 1 must be subtracted from the user's response for odd-numbered items, as well as subtracting the value of the user's response from 5 for even-numbered items. Then, the score obtained for each item was added, and the result is multiplied by 2.5 [23].

The evaluation of the 10 items that comprise the SUS questionnaire [23] was established by the final score, ranging from 0 to 100. Thus, the classification was established as: 0 to 20.5; 21 to 38.5; 39 to 52.5; 53 to 73.5; 74 to 85.5; and 86 to 100 [27, 28].

Additionally, two optional open-ended questions about user experience were added after the SUS scale was completed by participants: Do you have any suggestions for improvement? What were your biggest difficulties when using the application?

3.6 Statistical Procedures

For the statistical analysis of the variables, the data were organized in electronic spreadsheets in the Microsoft Excel® program and subsequently analyzed using the Stata® 14.0 software (StataCorp, Texas, USA).

To determine the best way to describe the SUS score values, the Shapiro-Wilk test was performed to evaluate the distribution of data in relation to normality. This test is especially indicated for samples smaller than 30 [29]. Only descriptive statistics were applied. The boxplot was analyzed to identify outliers and frequency, median, interquartile range with maximum and minimum values were described.

In the Stata® 14.0 statistical software, the command used to generate the SUS Score for each participant was:

*gen escoretotal = (2.5 * (abs(q1–1) + abs(q2 - 5) + abs(q3–1) + abs(q4 - 5) + abs(q5–1) + abs(q6 - 5) + abs(q7–1) + abs(q8 - 5) + abs(q9–1) + abs(q10 - 5))).*

Where "escoretotal" is a new variable generated, which was assigned to each of the 18 observations and q1 to q10 are questions 1 to 10 of the SUS questionnaire.

3.7 Ethical Considerations

This study was based on ethical principles, based on Resolution No. 466 of December 12, 2012, of the National Health Council [30], which incorporates, from the perspective of the individual and communities, the four basic references of bioethics: autonomy, non-maleficence, beneficence, and justice, among others, aiming to ensure the rights and duties that concern the scientific community, research subjects, and the State. This project was also approved by the Ethics Committee.

All volunteer participants were invited to read and sign the Free and Informed Consent Form, ensuring the confidentiality of their identity and the information provided solely for the purposes of the research, as well as the right to withdraw from the study at any time, without any harm to the participant.

Lastly, this work followed all the foundations of the second article of Law No. 13,709/2018, known as the General Data Protection Law [31].

3.8 Data Security

Due to the sensitive nature of personal health information involved, the application complied with Brazil's strict data protection laws (LGPD), guaranteeing that all patient data is encrypted while it's in use and while it's in transit.

4 Results

4.1 Application Prototype

Fifty-one screens were developed in common with patients and healthcare professionals. Figure 2 represents the "Welcome" screens, where professionals and patients can register and log in. In addition, both have access to the privacy policy. The "Quality of Life" screens has the contents Life's Essential 8 [32] with tips on how to eat better, be more active, quit smoking, have healthy sleep, take care of your weight, control your cholesterol, take care of sugar levels and blood pressure; explanatory videos about cardiac procedures that function as a link to the American Heart Association website; Podcasts with experts; depression, pain, quality of life and mood scales; recovery tips at home and finally, the references used to build the mobile application.

Eleven screens were developed just for patients. The initial screen after patients log in ("Home"), they must choose which cardiac procedure they underwent to carry out: register medications, add a photo of the wound, register signs and symptoms, and record exams. About registering a photo of the wound, patients can download the photo, enter the date it was taken, and indicate what was observed about the wound (pain, local heat, redness and/or secretion, pus) (Fig. 3). Regarding signs and symptoms, the patient can select between some pre-existing options (chest pain, shortness of breath, bleeding, fever, weakness/abnormal tiredness) or write in a space designated for "another" sign/symptom (Fig. 3); they must also record the date. Finally, regarding exam registration, in addition to the registration date, the patient can take a photo and attach it to the application, attach a file download, and/or write the result of an exam (e.g., prothrombin time control).

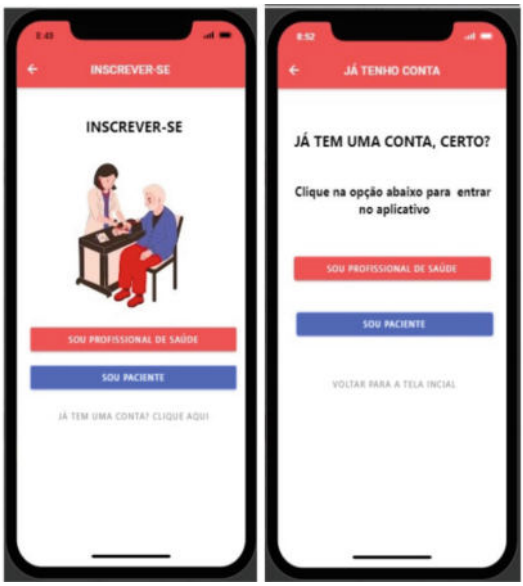


Fig. 2. Patients and healthcare professionals' screens

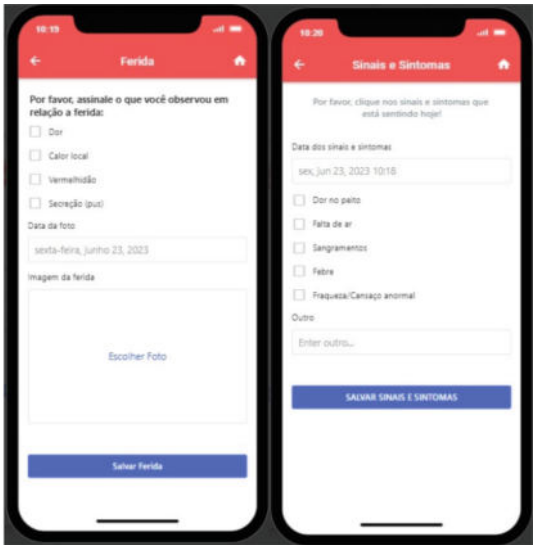


Fig. 3. Patient's screens

Eighteen screens were developed just for professionals. The initial screen after professionals log in ("Home"), they must choose which cardiac procedure to consult to obtain a list of patients. By clicking on the selected patient, they will access medications in use, photos and details of the wound, and signs and symptoms (Fig. 4). In the medicines

in use, the professional can quickly see the medications registered by the patient with the name and how many milligrams (e.g., Rosuvastatin 20) (Fig. 4); by clicking on the arrow next to the medication, the professional can view in more detail the number of tablets used and the frequency of use of that medication. Regarding the photo of the wound (on the right of Fig. 4), the professional has access to all the photos added by that patient; when clicking on the arrow, he will have more details, such as how many days ago the photo was attached, and what the patient pointed out about it (local pain, local heat, redness, secretion/pus). Regarding signs and symptoms, the professional has access to a list of records about how many days ago they were recorded (e.g., 24 days ago, 17 days ago); by clicking on the arrow on the right of the screen, professionals have access to the details of the recorded signs and symptoms.

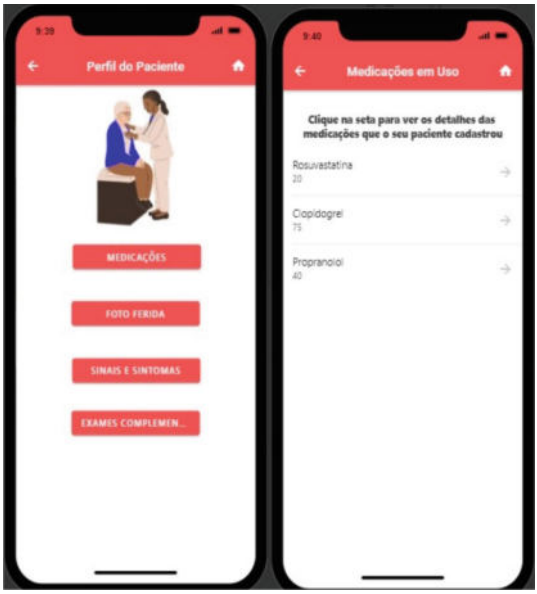


Fig. 4. Healthcare professionals' screens

4.2 SUS Score

According to the sociodemographic data, the population of this research was predominantly female, 55.6% (10); mean age of 38.2 (±9) years; 50% work as health professionals and the rest in the technology area (among them: systems analysis and development, software development, computer science, data science, computer engineering and user experience design), on average they had 12.2 (±9) years of experience; 61.1% usually use mobile applications related to health and 66.7% have never participated in scientific research before.

After participants completed the SUS questionnaire, each participant's final score was calculated, and the SUS classification level was assigned to each value obtained

(Table 2). Evaluation of usability has shown to be crucial for confirming how users and the system interact. After performing the Shapiro-Wilk test, it was found that the final score data did not pass normality. According to the box-plot analysis, this situation is justified since participant P11 was an outlier. However, it was decided to keep the participant due to the small sample size. The median SUS final score was 95 (IQ 90;97.5), ranging from a minimum score value of 62.5 to a maximum of 100. In summary, participants provided an overall application rating that reflects high satisfaction.

Table 2. SUS global classification according to each participant's final score

Participant	Final Score
P1	97,5
P2	90
P3	97,5
P4	92,5
P5	87,5
P6	77,5
P7	100
P8	100
P9	95
P10	95
P11	62,5
P12	100
P13	85
P14	97,5
P15	90
P16	95
P17	100
P18	97,5

4.3 Non-mandatory Open-Ended Questions

Two questions followed the completion of the SUS scale. "Do you have any suggestions for improving the application?" (Table 3) was the first question. This inquiry seeks to ascertain the areas in which the application requires improvement.

It was observed that, except for participant P17 (nursing), no other suggestions for improving the content were offered by the health professionals involved. This may have occurred because they did not belong to the cardiology area.

Participant P8 suggested that urgent alerts be added. This functionality was not tested, as at this stage of the prototype's development, there was no mutual interaction between healthcare professionals and the patient.

Participant P16 suggested several improvements. He also pointed out that the healthcare professional could carry out depression, pain, quality of life, and mood scales. However, we intentionally kept this option open to allow professionals to explore and understand how this feature how this feature works within the patient app.

Table 3. Suggestions obtained to the first non-mandatory question "Do you have any suggestions for improving the application?"

Participant	Suggestions
P5	Suggests starting content after login with the "General" menu and adding a "Cardiac Procedures" menu
P8	Suggests the inclusion of an urgency alert when identified
P9	Suggests the option to view the password on the login screen
P11	Suggests improvements in the visual part and in some fields. However, he/she does not specify which
P13	Mentions that he/she does not remember the system authorizing access only after receiving a confirmation email after registration. Participant also observed that some checkbox-type inputs behaved like radio-type inputs during initial tests, but he/she doesn't remember which screen this occurred on
P15	Suggests reviewing navigability
P16	a) Suggests that the link to accept the terms is clear and visible when starting the application b) Questions the need to request the number of Personal National Identification, suggesting that the email may be sufficient as identification, especially considering the use of Android and iPhone c) Mentions that, when accessing as a professional, it is also possible to take the "test" to measure the scales (depression, pain, quality of life and mood). Furthermore, it suggests that these tests can be carried out on the home screen, inviting the user to carry them out d) Suggests that a list be displayed with patients who had changes, indicating with colors (green for positive changes and red for negative changes) and a score as a metric for professionals e) Proposes improvements in the "Patient Profile" sections, such as adding start and end dates for medication history, including the number of days with reference to the surgical procedure in images for healing analysis, associating information on days after the procedure in Signs and Symptoms and exams, and extract information from exam images using Natural Language Processing to automatically identify abnormal values and alert the medical doctor
P17	Suggests the inclusion of a control table for the International Normalized Ratio result and anticoagulant adjustment
P18	Suggests adjustments to the app's appearance (without specifying which ones)

Finally, it is worth highlighting that of the 14 answers obtained to the question: "What were your biggest difficulties when using the application?" (Table 4), none reported difficulties regarding usage of the application. Two of the fourteen answers to the question that addressed additional topics are shown in Table 4.

Table 4. Additional topics obtained to the non-mandatory question "What were your biggest difficulties when using the application?"

Participant	Answer
P5	The participant mentions that, as he was not from the health area, he/she had difficulty understanding the meaning of cardiac procedures. However, he/she did not mention difficulties using the application
P11	No difficulties were reported, however, the participant reports that the graphic part is not attractive, and the reports could have time-evolution graphs, the scales could be more attractive to choose from. He/she states that the statements mix with the titles, there is a lack of a share button, a lack of FAQ or HELP. Finally, he/she suggests using a conversational agent to capture the user's attention

5 Discussion

Regarding the developed application prototype's usability, the research provided positive results (Median 95; IQ 90; 97.5). The findings show that participants accepted the system well, with the majority of reviews (83.3% gave it high rates).

The implementation of application pilots, as demonstrated in this research, is a widely endorsed approach during the development and testing phases, as supported by literature references [33, 34]. This study conducted a thorough usability assessment of the mobile application designed for healthcare professionals. It employed various strategies to identify and rectify issues in user-system interactions, aiming to enhance the application's functionality. The feedback from participants was indicative of their deep engagement and interest in the ongoing refinement of the application. Their recommendations primarily focused on visual improvements and interface adjustments, including design enhancements and modifications to input fields. Additionally, they proposed functional enhancements such as the inclusion of urgency alerts, password visibility options, testing of scales on the home screen, and the display of metrics for professional use. These suggestions reflect the participants' concern with usability and improving the user experience, highlighting the importance of intuitive and functional interface design. However, it is essential to highlight that the application was not tested in real-life scenarios, particularly during professional-patient interactions. Consequently, professionals couldn't observe alerts and consultation requests in real-time. These aspects were only evaluated through interactions with two "model patients," where fictitious data had been included for simulated tests.

The majority of participants in the survey reported a seamless experience with the application, expressing favorable views on its usability. This outcome is significant, as

it suggests that the application was crafted with an emphasis on intuitiveness and user-friendliness, thus enhancing user engagement and ensuring a positive user experience. This ease of use appears to transcend educational backgrounds and prior experience in scientific research, indicating that the application's design elements contributed to its simplicity and accessibility. Moreover, the absence of reported challenges may be seen as an indication of the effectiveness of the application's instructional interfaces, which offer clear directions and suitable guidance to users. This is especially relevant since 50% of the participants were not in the healthcare industry and may not have known much about cardiac procedures in the past. When an application is difficult to use, unclear, or counterintuitive, users may become discouraged and frustrated, which can lead to low adoption and abandonment [35].

Regarding the existence of mobile applications already developed with the same theme as this work, the literature does not present articles, to date, based on the PSD principles of Oinas-Kukkonen & Harjumaa [12]. The applications found after literature review [15–20] presented some features described in the PSD, however, without reference to persuasive technology, indicating that they were not designed with this purpose in mind, different from us. Therefore, the application developed in this study stands out as an innovative contribution in the research context at hand.

While there are existing studies on cardiac rehabilitation programs using Persuasive System Design (PSD), Salvi et al. [36] developed a notable mobile health system aimed at encouraging patients to participate in cardiac rehabilitation post-coronary artery disease. This system features functionalities for exercise tracking, guidance, motivational feedback, and educational materials. They conducted a randomized controlled trial to compare this mobile rehabilitation approach with standard care. The patient interface of the system included sections like "home," "messages," "calendar," "exercise," and "learning," while the professional interface facilitated initial assessments, progress tracking, and alert generation for complications. Despite encountering some technical challenges, the study revealed high user acceptance, perceived usefulness, and enhanced educational outcomes. However, the authors emphasized the need for further research to validate these findings and to affirm the efficacy of the design methodologies employed.

A systematic scoping review carried out by Ramachandran et al. [43] examines the acceptance of technology in cardiac telerehabilitation programs in patients with coronary artery disease. Although the article's objectives differ from ours, there is direct relevance to this research, as both studies address the use of technology in the context of cardiac care. Furthermore, both seek to understand the acceptance and usability of applications by users. The findings highlight the importance of technology acceptance by users, as well as the perception of usefulness and resulting educational benefits. These aspects were also evaluated in the present research, in which the results indicated a favorable acceptance of the application by professionals. Participants also demonstrated a high level of involvement, offering suggestions for improvement and showing a significant interest in the continuous improvement of the tool. Thus, both studies emphasize the relevance of technology acceptance and application usability in cardiac care. Furthermore, they also addressed the usability issue. The authors stress the significance of creating user-friendly interfaces that are easy to use and promote pleasant user interaction. This

strategy is consistent with our research, which also assessed the developed application's usability. To improve patient adherence and engagement outcomes, it is crucial to consider technology acceptance and usability when developing and implementing applications in cardiac care, as demonstrated by the results of this study and those of Ramachandran et al. [37].

Notably, not every intended PSD feature could be implemented during the application development process. It was not possible to create a ranking system, send out reminders, or integrate gamification for patients using stars and badges. These limitations arose because the focus of usability tests was aimed at professionals at this stage. Regarding what was intended in the methodology, the principles of supporting the primary task and supporting the system's credibility were thoroughly covered. Nonetheless, it is recommended that the postoperative phase of cardiac procedures be tailored into the categories of young and elderly adults to provide pertinent functionality for these various groups, which have different profiles, to further improve the persuasive components.

The necessity for a more detailed evaluation of adherence to the developed application in future research is crucial. While the current results indicate positive user reception and significant engagement with the application, a thorough examination of its effectiveness in enhancing therapeutic adherence, particularly using patient samples, is needed. In this vein, the study by Al-Arkee et al. [38] underscores the effectiveness of mobile applications in improving medication adherence for cardiovascular diseases. Their review, which included 16 randomized controlled trials, found that nine trials showed a statistically significant increase in medication adherence in the intervention group. Moreover, a meta-analysis of six trials indicated that mobile app-based interventions had a positive and significant impact on medication adherence, though no significant correlation was found between the duration of application use and therapeutic adherence outcomes. This parallels the findings of the current study, both underscoring the critical role of therapeutic adherence in cardiovascular diseases and the potential of mobile applications to enhance it. These studies lay a strong foundation for future research aimed at assessing the effectiveness and adherence to the developed application, as well as exploring additional methods to increase therapeutic adherence and improve cardiovascular treatment outcomes. Additionally, this research contributes to the growing body of scientific evidence supporting the use of persuasive technology and Persuasive System Design (PSD) in facilitating behavior change, underscoring their value and necessity in healthcare interventions.

Regarding the study's limitations, it is stated that the initial pilot application will require future improvements, as the needs of professionals and patients may change over time. Furthermore, this study did not evaluate usability from the patient's point of view, and future work is recommended. Additionally, since this study was developed in Brazil, there might be cultural variations in the application's content and user interface. Finally, the selection technique used to obtain the sample was non-probabilistic for convenience, resulting in a small sample size ($n = 18$).

6 Conclusion

The noteworthy research findings show that users were highly satisfied with the developed application, which prioritized usability and interface design.

The outcomes of this study show that success in prioritizing usability and user experience can be achieved when developing health-related mobile applications using a user-centered persuasive systems design approach.

Finally, it is desirable to conduct clinical trials to investigate the application's potential benefits in reducing postoperative complications and improving health outcomes. These studies will allow for a more in-depth assessment of the clinical impact of the application, providing robust scientific evidence and contributing to its validation as an effective tool in the context of postoperative cardiac procedures. Based on the results obtained, it will be possible to identify new opportunities for improvement and development of additional resources, consolidating the application as a comprehensive and efficient solution for patients and healthcare professionals

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Virani, S.S., et al.: American heart association council on epidemiology and prevention statistics committee and stroke statistics subcommittee. heart disease and stroke statistics-2021 update: a report from the american heart association. Circulation **43**(8), e254-e743 (2021)
2. Pan American Health Organization. Cardiovascular disease burden in the Region of the Americas, 2000–2019. ENLACE data portal. Pan American Health Organization. 2021. Homepage, https://www.paho.org/en/enlace/cardiovascular-disease-burden. (Accessed 18 Nov 2022)
3. Oliveira, G.M.M., Brant, L.C.C., Polanczyk, C.A., Malta, D.C., Biolo, A., Nascimento, B.R., et al.: Cardiovascular Statistics - Brazil 2021. Arq. Bras. Cardiol. **118**(1), 115–373 (2022)
4. American Society of Anesthesiologists. Heart Surgery. 2023. Homepage, https://www.asahq.org/madeforthismoment/preparing-for-surgery/procedures/heart-surgery/, (Accessed 17 Feb 2023)
5. Pahwa, S., Bernabei, A., Schaff, H., Stulak, J., Greason, K., Pochettino, A., et al.: Impact of postoperative complications after cardiac surgery on long-term survival. J. Card. Surg. **36**(6), 2045–2052 (2021)
6. Akinosun, A.S., Polson, R., Diaz-Skeete, Y., De Kock, J.H., Carragher, L., Leslie, S., et al.: Digital technology interventions for risk factor modification in patients with cardiovascular disease: systematic review and meta-analysis. JMIR Mhealth UhealthMhealth Uhealth **9**(3), e21061 (2021)
7. Cruz-Ramos, N.A., Alor-Hernández, G., Colombo-Mendoza, L.O., Sánchez-Cervantes, J.L., Rodríguez-Mazahua, L., Guarneros-Nolasco, L.R.: MHealth apps for self-management of cardiovascular diseases: a scoping review. Healthcare (Basel) **10**(2), 322 (2022)
8. Spring, B., et al.: American heart association behavior change committee of the council on epidemiology and prevention, council on lifestyle and cardiometabolic health, council for high blood pressure research, and council on cardiovascular and stroke nursing. better population health through behavior change in adults: a call to action. Circulation **128**(19), 2169–76 (2013)
9. https://www.heart.org/-/media/Files/Health-Topics/Heart-Attack/5-Ways-To-Lower-Your-Risk-of-Second-Heart-Attack-infographic.pdf. AHA 2022
10. McLean, A.: MHealth apps as effective persuasive health technology: contextualizing the "necessary" functionalities. JMIR Nursing **3**(1), e19302 (2020)

11. Fogg, B.J.: Persuasive technology: using computers to change what we think and do. Ubiquity 2002, December, Article 5 (1 December - 31 December 2002), 32 pages. https://doi.org/10.1145/764008.763957
12. Oinas-Kukkonen, H., Harjumaa, M.: Persuasive systems design: key issues, process model, and system features. Commun. Assoc. Inf. Syst. **24**(1), 28 (2009)
13. Markkanen, J.O., et al.: Mobile health behaviour change support system as independent treatment tool for obesity: a randomized controlled trial. Int. J. Obes., 1–8 (2023)
14. World Health Organization. Cardiovascular diseases (CVDs). 2021. Homepage, https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), (Accessed 12 Sep 2022)
15. Lowres, N., et al.: Self-monitoring for atrial fibrillation recurrence in the discharge period post-cardiac surgery using an iPhone electrocardiogram. Eur. J. Cardiothorac. Surg. **50**(1), 44–51 (2016)
16. Biersteker, T.E., et al.: Mobile health vs. standard care after cardiac surgery: results of The Box 2.0 study. Europace. **25**(1):49–58 (2023)
17. Yu, C., et al.: MISSION-2 Collaborative Group. Smartphone-based application to improve medication adherence in patients after surgical coronary revascularization. Am. Heart J. **228,** 17–26 (2020)
18. Lamberigts, M., et al.: Remote heart rhythm monitoring by photoplethysmography-based smartphone technology after cardiac surgery: prospective observational study. JMIR Mhealth UhealthMhealth Uhealth **9**(4), e26519 (2021)
19. Atilgan, K., et al.: Remote patient monitoring after cardiac surgery: The utility of a novel telemedicine system. J. Card. Surg. **36**(11), 4226–4234 (2021)
20. Aydın, A., et al.: Respiratory function and drug compliance postoperative monitoring with a mobile application after coronary artery bypass graft surgery: a randomized controlled trial. Surgical and Vascular Nursing. **16**(2), 53–59 (2022)
21. Fogg, B.J.: A behavior model for persuasive design. In: Persuasive 2009: Proceedings of the 4th International Conference on Persuasive Technology, vol. 40, pp. 1–7 (2009)
22. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design science research methodology for information systems research. J. Manag. Inf. Syst. **24**(3), 45–77 (2007)
23. Brooke, J.: SUS: A quick and dirty usability scale. Usability Eval. Ind. **189** (1995)
24. Hartson, R.; Pyla, P.S.: The UX Book: Process and Guidelines for Ensuring a Quality User Experience, p. 973. Morgan Kaufmann, São Francisco (2012)
25. Martins, A.I., Rosa, A.F., Queirós, A., Silva, A., Rocha, N.P.: European Portuguese validation of the system usability scale (SUS) Procedia Comput. Sci. **67**, 293–300 (2015)
26. Lourenço, D.F., Carmona, E.V., Lopes, M.H.B.M.: Translation and cross-cultural adaptation of the system usability scale to Brazilian Portuguese. Aquichan. **22**(2), e2228 (2022)
27. Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: adding an adjective rating scale. J. Usability Stud. **4**, 114–123 (2009)
28. Padrini-Andrade, L., et al.: Evaluation of usability of a neonatal health information system according to the user's perception. Revista paulista de pediatria:orgao oficial da Sociedade de Pediatria de Sao Paulo **37**(1), 90–96 (2019)
29. Miot, H.A.: Avaliação da normalidade dos dados em estudos clínicos e experimentais. J Vasc Bras. **16**(2), 88–91 (2017)
30. Ministério da Saúde. Resolução Nº 466, de 12 de Dezembro de 2012. Homepage. https://conselho.saude.gov.br/resolucoes/2012/Reso466.pdf, (Accessed 10 Oct 2022)
31. Brasil. Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet).Brasília, DF: Presidência da República (2018)
32. American Heart Association. Life's essential 8. 2022. Homepage, American Heart Association. Life's essential 8 (Accessed 17 Oct 2022)

33. Middleton, B., et al.: American Medical Informatics Association. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. J Am Med Inform Assoc. **20**(e1), e2–8 (2013)
34. Ratwani, R.M., Zachary, H.A., Kosydar, A., Fairbanks, R.J., Hodgkins, M.L.: A framework for evaluating electronic health record vendor user-centered design and usability testing processes. J. Am. Med. Inform. Assoc. **24**(e1), e35–e39 (2017)
35. Blue Sky Graphics. How Important Is the User Experience When Designing Applications? Homepage,https://blueskygraphics.co.uk/how-important-is-the-user-experience-when-designing-applications/ , (Accessed 23 June 2023)
36. Salvi, D., et al.: An m-Health system for education and motivation in cardiac rehabilitation: the experience of HeartCycle guided exercise. J. Telemed. Telecare **24**(4), 303–316 (2018)
37. Ramachandran, H.J., Jiang, Y., Teo, J.Y.C., Yeo, T.J., Wang, W.: Technology acceptance of home-based cardiac telerehabilitation programs in patients with coronary heart disease: systematic scoping review. J. Med. Internet Res. **24**(1), e34657 (2022)
38. Al-Arke, S., et al.: Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet **23**(5), e24190 (2021)

Effectiveness of Robot-Assisted Lower Limb Rehabilitation on Balance in People with Stroke: A Systematic Review, Meta-analysis, and Meta-regression

Riku Yli-Ikkelä[1], Aki Rintala[2], Anna Köyhäjoki[3], Harto Hakonen[4], Hilkka Korpi[1,5,6], Mirjami Kantola[1], Sari Honkanen[1], Outi Ilves[1,7], Tuulikki Sjögren[1], Juha Karvanen[8], and Eeva Aartolahti[9(✉)]

[1] Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland
[2] Physical Activity and Functional Capacity Research Group, Faculty of Health Care and Social Services, LAB University of Applied Sciences, Lahti, Finland
[3] Assistive Equipment Center, Central Ostrobothnia Wellbeing Service County "Soite", Kokkola, Finland
[4] Jamk University of Applied Sciences, LIKES, Jyväskylä, Finland
[5] Well-Being and Culture Unit, Oulu University of Applied Sciences, Oulu, Finland
[6] Social and Health Care Unit, Vaasa University of Applied Sciences, Vaasa, Finland
[7] Department of Sports and Rehabilitation, South-Eastern Finland University of Applied Sciences, Savonlinna, Finland
[8] Faculty of Mathematics and Science, University of Jyväskylä, Jyväskylä, Finland
[9] Institute of Rehabilitation, Jamk University of Applied Sciences, Jyväskylä, Finland
eeva.aartolahti@jamk.fi

Abstract. The objective of this study was to evaluate the effectiveness of robot-assisted lower-limb rehabilitation on balance in stroke patients and to explore the covariates associated with these effects.

A systematic literature search was carried out in four databases (MEDLINE (Ovid), CINAHL, PsycINFO, and ERIC) for studies published from inception to 25th of March 2022. Studies on robot-assisted lower-limb rehabilitation with a randomized controlled trial (RCT) design, participants with stroke, a comparison group with conventional training, and balance-related outcomes were included. Studies were assessed for Cochrane Risk of Bias 2 and quality of evidence. Meta-analysis and meta-regression were performed.

A total of 48 (RCT) with 1472 participants were included. The overall risk of bias in the included studies was unclear (n = 32), high (n = 15) or low (n = 1). Compared to conventional rehabilitation, robot-assisted lower-limb rehabilitation interventions were more effective for balance improvement (Hedges' g = 0.25, 95% CI: 0.10 0.41). In meta-regression, a relationship between the training effect was observed with the time since stroke, explaining 56% of the variance (p = 0.001), and with the ankle robots, explaining 16% of the variance ($p = 0.048$). No serious adverse events related to robot-assisted training were reported.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-59091-7_7.

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 101–116, 2024.
https://doi.org/10.1007/978-3-031-59091-7_7

Robot-assisted lower-limb rehabilitation may improve balance more than conventional training in people with stroke, especially in the acute stage. Robot-assisted lower-limb rehabilitation seems to be a safe rehabilitation method for patients with stroke. To strengthen the evidence, more high-quality RCTs with adequate sample sizes are needed.

Keywords: Robotics · Lower extremity · Exercise · Stroke Rehabilitation · Postural Balance · Meta-Analysis

1 Introduction

Stroke is one of the main causes of disability [1], with motor impairment being the most common [2]. Since stroke may affect the visual, vestibular, and somatosensory systems, balance impairments are common after stroke [3]. These impairments increase the risk of falls, with 73% of patients with stroke falling in the first year [4]. Poor balance impairs independent living and daily activities and increases the fear of falling in patients with stroke [3].

Robot-based neurorehabilitation is a rapidly growing field that uses robots to treat neurological injuries [5]. Systematic reviews have indicated that robot-assisted rehabilitation has more positive outcomes for stroke patients in improving walking and motor recovery than conventional training [6, 7]. Robotic devices can be classified into exoskeletons, end effectors, and upper- and lower-limb robots [8]. Exoskeletons (e.g. Lokomat) uses programmable drives or passive elements to move a patient's knees and hips during gait [9]. End-effector-type devices (e.g., Gait Trainer) have footplates that mimic the stance and swing phases of gait [9]. Other robots for lower-limb rehabilitation include ankle robots [10] and robotic mobile devices [11, 12].

Previous review studies on robot-assisted lower-limb stroke rehabilitation have focused on gait outcomes. A Cochrane review [7] found that combining automated electromechanical and robot-assisted gait training with conventional physiotherapy increased the odds of independent walking and walking speed, but not walking capacity in a 6-min walk. This type of training may be beneficial during the acute rehabilitation phase [7].

A few review studies [13–18] have examined the outcomes of balance. Zheng et al. [13] studied the effects of robot-assisted therapy on the balance of patients with stroke. Separate meta-analyses of the Berg Balance Scale (BBS) and Fugl-Meyer balance scale scores showed that robot-assisted therapy was more effective than conventional treatment, with the exception of the Timed Up and Go test (TUG). The effects were not influenced by the type of robotic device or if robot-assisted therapy was combined with other interventions. A single meta-analysis that combined all balance outcomes was not conducted.

The most recent systematic review and meta-analysis by Loro et al. [14] found that compared with conventional training, the Berg Balance Scale results improved more in patients who received robot-assisted gait training, but TUG test results did not differ between groups. The results of different balance measures were not included in the analyses. Meta-regression was restricted to intervention-related factors, and a longer treatment duration was associated with better balance (TUG).

To the best of our knowledge, in studying the effects of robot assisted training on balance in people with stroke, no previous meta-analysis of randomized controlled trials (RCT) has investigated the effects of all kinds of lower-limb robotic training but limited mainly to gait training only. Previous meta-analyses did not pool different balance measures in the same analysis, leading to a limited number of studies included in the main analysis. Therefore, the aim of this systematic review and meta-analysis was to provide new and more extensive information about the effectiveness of robot-assisted lower-limb rehabilitation on balance in people with stroke and explore the association of covariates with this effect.

The following questions were addressed: 1) Does the effect of robot-assisted lower-limb rehabilitation differ from that of conventional rehabilitation on outcomes measuring balance in persons with stroke? 2) Are study factors, such as personal, clinical, or intervention characteristics, associated with the effects of robot-assisted rehabilitation on balance?

2 Methods

The protocol for this systematic review was registered in PROSPERO (CRD 42022319241). Reporting followed the PRISMA guidelines [19] (Supplementary Material).

2.1 Data Sources and Searches

The first phase of a systematic literature search was conducted in a larger project that studied the effectiveness and meanings of robotics, virtual reality, and augmented reality in medical rehabilitation [20]. The National Library of Medicine (MEDLINE), Cumulative Index to Nursing and Allied Health Literature (CINAHL), Psychological Information Database (PsycINFO), and Education Resources Information Center (ERIC) databases were searched from inception to November 12, 2019. An updated search was conducted after this review was registered from the same databases for studies published between August 2019 and March 25, 2022. The search strategy used either MeSH or keyword headings related to therapies and rehabilitation, robotics, robotic devices, and RCT study design. The search strategy for the Ovid MEDLINE database is presented in Supplementary Material. In addition, reference lists of previously published systematic reviews were searched to identify potential publications not included in the database search.

2.2 Study Selection

The screening for this review was performed in two phases. The first phase served the larger project with wider scope [20] and included a screening of potential studies using the PICOS (patient, intervention, comparison, outcome, study design) framework as follows: P) Adults or children requiring medical rehabilitation, I) Any type of robotic device designed for rehabilitation purposes, C) Conventional rehabilitation, wait-list-control, or other training modality different from experimental group, O) Body functions and

structures, activities, or participation according to International Classification of Functioning, Disability and Health (ICF), or quality of life, and S) RCT or crossover RCTs. The second phase was carried out after the updated search with more specified PICOS criteria to identify eligible studies of interest in this particular review: P) Adults (18 years of age or older) with stroke requiring medical rehabilitation, I) Any type of rehabilitation and physiotherapy intervention including lower-limb robotic device designed for rehabilitation purposes, C) Conventional rehabilitation and physiotherapy intervention without the use of a robotic device, O) Validated and standardized measures of balance, S) RCTs and crossover RCTs. Studies focusing on patients with other neurological disorders, comparing robot-assisted interventions with other robotic training modalities, and studies reporting only self-reported measures of balance (e.g., balance confidence) were excluded.

Two researchers (AK, SH, RY, MK, OI, and EA) independently screened the study titles and abstracts according to the inclusion criteria using Covidence [21]. After the completion of title and abstract screening, two researchers (AK, MK, SH, RY, EA, and OI) independently evaluated potential studies in the full-text phase by applying the inclusion criteria and reporting the reasons for exclusion of ineligible studies. A third reviewer (EA) evaluated the studies in case of disagreement.

2.3 Data Extraction and Quality Assessment

Data extraction was performed in Covidence according to the pre-determined format to report participants, interventions, and outcomes of the studies included in the review (Supplementary Material). Twelve original researchers were approached via email because of inadequate outcome data (emails were sent no more than three times), of which six researchers responded and provided the requested outcome data (Supplementary Material).

The Cochrane Risk of Bias 2 tool (RoB 2) [22] was used for the quality assessment of the included studies. Two researchers (RY, AK, MK, and OI) independently performed data extraction and quality assessment, and a third reviewer (EA) evaluated the studies in case of disagreement. If applicable, the previously published protocols and registry records of the included studies were retrieved to assess the risk of bias.

2.4 Data Synthesis and Analysis

The results of all eligible studies were pooled in a meta-analysis to provide an overall estimate of the effect of robot-assisted lower limb rehabilitation. Balance improvement was the primary measure of the treatment effect. The outcomes of balance were prioritized according to validity and reliability to combine the results from the studies in the analysis. A priority list of the chosen balance outcomes and the rationale thereof are provided in Supplementary Material. If the direction of the values differed, the values of each outcome variables were multiplied by -1 when needed so that the higher values reflected in the same direction in the analyses [23]. Only the first part of the trial was analyzed if the study used a randomized controlled crossover design. In the meta-analysis, the mean and standard deviation (SD) post-treatment values of continuous outcomes were obtained to calculate the intervention effect size (Hedges' g) and 95%

confidence intervals (CI) between the groups. The scale of Hedges' g was interpreted as follows: 0.20 to less than 0.50 was considered a small effect, 0.50 to less than 0.80 was considered a medium effect, and >0.80 was a large effect [24]. A random-effects model with restricted maximum-likelihood estimation was used in the meta-analysis because effect sizes are independent across studies, and it was hypothesized that effect size would vary across the populations tested. We computed the test of heterogeneity using a Q-test to confirm that the study effect size varied across samples, and the I^2 index was used to compute the variance explained by this heterogeneity. Bias caused by selective publication within studies was evaluated by assessing the funnel plot of the trial mean differences for asymmetry [25]. Effect sizes, corresponding variances and funnel plots were computed with Metafor package for R [26] and forest plots with forest plot package for R [27].

Meta-regression analysis was conducted with Metafor package for R. We computed Univariate Mixed effects model with intercept and restricted maximum-likelihood estimation to investigate whether certain study or clinical characteristics explain the proportion of the variance in the observed effect in the meta-analysis. Overall, 10 different covariates were analysed in relation to the quality of the study (risk of bias), content of intervention (study duration, number of training sessions per week, time of one training session, weekly total intensity of training, type of limb robotic device, robotic training alone or combined with conventional training), and clinical characteristics of rehabilitees (age, female sex in percent, time since stroke in months, and mean score of baseline Berg Balance Scale). Heterogeneity accounted for by the covariates was measured using (pseudo) R^2 [28].

The certainty of evidence according to the outcomes and meta-analysis was evaluated using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) guideline [29]. The quality of evidence was classified as high (i.e., further research is unlikely to change our confidence in the effect estimate), moderate (i.e., further research is likely to have an important effect on our confidence in the effect estimate), low (i.e., further research is highly likely to have an important effect on our confidence in the effect estimate), or very low (i.e., any estimate of the effect is highly uncertain).

3 Results

3.1 Study Selection

Database searches generated 2099 studies (Fig. 1). After the removal of duplicates and exclusion of irrelevant studies in two phases, first, according to the PICOS criteria of the larger project and second, the PICOS criteria of this review, 48 RCTs were included in the review and 41 were included in the meta-analysis. A list of excluded studies is provided in Supplementary Material with references and justifications for exclusion. All the included studies were published in English between 2006 and 2021. Detailed characteristics of the studies in the narrative synthesis (n = 48) are provided in Supplementary Material.

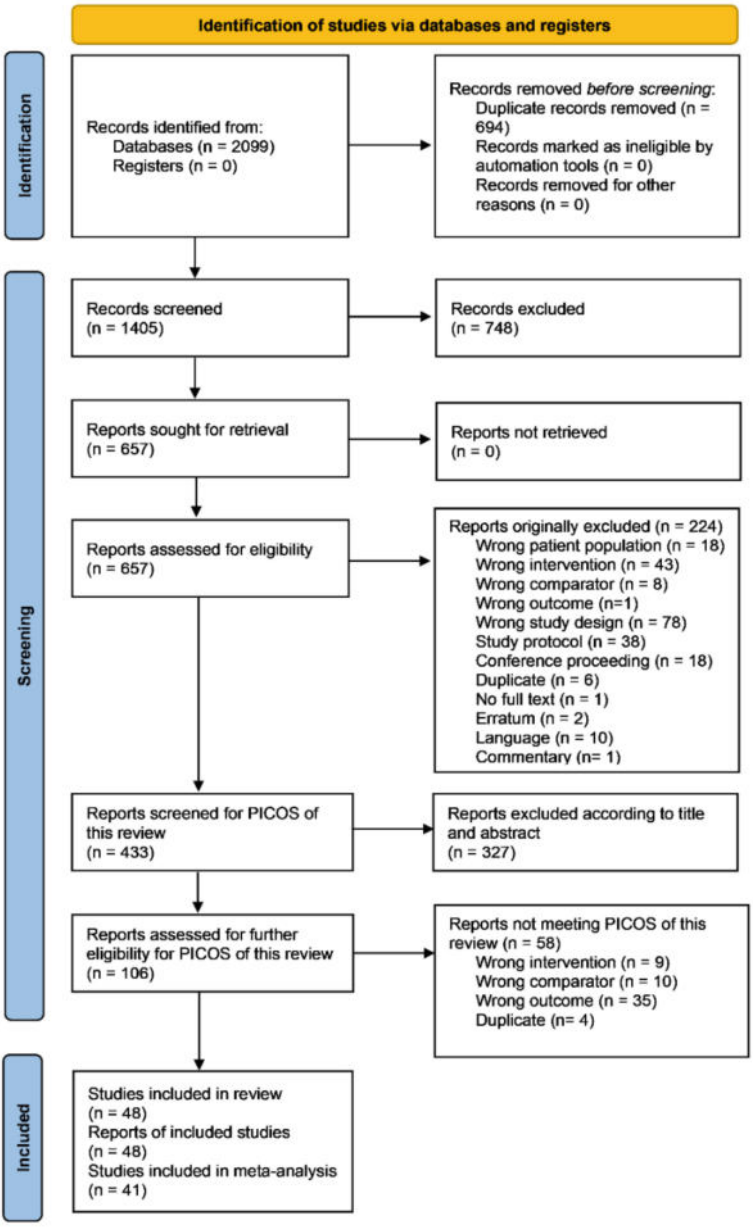


Fig. 1. Prisma flow diagram

3.2 Study Characteristics

Participants. A total of 1472 participants were involved in included studies. The sample sizes ranged from 6 to 37 in the experimental groups (mean: 15.39 $\pm$ 6.55 participants)

and from 6 to 30 in the comparison groups (mean: 14.96 ± 5.77). The mean age of the participants ranged from 44 to 76 years (mean: 59.87 ± 6.01 years). The percentage of women in the study group ranged from 0 to 64%. The mean time since stroke ranged from 11 days to 10 years (mean: 24.66 ± 33.67 months). More participants with ischemic stroke (67%) than with hemorrhagic stroke were involved. The type of stroke was not reported in 7 studies. Participants' functional ability varied at baseline, with some studies showing that all participants were able to walk and ambulate independently, while others had no participants who could walk without personal assistance.

Interventions. The duration of interventions ranged from 2 weeks to 20 weeks (mean: 5 ± 3 weeks). The most frequently used intervention duration was 4 weeks. The frequency of training ranged from twice a week to seven times a week (mean: 4 ± 1 times a week) and one session lasted 20–105 min (mean: 51 ± 25 min). The interventions were carried out in rehabilitation units and clinics (n = 17), hospitals and medical centres (n = 24), outpatient clinics (n = 3), or at home (n = 1). In three studies, the settings were not designated. Exoskeleton-type robotic devices were used in 34 studies, with Lokomat used in 13 studies. The other robotic devices used were end-effectors (n = 5), robotic ankle (n = 6), and robotic mobile devices (n = 3). In 21 studies, robot-assisted lower limb rehabilitation was offered with conventional rehabilitation or physiotherapy. Twenty-three studies had a follow-up period after intervention. Descriptions of the robot-assisted training protocols and devices are provided in Supplementary Material.

Comparisons. In 47 studies, the comparison groups underwent regular physiotherapy or conventional gait training. In one study, the control group underwent exercise training at home. Most often, comparison groups focused on gait training; however, in some studies, stretching and functional training were used for comparison. In all the studies, the training amount was similar between the comparison and intervention groups.

Outcomes. Balance was assessed using several measures. The results of the Berg Balance Scale (n = 35), Tinetti Balance Test (n = 3), and Timed Up and Go test (n = 9) were used in the studies included in the meta-analysis. One study assessed balance using the Sensory Organization Test (SOT), but this study was not included in the meta-analysis as balance was measured with laboratory devices. Thus, combining SOT with other balance measures was not considered appropriate for this meta-analysis.

3.3 Methodological Quality

The overall risk of bias in the included studies was unclear (n = 32), high (n = 15), or low (n = 1) (Fig. 2). The risk of bias in selective reporting was unclear in 41 studies (85%), high in five studies (11%), and low in two studies (4%). A funnel plot (Fig. 3) did not show any clear evidence of publication bias. The certainty of the evidence estimated using GRADE was considered low. The GRADE level was downgraded due to several studies with unclear or high risk of bias and inconsistency related to the heterogeneity of the original studies. The GRADE assessment is presented in 'Summary of Findings' -table. The risk of bias assessment of each included study as well as the 'Summary of Findings' -table are provided in Supplementary Material.

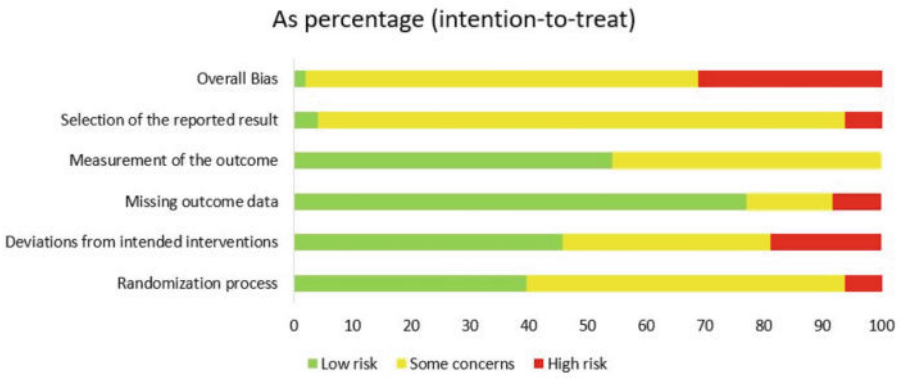


Fig. 2. Risk of Bias (%, intention-to-treat)

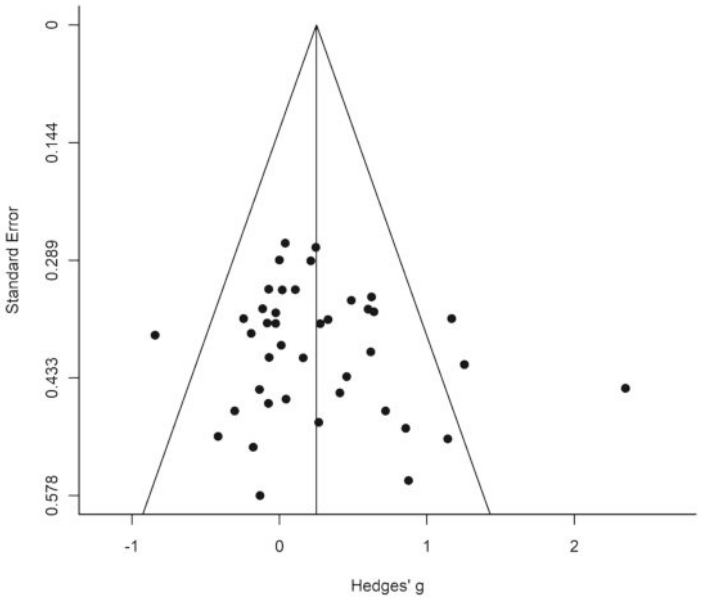


Fig. 3. Funnel plot

3.4 Effectiveness of Robot-Assisted Lower-Limb Rehabilitation on Balance

When comparing robot-assisted lower-limb rehabilitation to conventional rehabilitation methods in stroke patients, robot-assisted training showed a significant effect on the improvement of balance (Hedges' g = 0.25, 95% CI 0.10 to 0.41, 1192 participants, 41 studies; low-quality evidence) (Fig. 4). The level of statistical heterogeneity in the overall analysis was moderate ($I^2 = 42.7\%$).

3.5 Meta-regression

In meta-regression (Table 1), a relationship between the robot-assisted training effect and time since stroke was observed, explaining 56.0% of the variance of observed balance and indicating that the shorter the time in months since the stroke event, the greater the improvement in balance scores in the robot-assisted group compared with the conventional training group (point estimate −0.007; 95% CI, −0.012 to −0.003; p = 0.001) (Fig. 5). A relationship was also observed between the robot-assisted training effect and ankle robots, explaining 16.3% of the variance of observed balance (point estimate 0.552, 95% CI: 0.004 to 1.100, p = 0.048), indicating that more improvement in balance was achieved with ankle robots than with other types of robotic devices.

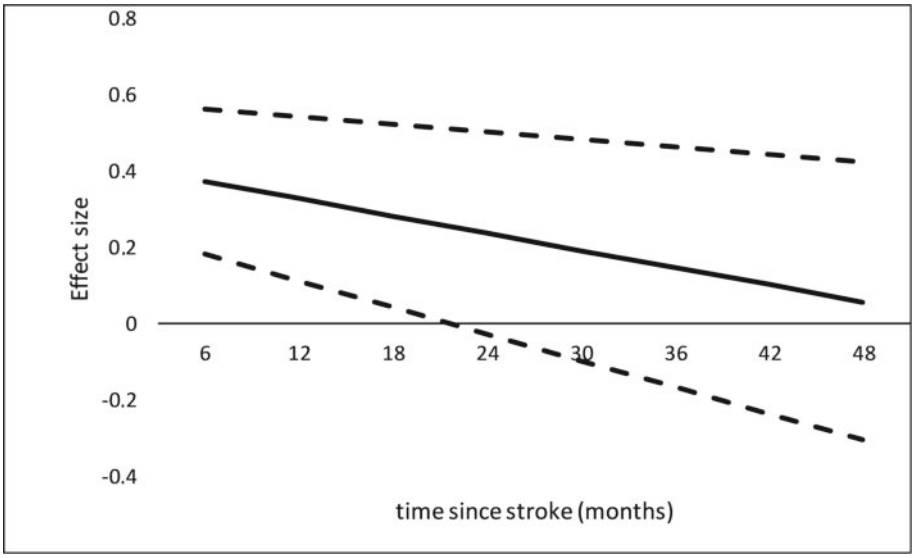


Fig. 4. Estimated effect of robot-assisted lower-limb rehabilitation (solid line) with 95% CI (dashed lines) according to time since stroke in months.

3.6 Adverse Events

Six of the 48 studies reported adverse events related to robot-assisted lower-limb rehabilitation. These events were mild, and no study reported serious adverse events. In a study by Calabro et al. [30], seven out of 20 patients in the experimental group experienced mild skin irritation and shank strap locations at the thigh. Hornby et al. [31] reported that two participants out of 24 dropped out because of leg pain during robotic training and one participant experienced pitting edema. In a study by Sczesny-Kaiser et al. [32], one of nine patients discontinued robot-assisted intervention due to intensive fatigue after each training session. In a study by Kang et al. [33], one person in the robot training group had recurrent skin problems and experienced skin abrasion in the tribal

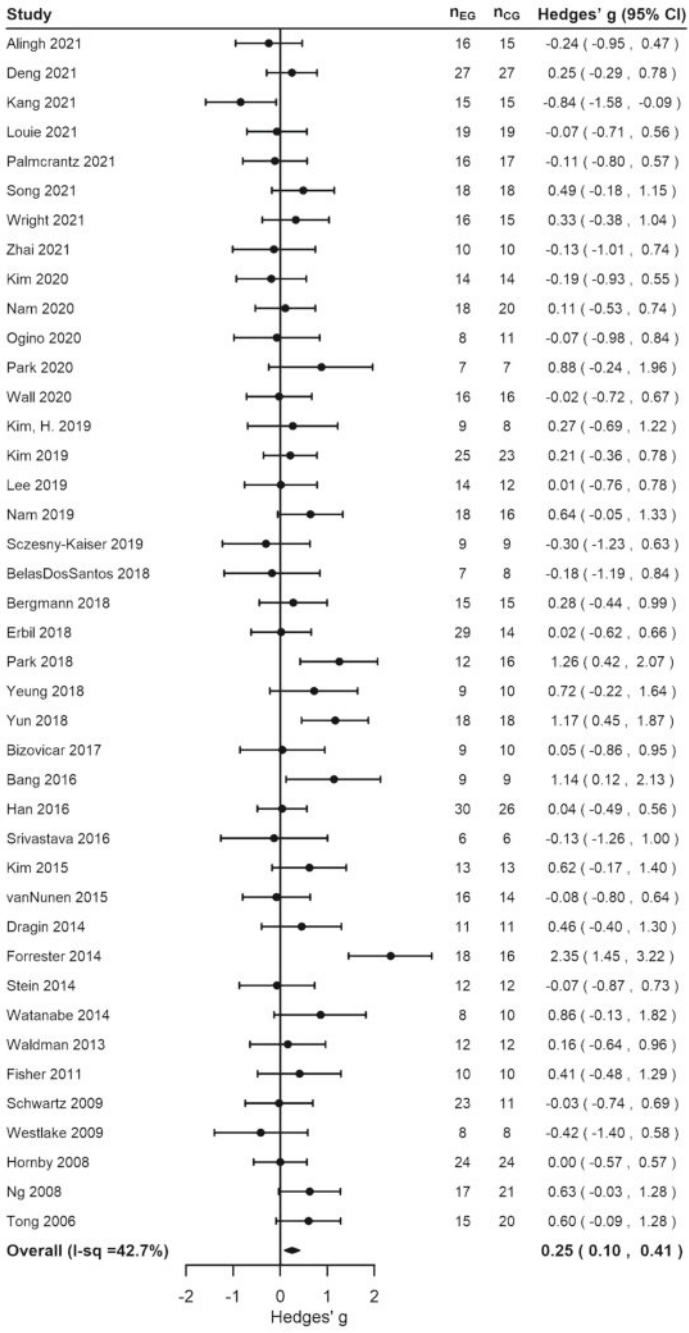


Fig. 5. Forest plot

Table 1. Results of the Meta-regression Analysis on Covariates Concerning the Study Factors and the High Risk of Bias domains

Covariates		Estimated Effect Size	SE	Lower CI	Upper CI	*P*	R^2 (%)	
Study factors								
Age (years)		0.018	0.013	-0.008	0.044	0.181	4.6	
Female (%)		0.012	0.006	-0.001	0.024	0.068	22.0	
Time since stroke (months)		-0.007	0.002	-0.012	-0.003	0.001	56.0	**
Intervention duration (weeks)		-0.040	0.030	-0.098	0.019	0.182	4.5	
Number of sessions per week		0.041	0.070	-0.096	0.178	0.557	0.0	
Session duration (min/session)		0.000	0.003	-0.006	0.007	0.954	0.0	
Intervention volume (min/week)		0.000	0.001	-0.001	0.001	0.943	0.0	
Type or interv.	Robotic exercise	Ref.						
	Robotic in addition to other exercise	-0.034	0.162	-0.351	0.283	0.834	0.0	
Type of robots	Other robotic types	Ref.						
	Exoskeleton	-0.285	0.181	-0.639	0.070	0.115	10.1	
	End-effector	0.120	0.230	-0.332	0.571	0.603	0.0	
	Ankle	0.552	0.280	0.004	1.100	0.048	16.3	*
	Robot mobile devices	-0.211	0.576	-1.340	0.917	0.714	0.0	
Berg Balance Scale at baseline (points)		-0.011	0.007	-0.024	0.002	0.099	9.4	
High Risk of Bias								
Overall		-0.285	0.156	-0.590	0.020	0.067	13.2	
Randomization process		-0.028	0.309	-0.635	0.578	0.927	0.0	
Deviations from intended interventions		-0.177	0.181	-0.532	0.178	0.330	0.0	
Missing outcome data		-0.259	0.216	-0.683	0.165	0.231	1.6	
Measurement of the outcome		-	-	-	-	-	0.0	
Selection of the reported result		-0.327	0.314	-0.942	0.288	0.298	0.0	

SE: Standard Error; CI: 95% Confidence Interval
* <0.05; ** <0.01

area. Palmcrantz et al. [34] reported that six persons experienced adverse events related to robotic training, including occasional transient redness or abrasion of the skin and discomfort or pain related to pressure from the HAL suit, attached shoes, or electrodes. In a study by Louie et al. [35], one person reported knee pain while wearing the robotic device, and three experienced transient pain or discomfort while using the exoskeleton. It did not affect their intervention adherence and could be resolved through device sizing adjustments. No adverse events occurred during robot-assisted training in 24 studies, and information related to adverse events was not provided in 18 studies.

4 Discussion

This systematic review and meta-analysis assessed the effectiveness of robot-assisted lower-limb training interventions on balance compared to conventional rehabilitation protocols in people with stroke. Robot-assisted interventions had a small, significant effect on balance (Hedges' g 0.25) [24]. The methodological quality of the studies was unclear and the evidence (GRADE) was low. Meta-regression showed that robot-assisted lower-limb rehabilitation was most effective in the early stages of rehabilitation and the sooner after stroke onset, the more significant the improvements. Ankle training robots may be the most beneficial, but the evidence is uncertain due to the small number of studies (n = 6). To the best of our knowledge, this is the largest review on this topic, with 41 RCTs and 1192 participants in the quantitative synthesis. We studied factors related to interventions, population, and study quality. Our intervention included all types of lower-limb robotic training and balance assessment combined with different measures based on a priority list. This study also reported adverse events from lower-limb robotic training and used GRADE to assess the evidence certainty.

Previous studies have shown that improvements in balance performance using conventional balance training interventions can be achieved at all stages of stroke rehabilitation, even more than 10 years post-stroke [36]. Based on our results, it appears that beneficial changes in the balance of stroke patients can be achieved with robot-assisted interventions. Similar findings were also reported in the Cochrane review [7] where the authors found that robot-assisted gait training may be more effective in enhancing walking-related outcomes in the acute stages of stroke rehabilitation in comparison to the chronic stages. Furthermore, the study revealed that using end-effector type devices for training exclusively resulted in significant improvements in gait speed and walking capacity for stroke patients, as opposed to conventional training methods [7]. Our review's findings suggest that the effect size for was higher also for balance outcomes of patients in the acute phase compared to those in the chronic phase. However, the effect sizes were not affected by whether the exoskeleton or end-effector type of robot was used.

The results of our study are mainly in line with those of other studies [13, 14] assessing the effectiveness of robot-assisted lower-limb rehabilitation on balance in patients with stroke, indicating that significant improvements in balance outcomes can be achieved with robot-assisted lower-limb training protocols. In our study, all suitable outcomes were included in the same meta-analysis; therefore, the results are not fully comparable with those of earlier studies [13, 14], which conducted independent meta-analysis for different outcomes of balance. However, the interpretations appear similar.

We conducted a meta-regression analysis to explore the association with the effects of covariates that were not investigated in previous studies. Our study differed from the previous meta-regression on this subject by Loro et al. [14], which focused on factors related to intervention protocols. They found an association between TUG results and treatment duration, indicating that the longer the robotic treatment phase, the greater the improvement in the TUG results. However, in our study, the effect sizes were not affected by any of the intervention-related covariates, such as treatment duration. It should be noted that Loro et al. [14] included only patients recovering from their first-ever stroke event, whereas our study also included recurrent events. Nevertheless, the results of our study did not provide clear answers regarding which training duration or intensity was the best for balance improvement. Future research should investigate the effects of different robotic training protocols to provide professionals important information for clinical settings.

Our study showed that the shorter the time since stroke, the greater the improvement in balance scores in the robot-assisted group than in the conventional training group. This suggests that robot-assisted training could be a useful alternative for individuals in the early phases of rehabilitation. Our meta-regression results also showed that baseline balance test scores and participant age did not affect the effect sizes of the analysis, indicating that robot-assisted training may be an option for people of different ages and levels of balance ability. Additionally, the results did not differ whether robotic training was the only training method or combined with some conventional training methods, contradicting the statement by Loro et al. [14] that a combination of robot-assisted gait training and conventional training is the most effective. Our findings suggest that the level of risk of bias in the original studies did not influence the results.

Robot-assisted training may be more beneficial than conventional training in improving balance in persons with stroke because it enables higher-intensity training, especially for most disabled patients [5]. Additionally, it may be beneficial for early retraining after stroke, when there is maximum plasticity and potential for recovery [37]. This may explain why robot-assisted training seems to improve balance and gait abilities more than conventional training, especially during the acute phase of recovery [7].

Previous reviews of robotic lower-limb rehabilitation in stroke patients have not reported any adverse events. Our study found that these events were mild and rarely reported, indicating that robotic lower-limb devices are generally safe for most patients with stroke. No adverse events occurred in 24 of the 48 studies included in the review. Six studies reported adverse events mostly related to skin irritation, discomfort, or pain during training. Many of these issues can be prevented by properly adjusting the robotic device before training.

4.1 Study Limitations

This review has some limitations that should be considered when interpreting the results and generalizability of the evidence. The quality assessment revealed several ratings of unclear and high risk of bias in the included studies, which led to the downgrading of the evidence quality. Almost all the studies had some concerns in the selection of the reported results' domain in the RoB 2 tool due to the lack of registration of the original study. One potential methodological limitation in these studies is the inability to blind participants and therapists, which may lead to performance bias [3]. Funnel plots were visually inspected and no clear evidence of publication bias was observed.

Statistical heterogeneity was present in the meta-analysis conducted for this review, which was one reason for downgrading the GRADE quality of evidence. There were many potential sources of heterogeneity among the included studies. Participant characteristics such as age, time since stroke event, stroke type, and severity of impairment at baseline differed widely across the studies. In addition, the use of robotic devices, duration and intensity of interventions, comparison training procedures and settings varied, and whether robot-assisted training was combined with other training. One limitation is the small sample sizes in the included studies; in many cases, the sample sizes were less than 20 in the experimental and comparison groups. Because of some methodological limitations and heterogeneity in the included studies, the certainty of the evidence is low, and more high-quality RCTs with adequate sample sizes are needed to improve the quality of evidence on this subject.

The challenge in studying the effects of robotic training is the rapid development of technology in relation to the slow production of effectiveness information. Newer technology may have different effects or disadvantages. On the other hand, it can be assumed that with the development of technology and professional competence, the goal is to improve the rehabilitation process.

4.2 Conclusions

The results of this systematic review and meta-analysis show that robot-assisted lower-limb rehabilitation may improve balance more than conventional training in stroke

patients, especially in the early stages of rehabilitation. Robot-assisted lower-limb rehabilitation also seems to be a safe rehabilitation method for stroke patients. This evidence suggests that physiotherapists and other rehabilitation professionals may consider robot-assisted lower limb rehabilitation as a useful rehabilitation method for improving balance in patients with stroke. However, more high-quality RCTs are required to strengthen this evidence.

Acknowledgements. Thank you to all who were part of the ROVA project at the University of Jyväskylä and contributed to this study.

Conflict of Interest Statement. The Authors declare that there is no conflict of interest.

Funding. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Social Insurance Institution of Finland (Kela) [grant number 50/26/2019].

Data Availability Statement. The data that support the findings of this study are available from the corresponding author, (EA), upon reasonable request.

References

1. World Health Organization. The top 10 causes of death (2020). https://www.who.int/news/room/fact-sheets/detail/the-top-10-causes-of-death. Accessed 18 May 2023
2. Kelly-Hayes, M., Robertson, J.T., Broderick, J.P., et al.: The American Heart Association stroke outcome classification. Stroke **29**, 1274–1280 (1998). https://doi.org/10.1161/01.STR.29.6.1274
3. Winstein, C.J., Stein, J., Arena, R.: Guidelines for adult stroke rehabilitation and recovery. Stroke **47**, 98–169 (2016). https://doi.org/10.1161/STR.0000000000000098
4. Denissen, S., Staring, W., Kunkel, D., et al.: Interventions for preventing falls in people after stroke. Cochrane Database Syst. Rev. **10** (2019). Art. No.: CD008728. https://doi.org/10.1002/14651858.CD008728.pub3
5. Morone, G., Paolucci, S., Cherubini, A., et al.: Robot-assisted gait training for stroke patients: current state of the art and perspectives of robotics. Neuropsychiatr. Dis. Treat. **13**, 1303–1311 (2017). https://doi.org/10.2147/NDT.S114102
6. Hobbs, B., Artemiadis, P.: A review of robot-assisted lower-limb stroke therapy: unexplored paths and future directions in gait rehabilitation. Front. Neurorobot. **14** (2020). Article 19. https://doi.org/10.3389/fnbot.2020.00019
7. Mehrholz, J., Thomas, S., Kugler, J., et al.: Electromechanical-assisted training for walking after stroke. Cochrane Database Syst. Rev. **10** (2020). Art. No.: CD006185. https://doi.org/10.1002/14651858.CD006185.pub5
8. Iosa, M., Morone, G., Fusco, A., et al.: Seven capital devices for the future of stroke rehabilitation. Stroke Res. Treat. (2012). Article ID 187965. https://doi.org/10.1155/2012/187965
9. Hesse, S., Waldner, A., Tomelleri, C.: Innovative gait robot for the repetitive practice of floor walking and stair climbing up and down in stroke patients. J. Neuroeng. Rehabil. **7**, 30 (2010). https://doi.org/10.1186/1743-0003-7-30

10. Shi, B., Chen, X., Yue, Z.: Wearable ankle robots in post-stroke rehabilitation of gait: a systematic review. Front. Neurorobot. **13**, 63 (2019). https://doi.org/10.3389/fnbot.2019.00063
11. Bizovičar, N., Matjačić, Z., Stanonik, I., et al.: Overground gait training using a motorized assistive device in patients with severe disabilities after stroke. Int. J. Rehabil. Res. **40**, 46–52 (2017). https://doi.org/10.1097/MRR.0000000000000199
12. Dragin, A.S., Konstantinovic, L.M., Veg, A., et al.: Gait training of poststroke patients assisted by the Walkaround (body postural support). Int. J. Rehabil. Res. **37**, 22–28 (2014). https://doi.org/10.1097/MRR.0b013e328363ba30
13. Zheng, Q., Ge, L., Wang, C.C., et al.: Robot-assisted therapy for balance function rehabilitation after stroke: a systematic review and meta-analysis. Int. J. Nurs. Stud. **95**, 7–18 (2019). https://doi.org/10.1016/j.ijnurstu.2019.03.015
14. Loro, A., Borg, M.B., Battaglia, M., et al.: Balance rehabilitation through robot-assisted gait training in post-stroke patients: a systematic review and meta-analysis. Brain Sci. **13**, 92 (2023). https://doi.org/10.3390/brainsci13010092
15. Maranesi, E., Riccardi, G.R., Di Donna, V., et al.: Effectiveness of intervention based on endeffector gait trainer in older patients with stroke: a systematic review. JAMDA **21**, 1036–1044 (2019). https://doi.org/10.1016/j.jamda.2019.10.010
16. Cho, J., Yoo, J.S., Kim, K.E., et al.: Systematic review of appropriate robotic intervention for gait function in subacute stroke patients. Biomed. Res. Int. (2018). Article ID 4085298. https://doi.org/10.1155/2018/4085298
17. Haarman, J.A.M., Reenalda, J., Buurke, J.H., et al.: The effect of 'device-in-charge' versus 'patientin-charge' support during robotic gait training on walking ability and balance in chronic stroke survivors: a systematic review. J. Rehabil. Assist. Technol. Eng. **3**, 1–16 (2016). https://doi.org/10.1177/2055668316676785
18. Swinnen, E., Beckwée, D., Meeusen, R., et al.: Does robot-assisted gait rehabilitation improve balance in stroke patients? A Systematic Review. Top. Stroke Rehabil. **21**, 87–100 (2014). https://doi.org/10.1310/tsr2102-87
19. Page, M.J., McKenzie, J.E., Bossuyt, P.M., et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ **372**, n71 (2020). https://doi.org/10.1136/bmj.n71
20. Ilves, O., Korpi, H., Honkanen, S., Aartolahti, E.: Effectiveness and meanings of robots, virtual and augmented reality in rehabilitation. Systematic literature reviews. Studies Soc. Secur. Health **159** (2022)
21. Veritas Health Innovation: Covidence systematic review software (2022). https://www.covidence.org/. Accessed 23 Jan 2024
22. Sterne, J.A.C., Savović, J., Page, M.J., et al.: RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ **366**, l4898 (2019). https://doi.org/10.1136/bmj.l4898
23. Higgins, J.P.T., Green, S.: Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration (2011). www.handbook.cochrane.org. Accessed 23 Jan 2024
24. Cohen, J.: A power primer. Psychol. Bull. **112**, 155–159 (1992). https://doi.org/10.1037/0033-2909.112.1.155
25. Sterne, J.A.C., Sutton, A.J., Ioannidis, J.P.A., et al.: Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ **343**, 1–8 (2011). https://doi.org/10.1136/bmj.d4002
26. Viechtbauer, W.: Conducting meta-analyses in R with the metafor package. J. Stat. Softw. **36**, 1–48 (2010). https://doi.org/10.18637/jss.v036.i03
27. Gordon, M., Lumley, T.: forestplot: advanced forest plot using 'grid' graphics. R package version 2.0.1. (2021)

28. Raudenbush, S.W.: Analyzing effect sizes: random effects models. In: Cooper, H., Hedges, L.V., Valentine, J.C. (eds.) The Handbook of Research Synthesis and Meta-Analysis, 2nd edn., pp. 295–315. Russell Sage Foundation, New York, NY, USA (2009)
29. Higgins, J., Thomas, J., Chandler, J., et al.: Cochrane Handbook for Systematic Reviews of Interventions version 6.0. Cochrane (2019). www.training.cochrane.org/handbook. https://doi.org/10.1002/9781119536604. Accessed 23 Jan 2024
30. Calabro, R.S., Naro, A., Russo, M., et al.: Shaping neuroplasticity by using powered exoskeletons in patients with stroke: a randomized clinical trial. J. Neuroeng. Rehabil. **15**, 35 (2018). https://doi.org/10.1186/s12984-018-0377-8
31. Hornby, T.G., Campbell, D.D., Kahn, J.H., et al.: Enhanced gait-related improvements after therapist- versus robotic-assisted locomotor training in subjects with chronic stroke a randomized controlled study. Stroke **39**, 1786–1792 (2008). https://doi.org/10.1161/STROKEAHA.107.504779
32. Sczesny-Kaiser, M., Trost, R., Aach, M., et al.: A randomized and controlled crossover study investigating the improvement of walking and posture functions in chronic stroke patients using HAL exoskeleton – the HALESTRO study (HAL-exoskeleton STROke study). Front. Neurosci. **13**, 259 (2019). https://doi.org/10.3389/fnins.2019.00259
33. Kang, C.J., Chun, M.H., Lee, J., et al.: Effects of robot (SUBAR)-assisted gait training in patients with chronic stroke: randomized controlled trial. Medicine **100**, e27974 (2021). https://doi.org/10.1097/MD.0000000000027974
34. Palmcrantz, S., Wall, A., Vreede, K.S., et al.: Impact of intensive gait training with and without electromechanical assistance in the chronic phase after stroke–a multi-arm randomized controlled trial with a 6 and 12 months follow up. Front. Neurosci. **15**, 660726 (2021). https://doi.org/10.3389/fnins.2021.660726
35. Louie, D.R., Mortenson, W.B., Durocher, M., et al.: Efficacy of an exoskeleton-based physical therapy program for non-ambulatory patients during subacute stroke rehabilitation: a randomized controlled trial. J. Neuroeng. Rehabil. **18**, 149 (2021). https://doi.org/10.1186/s12984-021-00942-z
36. Lubetzky-Vilnai, A., Kartin, D.: The effect of balance training on balance performance in individuals poststroke: a systematic review. J. Neurol. Phys. Ther. **34**, 127–137 (2010). https://doi.org/10.1097/NPT.0b013e3181ef764d
37. Rao, N., Zielke, D., Keller, S., et al.: Pregait balance rehabilitation in acute stroke patients. Int. J. Rehabil. Res. **36**, 112–117 (2013). https://doi.org/10.1097/MRR.0b013e328359a2fa

Virtual Reality in Rehabilitation of Executive Functions in Children (VREALFUN) – Study Protocols for Randomized Control Trials

Merja Nikula[1,2(✉)], Mirjami Mäntymaa[1,3], Steven M. LaValle[4], Ari Pouttu[4], Julia Jaekel[5,6,7,8], Eeva T. Aronen[9], Tytti Pokka[1,10], Juha Salmi[11], and Johanna Uusimaa[1,2]

[1] Research Unit of Clinical Medicine, Faculty of Medicine, University of Oulu, Oulu, Finland
merja.nikula@oulu.fi
[2] Paediatric Neurology Unit, Oulu University Hospital, Oulu, Finland
[3] Child Psychiatry Unit, Oulu University Hospital, Oulu, Finland
[4] Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland
[5] Faculty of Education and Psychology, University of Oulu, Oulu, Finland
[6] Department of Paediatrics I, Neonatology, Paediatric Intensive Care, Paediatric Neurology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany
[7] Department of Psychology, University of Warwick, Conventry, UK
[8] Public Health Unit, Finnish Institute for Health and Welfare (THL), Helsinki, Finland
[9] Child Psychiatry, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[10] Research Service Unit, Oulu University Hospital, Oulu, Finland
[11] The Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland

Abstract. Children with attention and executive function disabilities often have a long-lasting need for rehabilitation to support their functional ability. Yet the availability of rehabilitation services is insufficient, regionally unevenly distributed, and unequal in terms of access to rehabilitation. There is a need for easily accessible services. In this paper, we present the VREALFUN project where the major aim is to develop a novel Virtual Reality (VR) rehabilitation method for children with deficits in attention and executive functions. This ongoing Randomized Control Study (RCT) includes two arms, one in children with attention deficit hyperactivity disorder (ADHD) and the other in children with mild to moderate traumatic brain injury (TBI).

Keywords: Virtual Reality · Attention · Executive Functions · Rehabilitation · Child

1 Introduction

Children with attention deficit hyperactivity disorder (ADHD) and traumatic brain injury (TBI) often have deficits in their attention and executive functions causing disability in daily living. Many of these children need long-lasting rehabilitation, and there is a

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 117–124, 2024.
https://doi.org/10.1007/978-3-031-59091-7_8

growing demand for effective, cost-efficient, and feasible rehabilitation interventions where the training is targeted to support the everyday life functional ability of these children. Combining metacognitive skills and strategy training with skill practice seems to support everyday performance better [1]. Also, teaching parents to interact with their children more positively seems to promote their children's self-regulating ability [2, 3].

There is accumulating evidence that virtual reality (VR) can be effective in the rehabilitation of cognitive functions in children with ADHD [4, 5]. Also, VR in the rehabilitation of cognitive functions in adult patients with TBI has given some support [6], but in children with TBI the effectiveness of such a treatment method is still poorly understood. VR offers opportunities to build digitalized environments emulating situations where daily life attention and executive function deficits are manifested and to train skills helping to manage such challenging situations. In VR, it is also possible to expose the child to many repetitions in a highly motivating way which boosts learning of new skills.

In this registered (ClinicalTrials.gov, trials 206/2021 and 206/ 2021) randomized control study, the major aim is to develop and assess the feasibility of a novel rehabilitation method for children with deficits in attention, activity control and executive functions by using a virtual environment that corresponds to typical everyday life situations. Head-mounted displays (HMD) are used to present the tasks, and the levels of difficulty are adjusted according to the child's progress. The VREALFUN project consists of two RCT studies related to VR rehabilitation of attention and executive function deficits: S1) one conducted in children with ADHD and; S2) the other in children with mild to moderate TBI. After these two pilot studies, a national multicenter study with larger study groups will be set up.

We expect that; 1. Intensive training improves the attention regulation, activity control skills, and executive functions of the children in the intervention group; 2. Training of executive skills with motivating tasks in a virtual environment that is built to meet challenging everyday life situations transfers to the child‘s everyday life and; 3. The duration of the training effect does not depend on the success of the VR training itself, but on how well the child adopts new strategies that make everyday life easier and how the guardian is able to support the child's positive behavior in everyday life.

2 Subjects and Methods

This is a multidisciplinary project performed in collaboration with three faculties of the University of Oulu (Faculties of Medicine, Education and Psychology, and Information Technology and Electrical Engineering), two clinics of the Oulu University Hospital (Paediatric Neurology and Child Psychiatry Units), and researchers from the University of Helsinki, Helsinki University Hospital and Aalto University.

Two VREALFUN studies on VR rehabilitation in 8–12-years old Finnish-speaking children with ADHD and TBI are initiated in January 2024 and will be completed by the end of 2025. The Northern Ostrobothnia Regional Ethics Committee has approved the project plan on 30th August 2022 (EETTMK: 64/2021). The Wellbeing Services County of North Ostrobothnia has admitted research permission for this study on 23rd June 2023.

2.1 Participants

Eighty-eight children from the pediatric neurology and child psychiatry units of the Oulu University Hospital will be recruited for the ADHD study (S1) based on the informed consent of the guardian and the child, and randomized in three parallel intervention groups A, B, and C, and a treatment-as-usual (TAU) control group D (see below for a detailed description of the groups) with an allocation ratio 1:1:1:1. For the TBI study (S2) 44 children will be recruited, including one intervention group A and a TAU control group D, each of 22 children (an allocation ratio 1:1). Randomization will be performed by sealed envelopes and the randomization code will be released only after a baseline measure. The number of participants is based on sample size calculations for limited efficacy and effectiveness testing, with an expected meaningful training-induced behaviour change of 1 standard deviation mean difference between groups. The mean total score in the Behaviour Rating Inventory of Executive Function, Second Edition (BRIEF 2), is 50 and the standard deviation is 10 [7]. When considering that 10 would be a clinically significant difference in means, with a power of 80% and a statistical significance threshold of 0.05 the estimated sample size was 17 children per group. To consider possible dropouts of 20%, a total of 22 children will be recruited per group.

Patients in the control group are stratified to those in the intervention group for diagnosis (ADHD or TBI), age, sex, and very preterm birth status (< 32 weeks of gestation). Inclusion criteria in the ADHD group are the diagnosis of ADHD (ICD-10 F90.0) and methylphenidate medication; and in the TBI group, mild to moderate traumatic brain injury (ICD-10: S06.0-S06.6 and S06.8-S06.9, and criteria defined in the Current Care Recommendation 2021), and the challenges of attention and executive functions identified in the assessment of a neuropsychologist/experienced psychologist, as well as age 8–12 years and Finnish as a native language in both groups. The exclusion criteria in both groups include sensitivity to flashing light, epilepsy (ICD-10 G40), mental retardation (ICD-10 F70-F79), pervasive developmental disorders (ICD-10 F84), inflammatory diseases of the central nervous system (ICD-10 G00-G09), severe CP syndrome (ICD-10 G80), brain tumour, and twins/triplets, etc. In the ADHD group, TBI is also an exclusion criterion.

2.2 Procedure

The VREALFUN study set-up and methods are presented in Fig. 1 including four research visits: Baseline measure and three follow-up measures at 4–6 weeks, 6 months, and 12 months. The research methods are similar in ADHD and TBI cohort studies of which the ADHD cohort is randomized into four groups. In group A, children play the HMD-EPELI game and guardians get guidance to positive behavioural support and the introduction of a reward system (parental guidance); in group B the intervention is parental guidance only and no HMD-EPELI game is deployed; in group C children play the HMD-EPELI game only and no parental guidance is deployed. These interventions are described in more detail in Sect. 2.3 Intervention. Group D is a control group where children follow their rehabilitation plan drawn up in specialized medical care (treatment-as-usual). The TBI cohort with a smaller number of eligible patients is randomized into two groups (group A: HMD-EPELI game and parental guidance, and

group D: treatment -as usual). In all intervention groups (A, B, C) children may also receive conventional forms of rehabilitation or treatment, as planned in their rehabilitation plan, and no therapy or rehabilitation is discontinued because of the intervention offered in this VREALFUN study.

At the pretest and each follow-up visit, the neuropsychologist conducts the following neuropsychological examinations on children of all groups: The Conners Continuous Performance Test 3rd Edition (Conners CPT3) [8] which is a computerised attention task; n-back-test for working memory, and a virtual Executive Performance in Everyday Living (EPELI) task [9, 10] to measure the effectiveness of executive functions, time management, behaviour, task planning, memorization, and sensitivity to distractions. Furthermore, at each visit, the children in all research groups (A, B, C, D) are requested to fill out questionnaires about the functional ability of the child and the amount of positive feedback given by the guardians to the child (EPELI Questionnaire- child report, drawn up for this research), satisfaction with the assessment, and feelings of nausea and presence after playing the EPELI task using HMD and a hand controller as well as a questionnaire for measuring quality of life (KINDL-R Questionnaire for Measuring Health-Related Quality of Life in Children and Adolescents Revised Version, self-report) [11]. Guardians in groups A, B, C, and D are also requested to fill out the EPELI Questionnaire (parent report) and KINDL-R (parent version). Furthermore, they are asked to fill out questionnaires regarding the child´s executive function (BRIEF 2, parent form [7]) and ADHD symptoms (ADHD-rating scale IV (ADHD-RS), parent report) [12]. After each research visit based on the consent of the guardian the child´s teacher is sent BRIEF 2 (teacher form) and Concentration questionnaire [13] (in Finnish: Keskittymiskysely) which is an assessment of attention and executive function difficulties.

2.3 Intervention

There are three different intervention groups in the ADHD study where either the child plays the HMD-EPELI game (group C), the guardian gets guidance on positive behavioural support and use of a reward system (Group B), or both (Group A). In the TBI study there is one intervention group where the guardian gets parental guidance, and the child plays the HMD-EPELI game (group A) (Fig. 1). During the intervention period, guardians fill out a rehabilitation diary prepared for this study.

During the first research visit guardians in intervention groups A and B get guidance from the neuropsychologist on the use of self-care programs on the Health Village website on children's challenging behaviour, positive behavioural support, and the introduction of a reward system (https://www.mielenterveystalo.fi/fi/omahoito/lasten-haastavan-kaytoksen-omahoito-ohjelma). On this website, there is information on how to support the child‘s positive behaviour and use a reward system as well as rehearsals for the guardians on these topics. The guardians can return to these pages anytime they want to. A rehabilitation diary developed for this study will be introduced including four rehabilitation goals that are the same for all children of intervention groups of ADHD (groups A, B, and C) and TBI (group A) (Fig. 1), as well as two individual rehabilitation goals, which are defined together with the child and the guardian during the first research visit. Guardians fill in the rehabilitation diary daily for the first four weeks, and then once a week for five months, recording the actual activity of the child in accordance

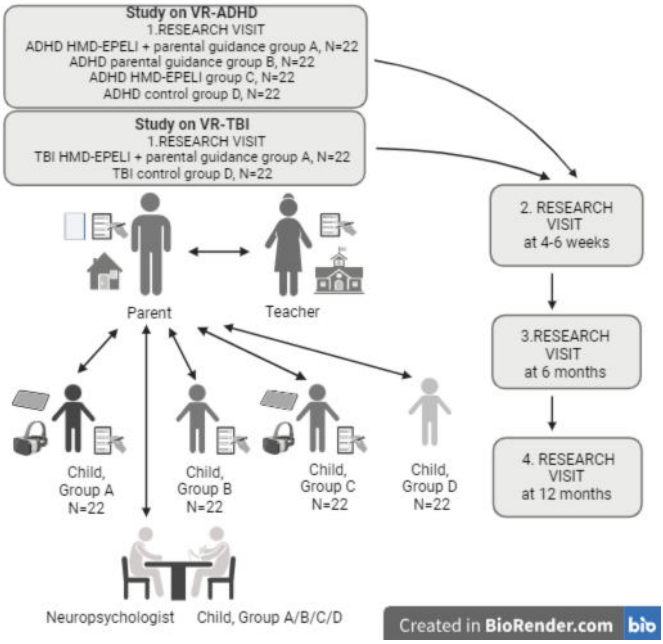


Fig. 1. The VREALFUN study set-up methods

with the rehabilitation goals, the amount of positive feedback given to the child, and the implementation of the rehabilitation plan drawn up in specialized medical care (e.g., occupational therapy sessions and any other rehabilitations, use of medication and possible changes in it). In addition, during the first research visit, the children of intervention groups A and C in the ADHD study and group A in the TBI study will be instructed in the use of the HMD-EPELI game under the supervision of the guardian for four weeks (five days a week, 30 min/day). This game is a VR rehabilitation program where the child rehearsals daily skills like homework and evening tasks in a virtual home environment instructed by a virtual character. After each rehearsal session, the child gets feedback in the form of a scoreboard. According to the child‘s progress, the level of difficulty is adjusted to keep rehearsal motivating. During the intervention period guardians also record in the diary how the gaming succeeded on those days the child has played. Gaming compliance is monitored through the server and guardians are reminded if necessary.

2.4 Research Ethical Considerations

In this study, research ethical principles are followed. Participation is based on voluntariness. The guardian and the child will be informed about the VREALFUN study in written form, and they have the possibility to ask questions concerning this study. Participation is based on informed consent and the participants can withdraw from this study at any time without any reason. Withdrawal does not influence on child‘s treatment. Information collected during research is confidential. This study does not cause any

pain or harm to the child, nor does it disturb their daily life. If the child feels a headache or nausea during the VR rehearsal, the child is instructed to stop playing. In the intervention groups participation in this study requires the guardian's commitment to home rehabilitation and time resources when they are following the child's rehearsal and/or filling the rehabilitation diary. In the intervention groups participation in this study might have some benefit, but not in the control groups during the research period. After the last follow-up visit the guardians in the control groups have a chance to get guidance on the use of self-care programs on the Health Village website about children's challenging behaviour, positive behavioural support, and the introduction of a reward system.

3 Results

As a part of this research project a new HMD-EPELI rehabilitation program has been developed (implemented by Peili Vision Company, which is not a collaborative partner of the research (http://www.peilivision.fi/)) as well as a rehabilitation diary to follow changes in each child´s functional ability during the intervention period. The EPELI diary will later be adapted into a mobile application for a digital care pathway. This is an ongoing study, the recruitment process has been initiated, and the data will be gathered by the end of 2025. Patients in the control group are stratified to those in the intervention group for diagnoses, age, sex, and very preterm birth status (<32 weeks of gestation) to control differences between groups other than the intervention. Other confounding factors (individual rehabilitation and pharmacotherapy) are considered by standardizing their effect in the analyses. The guardians of children in all intervention groups record in the rehabilitation diary daily the time of taking medication to support the child's attention and the individual rehabilitation received by the child (e.g. occupational therapy or neuropsychological rehabilitation). The results of RCT studies on ADHD and TBI cohorts will be published in high-quality international scientific journals.

4 Discussion

The prevalence of ADHD among children and young people in the general population is substantial, with 5–10% worldwide [14, 15]. However, the availability of rehabilitation services is insufficient, regionally unevenly distributed, and unequal in terms of access to rehabilitation. Attention difficulties, activity control and executive functions impair the functional capacity of these children and significantly increase their risk of social exclusion, academic underachievement, substance use, and co-morbidities including psychiatric disorders. A thesis on 200 families with ADHD children [16] revealed that families had unequal opportunities in getting support from educational, social and health sectors, and there was one socially excluded person or someone at risk of social exclusion in every third family. Furthermore, the Ministry of Education in Finland has estimated that every socially excluded young person costs EUR 1.2 million € for society and in Finland there are about 50,000 socially excluded people in the age groups between 15–29 years [17].

This VREALFUN study provides a potential to develop a novel VR-rehabilitation method with lower requirements for wireless communication (merely a wireless

consumer-grade Pico Neo 3 Pro Eye) enabling broader applicability and affordability. This project is linked to the strategic profiling 6 program of the University of Oulu, 6G-Enabling Sustainable Society (https://www.6gflagship.com/6gess/) that is based on 6G Flagship´s technological expertise to develop novel digital solutions for preventive healthcare and evaluate their feasibility, cost-effectiveness and impact on people´s health, lifestyles, and quality of life, and to develop virtual health care services involving digital care pathways, and implementation of novel health-related technologies. Thereby this VREALFUN project will promote long-term sustainability by supporting four of the sustainable development goals (SDGs) and their specified targets set by the United Nations: SDG3, good health and well-being (3B); SDG4, quality education (4.3); SDG5, gender equality; and SDG10, reduced inequalities.

Acknowledgments. This project is supported by the Arvo and Lea Ylppö Foundation, Helsinki, Finland, the Alma and K.A. Snellman Foundation, Oulu, Finland, the Finnish Brain Foundation, Helsinki, Finland, Academy of Finland (grant #352981 to Juha Salmi), Profiling 6 program 6G-Enabling Sustainable Society (6GESS), University of Oulu 6, Finland and FUWIRI (Research Infrastructure for Future Wireless Communication Networks), Academy of Finland.

Disclosure of Interests. The authors have no conflicts of interest relevant to the content of this article.

References

1. Cicerone, K.D., et al.: Evidence-based cognitive rehabilitation: Systematic review of the literature from 2009 through 2014. Arch. Phys. Med. Rehabil. **100**(8), 1515–1533 (2019). https://doi.org/10.1016/j.apmr.2019.02.011
2. Kaminski, J.W., Valle, L.A., Filene, J.H., Boyle, C.L.: A meta-analytic review of components associated with parent training program effectiveness. J. Abnorm. Child Psychol. **36**(4), 567–589 (2008). https://doi.org/10.1007/s10802-007-9201-9
3. Leijten, P., et al.: Meta-analyses: Key parenting program components for disruptive child behaviour. J. Am. Acad. Child Adolesc. Psychiatry **58**(2), 180–190 (2019). https://doi.org/10.1016/j.jaac.2018.07.900
4. Corrigan, N., Păsărelu, C.-R., Voinescu, A.: Immersive virtual reality for improving cognitive deficits in children with ADHD: a systematic review and meta-analysis. Virtual Reality **27**, 3545–3564 (2023). https://doi.org/10.1007/s10055-023-00768-1
5. Romero-Ayuso, D., et al.: Effectiveness of virtual reality-based interventions for children and adolescents with ADHD: A systematic review and meta-analysis. Children **8,** 70, (2021). https://doi.org/10.3390/children8020070
6. Alashram, A.R., Annino, G., Padua, E., Romagnoli, C., Mercuri, N.B.: Cognitive rehabilitation post traumatic brain injury: A systematic review for emerging use of virtual reality technology. J. Clin. Neurosci. **66**, 209–219 (2019). https://doi.org/10.1016/j.jocn.2019.04.026
7. Gioia, G.A., Isquith, P.K., Guy, S.C., Kenworthy, L.: Behavior Rating Inventory of Executive Function, Second Edition (BRIEF 2). Lutz, FL: PAR Inc. (2015)
8. Rosvold, H.E., Mirsky, A.F., Sarason, I., Bransome Jr., E.D., Beck, L.H.: A continuous performance test of brain damage. J. Consult. Psychol. **20**(5), 343–350 (1956). https://doi.org/10.1037/h0043220

9. Seesjärvi, E., et al.: Quantifying ADHD symptoms in open-ended everyday life contexts with a new virtual reality task. J. Atten. Disord. **26**(11), 1394–1411 (2022). https://doi.org/10.1177/10870547211044214
10. Seesjärvi, E., et al.: EPELI: A novel virtual reality task for the assessment of goal-directed behavior in real-life contexts. Psychol. Res. **87**, 1899–1916 (2023). https://doi.org/10.1007/s00426-022-01770-z
11. Ravens-Sieberer, U., Bullinger, M.: Assessing health-related quality of life in chronically ill children with the German KINDL: first psychometric and content analytical results. Qual. Life Res. **7**(5), 399–407 (1998). https://doi.org/10.1023/A:1008853819715
12. Du Paul, G.J., Power, T.J., Anastopoulos, A.D., Reid, R.: ADHD rating Scales IV: Checklist, norms, and clinical interpretation. Guildford Press, New York and London (1998)
13. Klenberg, L., Jämsä, S., Häyrinen, T., Korkman, M.: Keskittymiskysely. Psykologien Kustannus Oy, Helsinki (2010)
14. Smalley, S.L., et al.: Prevalence and psychiatric comorbidity of attention-deficit/hyperactivity disorder in an adolescent Finnish population. J. Am. Acad. Child Adolesc. Psychiatry **46**(12), 1575–1583 (2007). https://doi.org/10.1097/chi.0b013e3181573137
15. Sayal, K., Prasad, V., Daley, D., Ford, T., Coghill, D.: ADHD in children and young people: prevalence, care pathways, and service provision. Lancet Psych. **5**(2), 175–186 (2018). https://doi.org/10.1016/S2215-0366(17)30167-0
16. Sandberg, E.: ADHD in the family - The support provided by the educational, social and health sectors, and the experienced impact. [Doctoral Thesis, University of Helsinki] HELDA University of Helsinki Open Repository (2016). http://hdl.handle.net/10138/161374
17. Alatupa, S. (Ed.), Karppinen, K., Keltikangas-Järvinen, L., Savioja, H.: Koulu, syrjäytyminen ja sosiaalinen pääoma - Löytyykö huono-osaisuuden syy koulusta vai oppilaasta? Sitra Report 75. Edita Prima Oy, Helsinki (2007)

Novel Sensors and Bioinformatics

A Distributed Framework for Remote Multimodal Biosignal Acquisition and Analysis

Constantino Álvarez Casado[1(✉)], Pauli Räsänen[2], Le Ngu Nguyen[1], Arttu Lämsä[2], Johannes Peltola[2], and Miguel Bordallo López[1,2]

[1] Center for Machine Vision and Signal Analysis, University of Oulu (CMVS), University of Oulu, Oulu, Finland
{constantino.alvarezcasado,miguel.bordallo}@oulu.fi
[2] VTT Technical Research Centre of Finland, Oulu, Finland

Abstract. In recent times, several studies have presented single-modality systems for non-contact biosignal monitoring. While these systems often yield estimations correlating with clinical-grade devices, their practicality is limited due to constraints in real-time processing, scalability, and interoperability. Moreover, these studies have seldom explored the combined use of multiple modalities or the integration of various sensors. Addressing these gaps, we introduce a distributed computing architecture designed to remotely acquire biosignals from both radars and cameras. This architecture is supported by conceptual blocks that distribute tasks across sensing, computing, data management, analysis, communication, and visualization. Emphasizing interoperability, our system leverages RESTful APIs, efficient video streaming, and standardized health-data protocols. Our framework facilitates the integration of additional sensors and improves signal analysis efficiency. While the architecture is conceptual, its feasibility has been evaluated through simulations targeting specific challenges in networked remote photoplethysmography (rPPG) systems. Additionally, we implemented a prototype to demonstrate the architectural principles in action, with modules and blocks operating in independent threads. This prototype specifically involves the analysis of biosignals using mmWave radars and RGB cameras, illustrating the potential for the architecture to be adapted into a fully distributed system for real-time biosignal processing.

Keywords: Non-contact biosignal monitoring · Distributed computing architecture · Interoperability · mmWave radars · RGB cameras

1 Introduction

The development of digital health platforms has significantly altered the healthcare industry. These platforms leverage advanced technologies to improve patient care, make medical workflows more efficient, and boost diagnostic accuracy [38].

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 127–146, 2024.
https://doi.org/10.1007/978-3-031-59091-7_9

The rise of mobile health (mHealth) applications, powered by sophisticated algorithms, represents a major advancement in the domain of real-time health monitoring. Integrated with wearable technologies, these applications are effective in collecting, processing, and transmitting biosignal data continuously, initiating a new phase in patient monitoring [18]. Additionally, telemedicine platforms have begun to incorporate biosignal data, offering healthcare professionals augmented data sources that enhance the accuracy of remote consultations and mitigate the challenges posed by geographical distances [37,52]. The advent of wearable devices with advanced sensors has broadened the scope of mHealth applications, enabling comprehensive monitoring of vital signs, such as ECGs [25]. The application of machine learning algorithms has further improved the early detection of conditions, for example, atrial fibrillation [57]. The introduction of Remote Patient Monitoring (RPM) systems has additionally benefited patient care by providing healthcare professionals with real-time data, facilitating prompt clinical responses [52]. Despite these advancements, the integration of multimodal biosignal data into a coherent, scalable, and interoperable system remains a significant challenge.

We introduce a distributed computing architecture designed to bridge the gap in integrating multimodal biosignal data, crucial for advancing healthcare technologies. This solution is crafted to efficiently manage the complexities and variances of biosignal data, ensuring scalability and interoperability, key aspects for the evolution of healthcare systems. At the heart of modern healthcare is the need for effective medical data exchange. Health Information Exchanges (HIEs), leveraging advanced protocols, are instrumental in enabling the smooth flow of information across various healthcare environments, promoting accessibility and consistent interpretation of data [3,58]. Protocols such as HL7 and FHIR are critical in enhancing interoperability, ensuring that healthcare systems can communicate seamlessly and without data discrepancies [47]. The increasing volume and complexity of patient data require robust healthcare systems capable of secure, fast processing, and storage. Cloud-based solutions offer scalable storage and rapid data processing capabilities, ensuring data security and accessibility [49]. For organizations prioritizing data privacy, on-site storage presents a viable alternative [8]. Electronic Health Records (EHRs) facilitate patient health history management, supported by strong security measures such as encryption [51,63]. Anonymization algorithms enable the safe utilization of patient data in large-scale studies, preserving individual privacy [24]. Our architecture aims to refine user experience across the healthcare spectrum, making systems more intuitive for everyone from patients to professionals [20]. By proposing this architecture, we not only align with current technological advancements in healthcare but also seek to enhance the functionality of mHealth and RPM systems, potentially transforming a wide range of healthcare services. Our contributions are detailed as follows:

- We introduce a distributed computing architecture for remote acquisition of multimodal biosignals, utilizing cameras and radars.

- We propose an interoperable system with RESTful APIs, efficient video streaming, and adherence to standardized health-data protocols.
- We validate our architecture's effectiveness in remote health monitoring through a remote photoplethysmography (rPPG) subsystem evaluation, demonstrating its scalability and adaptability for improving mHealth and RPM systems in networked healthcare environments.

In the following sections, we detail this architecture, emphasizing its role in advancing remote health monitoring and addressing the current limitations in data integration and interoperability.

2 Related Work

Recent advancements in biosignal monitoring and analysis systems have significantly influenced the creation of healthcare software and hardware architectures. A comprehensive survey by [43] provides an overview of wearable sensor-based systems for health monitoring. Key research has focused on developing innovative solutions for efficient and real-time monitoring of physiological signals, utilizing technologies such as FPGA, wireless DSP architectures, cloud computing, IoT, and wearable sensors. FPGA-based systems, like those presented by [32], emphasize highly integrated hardware designs but may lack in modularity and scalability. Wireless DSP architectures, as discussed by [45], focus on biosignal recording and monitoring using ARM-based Bluetooth wireless systems. While effective in biosignal recording, they fall short in offering comprehensive, multimodal integration. The *BioStream* system by [9] represents a leap in real-time physiological signal monitoring and emphasizes multipatient monitoring capabilities, an area where our proposed architecture innovates by providing a more unified and interoperable platform.

On the other hand, cloud computing applications in biosignal analysis have been explored by [53], proposing architectures for seamless integration with health information systems, yet these solutions often do not address the challenge of real-time, multimodal data processing. Smart sensor architectures, such as those researched by [44], highlight the importance of in-sensor processing to enhance usability and reduce power consumption, but do not integrate smart sensing capabilities within the broader cloud-connected framework. Affordable and open-source platforms for biosignal measurements, like the *Biosignal PI* developed by [2], demonstrate the potential for compact and medically safe systems, but they are not integrated into distributed healthcare monitoring systems. Wireless-enabled processor modules, as shown in studies by [19,55], focus on real-time signal acquisition and transmission, without focusing in scalability and interoperability across different systems. IoT-based wearable systems, as developed by [26,61], have shown improved performance in patient monitoring. Specialized biosignal extraction devices, such as those by [4,23], highlight the need for accurate signal processing. Multimodal wearable systems for emergency

applications, like the one developed by [33], integrate various monitoring technologies for potential use in critical care settings. Advances in wearable electronics, as discussed by [17], highlight the role of advanced materials and low-power consumption in biosignal monitoring.

While these studies collectively indicate significant progress in biosignal monitoring and analysis systems, underscoring the importance of innovative technological integration for enhanced patient care and monitoring, our architecture built upon these advances offers a novel, scalable, and interoperable solution that addresses the complexities of modern healthcare monitoring.

3 Key Elements in Physiological Signal Devices

The block diagram in Fig. 1 depicts the essential components of medical devices used for physiological signal monitoring. Biosensors, the core components, convert physiological activity into electrical signals. These are then digitized by an Analog-to-Digital Converter (ADC), making them suitable for digital processing.

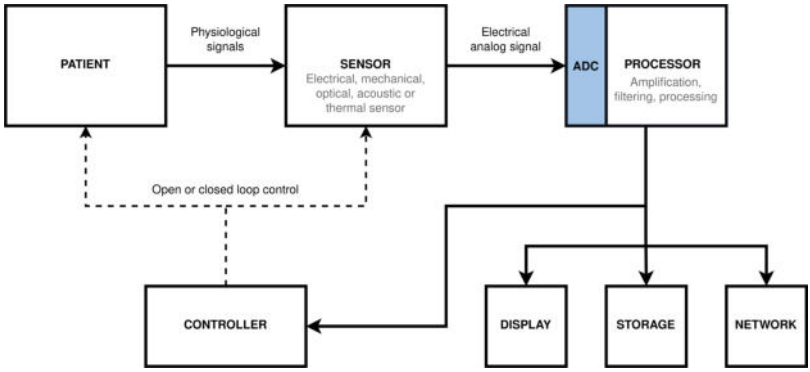


Fig. 1. Simplified block diagram illustrating the core components of medical devices for physiological signal monitoring: sensors, ADC, processor, and controller, with functionalities extending to display, storage, and network integration, and possibly therapeutic intervention based on the signal analysis.

Digitization is governed by the Nyquist-Shannon sampling theorem to prevent aliasing, ensuring the sampling rate is at least twice the maximum signal frequency. The bit resolution of the ADC affects the precision of the signal's digital representation. Post-digitization, the processor amplifies and filters the signal, preparing it for analysis, display, recording, and potential therapeutic actions. The integration of these devices with digital healthcare systems enhances patient monitoring by allowing for real-time interventions, long-term data storage, and telehealth capabilities, facilitating better care coordination. Device controllers maintain operational efficiency, while displays provide user interfaces. Effective device design prioritizes security, power efficiency, and user accessibility, all of which are key for the successful adoption of technology in enhancing patient care and health monitoring.

4 Key Challenges of rPPG Acquisition Systems

Physiological signal acquisition systems pose significant challenges in terms of accuracy and reliability of the extracted signals [5,27]. For rPPGs, most of the studies in the literature address issues such as motion artifacts, skin tone variations, noise, occlusions, and illumination variations, which can degrade accuracy [48]. These studies included stabilizing the region of interest with optical flow and tracking [12,29], tracking and aligning the face to remove head and face movements [15], using bandpass, adaptive, detrending, or LSTM-based filters to normalize the PPG signals and remove noise and motion artifacts [12,14,35], signal separation methods such as PCA, ICA or OMIT [15], signal separation methods based on skin reflection models such as CHROM [16] or POS [60], and correlating signals using a normalized reference waveform or a noise reference signal [31,59]. A few studies have also focused on tackling illumination variations [31,42], and skin tone variations [30].

However, challenges related to network and computing constraints in remote and streaming PPG systems remain underexplored, potentially impacting their performance. These challenges encompass limited bandwidth, packet loss, latency, video compression algorithms, resolution, and computing resources [40,48]. Hardware constraints involve computational limitations and other factors affecting signal quality, such as sensor capabilities or algorithmic complexity impacting system performance. Therefore, careful design is a critical aspect to consider, ensuring that the hardware can adequately support the algorithmic demands for accurate and efficient signal processing [27]. Previous research has focused on the impact of video compression on the quality of the recovered BVP signal [7,21,36,50]. Researchers have found that different compression schemes and codecs can lead to small quality losses in the extracted signal, which can significantly affect the signal's features and morphology. Some studies have proposed methods to address the issue of video compression artifacts, such as using image filtering or end-to-end deep learning-based methods [64] to improve video quality and reduce file sizes or to use singular spectrum analysis to reconstruct and select signal components [66]. Other studies have examined the effect of reduced frame rate and image resolution on heart rate estimation. However, they have generally found no significant differences in mean absolute error or error distributions resulting from reduced frame rates if they are kept in typical values (15–20fps) [10,56]. Several strategies have also been studied for the efficient codification of PPGs, but they require a particular learning strategy and architecture [65]. Álvarez et al. [7] investigate the impact of network and hardware constraints on the performance of the rPPG systems. Their approach included simulations and experimentation to develop mitigation strategies. These strategies specifically addressed challenges related to frame dropping, frame resolution, and frame rate in the rPPG system.

5 Proposed Architecture for rPPG and rBSG Acquisition

Many non-contact health monitoring systems align well with clinical devices in controlled studies but face challenges in real-world application, often neglecting real-time processing or distributed computation, scalability, and interoperability. To address these limitations, we propose a distributed computing architecture designed to overcome these challenges by supporting remote biosignal acquisition through both radar and camera technologies. It features a modular framework that organizes tasks related to sensing, computation, data management, and analysis. The system's interoperability, facilitated by RESTful APIs, efficient video streaming, and standard health-data protocols, enables the seamless addition of new sensors and signal analysis tasks. This feature allows the system to adapt to a range of applications and technologies, enhancing its versatility.

Conceptual demonstration of the system is provided via the implementation of a real-time and networked biosignal acquisition and analysis system using RGB cameras, and its possible extensibility to other sensors. This not only showcases the system's robust capabilities in handling different sensor types, but also its potential in harmonizing these diverse data streams for comprehensive biosignal analysis. With this, our proposed architecture results a in scalable, interoperable, and multi-modal remote biosignal monitoring system.

5.1 Overview of the Proposed Architecture

The proposed architecture, depicted in Fig. 2, consists of five main categories of blocks: sensing, processing, communications, data storage, and user interaction. These blocks communicate over a network, which can be a Local Area Network (LAN) or an Internet of Things (IoT) network. The design aims to seamlessly integrate various sensors, supporting diverse applications, and incorporates signal analysis and feature extraction methods within the processing blocks. This

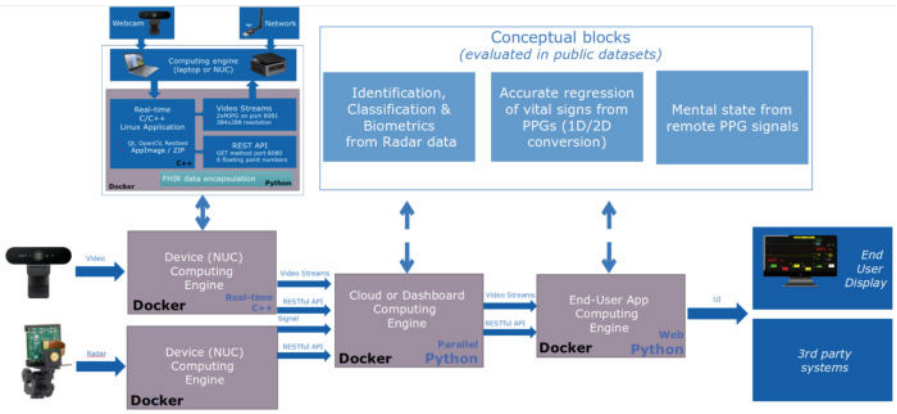


Fig. 2. Modular Architecture for Multimodal Biosignal Acquisition and Analysis

architecture follows the principles of a standard microservices architecture, where each processing procedure is designated as a "service," representing a software component. These components can run on a cluster of computers or a microcomputer with sufficient resources. They can be developed using established microservices frameworks like *Docker Compose*.

Figure 2 shows the proposed distributed computing architecture's block diagram, where each component is network-linked, facilitating data exchange. Key interconnected blocks, such as sensing, signal analysis, storage, and visualization, enable two-way communication, while others are designed for one-way data flow. Sensing blocks capture various data types, like video and radio frequency (RF) signals, and apply signal processing to yield processed data. This data is then standardized by an aggregation module into formats like Fast Healthcare Interoperability Resources (FHIR) [1], ready for storage in databases or cloud services. Analysis modules access this data, allowing any data analysis to work with the standardized inputs. Interface modules serve end-users by retrieving processed data for visualization. The architecture's modularity ensures components can be updated or replaced without disrupting the system's overall functionality.

5.2 Interoperability

Our architecture ensures interoperability, allowing various software components and systems to communicate and collaborate effectively. This facilitates the exchange of data between elements, enhancing system efficiency and flexibility. This is especially critical in digital health, where various devices and software must interact without creating data silos, promoting comprehensive patient care and more effective healthcare delivery.

The architecture integrates a wide range of **hardware and software libraries** critical for functions spanning from data acquisition to advanced analysis and visualization. These libraries enable the use of various sensors and the execution of complex data processing. For example, hardware libraries might include those necessary for interfacing with cameras and radars, while software libraries could encompass those for data processing, machine learning, and visualization.

Containerization, a lightweight alternative to full machine virtualization, is a core strategy in our architecture, instrumental in achieving software scalability and portability. By encapsulating a software component with its dependencies into a single, self-contained unit (container), it ensures consistent execution across varying computing environments, from a developer's local machine to testing environments and production servers. This approach minimizes the infamous problem of "it works on my machine" and significantly simplifies software deployment and scaling. A potential drawback could be added complexity in managing multiple containers, but this is generally mitigated by using container orchestration tools. The architecture adopts **RESTful APIs** (Representational State Transfer APIs) to facilitate smooth and efficient data exchange between blocks. RESTful APIs are widely recognized for their simplicity and scalability, which contribute to enhancing system reliability. Supporting CRUD operations

(Create, Read, Update, Delete), they facilitate straightforward communication and interoperability. However, for real-time communication, protocols like *WebSocket* may be preferable due to REST's request/response nature. Our architecture integrates **video and signal streaming** for real-time data transmission, vital in remote patient monitoring for immediate processing and feedback. While advantageous for real-time applications, it may pose challenges in network bandwidth management and necessitate efficient compression algorithms to handle large data volumes. **Extensibility**, other fundamental aspect of our architecture, enables the integration of new features and functionalities over time. Its modular design facilitates the incorporation of additional sensors, processing methods, or subsystems, ensuring adaptability to evolving needs and technologies. While offering future-proofing and customization benefits, this approach may involve initial design complexity to ensure smooth integration of future enhancements.

While **security and compliance** are not the core focus of this article, they are nevertheless essential considerations in the design of our digital health platform. Our architecture could be implemented to operate under the regulatory framework of the European Union (EU), which lays a foundation upon which robust security measures can be implemented.

5.3 Sensing and On-Device Computing

In our architecture, sensing and on-device computing are crucial components. They are responsible for capturing biosignals and conducting initial data processing, which forms the basis for further analysis. While the main focus is on standard cameras for sensing, the modular design enables the integration of additional sensors, like radars, to enhance the versatility of the system.

Camera-Based Subsystems rely on standard RGB webcams, as illustrated in Fig. 3, and optionally, other devices like thermal cameras. RGB webcams, known for their widespread availability and user-friendliness, offer a cost-effective means to capture high-quality visual data for various applications, from tracking body movements to facial analysis. Additionally, integrating a thermal camera enables infrared radiation detection, facilitating non-contact body temperature measurement, crucial in many health contexts [34]. These subsystems prioritize versatility and resilience, functioning effectively across different lighting conditions and accommodating multiple subjects in the frame. By converting visual data into standardized digital formats and supporting video transmission via MJPEG streams [54], they seamlessly integrate with the system, efficiently communicating collected data through RESTful APIs. Thus, the integration of RGB cameras is central to our architecture's monitoring capabilities, enabling comprehensive remote health monitoring across diverse conditions and scenarios.

Alternatively to the camera subsystems, our platform allows the use of other non-contact sensors such as radars [41]. As an example platform, we have integrated a Texas Instruments IWR1443 mmWave FMCW radar system operating in the 76–81 GHz frequency range, including four receivers and three transmitters. This mmWave radar can monitor vital signs up to 1.3 m away, exploiting

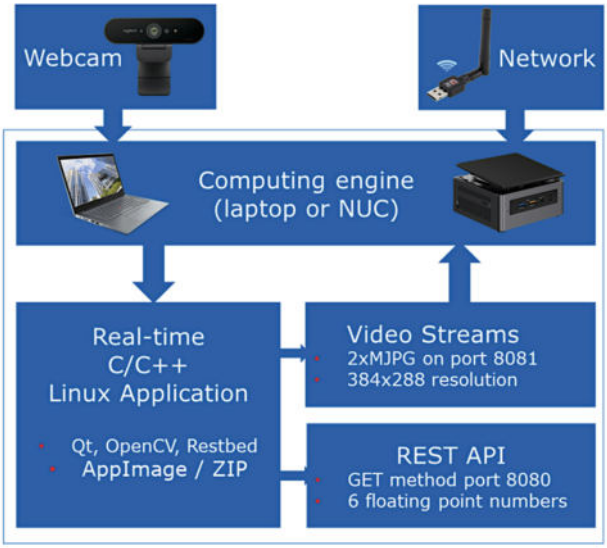


Fig. 3. Real-time Camera based subsystem to compute biosignals and vital signs

the Doppler effects to measure bodily movements induced by respiratory and cardiovascular activities, a technique known as remote ballistography (rBSC) [5]. Similarly to the camera, the signal processing involves a 16-second running window, updated every second. Raw signals and computed values of breath rate, heart rate, and distance are transmitted via network packets for real-time monitoring.

Real-time computation is crucial in our architecture for applications requiring immediate feedback. On-device computation offers benefits such as reducing data transmission, saving bandwidth, and decreasing communication costs. It also enhances system robustness by reducing reliance on continuous connectivity, conserves energy, and safeguards privacy by minimizing raw data exposure. In the proposed architecture, camera-based or radar-based subsystems collect raw data in our architecture, followed by real-time pre-processing at the device level. This step refines biosignals by filtering them within the relevant frequency band, typically associated with heart or respiration dynamics. Pre-processing involves techniques like noise filtering, normalization, and signal enhancement for high-quality biosignals. On-device computation also calculates key physiological indicators such as Heart Rate, Respiration Rate, and Heart Rate Variability (HRV) parameters, using advanced digital signal processing techniques. This preliminary computation significantly enhances system efficiency, reducing network data load and feedback latency to healthcare professionals.

5.4 Aggregation, Storage, and Standardization

In our proposed architecture, the Aggregation, Storage, and Standardization phase serves as the central processing hub, providing integration between the data collection devices and the end-users. This stage encompasses the receipt of raw and processed data from various end devices, its subsequent storage, and the generation of standardized health data messages as depicted in Fig. 4.

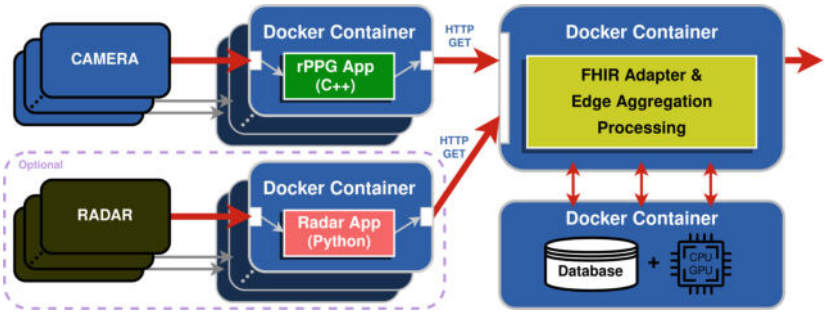


Fig. 4. Aggregation and data storage modules in the proposed platform. The system can be extended to other complementary sensors, such as radars (purple). (Color figure online)

The Aggregator, potentially hosted in a Docker container, serves as a central hub in our biosignal acquisition architecture. It receives heterogeneous data from sensor devices and converts it into standardized formats like Fast Healthcare Interoperability Resources (FHIR). FHIR, introduced by HL7 in 2014, facilitates electronic healthcare information exchange via web-based data exchange methodologies. FHIR's "resources" provide fundamental units of healthcare information adaptable to diverse applications, enabling data exchange in XML and JSON formats through RESTful APIs. Its strengths include extensive data definitions, adaptable exchange protocols, and widespread open-source tool support, ensuring consistent and secure healthcare information exchange [39]. Once converted to the FHIR standard, data is securely stored for further processing. InfluxDB, a time-series database, is recommended for our architecture due to its suitability for real-time biosignal and vital sign data handling. Alternatively, other solutions may be considered, especially if raw video storage is required. InfluxDB ensures rapid read/write operations, enabling processing to match real-time data capture. Operating within a Docker container, it offers scalability and easy deployment across various computing environments. This setup allows efficient access to biosignals and vital signs for subsequent processing stages, including AI model training.

5.5 Data Analysis and AI Computation

The data analysis and AI component of the proposed biosignal processing system is critical for interpreting biosignals and executing advanced computations. It includes sub-modules for cloud-based analysis, machine learning algorithms, and multimodal data integration. Immediate metrics like heart and respiration rates are processed on-device. Complex analyses are performed in the cloud, where features from the biosignals are extracted to train machine learning models. These models predict health indicators such as stress, depression levels, SpO2, and respiratory conditions with high accuracy, thanks to cloud computing's capacity to handle extensive computations.

Model training uses a comprehensive dataset of patient information, including medical histories and diagnoses, in a supervised learning framework. This ensures high-quality, accurately labeled data. The diversity of patient data increases the models' reliability and applicability. Transfer learning tailors these models to individual patients' needs, improving personalization. Multimodal data processing is achieved by fusing features from various sensors, enhancing the system's ability to provide a thorough analysis of the biosignals. This not only improves data richness but also the precision of the health assessments generated by the AI models.

5.6 Interactive User Interface

The user interface (UI) acts as a crucial link between users and underlying technology. Our proposed interactive UI manage biosignal measurements being designed for collaborative users. It enables measurement initiation by user positioning, omitting conventional buttons for a more intuitive interaction paradigm [13]. Figure 5 displays an example.

When users position themselves, the UI immediately starts measurements, recording and transmitting video and biosignals simultaneously for real-time sync. Stable network connection is crucial for video transmission, while device-level computation handles real-time vital sign calculations. The UI adds quality measurements to the data stream, confirming signal stability and assessing movement intensity to address facial motion artifacts. It detects when users leave the camera's view, signaling the end of the session. These metrics are sent via RESTful API for data interpretation and system optimization. To ensure real-time operation, the UI minimizes latency and integrates into a networked environment. Leveraging our unsupervised biosignal acquisition methods [15], the UI enables accurate and swift vital sign computation, supporting integration with medical monitoring processes.

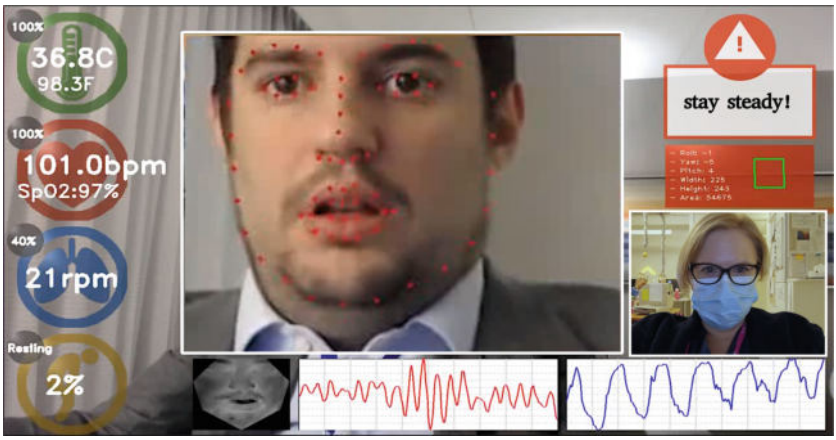


Fig. 5. A figure depicting an example interactive user interface. When the user moves to the center of the camera's FoV, the biosignal measurement is initiated.

6 Evaluation of the Camera-Based rPPG Component

We evaluate our architecture under two operational setups, highlighting the networked design. The first setup is cloud-based, using deep learning for facial analysis and algorithms to extract BVP signals from RGB data, ideal for environments with robust network support that enables high-quality video transmission and advanced processing. The second setup is designed for real-time operations on embedded systems, improving processing speed and focusing on efficient biosignal processing directly on the device, through algorithmic and implementation optimizations. This approach also addresses privacy concerns by potentially eliminating the need for video transmission. The real-time model demonstrates how components can work efficiently within a networked system.

We assessed the performance of both configurations in terms of speed and accuracy to determine the architecture's effectiveness. Both systems use a design that splits tasks across multiple threads, enhancing efficiency and facilitating their fit into a broader distributed framework. This strategy enables seamless operation and data sharing among components, applying mitigation strategies to the effects of network or hardware limitations [7]. These solutions are critical for ensuring the reliability of biosignal extraction and processing, regardless of the operational framework.

6.1 Experimental Setup and Configurations

The experimental setup for the evaluation of our camera-based rPPG component is based on two configurations, both based on the Face2PPG pipeline [15]. Face2PPG-RT is designed for real-time processing on embedded systems, while Face2PPG-Server is optimized for higher accuracy but with increased computa-

tional demands. Referenced in Table 1, these setups facilitates their comparison across different configurations.

The *Face2PPG-RT* configuration uses a RESTful API with the Restbed library for handling HTTP requests. It initiates an HTTP server in the **MainWindow** constructor to manage requests and responses on a specified port. Data is transmitted to the HTTP server every 33 ms as an MJPEG stream via a dedicated function, ensuring continuous data flow. Multithreading supports parallel operation of the camera, HTTP server, RESTful API, and GUI, improving system responsiveness and stability for real-time monitoring of patient bio-signals. The prototype incorporates optimization libraries like *Lapack*, *BLAS*, *OpenMP*, and *FFTW3*. The test hardware includes an Intel® Core i7-6700HQ CPU with HD Graphics 530, and 8GB of RAM, running Linux Ubuntu 18.04.6. Face detection uses YuNet CNN [62], and face alignment and skin segmentation is based on ERT-GTX models [6]. The segmentation method [15], focuses on the cheeks and forehead, and converts to CIE Lab color space.

The *Face2PPG-Server* prototype adopts a Python-based implementation using a deep learning-based face detection method with a Single Shot Multibox Detection (SSD) network [15]. Faces are aligned using the Deep Alignment Network (DAN) [28], These 85 landmarks construct a mesh of 131 triangles and selects the optimal facial regions for extracting raw RGB signals dynamically [15]. Testing was conducted on a high-performance setup featuring an AMD® Ryzen(TM) 3700X 8-core processor at 3.6GHz, with 64 Gigabytes of RAM, 4 terabyte SSD and two NVIDIA GeForce® RTX(TM) 2080.

6.2 Benchmark Datasets, Protocol and Metrics

We conducted an extensive evaluation of the rPPG systems performance across four publicly available datasets, each comprising videos and reference PPG signals obtained using medically-graded oximeters, alongside videos captured with user-grade webcams, three with uncompressed videos (1.5 GB per video) and one with highly compressed videos (2 MB per video). The datasets are: **COHFACE** [22], comprised of 160 highly compressed videos of 40 subjects, recorded at 20 Hz and a resolution of 640×480 pixels. **LGI-PPGI-Face-Video-Database**, contains 24 videos from 6 users across 4 scenarios [46], recorded at 640×480 pixels at 25 fps. **UBFC-RPPG**, consists of two subsets: UBFC1 or simple and UBFC2 or realistic [11]. UBFC1 contains 8 videos where participants remained seated in an office room under natural light conditions, while UBFC2 includes 42 videos recorded under constrained conditions. The videos are captured at 640×480 pixels at 30 fps.

To evaluate the accuracy of our rPPG measurements, we followed the assessment protocol described in [15] that compares the estimated heart rates extracted from the video streams to the reference (ground-truth) contact-based PPG signals. We employ three well established metrics: Mean Absolute Error (MAE), Root-Mean-Square Error (RMSE), and Pearson Correlation Coefficient (PCC) of the heart-rate envelope. In addition, we measure the processing time required to estimate heart rate per window (12-second), and we calculate the time per

frame expended for each module within each pipeline. We also compute the total frame rate of each configuration in frames per second (FPS).

6.3 Speed Performance

Table 1 illustrates the computational performance of the different modules within our Face2PPG pipelines across the two proposed configurations: Embedded (targeted for real-time operation) and Cloud (for unconstrained conditions, typically found in server environments).

Table 1. Configuration Setups and performance speed for individual modules in both rPPG systems, as well as the overall speed per face and frame.

Module	Face2PPG-RT		Face2PPG-Server	
	Embedded Configuration	Speed ms/frame	Cloud Configuration	Speed ms/frame
Face Detection	OpenCV YuNet	1.27	OpenCV SSD	20.35
Face Alignment	Dlib ERT GTX	3.74	DAN	82.52
Face Normalization	Mesh	4.37	Mesh	12.48
Skin Segmentation	Fix Patches	0.35	Best Regions	105.41
RBG to BVP	Lab	0.46	POS	0.079
Filtering	BP FIR	0.01	BP FIR	0.01
Spectral Analysis	FFT	0.21	Welch	1.02
Language	C++	–	Python	–
Total Time	–	10.41	–	221.87
Total FPS	–	96.1	–	4.5

The performance evaluation of *Face2PPG-RT* indicates that it processes each frame in about 10.41 milliseconds (ms), with 'Face Normalization' being the most time-consuming module at 4.37 ms per frame. Despite the computational demands, it maintains a high frame rate of 96 FPS, suitable for real-time operation. On the other hand, *Face2PPG-Server* shows a longer processing time of 221.87 ms per frame, (4.51 FPS). This increase is primarily due to the 'Face Alignment' and 'Skin Segmentation' highlighting the added complexity of multi-region analysis. Table 1 details the speed performance, demonstrating the versatility of the proposed distributed framework in handling both real-time and in-depth analytical tasks.

6.4 Accuracy Performance in Vital Signs Measurement

To evaluate the accuracy of the rPPG configurations, where we measure the heart rate accuracy using several databases. The results, summarized in Table 2, shows that embedded setup is just slightly less effective than the ones from server setup. These findings show that the impact of the performance of the real-time configuration is relatively minimal.

Table 2. Error comparison between the two configurations (Embedded and Cloud) of the proposed rPPG system in four different databases.

Pipeline	LGI-PPGI		COHFACE		UBFC1		UBFC2	
	MAE ± SD	PCC	MAE ± SD	PCC	MAE ± SD	PCC	MAE ± SD	PCC
Face2PPG-Server	4.5 ± 3.3	0.57	8.0 ± 4.4	0.06	0.9 ± 0.4	0.96	0.9 ± 0.9	0.98
Face2PPG-RT	5.9 ± 8.0	0.49	11.3 ± 7.3	−0.01	1.5 ± 1.2	0.83	6.7 ± 6.1	0.54

7 Conclusion

Advancements in wireless communications, particularly with 5G and the upcoming 6G, are significantly enhancing remote patient monitoring in healthcare by enabling faster data exchange and lower latency, crucial for real-time health monitoring. Our study introduces a distributed computing architecture that utilizes radar and camera technologies for efficient, real-time biosignal processing and analysis. This architecture integrates various sensors and ensures interoperability, marking a significant step forward in remote health data acquisition. The evaluation of the camera-based rPPG component within our framework confirms its utility in healthcare by showcasing its ability to manage biosignals in real-time within a networked environment, while emphasizing data security. These findings highlight the architecture's potential to enhance remote health monitoring and patient care. In essence, as demand for real-time, remote health monitoring grows, our research offers a robust, adaptable framework using the latest in communication technology. Future work will focus on expanding this framework and assessing its performance in clinical settings, aiming to further support remote healthcare delivery.

Acknowledgements. This research was supported by the Research Council of Finland (former Academy of Finland) 6G Flagship Programme (Grant Number: 346208) and PROFI5 HiDyn (Grant Number: 326291).

References

1. Fast healthcare interoperability resources. https://www.hl7.org/fhir/overview.html. Accessed 05 Jun 2023
2. Abtahi, F., Snäll, J., Aslamy, B., Abtahi, S., Seoane, F., Lindecrantz, K.: Biosignal PI, an affordable open-source ECG and respiration measurement system. Sensors **15**(1), 93–109 (2014)
3. Ademaj, G., Saenyi, B.: Systematic literature review: data standardization in health information systems. In: ECIS 2022 Research Papers, p. 138 (2022)
4. Al-Masri, S., Safi, A., EL-Farra, A.A., Shehada, M.: Designing a biosignal extraction device for measuring a blood pressure. In: 2018 International Conference on Promising Electronic Technologies (ICPET), pp. 108–112. IEEE (2018)
5. Álvarez Casado, C.: Biosignal extraction and analysis from remote video: towards real-world implementation and diagnosis support (2023)

6. Álvarez Casado, C., Bordallo López, M.: Real-time face alignment: evaluation methods, training strategies and implementation optimization. J. Real Time Image Process. **18**, 2239–2267 (2021)
7. Álvarez Casado, C., Nguyen, L., Silvén, O., Bordallo López, M.: Assessing the feasibility of remote photoplethysmography through videocalls: a study of network and computing constraints. In: Gade, R., Felsberg, M., Kämäräinen, J.K. (eds.) Scandinavian Conference on Image Analysis, pp. 586–598. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-31438-4_38
8. Banimfreg, B.H.: A comprehensive review and conceptual framework for cloud computing adoption in bioinformatics. Healthc. Analytics **3**, 100190 (2023). https://doi.org/10.1016/j.health.2023.100190
9. Bar-Or, A., Healey, J., Kontothanassis, L., Van Thong, J.: Biostream: a system architecture for real-time processing of physiological signals. In: The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 2, pp. 3101–3104. IEEE (2004)
10. Blackford, E., Estepp, J.: Effects of frame rate and image resolution on pulse rate measured using multiple camera imaging photoplethysmography. Progress Biomed. Opt. Imaging Proc. SPIE **9417**, 639–652 (2015). https://doi.org/10.1117/12.2083940
11. Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A., Dubois, J.: Unsupervised skin tissue segmentation for remote photoplethysmography. Pattern Recogn. Lett. **124**, 82–90 (2019). https://doi.org/10.1016/j.patrec.2017.10.017
12. Boccignone, G., et al.: pyVHR: a python framework for remote photoplethysmography. PeerJ Comput. Sci. **8**, e929 (2022)
13. Bordallo Lopez, M.: Designing for energy-efficient vision-based interactivity on mobile devices. Ph. D. thesis, University of Oulu Graduate School (2014). https://doi.org/10.13140/2.1.3356.0640
14. Botina-Monsalve, D., et al.: Long short-term memory deep-filter in remote photoplethysmography. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1242–1249 (2020). https://doi.org/10.1109/CVPRW50498.2020.00161
15. Álvarez Casado, C., Bordallo López, M.: Face2PPG: an unsupervised pipeline for blood volume pulse extraction from faces. IEEE J. Biomed. Health Inf. **27**(11), 5530-5541 (2023). https://doi.org/10.1109/JBHI.2023.3307942
16. De Haan, G., Jeanne, V.: Robust pulse rate from chrominance-based RPPG. IEEE Trans. Biomed. Eng. **60**(10), 2878–2886 (2013). https://doi.org/10.1109/TBME.2013.2266196
17. Deng, Z., Guo, L., Chen, X., Wu, W.: Smart wearable systems for health monitoring. Sensors **23**(5), 2479 (2023)
18. El-Rashidy, N., El-Sappagh, S., Islam, S.M.R., El-Bakry, H., Abdelrazek, S.: Mobile health in remote patient monitoring for chronic diseases: principles, trends, and challenges. Diagnostics **11**, 607 (2021). https://doi.org/10.3390/diagnostics11040607
19. Farshchi, S., Pesterev, A., Nuyujukian, P.H., Mody, I., Judy, J.W.: Bi-fi: an embedded sensor/system architecture for remote biological monitoring. IEEE Trans. Inf Technol. Biomed. **11**(6), 611–618 (2007)
20. Haleem, A., Javaid, M., Singh, R.P., Suman, R.: Telemedicine for healthcare: capabilities, features, barriers, and applications. Sensors Int. **2**, 100117 (2021). https://doi.org/10.1016/j.sintl.2021.100117
21. Hanfland, S., Paul, M.: Video format dependency of PPGI signals. In: Proceedings of the International Conference on Electrical Engineering, vol. 1, p. 2 (2016)

22. Heusch, G., Anjos, A., Marcel, S.: A reproducible study on remote heart rate measurement. CoRR **abs/1709.00962** (2017). http://arxiv.org/abs/1709.00962
23. Joaquinito, R., Sarmento, H.: A wireless biosignal measurement system using a SoC FPGA and bluetooth low energy. In: 2016 IEEE 6th International Conference on Consumer Electronics-Berlin (ICCE-Berlin), pp. 36–40. IEEE (2016)
24. Kaissis, G., Makowski, M., Rückert, D., Braren, R.: Secure, privacy-preserving and federated machine learning in medical imaging. Nat. Mach. Intell. **2**, 305–311 (2020). https://doi.org/10.1038/s42256-020-0186-1
25. Kang, H.S., Exworthy, M.: Wearing the future–wearables to empower users to take greater responsibility for their health and care: scoping review. JMIR Mhealth Uhealth **10**(7), e35684 (2022). https://doi.org/10.2196/35684
26. Kelati, A., Dhaou, I.B., Tenhunen, H.: Biosignal monitoring platform using wearable IoT. In: Proceedings of the 22st Conference of Open Innovations Association FRUCT, pp. 332–337 (2018)
27. Khanam, F.T.Z., Al-Naji, A., Chahl, J.: Remote monitoring of vital signs in diverse non-clinical and clinical scenarios using computer vision systems: a review. Appl. Sci. **9**(20), 4474 (2019). https://doi.org/10.3390/app9204474
28. Kowalski, M., Naruniec, J., Trzcinski, T.: Deep alignment network: a convolutional neural network for robust face alignment. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 88–97 (2017)
29. Kumar, M., Suliburk, J., Veeraraghavan, A., Sabharwal, A.: PulseCam: a camera-based, motion-robust and highly sensitive blood perfusion imaging modality. Sci. Rep. **10**, 4825 (2020). https://doi.org/10.1038/s41598-020-61576-0
30. Kumar, M., Veeraraghavan, A., Sabharwal, A.: DistancePPG: robust non-contact vital signs monitoring using a camera. Biomed. Opt. Express **6**, 1565–1588 (2015). https://doi.org/10.1364/BOE.6.001565
31. Li, X., Chen, J., Zhao, G., Pietikäinen, M.: Remote heart rate measurement from face videos under realistic situations. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 4264–4271 (2014). https://doi.org/10.1109/CVPR.2014.543
32. Lin, B.S., Yen, T.S.: An FPGA-based rapid wheezing detection system. Int. J. Environ. Res. Public Health **11**(2), 1573–1593 (2014)
33. Lin, C.T., Wang, C.Y., Huang, K.C., Horng, S.J., Liao, L.D.: Wearable, multimodal, biosignal acquisition system for potential critical and emergency applications. Emerg. Med. Int. **2021**, 9954669 (2021)
34. Lopez, M.B., del Blanco, C.R., Garcia, N.: Detecting exercise-induced fatigue using thermal imaging and deep learning. In: 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 1–6 (2017). https://doi.org/10.1109/IPTA.2017.8310151
35. Maity, A.K., Wang, J., Sabharwal, A., Nayar, S.K.: RobustPPG: camera-based robust heart rate estimation using motion cancellation. Biomed. Opt. Express **13**(10), 5447–5467 (2022). https://doi.org/10.1364/BOE.465143
36. McDuff, D., Blackford, E., Estepp, J.: The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography. In: IEEE International Conference on Automatic Face & Gesture Recognition, pp. 63–70 (2017). https://doi.org/10.1109/FG.2017.17
37. Michaud, A., Vadeboncoeur, A., Cloutier, L., Goupil, R.: The feasibility of home self-assessment of vital signs and symptoms: a new key to telehealth for individuals? Int. J. Med. Inform. **155**, 104602 (2021). https://doi.org/10.1016/j.ijmedinf.2021.104602

38. Mitchell, M., Kan, L.: Digital technology and the future of health systems. Health Syst. Reform **5**, 113–120 (2019). https://doi.org/10.1080/23288604.2019.1583040
39. Nan, J., Xu, L.Q.: Designing interoperable health care services based on fast healthcare interoperability resources: literature review. JMIR Med. Inform. **11**, e44842 (2023). https://doi.org/10.2196/44842
40. Narayanan, A., et al.: A variegated look at 5G in the wild: Performance, power, and QoE implications. In: Proceedings of the 2021 ACM SIGCOMM 2021 Conference, pp. 610-625. SIGCOMM 2021, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3452296.3472923
41. Nguyen, L., Casado, C.Á., Silvén, O., López, M.B.: Identification, activity, and biometric classification using radar-based sensing. In: 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–8. IEEE (2022)
42. Nowara, E.M., McDuff, D., Veeraraghavan, A.: The benefit of distraction: denoising camera-based physiological measurements using inverse attention. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4935–4944 (2021). https://doi.org/10.1109/ICCV48922.2021.00491
43. Pantelopoulos, A., Bourbakis, N.G.: A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Trans. Syst. Man Cybern. Part C **40**(1), 1–12 (2009)
44. Pfundt, B., Reichenbach, M., Eskofier, B., Fey, D.: Smart sensor architectures for embedded biosignal analysis. In: 2013 Conference on Design and Architectures for Signal and Image Processing, pp. 174–181. IEEE (2013)
45. Piccini, L., Arnone, L., Beverina, F., Cucchi, A., Petrelli, L., Andreoni, G.: Wireless DSP architecture for biosignals recording. In: Proceedings of the Fourth IEEE International Symposium on Signal Processing and Information Technology, 2004, pp. 487–490. IEEE (2004)
46. Pilz, C.S., Zaunseder, S., Krajewski, J., Blazek, V.: Local group invariance for heart rate estimation from face videos in the wild. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1335–13358 (2018). https://doi.org/10.1109/CVPRW.2018.00172
47. Pimenta, N., Chaves, A., Sousa, R., Abelha, A., Peixoto, H.: Interoperability of clinical data through FHIR: a review. Procedia Comput. Sci. **220**, 856–861 (2023). https://doi.org/10.1016/j.procs.2023.03.115, The 14th International Conference on Ambient Systems, Networks and Technologies Networks (ANT) and The 6th International Conference on Emerging Data and Industry 4.0 (EDI40)
48. Premkumar, S., Hemanth, D.J.: Intelligent remote photoplethysmography-based methods for heart rate estimation from face videos: a survey. Informatics **9**(3), 57 (2022). https://doi.org/10.3390/informatics9030057
49. Rai, V., et al.: Cloud computing in healthcare industries: opportunities and challenges. In: Singh, P.K., Singh, Y., Chhabra, J.K., Illés, Z., Verma, C. (eds.) Recent Innovations in Computing: Proceedings of ICRIC 2021, vol. 2, pp. 695–707. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-8892-8_53
50. Rapczynski, M., Werner, P., Al-Hamadi, A.: Effects of video encoding on camera based heart rate estimation. IEEE Trans. Biomed. Eng. **66**, 3360–3370 (2019). https://doi.org/10.1109/TBME.2019.2904326
51. Rhoades, C., Whitacre, B., Davis, A.: Higher electronic health record functionality is associated with lower operating costs in urban-but not rural-hospitals. Appl. Clin. Inform. **13**, 665–676 (2022). https://doi.org/10.1055/s-0042-1750415

52. Shaik, T., et al.: Remote patient monitoring using artificial intelligence: current state, applications, and challenges. WIREs Data Min. Knowl. Discovery **13**(2), e1485 (2023). https://doi.org/10.1002/widm.1485
53. Shen, C.P., et al.: Bio-signal analysis system design with support vector machines based on cloud computing service architecture. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 1421–1424. IEEE (2010)
54. Strukov, R., Athitsos, V.: Evaluation of video compression methods for network transmission on diverse data: a case study. In: Proceedings of the 16th International Conference on Pervasive Technologies Related to Assistive Environments, pp. 300–305 (2023)
55. Stuart, T., Cai, L., Burton, A., Gutruf, P.: Wireless and battery-free platforms for collection of biosignals. Biosens. Bioelectron. **178**, 113007 (2021)
56. Sun, Y., Hu, S., Azorin Peris, V., Kalawsky, R., Greenwald, S.: Noncontact imaging photoplethysmography to effectively access pulse rate variability. J. Biomed. Opt. **18**, 61205 (2013). https://doi.org/10.1117/1.JBO.18.6.061205
57. Tison, G., et al.: Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiol. **3**, 409–416 (2018). https://doi.org/10.1001/jamacardio.2018.0136
58. Vest, J., Gamm, L.: Health information exchange: persistant challenges & new strategies. J. Am. Med. Inform. Assoc. **17**, 288–294 (2010). https://doi.org/10.1136/jamia.2010.003673
59. Wang, W., den Brinker, A., Haan, G.: Discriminative signatures for remote-PPG. IEEE Trans. Biomed. Eng. **67**, 1462–1473 (2019). https://doi.org/10.1109/TBME.2019.2938564
60. Wang, W., den Brinker, A.C., Stuijk, S., de Haan, G.: Algorithmic principles of remote PPG. IEEE Trans. Biomed. Eng. **64**(7), 1479–1491 (2017). https://doi.org/10.1109/TBME.2016.2609282
61. Wu, J.Y., Wang, Y., Ching, C.T.S., Wang, H.M.D., Liao, L.D.: IoT-based wearable health monitoring device and its validation for potential critical and emergency applications. Front. Public Health **11**, 1188304 (2023)
62. Wu, W., Peng, H., Yu, S.: YuNet: a tiny millisecond-level face detector. Mach. Intell. Res. **20**, 656–665 (2023). https://doi.org/10.1007/s11633-023-1423-y
63. Yigzaw, K.Y., et al.: Chapter 14 - health data security and privacy: challenges and solutions for the future. In: Hovenga, E., Grain, H. (eds.) Roadmap to Successful Digital Health Ecosystems, pp. 335–362. Academic Press (2022). https://doi.org/10.1016/B978-0-12-823413-6.00014-8
64. Yu, Z., Peng, W., Li, X., Hong, X., Zhao, G.: Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 151–160 (2019). https://doi.org/10.1109/ICCV.2019.00024
65. Zhao, C., Chen, W., Lin, C.L., Wu, X.: Physiological signal preserving video compression for remote photoplethysmography. IEEE Sens. J. **19**(12), 4537–4548 (2019). https://doi.org/10.1109/JSEN.2019.2899102
66. Zhao, C., Lin, C.L., Chen, W., Li, Z.: A novel framework for remote photoplethysmography pulse extraction on compressed videos. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1380–138009 (2018). https://doi.org/10.1109/CVPRW.2018.00177

Passively Reconfigurable Antenna Using Gravitational Method

Thanatcha Satitchantrakul[1,2(✉)], Chanathan Manapreecha[1], Sukrit Phongpatrawiset[1], and Ping Jack Soh[2]

[1] King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
thanatcha.sat@gmail.com
[2] University of Oulu, Oulu, Finland

Abstract. This paper proposes a passive reconfigurable antenna by using liquid metal and gravity mechanism. When the antenna rotates to different angles, the liquid metal will flow to the lowest point of the container due to gravitational force, thereby acting as a reflector to redirect the main antenna radiation pattern at predefined angles at 5.8 GHz (ISM band). The upper section patterns can be maintained when antenna is tilted at 45°, 90°, 135°, 225°, 275° and 315° whereas the rotated antenna at 0° and 180° causes a forward radiation. More importantly, additional directors surrounding the main patch can be used to increase the directivity of these radiation patterns. Simulation results in terms of reflection coefficient, radiation patterns and gain indicate that the proposed antenna operated passively with a consistent bandwidth of approximately 400 MHz.

Keywords: Liquid Antennas · Reconfigurable Antennas · Gravitational Force · Passive Reconfiguration

1 Introduction

In the last decade, reconfigurable antennas have drawn significant attention from researchers due to the increase in demand for their functionality in modern communications. Reconfigurable antennas have the capability of changing or switching various characteristics i.e., frequency, radiation pattern and polarization. For that reason, they have been reported to be used in various applications such as communication systems, military devices, cognitive radio systems, and biomedical applications [1–3].

Reconfigurability techniques in radio components can be implemented by using either electrical, optical, mechanical, or material-based methods. Microelectromechanical switch (MEMs), PIN diode and varactor diode can be categorized as electrical methods. They generally offer ease of fabrication and cost efficiency, at the expense of more complex designs due to the need for biasing lines and circuits. On the other hand, photoconductive switching is typically integrated with optical or laser beams, which overcomes the need for biasing lines. However, this technique requires sufficiently powered and accurate beam generation to actuate photoconductive diodes in antennas and other radio components. Smart materials offer a low profile and potentially light-weight

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 147–154, 2024.
https://doi.org/10.1007/978-3-031-59091-7_10

solution. Depending on the type of materials, this solution is generally applicable as antenna substrates and therefore, are reconfigured on a larger scale with lower resolutions. Besides that, reconfiguration of these materials requires active devices to actuate changes in voltages, temperature, etc. On the other hand, mechanical reconfiguration techniques for antennas such as beam steering using motors also require active devices and may face wear-and-tear issues. Therefore, there is a need for a suitable and sustainable method without the need of integrating any active devices, and a good potential method for reconfiguring antenna properties is by changing its physical structure [1–3].

In general, main methods in actuating physical changes in antennas include using microfluidics, origami-based techniques, and rotating structures. A fascinating and potentially passive reconfiguration method is by using the gravitational force [4–6]. For instance, researchers in [4] characterized their own substance to build two non-miscible dielectric liquid layers to enable beam steering in a fixed upwards direction for a vehicle-mounted satellite communication system. On the other hand, liquid metal is used in [5] as the patch in rectangular patch antenna with frequency-reconfigurability to operate in 5G mobile network based on gravity and movement of the antenna.

In this research, a planar antenna which can be reconfigured passively using gravitational force is designed to operate in 5.8 GHz industrial, scientific, and medical (ISM) band using liquid metal. The Eutectic gallium-indium (EGaIn) will be selected as liquid metal due to low melting point and low toxicity [5]. Its main beam is directed towards a specific direction following the reflection of the liquid metal-filled cavity that surrounds the patch. Besides that, an additional set of directors strategically located at different adjacent positions to the patch is also considered in this investigation. Besides applying passive gravitational force for reconfiguration, this antenna is innovative from previous research as it offers higher directivity due to the use of directors to improve the directional pattern. The following section will present the design and operation of the proposed antenna, whereas the simulated behavior and performance of the antenna is discussed in Sect. 3. Our concluding remarks are presented in Sect. 4.

2 Antenna Design

2.1 Antenna Structure

Figure 1 depicts the structure of the proposed antenna. It is designed with a simple rectangular patch antenna with partial ground to ease radiation control. A coaxial feed method is used to minimize the effects of feeding on the reconfiguration of radiation patterns towards different directions. This concept is inspired by the concept of Yagi-Uda radiator-director coupling and is made more innovative using liquid metal. In this work, the liquid metal is EGaIn, which features a conductivity of $\sigma = 3.46 \times 10^6$ S/m, density = 6280 kg/m^3 and viscosity = 0.002 Pa•s A volume of 160.34 mm^3 is prefilled into a cavity that surrounds copper patch antenna, which is created on a 1 mm-thickness PDMS substrate ($\varepsilon_r = 2.75$) and located on a 1.6 mm-thick FR4 substrate ($\varepsilon_r = 4.3$, $tan\delta = 0.025$). Note that the patch and EGaIn-filled PDMS are completely isolated from each other. When the antenna rotates, the liquid will flow to the lowest point of this cavity due to gravity and act as a wave reflector.

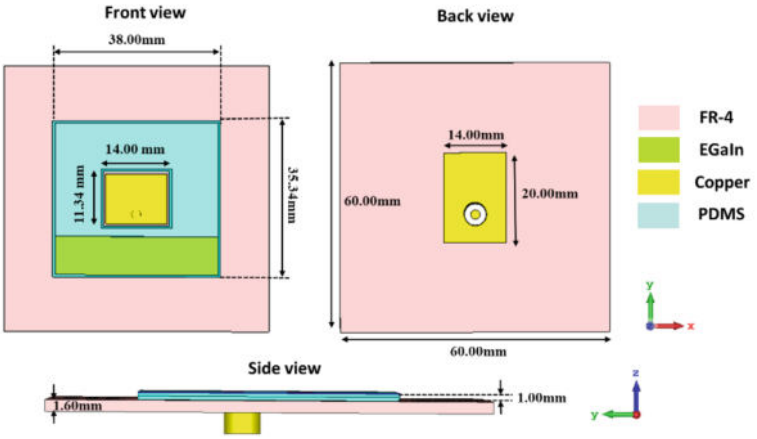


Fig. 1. The structure of the proposed reconfigurable antenna.

2.2 Directors

In this subsection, directors will be added adjacent to the structure to study the behavior of the main radiation patterns and the improvement in gain. Three main cases are investigated: without directors, with corner directors (90° apart) and with side and corner directors (45° apart), as illustrated in Fig. 2. In addition to that, antenna configurations with larger substrates (70 × 70 mm^2 instead of the 60 × 60 mm^2) and longer directors (10 mm instead of the proposed 6 mm) will also be studied.

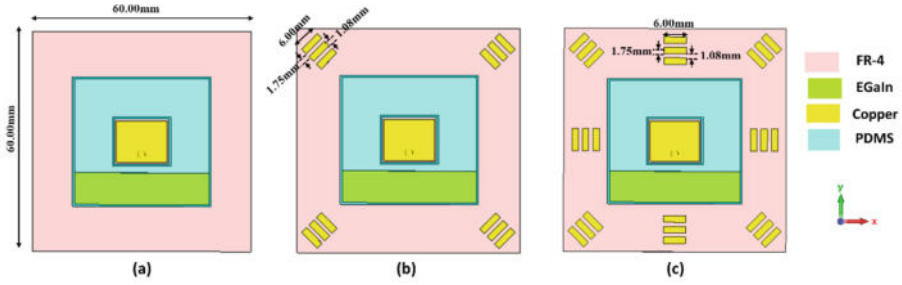


Fig. 2. The position of the directors of the proposed reconfiguration antenna; (a) without directors, (b) with corner directors (90° apart), (c) with side and corner directors (45° apart).

Throughout this work, the antenna is assumed to be operated when placed in a vertical direction (xy plane), as it would be in practice when mounted on a wall or attached to vehicle. This enables the liquid metal inside the cavity of the antenna to be reconfigured when tilted at eight angles, each at 0°, 45°, 90°, 135°, 180°, 225°, 270° and 325°. The performance of reconfiguration in terms of reflection coefficient (S_{11}), gain, bandwidth and radiation patterns will then be observed and analyzed. All simulations are performed using CST Microwave Studio software.

3 Results and Discussions

3.1 Without Directors

A summary of the reflection coefficients for the antenna rotated at different angles are shown in Fig. 3. It also includes the result of normal patch antenna without liquid metal. The normal patch antenna has resonance frequency at 5.74 GHz, which offers 2.8 dBi gain at only forward section. For the rotating antenna, due to the symmetry of the structure, its rotation at 45° and 315°, 90° and 275°, 135° and 225° resulted in the same S_{11}. On the other hand, the S_{11} from the rotation at degree 0° and 180° are not identical due to the asymmetry of the patch, ground plane, and coaxial feed. The operating frequencies are between 5.71 and 5.83 GHz with the central frequency of approximately 5.8 GHz. The bandwidths of this antenna are around 330–430 MHz. The highest gain of the structure, which is 4.13 dBi, is produced when the antenna is rotated at 45° and, 135°. All results are summarized in Table 1. The radiation patterns of the *xy* plane (E-plane) when rotated at each angle are shown in Fig. 4. It is observed that most of the patterns maintained radiation towards the upper section, even when the antenna is rotated. When the H-plane radiation is considered, the patterns are pointed in the upper-forward direction. However, the patterns at 0° and 180° showed the lowest gains compared to the other angles due to the asymmetric position of the feed and the effect of ground plane reflection. Only at these two angles, the patterns are stronger towards the forward section than the upper

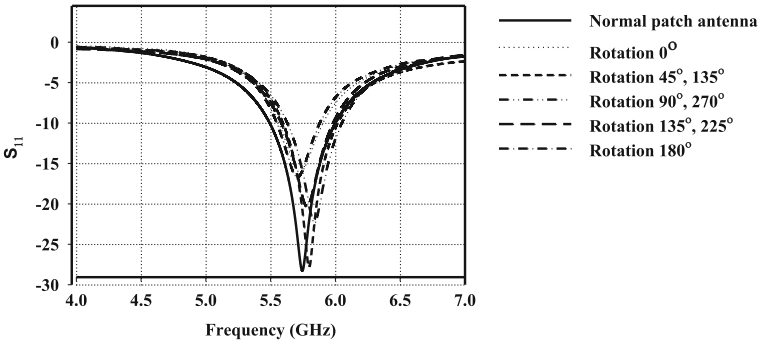


Fig. 3. The reflection coefficient of the antenna for various rotation angles.

Table 1. Properties of the proposed antenna without directors

Rotation	Gain (dBi)	Bandwidth (MHz)	Central frequency (GHz)
0°	2.15	340	5.73
45°, 315°	4.13	430	5.8
90°, 270°	3.66	420	5.83
135°, 225°	3.87	430	5.8
180°	2.05	330	5.71

section. To illustrate, Fig. 5 presents the example of H-plane radiation at the angles of 90o and 180°.

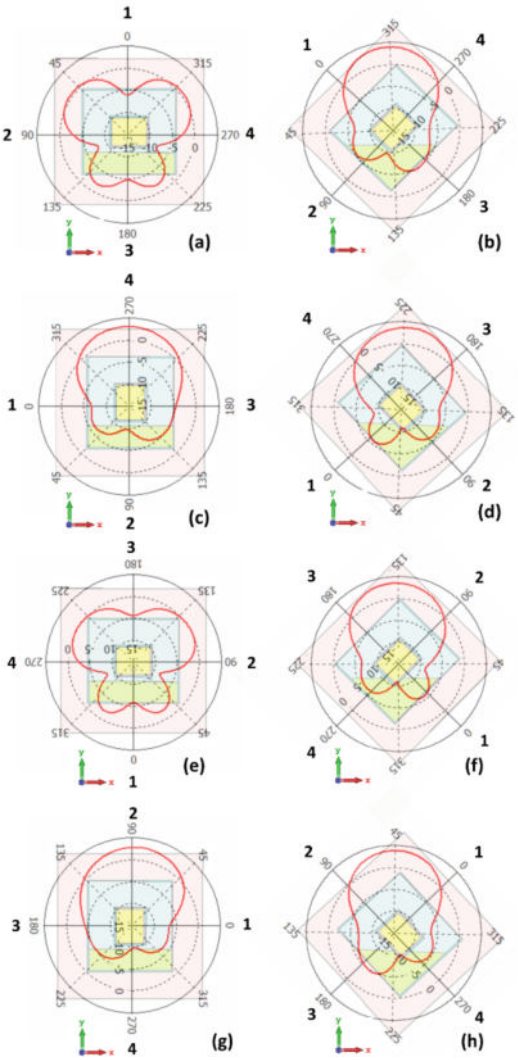


Fig. 4. Simulated radiation patterns at the E-plane (Phi (deg) vs dBi) of the proposed antenna for various rotation angles; (a) 0°, (b) 45°, (c) 90°, (d) 135°, (e) 180°, (f) 225°, (g) 270°, (h) 325°.

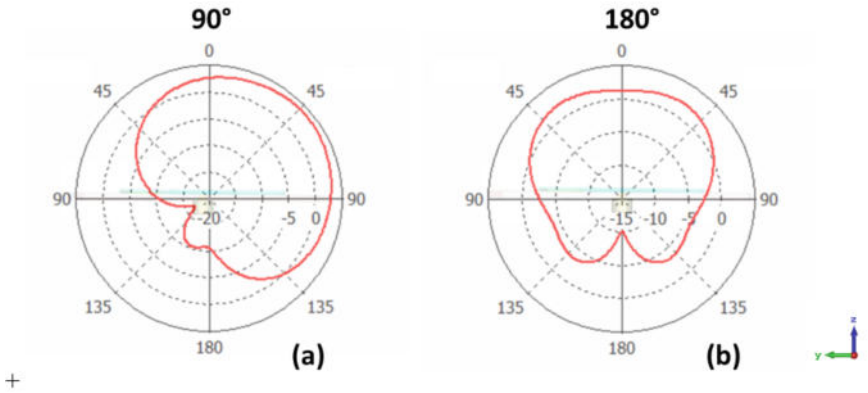


Fig. 5. Simulated radiation patterns at the H-plane (Theta (deg) vs dBi) of the proposed antenna for various rotation angles; (a) 90°, (b) 180°.

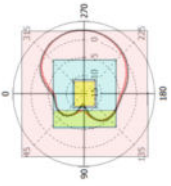
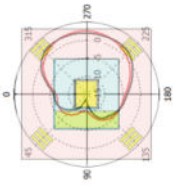
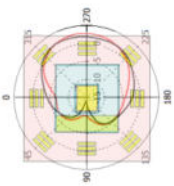
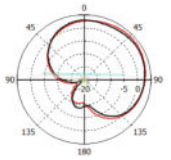
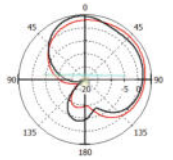
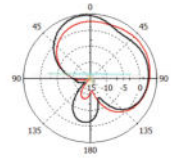
3.2 Effect of Directors

Table 2 displays the comparison of maximum gains between antennas without directors, with corner directors (at every 90°) and with corner and side directors (at every 45°) relative to the main radiator. It is seen that all three cases resulted in similar gain values. Thus, with this dimension, the directors are slightly affected. Then, the substrate dimensions have been changed to 70 × 70mm^2 with a longer (10 mm) director length. As an illustration, the example of an antenna rotating at 90° is chosen as a case study. When the substrate dimensions are larger, the resulting gains are higher, with directors changing the behavior of the corresponding patterns more significantly than a smaller substrate. This effect is shown in Table 3. However, with the addition of the directors, the major lobes of the radiation pattern in the *yz* plane (H-plane) become higher in the forward direction. Therefore, in such case, the antenna operating with only the cavity-filled liquid metal as a reflector is sufficient to perform beam reflection.

Table 2. Gain of the proposed antenna without directors and with different directors configurations.

Rotation	Without directors (dBi)	With corner directors (90° apart) (dBi)	With corner and side directors (45° apart) (dBi)
0°	2.15	2.25	2.16
45°, 315°	4.13	4.16	4.14
90°, 270°	3.66	3.53	3.55
135°, 225°	3.87	3.9	3.85
180°	2.05	2.13	2.03

Table 3. Comparison of the proposed antenna substrates.

Rotation 90°, 270°	Without directors (dBi)	With corner directors (90° apart) (dBi)	With corner and side directors (45° apart) (dBi)
Gain (dBi) substrate 60 x 60 mm^2	3.66	3.53	3.55
Gain (dBi) substrate 70 x 70 mm^2	3.28	4.25	4.56
E-plane Phi (deg) vs dBi			
H-plane Theta (deg) vs dBi			
			substrate 60 x 60mm² substrate 70 x 70mm²

4 Conclusions

A passively reconfigurable antenna using gravitational mechanism is proposed and studied in this paper. The antenna is designed on a low-cost FR4 and PDMS substrate, implementing the radiator-director concept in a Yagi-Uda antenna. Liquid metal EGain partially fills a cavity in the PDMS enclosure and is rotated to study its radiation behavior due to gravitational reconfiguration. Results indicate that the antenna operated with a consistent 330–430 MHz bandwidth centered at 5.8 GHz when rotated at eight different angles spaced at 45° when operated in a vertical configuration. Six of these eight rotation angles also produced a consistent radiation towards the upper direction despite these rotations. On the other hand, the remaining two angles of 0° and 180° mainly radiated towards the forward direction, similar as conventional patch antennas. Additional directors and a larger antenna substrate can be used to improve the directivity of these radiation patterns, but this expected improvement is not obvious due to the existence of the ground plane. Further experiments with actual implementation will be reported separately to study the practical aspects such as the flow of liquid metal on the antenna performance in the near future.

Acknowledgments. This research is supported in part by the Research Council of Finland (RCF)'s 6G Flagship Program (Grant no: 346208), RCF Academy Research Fellowship (Grant no. 355643) and the US Office of Naval Research (ONRG) (Grant no: N62909-22-1-2057). The authors also acknowledge King Mongkut's University of Technology North Bangkok for the workplace and equipment.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Mohamadzade, B., Simorangkir, R., Maric, S., Lalbakhsh, A., Esselle, K.P., Hashmi, R.M.: Recent developments and state of the art in flexible and conformal reconfigurable antennas. Electronics **9**(1375), 1–26 (2020)
2. Parchin, N.O., Basherlou, H.J., Al-Yasir, Y.A., Abdulkhaleq, A.M., Abd-Alhameed, R.A.: Reconfigurable antennas: switching techniques - A survey. Electronics **9**(336), 1–14 (2020)
3. Karthika, K., Kavitha, K.: Reconfgurable antennas for advanced wireless communications: a review. Wireless Pers. Commun.Commun. **120**, 2711–2771 (2021)
4. Song, C., Bennett, E.L., Xiao, J., Jia, T., Pei, R., Luk, K., Huang, Y. Passive beam-steering gravitational liquid antennas. IEEE Trans. Antennas Propagation (68), 3207–3212 (2020)
5. Qin, P., et al.: A gravity-triggered liquid metal patch antenna with reconfigurable frequency. Micromachines **12**(701), 1–14 (2021)
6. Huang, T., Zeng, L., Liu, G.B., Zhang, F.: A novel tailored coplanar waveguide circularly polarized antenna controlled by the gravity field. International Journal of RF and Microwave Computer-Aided Engineering 29 (10) (2019)
7. Gough, R.C., Morishita, A.M., Dang, J.H., Hu, W., Shiroma, W.A., Ohta, A.T: Continuous electrowetting of non-toxic liquid metal for RF applications. IEEE Access (2), 874–882 (2014)

A Skewness-Based Harmonic Filter for Harmonic Attenuation of Wearable Functional Near-Infrared Spectroscopy Signals

Hany Ferdinando[1(✉)], Martti Ilvesmäki[1], Janne Kananen[2], Sadegh Moradi[3], and Teemu Myllylä[1,3]

[1] Research Unit of Health Science and Technology, University of Oulu, Oulu, Finland
hany.ferdinando@oulu.fi
[2] Oulu Functional Neuroimaging (OFNI), University of Oulu, Oulu, Finland
[3] Opto-Electronics and Measurement Technique Research Unit, University of Oulu, Oulu, Finland

Abstract. Harmonics is an unavoidable phenomenon, even before we knew about digital circuits. In our sleep study, we found harmonic artefacts (HA) in our functional near-infrared spectroscopy (fNIRS) signal. Interestingly, it was neither device- nor subject-dependent. The fundamental frequency was around either 0.5 Hz or 1 Hz. It appeared to be very sharp peaks and they were within the band of interest, i.e., respiratory (0.1–0.6 Hz) and cardiac (0.6–5 Hz) bands. Since the exact location might change, we proposed a skewness-based harmonic filter (sbHF) to identify the fundamental frequency and attenuate HA. Since suppressing certain frequencies may change signal characteristic, spectral entropy was used to evaluate it based on Wilcoxon-test at a 0.05 significant level. 25 controls (6 females, age: 39.0 ± 8.5 years, height: 175.6 ± 8.0 cm, weight: 80.3 ± 10.8 kg) and 16 sleep apnea patients (1 female, age: 48.3 ± 12.4 years, height: 177.3 ± 6.0 cm, weight: 93.6 ± 17.1 kg) were recruited for our sleep study. sbHF showed good performance to identify fundamental frequency and attenuate HA from our raw fNIRS signals and 5% of the signal experienced changes in signal characteristics based on the spectral entropy analysis. Combining sbHF with a certain motion artefact reduction, we found that specific order of operation to get appropriate chromophore concentration was needed. This method is not only for problems in wearable fNIRS, but also can be modified for other problems by adjusting the suspected area or sweeping the frequency range to identify a fundamental frequency.

Keywords: skewness · fNIRS · harmonic artefacts · sbHF

1 Introduction

Harmonics are present anywhere in the world, especially when we are now living in digitized environments; hence, it was hard to avoid them completely. So, the only choice is to deal with them. Brillinger listed interesting physical examples that led to harmonics [1], which even appeared before digital system was introduced to our world. The

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 155–167, 2024.
https://doi.org/10.1007/978-3-031-59091-7_11

important thing to deal with harmonics is identifying the fundamental frequency. Once this frequency is identified, a simple notch filter can be applied accordingly to attenuate the harmonics.

Functional near-infrared spectroscopy (fNIRS) has been greatly improved and widely used in many applications, including health. Having cleaned signals without any artefacts is an ultimate hope for any signal analysis to provide reliable results. fNIRS, however, is not free from these artefacts either and different researchers may have different opinions about artefacts. For example, for those who are interested in hemodynamic response, cardiac beat is considered an artefact that must be removed prior to further analysis, but others may use the cardiac pulsation to extract features from fNIRS signals.

The trend in digital health motivates fNIRS miniaturization as a wearable device to enable location-free measurement in any situation, e.g., sleep, exercise, working, etc. With the advancement of secured wireless connections, it enables patient home monitoring. However, it is difficult to avoid interferences that may induce artefacts in these uncontrollable environments. A common example is power line interference in ECG and EEG signals [2, 3]. Hence, we need to make sure that the signals are usable with minimum artefacts if we cannot remove them completely.

We developed a wearable device able to measure hemodynamic and electrical activities of the neurons from the brain, specifically from the forehead area. It consists of light source and detectors for fNIRS to read hemodynamic response, EEG electrodes, and accelerometers.

From signals measured in our sleep study, we found a special artefact in raw fNIRS signal. This artefact appeared as harmonic frequencies clearly visible in the frequency domain, while signal shape in the time domain looked normal. In this case, no one can identify the presence of this artefact until the signals were viewed in the frequency domain. Since the artefact was found as harmonics in frequency domain, we call it harmonic artefact (HA).

Interestingly, the artefact was neither device- nor subject-dependent because we did not get HA in other measurements from different studies using the same device. Since the source of HA is still unknown, a similar scenario might occur somewhere else, especially when the wearable system is worn outside uncontrollable environments. Figure 1 showed examples of the signals from the same device measured on different days in our sleep study.

Baratta et al. addressed a similar phenomenon appeared in EMG signals [4]. They used baseline signals to estimate the background noise in frequency domain and subtracted it from the spectrum of the measured signals. The signals were reconstructed by applying inverse Fourier transformations. Although this method looked promising, it was not applicable to our problem because we cannot isolate the signals without any pulsation as the baseline. Hence, another approach is needed.

The fundamental frequency was around either 0.5 Hz or 1 Hz, which was within our band of interest in fNIRS signal analysis. If we compare them to fundamental frequency from power line interference, they are relatively low. In some cases, it is easy to identify power line interference in time-domain because the fundamental frequency sometimes can be visible. However, it is difficult to spot such a low fundamental frequency as in our case.

Having fundamental frequency around either 0.5 Hz or 1 Hz poses two challenges, i.e., firstly, we need an algorithm to detect the exact location of the fundamental frequency, and secondly, attenuating these harmonics should give minimum impact on the signal characteristic because the interested physiological pulsations are within this range. These two challenges were addressed in this present study.

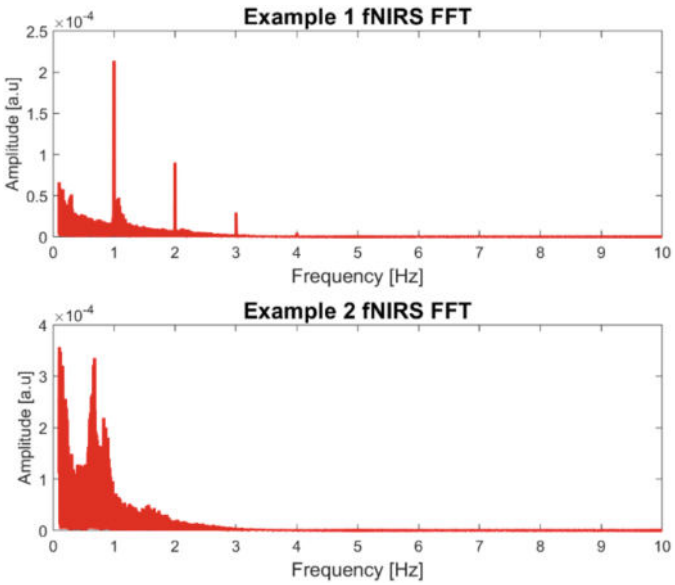


Fig. 1. Raw signals in frequency-domain from different subjects and measurement days using the same device.

2 Materials and Methods

2.1 Data Collection

The data collection followed the Declaration of Helsinki and the whole measurements were conducted at the Clinical Skills Centre Knoppi, Oulu University Hospital which is a collaboration environment between hospital and faculty of Medicine, University of Oulu. We recruited adult subjects for this study, and a consent letter was signed after the explanation of the measurement. Subjects might leave the measurement anytime during data collection.

Within the sleep study, the objective is to discern alterations in brain pulsations in sleep apnea patients when compared to a control population. To facilitate this investigation, we employed functional near Infrared Spectroscopy (fNIRS) and standard clinical night polygraphy equipment (NOX T3s) that will measure breathing movements, oxygen saturation, pulse, and respiratory flow. This approach allows for the correlation of fNIRS signals with precisely timed pauses in breathing, such as hypopneas or apneas,

which are used to calculate Apnea-hypopnea index (AHI). AHI is considered normal if the index is under 5. Additionally, in collaboration with Tampere University Hospital, signals were captured using the Emfit sleep sensor placed under the mattress, a modality that they are in TAYS region routinely utilize in clinical sleep apnea diagnostics. Furthermore, equipment with three different devices from Polar Electro Oy was used to measure physiological signals, allowing for comprehensive comparisons with signals obtained from other modalities. One device was placed on wrist and was comparable to commercial version of Polar Vantage, additionally two other devices, the light sensor to upper arm and impedance cardiography around chest wall, were used. Figure 2 shows the equipment used in this study. This methodology ensures accurate insights into the physiological pulsations of the brain, as well as the overall physiology of the body and the functioning of the nervous system in individuals with sleep apnea during the night.

Fig. 2. We put several devices on the subject and the one on the forehead was our fNIRS device (left). Measurement environment at Clinical Skills Centre Knoppi and some sensors (right)

The study included 47 measurements of which 41 measurements had valid NOX signals. The inclusion criteria for study subjects were following; Age between 18–68 years old, no neurological diseases, no smoking. Underlying diseases that were medicated and therefore in control e.g. hypertension or diabetes, were allowed. After classification 25 controls (AHI < 5, 6 females, age: 39.0 ± 8.5 years, height: 175.6 ± 8.0 cm, weight: 80.3 ± 10.8 kg) and 16 sleep apnea patients (AHI 20.5 ± 16.6 events/h, 1 female, age: 48.3 ± 12.4 years, height: 177.3 ± 6.0 cm, weight: 93.6 ± 17.1 kg) were included to further analysis. Subjects were recruited from the outpatient clinic at Oulu University Hospital or with email.

2.2 Wearable Device

The wearable device is a battery powered NIRS device, which is made of two different units: head and main units. They are connected to each other using a cable. Head unit is placed at forehead and consists of a 4-in-1 LED (980 nm, 830 nm, 810 nm, and 690 nm) and 2 photodiode sensors. The photodiodes are symmetrically placed at both sides of LED by a separation of 3 cm to measure left and right hemisphere hemodynamic. At the main unit, the rest of the circuits are embedded: demodulator front ends, analog filters, LED modulators and drivers, ADC, and microcontroller. The data can be transferred to PC via USB cable or be saved in SD card.

One of the major challenges in design of this device was head unit. People have different shape of the forehead, and it forced us to use flexible circuits for head unit. Another challenge is sensors attachment on the skin. Applying more pressure on the forehead guarantees better illumination of the brain tissue. At the same time, the photodiodes can capture more lights. However, it can be painful for the subjects and can even produce allergies reactions. On the other hand, with a looser head unit placement, LEDs need more current to produce more light, resulting in battery depletion as well as increasing light reflection noises. Moreover, the photon may propagate on the skin and arrive at the photodetector, inducing noise the hemodynamic signals. So, there is a tricky trade-off in head unit design for sleeping purposes that demand lots of trial and error to find the best pressure level and perfect sensor angling. We have designed and tested several head unit designs to realize our goal.

2.3 HA Attenuation Algorithm

We only knew that the fundamental frequency was around either 0.5 Hz or 1 Hz. For this reason, the search area can be defined easily by defining the fundamental frequency as its center and the width. Hence, the signals must be transformed to frequency domain prior to this detection.

The fundamental frequency is identified as a sharp spike at certain frequency. However, choosing a frequency with the highest power within the search area is not the answer. For signal without HA, we can always find such a frequency with the highest power, see Fig. 1(bottom). Thus, a metric should be used to identify if a frequency with the highest power is a fundamental frequency or not.

The simplest method was comparing the maximum to average power within the search area. Unfortunately, the ratio could not be generalized to get a certain threshold for all subjects to identify it as a fundamental frequency or not. Moreover, introducing new subjects may change the threshold, making the algorithm not adaptable to most of the scenarios. The main challenge was on the power around the fundamental frequency. Specific for fundamental frequency around 1 Hz, it overlaps with the common heartbeat at rest for adults. When the heartbeat changes from time to time, the average power within the search area can be quite high and it may fail to give an appropriate metric when the power of the fundamental frequency is not that high.

We studied the distribution of power within the search area and found that the presence of the fundamental frequency increases the skewness of the power distribution. Without the fundamental frequency in the search area, the skewness of the distribution

is less than one. So, skewness value can be used as an indicator to identify the presence of the fundamental frequency.

Based on the detected fundamental frequency, we applied a 2nd-order IIR-notch filter to attenuate HA. From our preliminary experiments, we found that using high Q values to get a narrow bandwidth was crucial. Fundamental frequency around 0.5 Hz also introduces a spike around 1 Hz, which will be found within the search area. For this reason, we need to set the algorithm to avoid double attenuation when the fundamental frequency is found around 0.5 Hz. The algorithm to detect and attenuate HA can be read from Fig. 3. Later, this algorithm is called skewness-based Harmonic Filter (sbHF).

```
1   s = input signal
2   s_cleaned = s // copy the signal
3   L = signal length // get the length of input signal for FFT operation
4   span = span to left and right from the fundamental freq.
5   Y = fft(s)
6   Calculate power from Y
7   isolate power within search area around 0.5 Hz → P1
8   isolate power within search area around 1 Hz → P2
9   If skewness(P1) > 1
10    get fundamental freq. around 0.5 Hz → fund_freq_tmp
11  elseif skewness(P2) > 1
12    get fundamental freq. around 1 Hz → fund_freq_tmp
13  if fund_freq_tmp exist
14    n = 1 // set multiplier for the harmonic
15    // reduce harmonics up to 10 Hz
      while (n*fund_freq_tmp) < 11
16      // refine the detected fundamental freq.
        find max power between fund_freq_tmp-span and fund_freq_tmp+span
17      fund_freq is found based on the max power
18      calculate coefficient of IIR-notch filter at fund_freq
19      apply 2nd-order IIR-notch filter at fund_freq to s_cleaned
20      n = n + 1;
```

Fig. 3. Pseudocode of the skewness-based Harmonic Filter (sbHF)

The algorithm starts with copying input signal to the cleaned version (line 2). A very narrow search area is defined based on the span variable. Since it works in the frequency domain, FFT is employed (line 5). We isolate power from two bands of interest around 0.5 Hz and 1 Hz with the bandwidth of `2*span` (line 7–8). Next, sbHF evaluates skewness of the isolated power and compares it with the threshold value to get fundamental frequency (line 9–12). The exact fundamental frequency and its harmonics were not at precise locations. So, we need to refine it to get the real fundamental and harmonic frequencies (line 16–17). The harmonic frequencies were removed by applying IIR-notch filter (line 18–19).

Attenuating certain frequencies in the signal changes the signal characteristics. We must evaluate if our proposed algorithm did not change them. Since the attenuation involves oscillations at certain frequencies, we used spectral entropy value to evaluate before and after applying the algorithm. Spectral entropy measures the flatness of the spectrum by ignoring the order of oscillation found in the signal [5]. So, changes in signal characteristic can be detected using this method.

The spectral entropy was applied to every 10-min segment of raw signals across the whole measurement. In this way, we collected spectral entropy scores from both original and cleaned signals. A Wilcoxon-test at 0.05 significant level was used to evaluate whether the algorithm changes the signal characteristic or not. Hence, for each subject we had eight p-value scores: two channels and four wavelengths from each channel. We only evaluated signal characteristic changes when sbHF was applied.

3 Results

Using original raw signals without HA, we corrupted them by adding artificial HA using fundamental frequencies of 0.5 Hz and 1 Hz. It was a series of sinusoidal signals with decaying amplitudes as the frequencies increase. We used these signals to test our proposed algorithm to detect and attenuate HA. For this experiment, we used Q factor of 500 for IIR-notch filter.

Figure 4 presents the performance of our algorithm using signal with artificial HA. The upper row shows the original raw signals in frequency domain, which have no HA. In the middle and bottom rows, sbHF detected (marked with blue dots) and attenuated artificial HA respectively. We use root mean square error (RMSE) to compare the cleaned signal to the original one (without HA), i.e., 1e−3 and 9e−4 for 0.5 Hz and 1 Hz fundamental frequencies respectively.

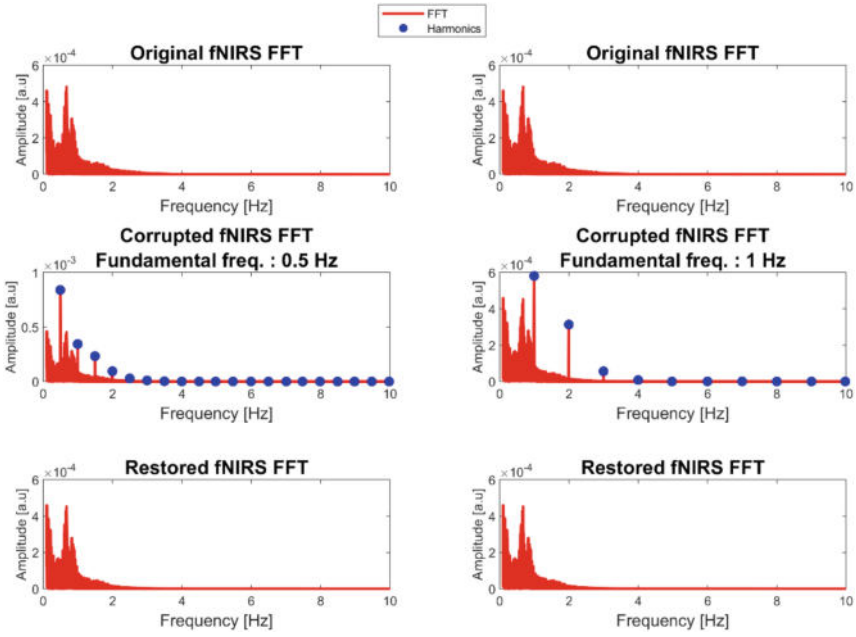


Fig. 4. The location of the fundamental frequency of 0.5 Hz (left column) and 1 Hz (right column) can be detected based on the skewness score of the distribution (middle row) and then removed (bottom row).

Figure 5 compares the spectrum before and after applying sbHF from three possible cases using real raw signals, i.e., no HA, HA with fundamental frequency around 0.5 Hz, and HA with fundamental frequency around 1 Hz in the raw signals. When sbHF found no harmonic artefact, the algorithm did not apply anything to the raw signals (Fig. 5 left column). In Fig. 5 middle and left, sbHF successfully detected and attenuated HA in the other cases.

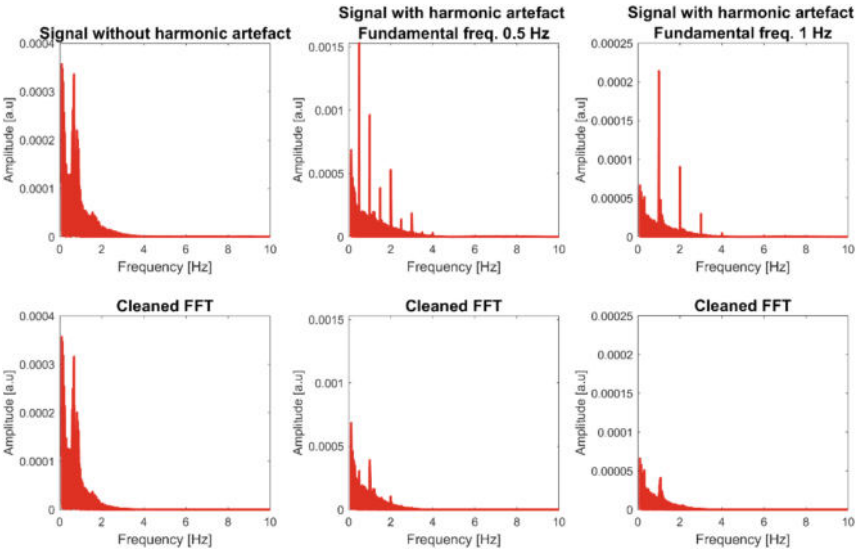


Fig. 5. sbHF performance in three different cases. On the left column, sbHF did not detect any fundamental frequency, hence sbHF was not applied to this signal. For middle and right columns, sbHF identified fundamental frequency at around 0.5 Hz and 1 Hz, respectively, and attenuate the harmonics, see results at the bottom row.

Figure 6 presents how signal changes in time-domain after applying sbHF. The shape of the individual pulse slightly changed as the harmonics were attenuated, see left (before) and right (after) of the top panel.

Our sleep data also contained motion artefacts, which cannot be avoided as subjects were free to move during sleeping. Among various proposed algorithms to reduce this artefact, we found kurtosis-based Wavelet Filtering (kbWF) [6] provided the optimum results. Then, the interesting question is about how to combine these two algorithms to process our sleep data until we get chromophore concentrations for further analysis.

To answer this question, we employed six scenarios based on the order of operation among kbWT, sbHF, and chromophore calculations. In each scenario, we processed raw signals until we got the chromophore concentration in different orders and see results from each scenario. Figure 7 shows result from each scenario (case 2–7) and the references (case 1), while Fig. 8 presents signals before and after the manipulation using case 7.

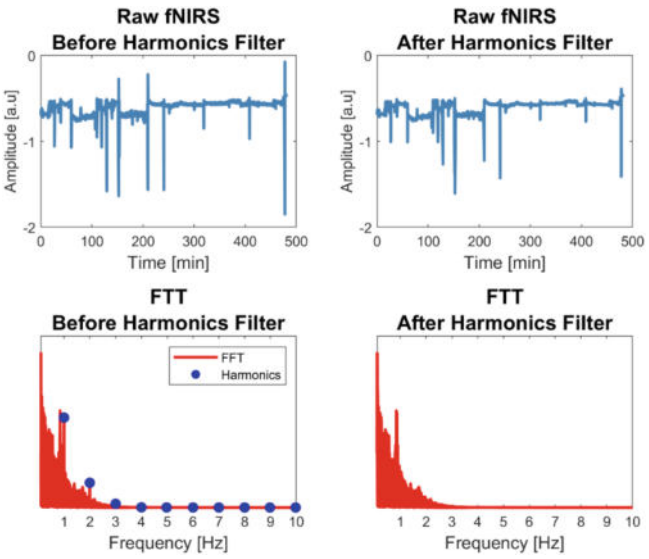


Fig. 6. Signal in time-domain before and after applying sbHF. It removed fundamental frequency around 1 Hz but kept the cardiac beat peak under 1 Hz. Signal shape in time domain was generally unaltered.

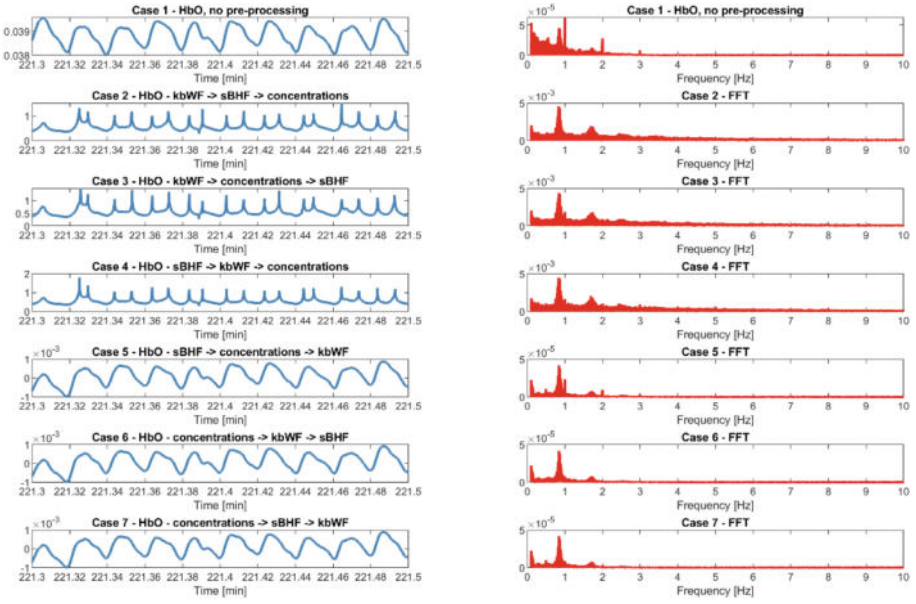


Fig. 7. Processing raw signals with different order produced different results.

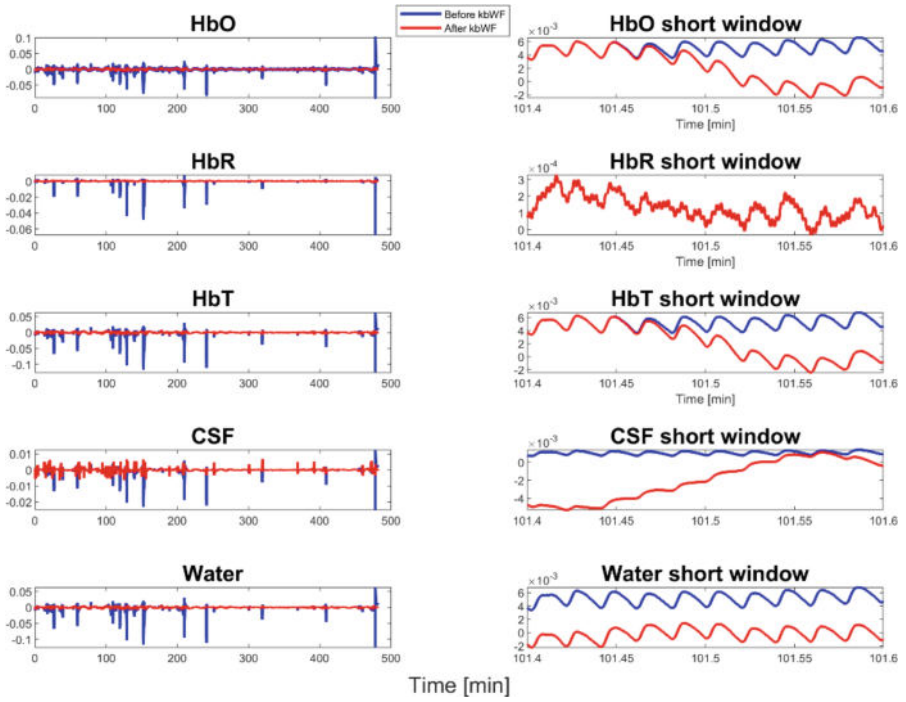


Fig. 8. Processing HbO using case 7.

4 Discussions and Conclusions

We proposed an algorithm to attenuate HA in fNIRS signal. Firstly, the algorithm detects the presence of the fundamental frequency based on the skewness value of the isolated power within a certain narrow band of interest. Then, 2nd order IIR-notch filter is applied repetitively up to 11 Hz to attenuate HA. Applying sbHF to raw signals successfully attenuated HA, while signals without HA were left as they were.

We evaluated sbHF using original signals corrupted with artificial HA, see Fig. 4. Based on the RMSE score, sbHF looked promising for our purpose as it could identify if input signals contain HA and attenuate it. Using the original signals, sbHF can detect the presence of the real HA and then attenuate it. Hence, the skewness of the distribution provides good information about the presence of the fundamental frequency.

Attenuating HA means reducing the amplitude of some frequencies. Consequently, it also changes signals in the time-domain as some of the components are attenuated, see Fig. 6. For this reason, we need to evaluate if sbHF changes the signal characteristics. If signal characteristic changes, then sbHF is useless. So, we compared SE scores before and after applying sbHF and found that 10% (13 out of 127) of the signals with HA had different SE scores mostly from cardiac band. The IIR-notch filter attenuated frequency components within its narrow bandwidth. Perhaps, we need narrower bandwidth to preserve most of the signal characteristics. When we doubled the Q value, 10% of the processed signals still had different SE scores.

We experimented with a high-order IIR-notch filter. The filter order was increased from 2 to 10 and 50 to get extremely different filter order. Unfortunately, no significant improvement was made in both scenarios. It looked filter order has no impact on the performance of sbHF.

We had been using the same Q for all harmonic frequencies. Consequently, the larger the frequency, the wider the bandwidth and it may affect the performance of sbHF. In this case, the larger harmonic frequency would have more frequencies being attenuated. So, we attempted to change the Q value accordingly such that all harmonics frequencies had the same bandwidth. Table 1 displays experiments by varying initial Q factor using 2nd-order IIR-notch filter.

Table 1. Performance of sbHF using the same bandwidth for all harmonics frequencies and 2nd-order IIR-notch filter

Initial Q	Proportion of signal with different SE Score (%)
500	10
1000	5
2000	4

Although the initial Q values in Table 1 were quite sparse, we could see here that by varying Q factor to get the same bandwidth for each harmonic frequency helped improving the performance. Perhaps the performance gradually improved until at a certain point it gave just an almost flat response with a little improvement.

Next, we investigated whether the filter order may change the performance under varying Q factor conditions. Unfortunately, changing the filter order did not improve the performance. High filter order has steeper transition than the low one. As the bandwidth around the harmonic frequencies was small enough, the transition slope was already steep and changing filter order did not give significant transition changes.

Apparently, Q factor affects the computational time, but it is not an important issue when sbHF is applied offline. The limit of this parameter is the computational power of the computer. Of course, it is important to optimize it, but it is beyond the scope of this present study. It will be considered as a future work.

Up to this point, we were still working on the raw signals without any chromophore calculation. Moreover, fNIRS signals from sleep study were not free from motion artefacts. So, it is important to evaluate how to combine sbHF with motion artefacts reduction and chromophore calculation. The examples of the experiments were shown in Fig. 7 for different cases.

Although signals in frequency domain looked fine, see the right column, the representation in time-domain, see left column, could be completely different in case 2, 3, and 4. These results indicated clearly that concentration calculation must be done before applying kbWF, see case 5, 6, and 7. In case 5, not all HA was attenuated, see the visible spikes at 1 Hz and 2 Hz. Case 6 and 7 looked similar, but case 7 presents better signal representation as it keeps the respiratory peak around 0.5 Hz; this respiratory peak is also visible in case 1.

Figure 8 displays the results after applying case 7 to all possible concentrations from our wearable fNIRS. The whole measurement, see left column, clearly shows that the motion artefact was reduced by kbWF. Cardiac pulsation, which is one of our interests, can be conserved well after applying sbHF, see right column.

The proposed algorithm defined the fundamental frequency at 0.5 Hz or 1 Hz. It can be extended to seek out fundamental frequency within specific range. Hence, it is more flexible for more general applications.

We can extend this method to deal with harmonic artefacts in photoplethysmography (PPG) and pulse oximetry signals. Both signals are harvested using the same principle as in fNIRS. Even more interesting, PPG signals from the wrist are commonly available in wearable devices, where the user are in uncontrollable environments.

The main part that needs improvement is the attenuation filter. IIR-notch filter attenuates the central frequency up to the minimum level. It means certain frequencies are reduced significantly. This phenomenon changes some of the signal characteristics, including the ones within our interest, e.g., respiratory band in 0.1–0.6 Hz. If we can adjust the attenuation gain such that the center frequencies are attenuated up to the level of their neighboring, perhaps we can preserve more information in the signal. This will be our future work in addition to optimizing the initial Q factor.

Acknowledgments. Special thanks to Jari Paunonen and Auri Myllylä for the hard work in conducting overnight sleep data recordings for this study. This work was supported by the Academy of Finland Profi6 programme: 6G-Enabling Sustainable Society (6GESS) and 6G Flagship programme, and Business Finland 6GBRIDGE - Next generation healthcare and wearable diagnostics utilizing 6G project (11146/31/2022).

Disclosure of Interests. The authors have no conflict of interest.

References

1. Brillinger, D.R.: Fitting cosines: some procedures and some physical examples. In: MacNeill, I., Umphrey, G., Bellhouse, D., Kulperger, R. (eds.) Advances in the Statistical Sciences: Applied Probability, Stochastic Processes, and Sampling Theory, pp. 75–100. Springer, Netherlands, Dordrecht (1987)
2. Sörnmo, L., Laguna, P.: EEG signal processing. In: Bioelectrical Signal Processing in Cardiac and Neurological Applications, pp. 55–179 (2005). https://doi.org/10.1016/B978-012437552-9/50003-9
3. Sörnmo, L., Laguna, P.: ECG signal processing. In: Bioelectrical Signal Processing in Cardiac and Neurological Applications, pp. 453–566 (2005). https://doi.org/10.1016/B978-012437552-9/50007-6
4. Baratta, R.V., Solomonow, M., Zhou, B.H., Zhu, M.: Methods to reduce the variability of EMG power spectrum estimates. J. Electromyogr. Kinesiol. **8**, 279–285 (1998). https://doi.org/10.1016/S1050-6411(97)00031-X
5. Ferdinando, H., Moradi, S., Korhonen, V., et al.: Spectral entropy provides separation between Alzheimer's disease patients and controls: a study of fNIRS. Eur. Phys. J. Spec. Top. **2022**, 1–8 (2022). https://doi.org/10.1140/EPJS/S11734-022-00753-W

6. Chiarelli, A.M., Maclin, E.L., Fabiani, M., Gratton, G.: A kurtosis-based wavelet algorithm for motion artifact correction of fNIRS data. Neuroimage **112**, 128–137 (2015). https://doi.org/10.1016/j.neuroimage.2015.02.057

Wearable Motion Sensors in the Detection of ADHD: A Critical Review

Jakov Basic[1], Johanna Uusimaa[2], and Juha Salmi[1](✉)

[1] Aalto University, Otakaari 24, 02150 Espoo, Finland
juha.salmitaival@aalto.fi
[2] University of Oulu, Aapistie 5, 90220 Oulu, Finland

Abstract. Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder with inattention, hyperactivity, and impulsivity as core symptoms. Current diagnostic methods of ADHD consisting of interviews and self-ratings come with a risk of subjective bias and are dependent on the limited availability of healthcare professionals. However, recent technological advances have opened new opportunities to develop objective and scalable methods for precision diagnostics. The present critical review covers the current literature concerning one of the promising technologies, the use of motion sensors or accelometers for detecting ADHD, particularly evaluating the related clinical potential. Several studies in this field, especially recent studies with advanced computational methods, have demonstrated excellent accuracy in detecting individual participants with ADHD. Machine learning methods provide several benefits in the analysis of rich sensor data, but the existing studies still have critical limitations in explaining the underlying cognitive functions and demonstrating the capacity for differential diagnostics is still underway. Clinical utility of sensor-based diagnostic methods could be improved by conducting rigorous cross-validation against other methods in representative samples and employing multi-sensor solutions with sophisticated analysis methods to improve interpretation of the symptom manifestation. We conclude that motion sensors provide cost-effective and easy-to-use solutions with strong potential to increase the precision and availability of ADHD diagnostics. Nevertheless, these methods should be employed with caution, as only a fraction of ADHD symptoms relate to hyperactivity captured by motion sensors. At best, this technique could complement the existing assessment methods or be used along with other digital tools such as virtual reality.

Keywords: ADHD · Diagnostics · Motion sensors · Accelometer · Machine Learning

1 Introduction

1.1 What is ADHD and How is It Currently Diagnosed?

ADHD (attention deficit hyperactivity disorder) is the most common childhood neurodevelopmental disorder with a prevalence of about 5–7.2% in children and adolescents and 2.5–6.7% in adults [1]. The relative number of ADHD diagnoses has been rapidly increasing in the Western countries: For instance, in US ADHD prevalence raised from 6.1% to

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 168–185, 2024.
https://doi.org/10.1007/978-3-031-59091-7_12

10.2% between 1997 and 2016 [2]. While the causes of 'ADHD epidemic' remain partially unclear, the increase in referrals to psychiatric care has resulted in global healthcare crisis where the resources are not matching the needs [3]. At the same time, various new opportunities for how technology could assist in related healthcare solutions have been found. One exciting potential landscape involves the use of motion sensor technology in ADHD diagnostics.

Out of the two broader ADHD symptom domains, inattention, and hyperactivity/impulsivity, especially the latter one that directly concerns physical movements of the body could potentially be objectively quantified with motion sensors. Hyperactivity/impulsivity symptoms include, for example, fidgeting or tapping with hands or feet, squirming in the seat, leaving when remaining seated would be appropriate, and running about in situations where one is expected to not do so. To detect inattention there are other technological solutions such as virtual reality [4], but it is good to keep in mind that the symptom domains are often highly correlated and capturing a single domain reliably may hence provide valuable information even on the broader scale. While large-scale initiatives have been recently made to improve the precision of psychiatric diagnostics [5, 6], the current ADHD diagnostics is still far from precise quantitative measurement relying instead on subjective evaluations gathered with structured interview and symptom screening questionnaires. Subjective experiences are sensitive to various biases [7] that are dependent on the awareness and reliability of the informants, interpretation of the questions, ability to scale the outcomes against others, and generally the longitudinal data provided by technologies with high sampling rates has benefits against scarce evaluations of one aspect in life over the months. Hence, some limitations of the current diagnostics could be potentially tackled with objective, quantitative sensor-based methods solidly grounded on the biobehavioral reality to which so many other fields in medicine rely on [8]. Could a machine do the assessment of hyperactive-impulsive patterns better than a man, and what would be required for fulfilling the high medical standards?

1.2 Movement Sensors in ADHD Assessment

Movement sensors such as an accelometer or gyroscope, have been popular in actigraphs employed for research purposes already for several decades [9]. During the past ten years related solutions have become common also in customer devices (e.g., mobile phones, smart watches, and other wearables), which has facilitated related technological developments even further. By gathering information about linear acceleration in multiple axes, microelectromechanical accelometers can reliably detect gross body movements but also provide signal for more detailed analysis of movement patterns and trajectories. Gyroscopes using gravity to determine orientation of the movement can detect also angular velocity or rotation of the moving object, which allows more comprehensive interpretation of the motion signals. A combination of these two sensor types hence gives the most precise picture of the movement features. Characterizing the type of movement signal is also affected by the number of sensors and sampling rates (e.g., 1–100 Hz), which considerably affects the battery consumption and sensitivity to measurement noise [10, 11]. Choosing the sensor types, temporal resolution, and the number of sensors depends on the need. For instance, detecting the overall activity level or even the quality of sleep does

not require high precision signals, while the capturing movement signals mimicking natural human behavior in its richness (e.g., sports) have different hardware requirements [12]. Regarding detecting and interpreting the type of hyperactivity in ADHD, such measurement standards remain to be carefully investigated.

Precise modelling of human movements requires not only high-quality input data but also benefits of advanced computational methods [5]. Machine learning methods such as convolutional neural nets (CNN) can detect regularities in movement trajectories signaling, for instance, different posture positions, movement types, or activities [13]. Such methods should be able to detect all possible variants within any single interpretable movement category at individual level. This is a big challenge in studying heterogeneous disorders like ADHD and yet another issue to tackle is the context where the movements take place that should be carefully considered when examining ADHD symptoms, as the symptoms essentially relate to whether the movement is appropriate in a specific context rather than whether the movement is appropriate as such. Here, determining the movement in relation to other individuals (inter-individual differences) and deriving the changes in the movement patterns of the same person in different contexts (intra-individual differences) come into play. More specifically, like human individuals, machine learning algorithms may learn to identify certain types of movements in a particular context (e.g., fidgeting with hands or feet), but this comes clinically interpretable only when the system has first characterized whether the movement signals maladaptive behavior in the specific measurement context (e.g., observed during a school class when one should stay still and concentrate). One approach that helps here is supervised learning: when reference data where classification has already been done is available and the training sample is representative and large enough, such methods provide powerful means [14]. Alternatively, when predesigned categorization information is not available, it is possible to use unsupervised learning where the algorithm categorizes the data according to the statistical regularities in the input [14]. This method can be powerful in detecting, for instance, inter-individual differences. Along with manual annotation, both of these approaches could provide higher interpretability than in the analysis of gross movement levels, which may also vary between individuals with *vs.* without ADHD [15].

1.3 The Present Study

This critical review will examine the existing literature concerning clinical utility of motion sensors measuring bodily movements in ADHD diagnostics, especially in the context of a) research quality (e.g., interpretability of the signals, contextual control) and b) diagnostic standards (e.g., representativeness of the study samples, observations in different contexts, length of the measurements or test-retest reliability). We hope that this paper raise questions helping to improve sensor-based research of ADHD and is able to guide development of future health care applications. As the current research in this field has used highly heterogeneous methods, it is important to raise up questions that would allow building high clinical quality standards. Moreover, for the clinicians it is currently difficult to evaluate the readiness state of the technology as such critical analysis is lacking. One important caveat here is how to interpret the performance of the methods.

For example, some of the studies have reported considerably high classification accuracies (e.g., >98%) and if such studies would consider all the relevant clinical aspects, one could easily argue that the method is ready for clinical use. However, it should be borne in mind that detection accuracies are highly dependent on the difficulty of the specific classification problem that in this case largely raises from sample characteristics. From the clinical point of view, the algorithm should be able to identify the status of every single individual that comes to the assessment. For this reason, population-wide representativeness of the training and testing samples would be utterly important. In most cases individuals with ADHD have also other problems. Indeed, challenges in the clinical assessment especially concern evaluating the severity of the problems near the diagnostic threshold and ruling out the possible other problems that may overlap with ADHD (e.g., autism [16], learning disabilities [17] as well as in mood, anxiety, and conduct disorders [18, 19]). Finally, it is worth noting that this review will not cover comparisons between motion sensors and other methods employing machine learning in detail as these were at focus in another recent review [5]. We will also focus more on the more recent and technologically advanced studies with higher clinical potential as older studies with standard methods have been carefully meta-analyzed by De Crescenzo and colleagues [15].

2 Methods

2.1 Study Selection

Two researchers (JB and JS) independently conducted the search and selection of the studies. To find the relevant studies, we employed PubMed and Scopus as the primary search engines. PubMed's comprehensive coverage of biomedical and life sciences publications provided a strong foundation for identifying relevant studies. Additionally, Scopus, with its multidisciplinary scope and strong representation in life sciences, complemented the search potentially capturing research studies not included in PubMed. For both search engines we used search words: ADHD AND (movement OR motion) AND sensor AND diagnostics. The initial abstract selection was based on whether the abstracts concerned detection or diagnostics of ADHD. Studies where the primary focus was on brain signals or other aspects associated with objective diagnostics (e.g., task performance) were excluded, as well as studies not published in international peer-reviewed English journals or not reporting quantitative results. We also excluded studies examining eye movements, as they are likely to reflect different aspects in the ADHD symptoms (i.e., shifting and focusing of attention) than bodily movements captured by the sensors. Scopus found 13 studies of which six were found to be eligible for the present purposes, while PubMed gave 15 hits of which seven were found to be eligible. Out of the seven eligible studies found by PubMed, five were the also given by Scopus, leaving us eight unique articles in total. Additional search words (hyperactivity, accelometer, accelerometer, gyroscope, IMU, wearable-sensors) and Google Scholar were used to complement the search procedure and previous meta-analyses and reviews were examined to further identify eligible studies. Altogether, 25 studies were selected for more careful inspection (see Table 1).

2.2 Study Participants

Out of the eligible studies, 23 had children and 2 had adults as participants (Table 1). The average age in pediatric studies has been 9 years and 33 in adult studies. On average about 80% of the participants were males, reflecting approximately the typical gender distribution of ADHD [20]. The information about the ADHD subtyping and examination of comorbid symptoms in the clinical group as well as the methodological standards for verifying that the controls do not have psychiatric or neurodevelopmental issues or how well they represent the general population (e.g., distribution of socio-economic status and education) varied across individual studies. However, in most cases participants with neurological or psychiatric disorders, other than ADHD in the clinical groups, had been excluded from the samples in the original studies.

2.3 Sensor Data Collection

The measurement devices have been actigraphs, smart-watches, VR-controllers, some of which contain only an accelometer and some also a gyroscope. The studies have used distinct types of sensors typically placed on hands (either a wrist monitor or hand controller), and sometimes also on ankles or waist (Table 1). The wrist measurement was sometimes done from the dominant and sometimes from the non-dominant hand, which depend on the situation. For instance, during a school class the dominant hand may be used more for writing or drawing, and other such activities and the movements of the non-dominant hand could therefore give information about 'irrelevant' movements. In experimental tasks that are performed with the dominant task, the motivation for the sensor placement might be different, although in both cases data could be collected from both hands, even just for the cross-validation. A few studies have used sensors simultaneously in multiple body parts, also including hand and leg [13, 21]. The sampling rates of the devices typically range between 1–30 Hz.

2.4 Experimental Designs

Experimental designs in ADHD studies collecting motion sensor data can be scarcely divided into naturalistic studies where the data is collected at home and/or school and laboratory studies where typically a specific task is being presented (Table 1). The design also influences the duration of the measurement: naturalistic data can be collected over several days (has been on average 18 h/day) and with a few sessions could potentially fulfil the criteria concerning durability of the symptoms. The laboratory measurements with experimental tasks, in turn, typically last for tens of minutes and maximum few hours (on average 60 min). The main trade-off in the selection of this experimental design is in sampling distribution and representativeness of the situations where the symptoms are manifested (naturalistic designs) vs. contextual control with a measurement situation that can be carefully interpreted and more reliably compared between individuals (laboratory tasks). For example, a person could able to inhibit hyperactivity to manifest during a few minute laboratory task as such behaviors are generally considered inappropriate and the

situation is new to the participant, but on the other hand, data collected at the classroom or home could be affected by numerous potential confounds related to what the activities in the measurement days have been exactly (what kind of teaching was arranged and how, what is the child's situation at home or school etc.). At the moment, the empirical evidence demonstrating the comparative benefits from single studies is limited. Average classification accuracies have been 95% in naturalistic studies and 86% in studies with laboratory tasks. Also, hybrid paradigms employing naturalistic tasks attempting to combine the benefits of the naturalistic and laboratory designs are becoming increasingly common. Such paradigms where motion sensor data is collected in a naturalistic situation that is emulated in a virtual reality laboratory task, have been developed for classroom [22] and home situations [23, 24]. A virtual classroom task that is commonly used in ADHD studies is a variant of the CPT (see Introduction section [24]) that is one of the most widely used experimental tasks in this domain overall. Finally, there are at least two studies dividing the measurement period into multiple different real-world and experimental measurement sessions that gives information about the influences of the measurement context [21, 25]. O'Mahony and colleagues collected sensor data when the participants were in the 1) waiting room with their parents, 2) in the waiting room with a supervisor, 3) with the psychiatrist in her/his office, 4) with the psychiatrist and parent, and 5) during performance of an experimental task [21]. Miyahara and colleagues collected movement data from rather small children for about two hours when they were performing multiple types of neuropsychological or computerized cognitive tasks [25].

2.5 Analysis Methods

The studies reviewed have used various analytical techniques to interpret the measurement results collected by the sensors (Table 1). These techniques include machine- and deep learning algorithms, as well as traditional statistical methods. The choice of method generally depended on the design and main objective of the study, as well as the nature of the data to be analyzed. Many of the reviewed studies used statistical tests, especially analysis of variance (ANOVA). Statistical tests are considered useful for hypothesis-driven research, as they allow for testing the significance of differences between groups (ANOVA), means (t-tests), proportions (chi-square tests), and correlations (Pearson or Spearman correlation tests). ANOVAs were typically used to examine the significance of group differences in factors related to overall activity or changes in activity during measurement periods. Similarly, the studies focusing on classification of the group status of single individuals often used various statistical tests to evaluate which features should be used in the process. While these tests are powerful for hypothesis testing, they come with limitations. One major limitation observed in study by Miyahara et al. [25], was that each test has its assumptions. For instance, ANOVA assumes homogeneity of variances and normal distribution in the populations being compared. These assumptions can sometimes be restrictive and not met by all data sets. Violation of these assumptions can lead to inaccurate results. Another limitation considering the aim of our study is the classification of an individual, as traditional statistical tests are mainly capable of describing differences and relations between features. However, there are classification

methods that rely on some of these tests like discriminant analysis, which was used in some of the studies [24, 25]. Discriminant analysis offers a rather simple and efficient classification method, but it is limited by the assumptions in the tests included. Similarly, machine- or deep learning algorithms have their own requirements for the input data, but as there is a wide range of classification algorithms for different data types, the violation of assumptions can usually be avoided. Studies reviewed showed use of different basic machine learning classification methods, such as support vector machine (SVM), logistic regression and decision tree, as well as more advanced deep learning methods like CNN. These methods were tailored for different types of input data and often provided accurate classification results. Considering all the methods, CNN offers a rather different approach for the analysis, as it uses image data as an input. The accuracy of CNN can also be affected by the number of convolutional layers. CNN was used only in two of the studies reviewed [13, 26]. Amado-Caballero et al. [26] further experimented on different combinations of convolutional layers and input window sizes to find the highest accuracy. The implementation of these methods usually requires expertise in the field, rather large data sets and more computing power compared to traditional statistical methods.

3 Results

The published results have consistently reported group differences in the movement of the participants with ADHD and neurotypical controls (Table 1). Classification accuracies for the detection of the group status in single participants range between around 70% and 99%. More specifically, there are many studies with acceptable to excellent discrimination rates (70–90%), but then a few studies with outstanding classification accuracies (>90% or even around 98–99%). Overall, the studies reported sensitivity values of 87% on average. Similarly, the average specificity reached an average level of 86%. Both values are, as expected, close to the corresponding classification accuracies. Overall, each individual study reported significant group differences in motion sensor data between individuals with or without ADHD. Some studies reported the results separately in multiple different experimental situations or with a comparative analytical method. For example, O'Mahony et al. [21] reported multiple accuracies, each obtained from different experimental situations. Accuracies of these situations ranged from 81% to 93%, and the final accuracy (95%) was obtained by combining the data in each independent situation. Similarly, Kam et al. [27] reported results of two models which used different situations (class, class + recess), but also differently implemented decision trees. These models showed differences of 1–2% units in discrimination accuracy. More dramatic differences were reported in the study by Amado-Caballero et al. [26], with accuracies ranging from 56% up to 99%. For each study, Table 1 presents the highest achieved accuracies, if reported. Otherwise, the most significant features by group difference are presented.

Table 1. Summary of included studies divided between naturalistic designs and laboratory tasks. Both categories are in ascending order by mean age of the participants.

Author(s) and Publication Year	Number and Gender Ratio of Subjects and Controls	Age range or mean (M)	Sensor Type, Placement and Duration	Analysis Method	Discrimination Accuracy / Effect size	Interpretation
Studies with Naturalistic Design:						
Lin et al. 2020 [28]	15 ADHD, 15 controls (total 27% females)	M: 6.83 SD: 1	Smartwatch IMUs on wrist 2 h for 3 consecutive days	Two sample t-test, ROC	AUC: 82.0% $p < 0.001$	Method is moderately accurate
Kam et al. 2010 [27]	10 ADHD, 132 controls	M: 7.44 SD: 0.62	Actigraph on non-dominant wrist Single 3h period	Decision tree	ACC: 99.3% SE: 100.0% SP: 99.2%	Method is extremely accurate
Langevin R et al. 2012 [29]	5 ADHD, 5 medicated ADHD, 5 healthy controls (total 13% females)	M: 8.13	Actigraph on non-dominant wrist Two periods of 5 consecutive days	Kruskal-Wallis test	Nocturnal movement, group difference by period: $p = 0.008$	Significant group differences
Lindhiem et al. 2022 [30]	15 ADHD, 15 healthy controls	6–11 yr	Smartwatch IMUs on wrist Single period of 2 days	SVM, Logistic regression, Random Forest	ACC: 89.0% SE: 93.0% SP: 86.0%	Method is highly accurate
Gruber R et al. 2011 [31]	11 ADHD (36.4% female), 32 healthy controls (37.5% female)	M: 8.70 SD: 1.30	Actigraphy Period of 6 consecutive nights	ANOVA, MANOVA	sleep efficiency: $F(1, 38) = 86.18$ $p < 0.001$	Significant group differences
Munoz-Organero et al. 2018 [13]	11 ADHD, 11 healthy controls (total 9% females)	6–12 yr	Accelometer on dominant wrist and ankle Single 24h period	CNN	ACC: 93.8% SE: 80.0% SP: 100.0%	Method is very accurate
Licht CA et al. 2009 [32]	9 ADHD (11% females), 9 healthy controls (22% females)	M: 9.22 SD: 1.09	Actigraph around the waist 24h for a full 7-day week	ANOVA	Group X Period quadratic effect: $F(1, 16) = 5.12$, $p = 0.04$	Significant group differences
Tsujii N et al. 2009 [33]	18 ADHD, 10 PDD, 18 Controls	M: 9.23 SD: 1.45	Actigraph on non-dominant wrist Single 1 week period	ANOVA, post-hoc Scheffét-test	Average and SD Activity during recess: $F = 8.84$ and $F = 12.11$ $p < 0.001$	Significant group differences

(*continued*)

Table 1. (*continued*)

Author(s) and Publication Year	Number and Gender Ratio of Subjects and Controls	Age range or mean (M)	Sensor Type, Placement and Duration	Analysis Method	Discrimination Accuracy / Effect size	Interpretation
Amado-Caballero et al. 2020 [26]	73 ADHD, 75 healthy controls	6–15 yr	Actigraph on dominant wrist Single 24 h period	CNN	ACC: 98.6% SE: 97.6% SP: 99.5%	Method is extremely accurate
Brandt et al. 2012 [34]	5251 ADHD	Initial: M: 7.00, Follow-up: M: 14.00	Actigraph around the hip 7 consecutive days during waking hours	Linear regression models	Model 2: $F = 477.07$ $p = 0.001$	Significant group differences
Faedda et al. 2016 [35]	44 ADHD, 42 healthy controls, 48 bipolar subjects	5–18 yr	Actigraph around the waist Continuous period of 3–5 days	ANCOVA	ADHD vs Control Activity levels: $p < 10^{-6}$	Significant group differences
Studies with Laboratory Task:						
Miyahara et al. 2014 [25]	93 ADHD (26% females), 76 healthy controls (33% females)	M: 3.72	Actigraph on waist and non-dominant ankle 2 visits of approx. 2 h each	Discriminant analysis	ACC: 69.8% $p < 0.001$	Satisfactory accuracy and significant group differences
Bhattacharyya et al. 2022 [36]	10 ADHD, 20 controls	3–5 yr	EEG on forehead and CCD Single period of 14 min	two-tailed t-test after Welch's correction	Hyperactivity index: $t = 8.836$, $p < 0.0001$	Significant group differences
Chang et al. 2023 [37]	31 ADHD, 31 controls (48% females for both groups)	M: 7.66 SD: 2.33	Smart chair Routine visit, approx. 7–15 min	KNN, SVM, XGBoost	ACC: 92.3% AUC: 98.0%	Method is highly accurate
O'Mahony et al. 2014 [21]	24 ADHD (29% females), 19 controls (53% females)	M: 9.00 SD: 1.37	IMUs on waist and dominant ankle Lab visit, approx. 1h	SVM	ACC: 95.1% SE: 94.44% SP: 95.65%	Method is very accurate
Rapport MD et al. 2009 [38]	12 ADHD, 11 Controls (no female)	M: 9.04 SD: 1.36	Actigraph on non-dominant wrist and both ankles Single 2.5 h period with 2–15 min breaks	ANOVA	Total activity by group: $F = 36.55$, $p < 0.001$	Significant group differences

(*continued*)

Table 1. (*continued*)

Author(s) and Publication Year	Number and Gender Ratio of Subjects and Controls	Age range or mean (M)	Sensor Type, Placement and Duration	Analysis Method	Discrimination Accuracy / Effect size	Interpretation
Dane AV et al. 2000 [39]	20 inattentive ADHD (15% female), 22 combined ADHD (18% female), 22 controls (36% female)	M: 9.17 SD: 1.40	Actigraphy Two 2h periods	Univariate analysis and ANOVA	ADHD vs. Control, by Session: $F(1,61) = 8.32$ $p < 0.01$	Significant group differences
Inoue K et al. 1998 [40]	20 ADHD 52 Controls (no female)	M: 9.36	Actigraph on waist Single 1–2 h period		First 10 min mean activity: SE: 75.0% SP: 62.0% $p < 0.01$	Significant group differences
Halperin JM et al. 1992 [41]	31 ADHD, 53 non-ADHD, 18 controls	M: 9.65 SD: 1.82	Actigraph on waist Single 1h period	ANOVA, ANCOVA	Activity level $F = 8.25$ $p = 0.001$	Significant group differences
Seesjärvi et al. 2022 [24]	38 ADHD (13% females), 38 healthy controls (21% females)	M: 10.54 SD: 1.08	EPELI VR simulation and controllers Total duration max 35 min	Discriminant analysis	Controller motion: AUC: 73.0% SE: 71.0% SP: 66.0%	Significant group differences
Merzon et al. 2022 [23]	37 ADHD (22% females), 36 healthy controls (42% females)	M: 10.59 SD: 1.08	Eye tracker in EPELI VR simulation Single period of 25–35 min	SVM	Controller motion: AUC: 70.0% $p = 0.0085$	Significant group differences
Wood AC et al. 2009 [42]	116 ADHD (10% female), 119 siblings of ADHD (51% female), 218 controls (22% female)	M: 11.87 SD: 2.62	Actigraph on waist and dominant leg Single 2h period with 25 min break	t-test, ROC	Leg & waist movement intensity: AUC: 79.0%	Method is moderately accurate
Halperin JM et al. 2008 [43]	98 ADHD, 85 controls (no female)	Initial M: 9.09 SD: 1.30 Follow-up: M: 18.40 SD: 1.63	Actigraph on non-dominant ankle and waist Interview and test battery	MANOVA	Ankle Activity, Cohen's d: 0.66 $p < 0.001$	Significant group differences
Delcour Jensen et al. 2021 [22]	10 ADHD 10 controls	Na (Adult)	VR-CPT system and controllers (head, leg, hand)	Two sample t-test	Overall activity: $t = 2.33$, $p = 0.039$	Significant group differences
Edebol et al. 2013 [44]	55 ADHD (54% female), 202 healthy controls (44% female), 84 ADHD-normative (44% female)	M: 33.16 SD: 9.82	QbTest-plus Single 20 min period	Fisher's exact test	SE: 86.0% SP: 83.0%	Method is moderately accurate

4 Discussion

The present critical review identified 25 studies examining the role of motion sensors in the clinical assessment of ADHD. Fifteen of these studies, focusing on overall daytime activity levels and not on detection of ADHD symptoms are not discussed here in detail because the research questions in these studies do not allow comprehensive discussion of the clinical interpretability and utility of the findings, these studies are not methodologically comparable to the novel studies, and a meta-analysis on these older studies already exists [15]. Overall, the reported classification accuracies or AUC's for identifying single participant status are highly varying across individual studies (Table 1), which could be due to several factors (e.g., different analysis methods, sample characteristics or measurement solution, variability in the measurement context). Such heterogeneity in the research in this emerging field should be carefully considered and one important issue to advance the clinical use of these methods would be to establish generally accepted research standards to this field. This paper raises some of the critical questions to improve sensor-based research of ADHD attempting to serve in this path toward future health care applications. Besides varying experimental designs and research methods, clinical interpretation of the findings is limited by the participant samples that rarely represent the true variability in the population especially lacking demonstrated cases of attention deficits below the diagnostic threshold (i.e., the groups may have included those with a diagnosis or individuals with no attention deficits whatsoever) and individuals with other neurodevelopmental disorders (e.g., learning disabilities, autism spectrum disorder, conduct disorder, mood and anxiety disorders). Finally, the work resulting to detailed understanding of the motion sensor signals as part of the manifestation of ADHD symptoms is still underway. It would be critical to carefully benchmark or cross-validate the sensor-based methods against other assessment methods and determine which individuals with ADHD can or cannot be detected by the accelometer data. Research addressing these topics is likely to determine how broadly and for which purposes sensor-based methods could be used at the clinic. In most cases the challenges ahead are such that at least in principle they can be solved even with the existing methods by running more extensive high-quality studies employing the current technologies (e.g., large-scale multi-center studies) along with other benchmarking methods and detailed contextual descriptions. In the following sections, we will go through these research quality issues and clinical aspects in more detail.

4.1 Critical Analysis of the Research Quality in Sensor-Based ADHD Studies

Evaluation of the quality of research findings here involves several issues starting from (1) the number and location of the sensors in the measurement concept, (2) sensor type and sampling rate, to (3) the measurement situation and contextual control (e.g., naturalistic *vs.* experimental) and (4) various choices made in the data analysis (e.g., simple frequentist statistics *vs.* advanced machine learning methods). The existing studies have mostly utilized a single sensor worn in the hand, leg, or waist (Table 1). Although a single sensor is probably able to detect the overall level of activity, it may not detect all the relevant hyperactivity symptoms that are essentially characterized by specific types of bodily movements when the participant is otherwise still [45]. For instance,

in the measurements at the school class, distinguishing movements of the hands and legs (fidgeting), torso rotations (e.g., talking to another student), moving from the seat (interruptions of the learning situation) and other such distinct behaviors signal very different issues and detecting such movement patterns could significantly improve clinical interpretability. For such an analysis, including at least four sensors would be critical. Inaccuracies in the detection of the symptoms could also relate to the sampling rate. Some studies have collected data at 5 Hz sampling rate [28] that could potentially limit detection of high frequency movement signals. However, overall, it is likely that the sampling rates of the existing commercially available sensors are sufficient for detailed enough movement analysis and the bottle necks could be in other factors coming from the sensor placement and data analytics [21, 27]. Some data loss has been taking place in the existing studies, but such problems are likely not going to be a key factor from the methods development side since there are many sufficiently reliable measurement solutions available and the data is generally exceptionally rich as compared with many other methods and a few percentages data loss could be easily tolerated in the measurements that may last for several days. We suggest that a more important factor instead would be to obtain more detailed data on the measurement situation to help to improve interpretability of the findings. Apparently, the accuracy to detect the symptoms may vary a lot even in the same group of participants within a study according to the measurement situation [25, 39]. It would also be important to acquire reference data from certain type of bodily movements to teach the algorithm to identify certain type of movement patterns (supervised learning), improve interpretability of the complex machine learning and deep learning methods that tend to be 'black boxes', and share the algorithms for transparent evaluation and testing across the datasets to increase the transparency of the research.

4.2 Evaluation of the Clinical Utility of the Sensor-Based Diagnostics

Most of the so far conducted studies have been pilot studies with small samples not representing the variability of attention deficits or psychiatric and neurodevelopmental disorders in the population, but a few refreshing examples with two clinical populations [33, 35] or impressive sample sizes [34] have already been conducted indicating that this research field is going to a direction where sufficient representativeness of the normative and clinical data maybe be evidently reached. Measurement durations are highly varying (Table 1), but as there are so many factors that differ between the individual studies (sensor solutions, analytical methods, experimental paradigms) apparently the amount of datapoints is not a critical factor in achieving excellent classification accuracy. Several studies have obtained outstanding classification accuracy already in short one session measurements. Hence there is no reason to assume that at least in those cases where the two populations clearly differ (no suspicion of other disorders, comorbid symptoms, or close cases near the diagnostic threshold) even a relatively short measurement is sufficient. The amount of data might be an issue to consider when extending to other types of populations and trying to fulfill the more stringent clinical criteria. The stringent clinical criteria require not only confirming the prolonged appearance of the symptoms (min. 6 months) and manifestation of the symptoms in multiple environments (e.g., home

and school), but also the influence of the measurement time (e.g., summer holiday vs. stressful school or work period) would be highly important.

Another factor that should be accounted for and has already been examined to some extent is the time of the day, which could influence participants with or without ADHD differentially and should be considered in short measurements. Especially the mid-range activity periods have been suggested to contribute to detecting ADHD [13, 27]. Due to the day-to-day differences in measurement outcomes multiple different measurement days might be the best option to assure reliable and representative results [25, 29]. It has been noted that during a multiple session study, activity levels in ADHD participants may not change over the study, but in the control group activity levels reduced after the first measurement day [25]. This could be due to adjusting to the study participation and normalization of the behavior after the study beginning [25]. Finally, the possible role of the medication in the clinical reference samples should be further examined. It is difficult to obtain medication naïve reference samples and recommendations for the washout period and knowledge of the history of drug use are varying. Generally, it can be expected that the history of stimulant use should not be a major confounding factor as most of the drugs that are at use have limited aftereffects and are the effect is relatively short lasting.

Factors that are increasingly commonly examined in ADHD studies are the influence of age and gender on the manifestation of the symptoms. Among the published studies, only a few have been conducted in adults [22, 44]. In general, hyperactivity/impulsivity symptoms are less often observed in adults and the symptoms may be milder. In a similar vein, symptoms in females are typically more on the inattention domain and therefore it could be that detection of females with ADHD might be more difficult based on motion sensor data. Based on the currently available data, it is difficult to make detailed inferences on the role of age and gender on the detection accuracy. However, a recent review included data on gender and age differences in ADHD symptoms among a cohort of 1,326 children and adolescents revealed significant negative associations between female gender and total, inattentive, and hyperactive/impulsive ADHD symptoms, and age was found to be significantly associated only with hyperactive/impulsive symptoms [46]. Until these issues are carefully examined, e.g., by providing reference samples and sensitivity/specificity values for different subgroups, the application of these methods should be handled with caution. Although especially in children and in boys hyperactivity/impulsivity are quite strongly correlated with inattention, the limited available data could lead into underdetection of individuals with particular type of symptom patterns. While the gender/sex bias of ADHD may result to underrepresented female populations in research, such factors could be taken into account e.g., by prescreening gender-balanced groups from a larger sample. With prescreening, also the possible influence of many other potential confounding factors (e.g., socioeconomic status, general abilities, academic performance) could be controlled. In many of the published studies, even the lack of neurodevelopmental disorders in control group participants has not been carefully examined. Researchers have raised the point that in larger representative samples with detailed background information many of these factors could be accounted for [27].

At the same time when the current literature gives still a limited window on what aspects in ADHD sensor-based methods detect and what the obtained results reflect, it is good to keep in mind that potential applications of these methods go way beyond examination of the core symptoms. Studies reporting several other use cases have already been conducted. More specifically, motion sensors could be useful even in detecting the aggressive episodes that are commonly observed as ADHD has high comorbidity with conduct disorders [47]. Motion sensors could also help in detecting comorbid coordination disorder [48], neurological soft signs [49]. Finally, they have considerable potential for monitoring the treatment such as examining the effects of stimulants (including the detailed data on the type of the stimulants and individual dosing) [50] or even quality of life [51]. Together these promising opportunities paint a rather positive landscape on the potential clinical applications of sensor-based methods. There are several aspects raising from the critical analysis of the published studies that could be considered in planning future research. To control for the contextual effects, researchers could consider measuring multiple participants simultaneously in a same adult supervised situation could help in interpreting the data from several participants in the same situation. With such a setting, it might be possible to get further information also regarding other clinically relevant aspects such as hyperactive behavior during social interaction in group situations. Besides contextual control, generally the number of participants within a study should be larger. Based on the reported studies, the role of the measurement duration in the accuracy in detecting ADHD remains unclear, as shorter studies have been stringently controlled laboratory studies while the longer ones are naturalistic studies including various potential confounding factors coming from the naturalistic measurement context. This is certainly one factor that could be considered, for instance, by developing experimental designs with somewhat comparable naturalistic and laboratory conditions (e.g., virtual vs. real school class or home situation). Due to the complementary nature of different technological advances aiming at objective diagnostics, integrative solutions potentially combining input data from multiple sources could give best results. Large-scale data pools with rich questionnaire, interview, neuropsychological, virtual reality, motion sensor, biosensor etc. data where advanced computational analyses can be performed could help clinicians to obtain reliable results. According to the present results, the opportunities of standard smart watches, rings or even mobile phone sensor signals for diagnostics could also be further examined. For example, smart watches may contain biosensors that could complement the data provided by the motion sensors [5]. In one potential scenario, smart watch users could download a medical app that would give data to healthcare service providers that could then be accounted for in the diagnostics process.

4.3 Conclusions

This article critically evaluated the research quality and clinical utility of the studies employing motion sensor data for ADHD assessment, discussing the current state of the research in this field as well as needs for future improvements. The motivation for this branch of research relates to the need for developing objective assessment methods being able to record manifestation of the symptoms even over long times in everyday life with relatively little effort from the participants. These features also make motion

sensor methods such that the participants will not pay too much attention to their presence and change their behavior in a way that might bias the results. Such methods could be relatively cheap, easy to use, and cost-effective, potentially saving limited healthcare resources and improving the quality of the assessment. Motion sensors hence provide multiple potential benefits as compared with current diagnostic methods such as interviews and questionnaires. Despite these promising features, this branch of research is still at the early stages considering large-scale clinical use. Majority of the studies covered in this review have limited sample sizes, underrepresentative populations, and especially the performance of these methods in diffential diagnostics remains largely unclear. The methods in the studies are heterogeneous, which makes rigorous quantitative assessment of factors contributing to clinical value of these methods difficult. Quality standards, some of which were introduced in the present manuscript, should be kept high to meet the medical regulation criteria and large enough studies with representative samples need to be conducted replicating the promising results. Especially the measurement concept (naturalistic vs. laboratory-based, and which type of measurement sessions/tasks) and annotation of the context, supervising the classifiers and such factors influencing the performance of the computational methods should be carefully examined. When meeting these criteria, motion sensor research may provide methods complementing ADHD diagnostics already in the near future.

Acknowledgments. JS received funding from the Academy of Finland (grants #325981, #328954, and #353518).

Disclosure of Interests. The Authors have no conflicts of interest relevant to the content of this article.

References

1. Song, P., Zha, M., Yang, Q., Zhang, Y., Li, X., Rudan, I.: The prevalence of adult attention-deficit hyperactivity disorder: A global systematic review and meta-analysis. Journal of Global Health 11 (2021)
2. Xu, G., Strathearn, L., Liu, B., Yang, B., Bao, W.: Twenty-year trends in diagnosed attention-deficit/hyperactivity disorder among US children and adolescents, 1997–2016. JAMA Netw. OpenNetw. Open **1**(4), e181471–e181471 (2018)
3. Faraone, S.V., et al.: Attention-deficit/hyperactivity disorder. Nat. Rev. Dis. Primers. **1**(1), 1–23 (2015)
4. Parsons, T.D., Duffield, T., Asbee, J.: A comparison of virtual reality classroom continuous performance tests to traditional continuous performance tests in delineating ADHD: a meta-analysis. Neuropsychol. Rev.. Rev. **29**, 338–356 (2019)
5. Loh, H.W., Ooi, C.P., Barua, P.D., Palmer, E.E., Molinari, F., Acharya, U.R.: Automated detection of ADHD: current trends and future perspective. Comput. Biol. Med.. Biol. Med. **146**, 105525 (2022)
6. Kalia, M., e Silva, J.C.: Biomarkers of psychiatric diseases: current status and future prospects. Metabolism **64**(3), S11-S15 (2015)

7. Murphy, K.R., Adler, L.A.: Assessing attention-deficit/hyperactivity disorder in adults: focus on rating scales. J. Clin. Psychiatry **65**, 12–17 (2004)
8. Gualtieri, C.T., Johnson, L.G.: ADHD: Is objective diagnosis possible? Psychiatry (Edgmont) **2**(11), 44 (2005)
9. Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., Pollak, C.P.: The role of actigraphy in the study of sleep and circadian rhythms. Sleep **26**(3), 342–392 (2003)
10. Migueles, J.H., et al.: Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med. **47**, 1821–1845 (2017)
11. Small, S., et al.: Impact of reduced sampling rate on accelerometer-based physical activity monitoring and machine learning activity classification. J. Measurement Phys. Behav. **4**(4), 298–310 (2021)
12. Van der Kruk, E., Reijne, M.M.: Accuracy of human motion capture systems for sport applications; state-of-the-art review. Eur. J. Sport Sci. **18**(6), 806–819 (2018)
13. Muñoz-Organero, M., Powell, L., Heller, B., Harpin, V., Parker, J.: Automatic extraction and detection of characteristic movement patterns in children with ADHD based on a convolutional neural network (CNN) and acceleration images. Sensors **18**(11), 3924 (2018)
14. Mohri, M., Rostamizadeh, A., Talwalkar, A.: Foundations of machine learning. MIT Press (2018)
15. De Crescenzo, F., et al.: The use of actigraphy in the monitoring of sleep and activity in ADHD: a meta-analysis. Sleep Med. Rev. **26**, 9–20 (2016)
16. Taurines, R., Schwenck, C., Westerwald, E., Sachse, M., Siniatchkin, M., Freitag, C.: ADHD and autism: differential diagnosis or overlapping traits? a selective review. ADHD Attention Deficit and Hyperactivity Disorders **4**, 115–139 (2012)
17. Tistarelli, N., Fagnani, C., Troianiello, M., Stazi, M.A., Adriani, W.: The nature and nurture of ADHD and its comorbidities: a narrative review on twin studies. Neurosci. Biobehav. Rev.. Biobehav. Rev. **109**, 63–77 (2020)
18. Cumyn, L., French, L., Hechtman, L.: Comorbidity in adults with attention-deficit hyperactivity disorder. Canadian J. Psychiatry **54**(10), 673–683 (2009)
19. Gnanavel, S., Sharma, P., Kaushal, P., Hussain, S.: Attention deficit hyperactivity disorder and comorbidity: a review of literature. World J. Clin. Cases **7**(17), 2420 (2019)
20. Biederman, J., et al.: Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic. Am. J. Psychiatry **159**(1), 36–42 (2002)
21. O'Mahony, N., Florentino-Liano, B., Carballo, J.J., Baca-García, E., Rodríguez, A.A.: Objective diagnosis of ADHD using IMUs. Med. Eng. Phys. **36**(7), 922–926 (2014)
22. Jensen, T.D., Korbutt, W.K., Nedelev, G.P., Bemman, B.: Towards diagnostic support of hyperactivity in adults with ADHD using a virtual reality based continuous performance test and motion sensor data. In: International Conference on Pervasive Computing Technologies for Healthcare, pp. 505–521. Springer, Cham (2021)
23. Merzon, L., et al.: Eye movement behavior in a real-world virtual reality task reveals ADHD in children. Sci. Rep. **12**(1), 20308 (2022)
24. Seesjärvi, E., et al.: Quantifying ADHD symptoms in open-ended everyday life contexts with a new virtual reality task. J. Atten. Disord.Disord. **26**(11), 1394–1411 (2022)
25. Miyahara, M., Healey, D.M., Halperin, J.M.: One-week temporal stability of hyperactivity in preschoolers with ADHD during psychometric assessment. Psychiatry Clin. Neurosci.Neurosci. **68**(2), 120–126 (2014)
26. Amado-Caballero, P., et al.: Objective ADHD diagnosis using convolutional neural networks over daily-life activity records. IEEE J. Biomed. Health Inform. **24**(9), 2690–2700 (2020)
27. Kam, H.J., Shin, Y.M., Cho, S.M., Kim, S.Y., Kim, K.W., Park, R.W.: Development of a decision support model for screening attention-deficit hyperactivity disorder with actigraph-based measurements of classroom activity. Appl. Clin. Inform. **1**(04), 377–393 (2010)

28. Lin, L.C., Ouyang, C.S., Chiang, C.T., Wu, R.C., Yang, R.C.: Quantitative analysis of movements in children with attention-deficit hyperactivity disorder using a smart watch at school. Appl. Sci. **10**(12), 4116 (2020)
29. Langevin, R., Ramdé, J.: Attention deficit hyperactivity disorder (ADHD) in children, seasonal photoperiods, nocturnal movements and diurnal agitation. J. Canadian Acad. Child Adolescent Psychiatry **21**(1), 53 (2012)
30. Lindhiem, O., et al.: Objective measurement of hyperactivity using mobile sensing and machine learning: Pilot study. JMIR Formative Res. **6**(4), e35803 (2022)
31. Gruber, R., Wiebe, S., Montecalvo, L., Brunetti, B., Amsel, R., Carrier, J.: Impact of sleep restriction on neurobehavioral functioning of children with attention deficit hyperactivity disorder. Sleep **34**(3), 315–323 (2011)
32. Licht, C.A., Tryon, W.W.: Are children diagnosed with the combined form of ADHD pervasively hyperactive? Behav. Modif.. Modif. **33**(5), 655–681 (2009)
33. Tsujii, N., Okada, A., Kaku, R., Kuriki, N., Hanada, K., Shirakawa, O.: Differentiation between attention-deficit/hyperactivity disorder and pervasive developmental disorders with hyperactivity on objective activity levels using actigraphs. Psychiatry Clin. Neurosci.Neurosci. **63**(3), 336–343 (2009)
34. Brandt, V., Patalay, P., Kerner auch Koerner, J.: Predicting ADHD symptoms and diagnosis at age 14 from objective activity levels at age 7 in a large UK cohort. European Child & Adolescent Psychiatry **30**, 877–884 (2021)
35. Faedda, G.L., et al.: Actigraph measures discriminate pediatric bipolar disorder from attention-deficit/hyperactivity disorder and typically developing controls. J. Child Psychol. Psychiatry **57**(6), 706–716 (2016)
36. Bhattacharyya, N., et al.: Integration of electroencephalogram (EEG) and motion tracking sensors for objective measure of attention-deficit hyperactivity disorder (MAHD) in preschoolers. Rev. Sci. Instruments **93**(5) (2022)
37. Chang, T.M., et al.: Objective diagnosis of ADHD through movement analysis by using a smart chair with piezoelectric material. Pediatr. Neonatol.. Neonatol. **64**(1), 46–52 (2023)
38. Rapport, M.D., Bolden, J., Kofler, M.J., Sarver, D.E., Raiker, J.S., Alderson, R.M.: Hyperactivity in boys with attention-deficit/hyperactivity disorder (ADHD): a ubiquitous core symptom or manifestation of working memory deficits? J. Abnorm. Child Psychol.Abnorm. Child Psychol. **37**, 521–534 (2009)
39. Dane, A.V., Schachar, R.J., Tannock, R.: Does actigraphy differentiate ADHD subtypes in a clinical research setting? J. Am. Acad. Child Adolesc. PsychiatryAdolesc. Psychiatry **39**(6), 752–760 (2000)
40. Inoue, K., et al.: Clinical evaluation of attention-deficit hyperactivity disorder by objective quantitative measures. Child Psychiatry Hum. Dev. **28**, 179–188 (1998)
41. Halperin, J.M., Matier, K., Bedi, G., Sharma, V., Newcorn, J.H.: Specificity of inattention, impulsivity, and hyperactivity to the diagnosis of attention-deficit hyperactivity disorder. J. Am. Acad. Child Adolesc. PsychiatryAdolesc. Psychiatry **31**(2), 190–196 (1992)
42. Wood, A.C., Asherson, P., Rijsdijk, F., Kuntsi, J.: Is overactivity a core feature in ADHD? familial and receiver operating characteristic curve analysis of mechanically assessed activity level. J. Am. Acad. Child Adolesc. PsychiatryAdolesc. Psychiatry **48**(10), 1023–1030 (2009)
43. Halperin, J.M., Trampush, J.W., Miller, C.J., Marks, D.J., Newcorn, J.H.: Neuropsychological outcome in adolescents/young adults with childhood ADHD: profiles of persisters, remitters and controls. J. Child Psychol. Psychiatry **49**(9), 958–966 (2008)
44. Edebol, H., Helldin, L., Norlander, T.: Measuring adult attention deficit hyperactivity disorder using the quantified behavior test plus. PsyCh J. **2**(1), 48–62 (2013)
45. Eaton, W.O., McKeen, N.A., Saudino, K.J.: Measuring human individual differences in general motor activity with actometers. Measuring movement and locomotion: From invertebrates to humans, 79–92 (1996)

46. Gilbert, M., et al.: Gender and age differences in ADHD symptoms and co-occurring depression and anxiety symptoms among children and adolescents in the BELLA study. In: Child Psychiatry & Human Development, pp. 1–11 (2023)
47. Park, C., et al.: Machine learning-based aggression detection in children with ADHD Using sensor-based physical activity monitoring. Sensors **23**(10), 4949 (2023)
48. James, M.E., King-Dowling, S., Graham, J.D., Missiuna, C., Timmons, B.W., Cairney, J.: Effects of comorbid developmental coordination disorder and symptoms of attention deficit hyperactivity disorder on physical activity in children aged 4–5 years. In: Child Psychiatry & Human Development, pp. 1–11 (2021)
49. Kaneko, M., Yamashita, Y., Iramina, K.: Quantitative evaluation system of soft neurological signs for children with attention deficit hyperactivity disorder. Sensors **16**(1), 116 (2016)
50. Sydenstricker, S., Moore, A., Nagao, K.: Comparison of Fidgeting in Adolescents with Attention-Deficit/Hyperactivity Disorder Between Before and After Stimulant Medication Intake. J. Child Adolescent Psychopharmacology **33**(4), 143–148 (2023)
51. Li, R., et al.: Mediating effect of motor competence on the relationship between physical activity and quality of life in children with attention deficit hyperactivity disorder. In: BioMed Research International (2021)

Influence of Arterial Vessel Diameter and Blood Viscosity on PTT in Pulsatile Flow Model

Aleksandra Zienkiewicz[1(✉)], Erkki Vihriälä[1], and Teemu Myllylä[1,2,3]

[1] Optoelectronics and Measurement Techniques Research Unit, University of Oulu, Oulu, Finland
aleksandra.zienkiewicz@oulu.fi
[2] Physics and Technology, Research Unit of Medical Imaging, University of Oulu, Oulu, Finland
[3] Medical Research Center, Oulu, Finland

Abstract. Modelling relation between Pulse Transit Time (PTT) and blood pressure (BP) is a critical step in BP estimation for wearable technology. Recognizing the limitation of assuming constant vessel and blood conditions, we developed a simplified pulsatile flow model to analyze how various factors affect PTT values. Our research focuses on the impact of mechanical characteristics, such as vessel diameter, wall thickness, blood viscosity, and pressure, on PTT measurements and subsequent BP estimation. Measurements were conducted using accelerometer sensors within a custom-designed mock circulatory loop. This setup allowed for the testing of a wide range of pressure values and pulsation rates, as well as the modification of viscosity in blood-mimicking liquids across different vessel models. We employed the Moens-Korteweg conversion model for pressure estimation, initially trained on PTT data from a specific setup parameter combination, and subsequently tested with data from varied setup parameters. We observed high correlation levels ($r = 0.93 \pm 0.09$) paired with high error (RMSE $= 163 \pm 100$ mHg), suggesting potential inaccuracies in pressure estimation. We present the recorded signals and discuss how alterations in physical conditions influence PTT values and the precision of BP estimation.

Keywords: blood pressure · body area sensor · heart rate · phantoms and simulation · pulse transit time

1 Introduction

The demand for continuous and non-intrusive health tracking has driven advancements in the field of blood pressure (BP) monitoring. While the conventional cuff-based approach remains the current standard for BP measurement, recommended by hypertension experts for clinical evaluation [1], its inherent limitations, including intermittent measurements, discomfort, and the need for user compliance have fueled a search for more innovative and user-friendly alternatives. Therefore, various approaches have been investigated over the last few decades, aiming to develop wearable and cuffless method of BP measurement. It is included in the latest generation of smart watches [2, 3], wristbands [4], armbands [5],

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 186–199, 2024.
https://doi.org/10.1007/978-3-031-59091-7_13

or rings [6, 7]; there are proposals to measure BP using smartphones [8] or diverse types of skin patches [9–11]. Despite of the fact that some of the devices are already available on the market, accuracy of the measurement is often questionable; consequently, global BP guidelines do not endorse the utilization of wearable devices for diagnostic and treatment decisions. [1, 12, 13]. Moreover, validation against cuff-based measured BP is being criticized as a reliable reference for this purpose, and thus changes in requirements are called for [14, 15]. Consequently, research on cuffless non-invasive BP measurement techniques is still ongoing.

One of the methods utilized in some of the aforementioned solutions is indirect estimation of BP based on measurement of Pulse Transit Time (PTT). PTT is the time between two pulse waves propagating on the same cardiac cycle from two separate arterial sites. It is assumed that PTT is inversely related to BP, since with increasing BP, increasing distending pressure and decreasing arterial compliance, pulse wave velocity (PWV) rises and thus PTT shortens [4]. The PTT-based methodology has garnered significant attention within the research community due to its potential for wearable applications and its apparent simplicity. From a hardware perspective, the non-invasive detection of pulse can be achieved relatively easily using various modalities and sensor types. However, the relationship between BP and PTT is intricate and difficult to model. Both values are influenced by interconnected factors, such as blood viscosity, vessel stiffness and diameter, wave reflections along the arterial tree, varying shear rates in blood flow, cardiac output (determined by heart rate and stroke volume), sympathetic system activity, and the overall condition of the vasculature. Consequently, the measured values of pulse delays can be difficult to interpret, showing little to no value for beat-to-beat BP estimation [16]. Thus, an understanding of the cardiac flow properties is crucial in development of new methods and tools for cardiovascular health monitoring and improvement of the PTT-BP conversion models.

1.1 Mechanical Properties of Blood Vessels, Flow, and Pressure

The characterization of blood flow presents inherent challenges due to the irregular structure of blood and the disruptive influence of red blood cells on viscosity. Additionally, vessel thickness varies throughout the arterial tree, introducing further complexity to the modeling process. Blood vessels endure forces from blood flow and surrounding tissues. The blood viscosity causes different levels of shear stress occurring tangentially to the vessel lumen, influencing the ease with which blood flows through the vessels and the resulting impact on vascular dynamics. Depending on factors such as the hematocrit levels, plasma viscosity, and the properties of red blood cells, blood viscosity is difficult to define with single number. Common agreement is that the normal range is between 3.5 and 5.5 cP, however the values can be different in the large arteries, the veins, and the microcirculation [17, 18]. Although 70% of blood vessel walls consist of water, the rest is a complex mix of collagen, elastic fibers, proteoglycans, and vascular cells organized in layers. These layers differ in thickness and composition across vessel types and diameters. Large arteries have a thick media layer with more elastin, while small arteries have more smooth muscle cells; veins have a thinner media layer and less elastic tissue [19]. The vessel wall is structured to withstand and transmit forces from

blood flow, pressure, and surrounding tissues. The composite characteristics of the vascular wall result in distinctive mechanical properties when responding to physiological forces. In the example of artery, the response depends on the pressure inside the artery: low pressure states (<80 mmHg) engage the response from soft elastin fibers, while higher pressures cause stiff collagen fibers to dominate the response in order to avoid vessel damage [19, 20].

Therefore, the placement of the sensors in PTT measurement will have an impact on PTT-BP way beyond just a distance. Commonly, ECG R-peak is used as a proximal point, paired with the slope of PPG pulse measured periphery, e.g. on the finger. There is no agreement on which endpoints are the best choice. On one hand, PTT estimation along central arteries seems to be promising, because of central arterial wall properties and little interference caused by vasomotion and wave reflection [21]. On the other hand, distal waveforms are measured often with satisfactory results, with measurement from heart to toes and fingers showing better correlation with cuff-based BP than from heart to earlobes [22, 23].

1.2 Mathematical Models to Estimate BP Based on PTT

Modeling of PTT-BP relation is based on the works by Moens and Korteweg on the flow in tubes. The velocity of the fluid wave was be determined as a function of vessel and fluid characteristics [24, 25]:

$$PWV = \frac{d}{PTT} = \sqrt{\frac{Eh}{2\rho r}} \tag{1}$$

where PWV – pulse wave velocity, d is distance between sensors, ρ is the blood density, r is the inner radius of the vessel, h is the vessel wall thickness, and E stands for Young's modulus describing the elasticity of the arterial wall. Young's modulus E is not a constant, but it varies nonlinearly with pressure [26]:

$$E(P) = E_0\, e^{\alpha P} \tag{2}$$

where E_0 is the zero-pressure modulus, α is a constant that depends on the vessel and P is pressure. Deriving a formula directly from Eqs. (1) and (2), we get logarithmic relation between PTT and BP [27]:

$$BP = \frac{1}{\alpha} \cdot \ln\left(\frac{d}{PTT}\right)^2 + \frac{1}{\alpha} \cdot \ln \frac{2r\rho}{hE_0} = k_2 - k_1 \cdot \ln PTT^2 \tag{3}$$

Alternative formulas have been proposed, defining the relation e.g. as linear [28] or polynomial [29]. However, there are inaccuracies in all models deriving from Moens-Korteweg equation, as they make multiple presumptions to simplify the complicated relationship. The primary presumptions are that smooth muscle contraction and viscous effects are negligible, arterial elasticity is not significantly modified by aging or disease during measurement, and the timing points of the distal pulse are free of wave reflections. Additionally, both approaches assume that the thickness and diameter of the vessel

remain constant for changing BP level, and that the vessel wall is thin and can be modeled as an unchanging thin shell. However, the thickness-to-radius ratio of human vessel is beyond the limit for a thin shell, and the change of the radius of a human artery can reach ∼30% due to BP changes [30].

In this study, we aim to assess the impact of variations in blood viscosity, vessel diameter, and vessel wall thickness on the error in BP prediction. We compare how the model trained on the same environmental parameters will respond to change in one of said parameters. Measurements of PTT were conducted within a simplified mock circulatory loop. This setup allows testing across a wide range of pressure values and pulsation rates, along with precise control over the viscosity of the blood-mimicking liquid and the parameters of the flow architecture elements. The controllable environment facilitates the analysis of components influencing the recorded signal characteristics, the derived PTT, and subsequent pressure estimation.

2 Methodology

2.1 Pulsatile Flow Simulation

Haemodynamic simulations were performed using a mock circulatory loop consisting of pulsatile flow generator, an artificial blood vessel, pressure control, pulse sensors and blood-mimicking liquid, see Fig. 1. The system employs a dosing pump (Injecta, Athena AT.MT4) with a pumping rate set to 1 Hz. The pump induces a circulation of fluid inside of the tube system, with a detectable pulsation corresponding to the pumping rate. For blood vessel emulation we used tubes made of latex, with the manufacturer determined hardness K = 40 (Shore A). System tubing was made with tubes of inner diameter 9 mm and wall thickness 2 mm. To study the influence of vessel dimensions and elastic properties on measured signals, system architecture included a replaceable test tube (see Fig. 1).

Four tubes with different parameters were used in the experiments, with their dimensions shown in Table 1. Pressure of the fluid in the system is controlled using disc pump (TTP Ventus). Two pressure sensors (Honeywell sensing and solutions, SSCDRRN005PDAA5) with measurement accuracy of ±2.5% (full scale span) were placed at both ends of the test tube. In order to determine propagation velocity of the pressure waves travelling along the tubes, we used accelerometers (ACM) (model LIS344ALH from STMicroelectronics), which allow detection of the exact moment of the pulse appearance in corresponding set locations. The distance between ACMs was 45 cm. ACMs are sensing the tube displacement caused by travelling pulse wave that can be easily distinguished as sharp acceleration responses enabling exact time determination. All sensors used in the setup are connected to NI USB – 6289 multifunction I/O device with data acquisition at 5 kHz sampling rate. The pump and pressure are controlled via custom-made LabView software.

Water-glycerol mixture of different proportions served to mimic the blood on different viscosity levels. The temperature of the glycerol mixture influences its viscosity and therefore it requires temperature control. If we assume the viscosity of human blood is approximately 4 cP [18], it is almost equivalent to the viscosity of a 40% glycerin solution. A temperature sensor was placed inside the liquid tank. Viscosity of used mixtures

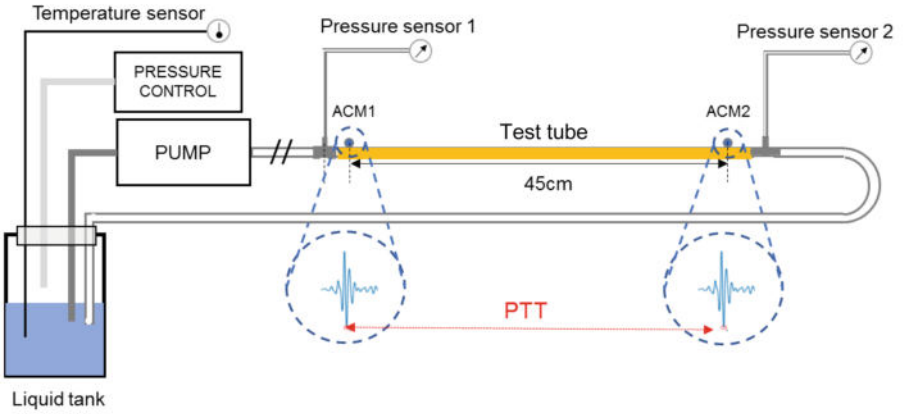


Fig. 1. Mock circulatory loop emulating pulsatile flow with known pulsation rate and variable pressure levels. Water-glycerol solutions of different concentrations were serving as a blood mimicking liquid. Latex test tubes of different dimensions were used for emulating the vessels.

Table 1. Parameters of elastic tubes used in the experiment.

	Inner diameter (mm)	Wall thickness (mm)	Outer diameter (mm)	Thickness-to-radius ratio h_0
Tube A	9	2	13	0.44
Tube B	9	2.5	14	0.56
Tube C	10	2	14	0.4
Tube D	10	3	16	0.6

were calculated using the standard formula [31, 32]. Tested concentrations together with their density and viscosity levels in measured temperature are shown in Table 2:

Table 2. Glycerol concentrations used in the study. The temperature of the liquid was in the range of 22 °C.

Glycerol concentration	Density [kg/m3]	Dynamic viscosity [Ns/m^2]	Dynamic viscosity [cP]
0%	997.61	0.0010	0.96
20%	1056.2	0.0019	1.88
30%	1085.2	0.0028	2.81
40%	1113.5	0.0044	4.48
45%	1127.3	0.0058	5.81
50%	1140.9	0.0077	7.71

2.2 Data Acquisition

Twenty-four scenarios were tested using four different tube diameters, each measured with six concentrations of aqueous-glycerol mixtures. Pressure levels in the system were continuously increasing in the range from 0 to 220 mmHg. Measurements for each combination were repeated three times in order to ensure results repeatability, with PTT value being an averaged value of three repetitions. An example of the signal recorded with ACM with the peaks detected for each pulse is shown on Fig. 2. Signal quality was good with clear separation between pulse complexes from consecutive pressure cycles. For this reason, raw signal was used directly. The offset value has been removed, so that the signal was oscillating around zero.

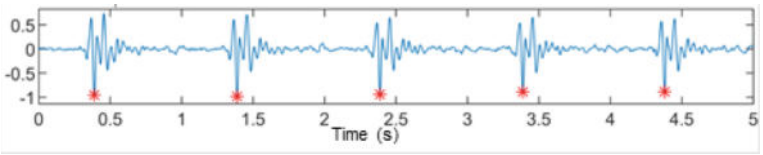


Fig. 2. Part of the signal recorded using ACM with detected pulses marked with red asterisks. Acceleration waveforms are sufficiently similar to typical waveforms when measuring heartbeat.

2.3 Statistical Analysis

Data are expressed as means ± standard deviations or percentages. Root means square error (RMSE), Pearson's correlation coefficient (R) and Bland–Altman analysis were used for the evaluation of agreement between the two methods of pressure estimation. Bland-Altman rates included RPC, reproducibility coefficient (±1.96 * SD values) and CV, coefficient of variation (SD of mean values in %).

3 Results

3.1 System Stability Testing

The stability test was conducted using a 50% glycerol liquid. Twelve different pressure levels were applied, and the dosing pump was operated at a frequency of 1 Hz. The experiment was repeated five times. For the analysis, three random pressure levels (40 mmHg, 80 mmHg, and 180 mmHg) were chosen, and individual random acceleration pulses and pressure pulses were examined. The Pressure 1 signal was utilized in determining the pressure levels. The analysis using longer signal periods was hindered due to the pump driving clock pulse exhibiting jitter up to 0.04 s, causing the pulses to be out of phase. Stability was assessed by identifying the maximum PTT errors (the differences between the highest and lowest PTT values) in each situation and calculating the mean and standard deviation of the errors. The pump driving clock pulses were precisely set to the same phase. Figure 3 shows the pressure and acceleration pulses in different pressure levels. The average maximum PTT error was 9 ms and the standard deviation 3.9 ms. The number of used acceleration pulses was 45.

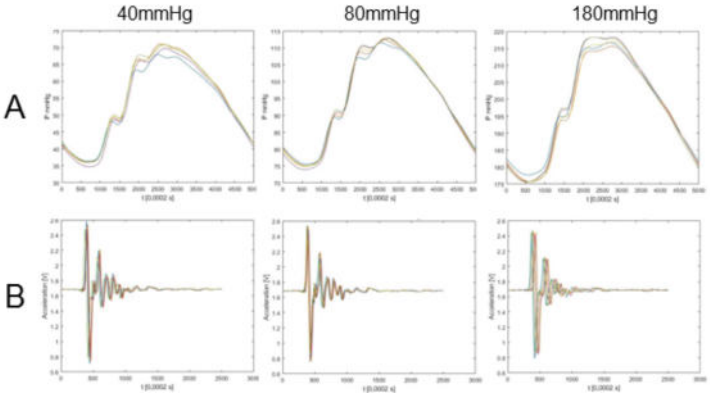


Fig. 3. Changes in pulse signal shapes when detected by pressure sensors (A) and acceleration sensors (B) for pressure levels of 40 mmHg, 80 mmHg and 180 mmHg and 50% water-glycerol solution.

The second part of the stability test was measurement of the single pulse signal shape, when liquids of different glycerol concentrations are used in the similar size of the test tube. Observed signals are displayed on Fig. 4:

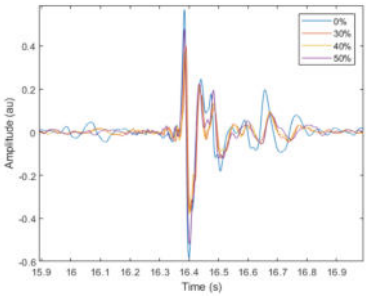


Fig. 4. Attenuation of the acceleration pulse amplitude, as the effect of viscosity change. Tested concentrations: 0%, 30%, 40%, 50%.

3.2 Influence of Liquid Viscosity and Tube Dimensions on PTT Levels

Figure 5A shows the variations in PTT levels across different viscosities of the blood mimicking liquid, all measured within the same tube. Numerical values for all tubes presented as a mean PTT change in comparison to reference concentration of 40% are shown in Table 3. Conversely, Fig. 5B displays the PTT variations observed across different tubes while maintaining a consistent viscosity of the blood-mimicking liquid. Changes of PTT values as compared to Tube A serving as a reference are shown in Table 4.

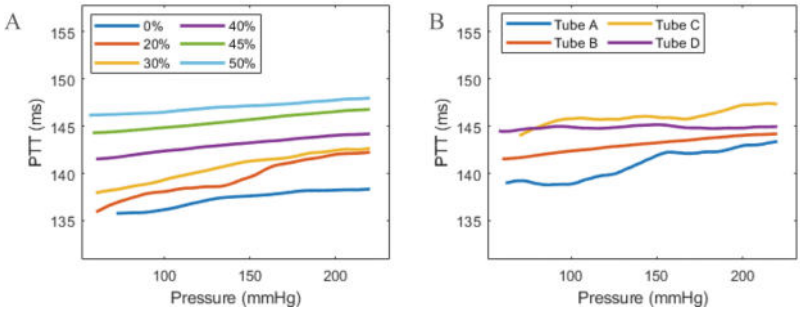


Fig. 5. Changes in PTT values for: A. different viscosities of glycerol solution in Tube A. B. different tube parameters, 40% glycerol solution.

Table 3. Variations in PTT levels (ms) caused by changes in liquid viscosity in reference to 40%.

	0%	20%	30%	40%	45%	50%
Tube A	-4.2 ± 0.5	-2.5 ± 0.6	-1.3 ± 0.5	0	2.7 ± 0.7	3.9 ± 0.9
Tube B	-5.8 ± 0.2	-3.7 ± 0.3	-2.4 ± 0.9	0	2.5 ± 0.1	4.0 ± 0.3
Tube C	-3.6 ± 1.0	-1.9 ± 0.9	-1.3 ± 0.9	0	1.6 ± 0.2	2.5 ± 0.8
Tube D	-7.2 ± 1.2	-0.9 ± 0.6	-2.3 ± 0.5	0	1.6 ± 2.1	2.0 ± 0.5

Table 4. Variations in PTT levels (ms) caused by changes in liquid viscosity in reference to Tube A.

Concentration	Tube A	Tube B	Tube C	Tube D
0%	0	0.5 ± 0.7	5.5 ± 0.2	0.8 ± 0.2
20%	0	0.9 ± 0.9	5.5 ± 0.9	5.6 ± 0.9
30%	0	1.0 ± 0.7	4.9 ± 0.6	3.1 ± 0.4
40%	0	2.1 ± 1.0	4.9 ± 1.3	4.0 ± 1.6
45%	0	1.9 ± 0.6	3.8 ± 0.4	2.9 ± 2.6
50%	0	2.2 ± 0.4	3.4 ± 0.4	2.1 ± 0.4

3.3 Regression Model Based on Moens-Korteweg Equation

A regression model was created based on PTT and BP levels measured for a reference viscosity 40% and Tube B, following the formula (3). Correlation plot and Bland-Altman plot are shown on Fig. 6A–B. Model parameters are shown on Fig. 6C.

Model parameters (Fig. 6C) were used to estimate pressure values based on the PTT obtained using data from different liquid viscosities and tube dimensions. Relative errors

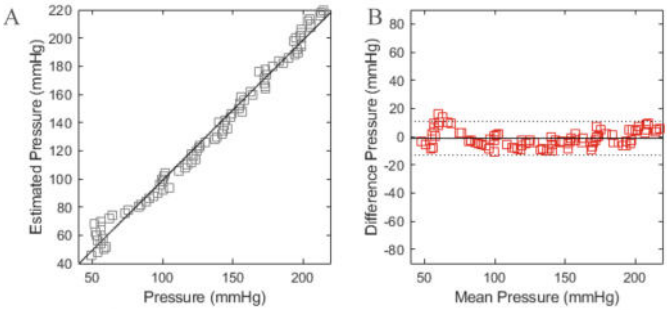


Fig. 6. A: Correlation plot between estimated and measured pressures, r = 0.96. B: Bland-Altman plot: mean difference = 0, RPC = 11.96, C: model parameters.

resulting from these estimations are shown in Table 5. Correlation coefficients are shown on Fig. 7A, RMSE values are shown on Fig. 7B.

Table 5. Relative error between estimated and measured values.

Concentration	Tube A	Tube B	Tube C	Tube D
0%	3.2 ± 1.6	2.8 ± 1.2	0.5 ± 0.5	2.5 ± 1.2
20%	2.4 ± 1.3	2.2 ± 1.8	-0.5 ± 0.5	-0.5 ± 0.4
30%	1.8 ± 1.0	1.4 ± 1.1	-0.8 ± 0.7	0.1 ± 0.1
40%	1.2 ± 0.9	-0.01 ± 0.1	-1.2 ± 0.5	-1.2 ± 1.0
45%	-0.1 ± 0.3	-1.3 ± 0.7	-2.1 ± 0.8	-2.2 ± 3.1
50%	-0.9 ± 0.5	-2.2 ± 1.3	-2.6 ± 1.3	-1.9 ± 0.8

4 Discussion

In this study, we used a system with simulated pulsatile flow in order to study PTT, vessel dimensions and viscosity relationship in a controlled environment, and without wave reflections. Four tubes made of the same material but with different thickness-to-radius ratio were used in the experiments, and six concentrations of water-glycerol solutions were mimicking blood of different viscosity levels. Pressure level in the system was gradually increased from 20 to 220 mmHg. Pulsations were detected using ACMs placed

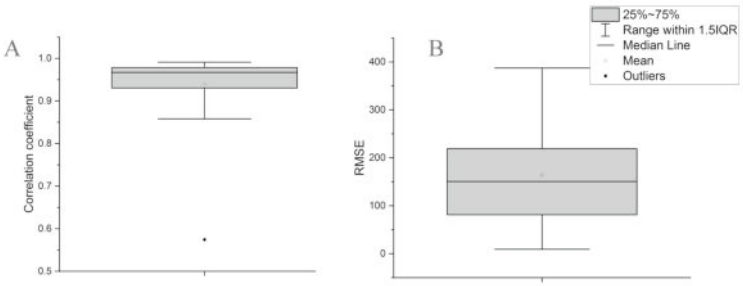


Fig. 7. A. Correlation coefficients B. RMSE between estimated and measured pressure values for all viscosities and tubes combinations.

on two ends of the test tube. Calculated PTT was then compared with simultaneously measured pressure.

We have tested the stability of the ACM sensor response for system pressure of 40 mmHg, 80 mmHg and 180 mmHg. The higher the pressure, the higher observed PTT error, with the maximum value of 9 ms. Since the measured PTTs were staying in the range between 100–150 ms, this potential value variation can be a significant source of error. However, we have decided to average the value of measurements repeated with the same setup, in order to reduce the impact of random errors. We have also measured the influence of liquid viscosity on pulse signal attenuation- the observed effect was negligible (Fig. 4). This conclusion corresponds to the study by Ikenaga et al. [33], in which authors were comparing the attenuation of a pressure wave in a phantom of human circulation.

When observing the levels of PTT in the single tube, the results met the expectations- more viscous liquid resulted in longer PTT (Fig. 5A). It can be seen in Table 3, that although the effect was visible for all tested tubes, the observed changes were in the small range. However, when comparing results between different tubes, there was no straightforward pattern visible (Fig. 5B, Table 4). The potential source of this effect might be e.g. a small difference in distance between sensors, when the new test tube was attached to the system. It might be possible that tube radius non-uniformities were present, e.g. due to tube stretching or displacement.

We have calculated parameters of the transit time-pressure conversion model, based on Moens-Korteweg formula shown in (3). Even when tested on the original viscosity-tube combination, it resulted in deviation of ± 11.96 mmHg, even though the correlation coefficient was very high ($r = 0.96$; Fig. 6). When applying similar model parameters to PTT values from other measurements, the correlation coefficients remained notably high ($r = 0.93 \pm 0.09$, Fig. 7A); however, the RMSE values were found to be exceedingly elevated (RMSE $= 163 \pm 100$ mHg, Fig. 7B). Analysis of relative error in Table 5 reveals positive error levels for concentrations lower than in the original model, and negative ones for the higher concentrations, when comparing the results for same tube as the original model (Tube B). It means, that for lower liquid viscosity the model was estimating too low pressure, whereas for higher viscosity the estimate was too high.

However, further studies are needed to confirm such patterns and the severity of the effect.

Modeling blood flow is a complicated task. We have observed possible sources of errors in the presented setup. Pressure sensors in the presented setup were connected to the system with tubes of different diameter than the test tube, which might have resulted in wave reflections. The conversion of the pressure levels to mmHg was done based on single values of atmospheric pressure, which might have differed for measurements done in separate days. Every measurement was done with steadily increased pressure; however, the system dynamics might be different when pressure is dropping. It also might be beneficial to increase the number of measurement repetitions, in order to eliminate random errors. Our next step is to increase the complexity of the phantom setup, in order to study e.g. effect of bifurcation. Furthermore, testing the tubes of different materials and wall thickness can be used e.g. to study the effect of vessel stiffening, resulting from aging. Another possibility is to test different position of the phantom (now it was always horizontal), which would correspond to different body positions. In terms of analysis, other transit time-pressure conversion models need to be tested.

Since blood viscosity and vessel diameter can dynamically change [34], it is important to understand the effect of these parameters on PTT levels and improve PTT-BP calibration process accordingly. Alternatively, the methods such as single point measurement as could be tested [35], which could potentially enable avoiding the calibration error coming from viscosity and vessel changes.

5 Conclusion

This study investigated PTT in relation to blood viscosity, vessel dimensions, and pressure levels using a simulated pulsatile flow system. Our experiments utilized four tubes with varying thickness-to-radius ratios and six water-glycerol solutions to mimic different blood viscosities, all tested in gradually increasing system pressure. PTT was measured with ACMs and compared against simultaneous pressure readings. Our findings suggest that liquid viscosity and tube dimensions impact PTT and following pressure estimation. The Moens-Korteweg formula-based model showed high correlation but large RMSE when applied across different conditions, indicating the need for further refinement. These results highlight the complexities in accurately modeling blood flow and the influence of factors like viscosity and vessel dimensions on PTT. Future work will focus on enhancing our model's accuracy and exploring effects like bifurcation and vessel stiffening. This study underscores the importance of considering dynamic changes in blood properties and vessel characteristics for effective PTT-based blood pressure monitoring.

Acknowledgments. AZ thanks Alfred Kordelin Foundation. This work was partly supported by the European Structural and Investment Funds - European Regional Development Fund (ERDF): EMUVALID, and the Academy of Finland Profi6 programme 6G-Enabling Sustainable Society (6GESS).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Stergiou, G.S., Palatini, P., Parati, G., et al.: 2021 European Society of Hypertension practice guidelines for office and out-of-office blood pressure measurement. J. Hypertens. **39**, 1293–1302 (2021)
2. Moon, J.H., Kang, M.-K., Choi, C.-E., et al.: Validation of a wearable cuff-less wristwatch-type blood pressure monitoring device. Sci. Rep. **10**, 1–9 (2020)
3. Lazazzera, R., Belhaj, Y., Carrault, G.: A new wearable device for blood pressure estimation using photoplethysmogram. Sensors **19**, 2557 (2019)
4. Ibrahim, B., Jafari, R.: Cuffless blood pressure monitoring from a wristband with calibration-free algorithms for sensing location based on bio-impedance sensor array and autoencoder. Sci. Rep. **12**, 1–14 (2022)
5. Zheng, Y.-L., Yan, B.P., Zhang, Y.-T., Poon, C.C.Y.: An armband wearable device for overnight and cuff-less blood pressure measurement. IEEE Trans. Biomed. Eng. **61**, 2179–2186 (2014)
6. Fortin, J., Rogge, D.E., Fellner, C., et al.: A novel art of continuous noninvasive blood pressure measurement. Nat. Commun. **12**, 1387 (2021)
7. Osman, D., Jankovic, M., Sel, K., et al.: Blood pressure estimation using a single channel bio-impedance ring sensor. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 4286–4290 (2022)
8. Frey, L., Menon, C., Elgendi, M.: Blood pressure measurement using only a smartphone. NPJ Digit. Med. **5**, 86 (2022)
9. Kireev, D., Sel, K., Ibrahim, B., et al.: Continuous cuffless monitoring of arterial blood pressure via graphene bioimpedance tattoos. Nat. Nanotechnol. **17**, 864–870 (2022)
10. Ion, M., Dinulescu, S., Firtat, B., et al.: Design and fabrication of a new wearable pressure sensor for blood pressure monitoring. Sensors **21**, 2075 (2021)
11. Wang, J., Zhu, Y., Wu, Z., et al.: Wearable multichannel pulse condition monitoring system based on flexible pressure sensor arrays. Microsyst. Nanoeng. **8**, 16 (2022)
12. Mukkamala, R., Yavarimanesh, M., Natarajan, K., et al.: Evaluation of the accuracy of cuffless blood pressure measurement devices: challenges and proposals. Hypertension **78**, 1161–1167 (2021)
13. Falter, M., Scherrenberg, M., Driesen, K., et al.: Smartwatch-based blood pressure measurement demonstrates insufficient accuracy. Front. Cardiovasc. Med. **9**, 958212 (2022). https://doi.org/10.3389/fcvm.2022.958212
14. Lee, H.Y., Lee, D.-J., Seo, J., et al.: Smartphone/smartwatch-based cuffless blood pressure measurement: a position paper from the Korean Society of Hypertension. Clin. Hypertens. **27**, 1–8 (2021)
15. Cheng, H.-M., Chen, C.-H.: Cuffless measurement of blood pressure: not good enough for diagnosis and treatment of hypertension. Pulse **10**, 52 (2022)
16. Hoshide, S., Yoshihisa, A., Tsuchida, F., et al.: Pulse transit time-estimated blood pressure: a comparison of beat-to-beat and intermittent measurement. Hypertens. Res. **45**, 1001–1007 (2022)
17. Nader, E., Skinner, S., Romana, M., et al.: Blood rheology: key parameters, impact on blood flow, role in sickle cell disease and effects of exercise. Front. Physiol. **10**, 1329 (2019)
18. Hall, J.E., Hall, M.E.: Guyton and Hall Textbook of Medical Physiology. Elsevier Health Sciences (2020)

19. Morin, C., Krasny, W., Avril, S.: Multiscale mechanical behavior of large arteries. arXiv preprint arXiv:191206052 (2019)
20. Singh, C., Wong, C.S., Wang, X.: Medical textiles as vascular implants and their success to mimic natural arteries. J. Funct. Biomater. **6**, 500–525 (2015)
21. Sola, J., Chételat, O., Sartori, C., et al.: Chest pulse-wave velocity: a novel approach to assess arterial stiffness. IEEE Trans. Biomed. Eng. **58**, 215–223 (2010)
22. Block, R.C., Yavarimanesh, M., Natarajan, K., et al.: Conventional pulse transit times as markers of blood pressure changes in humans. Sci. Rep. **10**, 1–9 (2020)
23. Di Rienzo, M., Avolio, A., Rizzo, G., et al.: Multi-site pulse transit times, beat-to-beat blood pressure, and isovolumic contraction time at rest and under stressors. IEEE J. Biomed. Health Inform. **26**, 561–571 (2021)
24. Moens, A.I.: Die Pulscurve. Brill (1878)
25. von Korteweg, D.J.: Ueber die Fortpflanzungsgeschwindigkeit des Schalles in elastischen Röhren. Ann. Phys. **241**, 525–542 (1878)
26. Hughes, D.J., Babbs, C.F., Geddes, L.A., Bourland, J.D.: Measurements of Young's modulus of elasticity of the canine aorta with ultrasound. Ultrason. Imaging **1**, 356–367 (1979). https://doi.org/10.1177/016173467900100406
27. Myllylä, T.S., Elseoud, A.A., Sorvoja, H.S.S., et al.: Fibre optic sensor for non-invasive monitoring of blood pressure during MRI scanning. J. Biophotonics **4**, 98–107 (2011)
28. Younessi Heravi, M.A., Khalilzadeh, M.A.: Designing and constructing an optical system to measure continuous and cuffless blood pressure using two pulse signals. Iran. J. Med. Phys. **11**, 215–223 (2014)
29. Gesche, H., Grosskurth, D., Küchler, G., Patzak, A.: Continuous blood pressure measurement by using the pulse transit time: comparison to a cuff-based method. Eur. J. Appl. Physiol. **112**, 309–315 (2012)
30. Fung, Y.-C.: Mechanical properties and active remodeling of blood vessels. In: Biomechanics, pp 321–391. Springer, New York (1993). https://doi.org/10.1007/978-1-4757-2257-4_8
31. Cheng, N.-S.: Formula for the viscosity of a glycerol-water mixture. Ind. Eng. Chem. Res. **47**, 3285–3288 (2008)
32. Volk, A., Kähler, C.J.: Density model for aqueous glycerol solutions. Exp. Fluids **59**, 1–4 (2018)
33. Ikenaga, Y., Nishi, S., Komagata, Y., et al.: Experimental study on the pressure and pulse wave propagation in viscoelastic vessel tubes-effects of liquid viscosity and tube stiffness. IEEE Trans. Ultrason. Ferroelectr. Freq. Control Freq. Control **60**, 2381–2388 (2013)
34. Çinar, Y., Şenyol, A.M., Duman, K.: Blood viscosity and blood pressure: role of temperature and hyperglycemia. Am. J. Hypertens. **14**, 433–438 (2001)
35. Zienkiewicz, A., Vihriälä, E., Seppälä, E., et al.: Wearable sensor system on chest for continuous measurement of blood pressure and other vital signs. In: 2022 IEEE 16th International Symposium on Medical Information and Communication Technology (ISMICT), pp 1–6 (2022)

Clinical Decision Support and Medical AI 2

A Hybrid Images Deep Trained Feature Extraction and Ensemble Learning Models for Classification of Multi Disease in Fundus Images

Jyoti Verma[1], Isha Kansal[2], Renu Popli[2], Vikas Khullar[2], Daljeet Singh[3], Manish Snehi[1], and Rajeev Kumar[2(✉)]

[1] Department of Computer Science and Engineering, Punjabi University Patiala, Patiala, Punjab, India

[2] Chitkara Institute of Engineering and Technology, Chitkara University, Jansla, Punjab, India
{isha.kansal,vikas.khullar}@chitkara.edu.in, rajeev.kumar@chitkara.edu

[3] Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland
daljeet.singh@oulu.fi

Abstract. Retinal disorders, including diabetic retinopathy and macular degeneration due to aging, can lead to preventable blindness in diabetics. Vision loss caused by diseases that affect the retinal fundus cannot be reversed if not diagnosed and treated on time. This paper employs deep-learned feature extraction with ensemble learning models to improve the multi-disease classification of fundus images. This research presents a novel approach to the multi-classification of fundus images, utilizing deep-learned feature extraction techniques and ensemble learning to diagnose retinal disorders and diagnosing eye illnesses involving feature extraction, classification, and preprocessing of fundus images. The study involves analysis of deep learning and implementation of image processing. The ensemble learning classifiers have used retinal photos to increase the classification accuracy. The results demonstrate improved accuracy in diagnosing retinal disorders using DL feature extraction and ensemble learning models. The study achieved an overall accuracy of 87.2%, which is a significant improvement over the previous study. The deep learning models utilized in the study, including NASNet-Mobile, InceptionResNetV4, VGG16, and Xception, were effective in extracting relevant features from the Fundus images. The average F1-score for Extra Tree was 99%, while for Histogram Gradient Boosting and Random Forest, it was 98.8% and 98.4%, respectively. The results show that all three algorithms are suitable for the classification task. The combination of DenseNet feature extraction technique and RF, ET, and HG classifiers outperforms other techniques and classifiers. This indicates that using DenseNet for feature extraction can effectively enhance the performance of classifiers in the task of image classification.

Keywords: Deep learning · ensemble learning · fundus · feature extraction · imaging

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 203–221, 2024.
https://doi.org/10.1007/978-3-031-59091-7_14

1 Introduction

Eye-related issues are critical for survival as they are an essential part of our sensory system, and any vision loss can significantly affect our quality of life. Visual impairment can make it challenging to perform everyday tasks, such as reading, writing, driving, or even recognizing faces. Ophthalmologists traditionally rely on manual screening processes to detect eye problems through fundus images and can identify some eye problems, including glaucoma, vision problems, and many other eye illnesses [1]. The present increase in patients and a dearth of skilled practitioners have made it challenging to offer patients proper care. Long patient wait times and a decline in the standard of care can both be caused by this shortage. An estimated 2.2 billion people worldwide have vision problems, either in their near or distant vision, as reported by the World Health Organization [2]. Advancements in technology, such as automated screening processes, can also help address the shortage of qualified ophthalmologists and improve the quality of care delivered to patients [3].

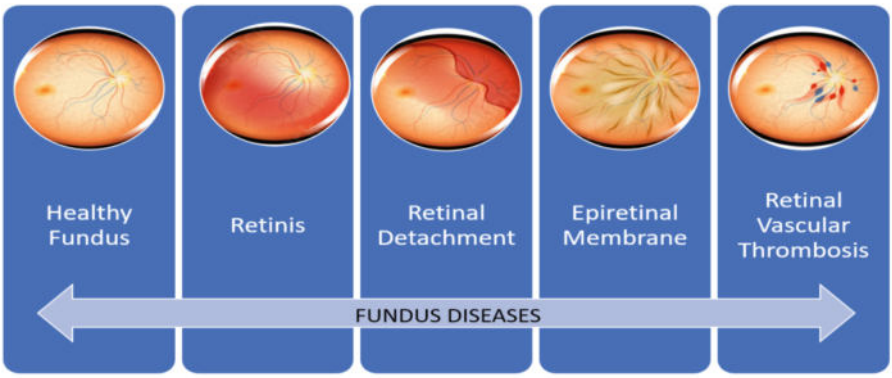


Fig. 1. Ocular Diseases [3]

Figure 1 shows Ocular diseases. There is a significant amount of diversity in the factors between nations and even within countries, and this variation is largely influenced by the accessibility of clinical services, the cost of those services, and the level of eye care awareness among the public [4, 5]. The technique is helpful to medical professionals in making a preliminary diagnosis, and it reduces the amount of time and effort required from patients as well [6, 7]. When one fundus image is evaluated in three separate color channels, it is possible to determine the presence of many diseases based on abnormalities in the fundus [8]. Because of the complexity and interdependence of ocular disorders, patients will typically develop various ocular diseases in each eye as they progress through their treatment [9]. Images of the left and right fundus may be seen in Fig. 1, which was derived from the ODIR dataset.

Medical diagnosis and therapy have employed deep learning techniques to classify photos and videos [10, 11]. The accomplishments of these models can be credited to the improved feature representation that was accomplished through the utilization of

multilayer processing architectures [12]. The goal of the ensemble learning approach is to combine several different models in such a way that the resulting model is superior to all the component models taken separately [13]. The paper is organized into several sections that cover different aspects of research related to ocular diseases. Section 2 focuses on previous work that utilizes ensemble, deep, and machine-based learning to address ocular diseases. Section 3 outlines the proposed work in this area. Section 4 presents the results, and Sect. 5 provides the discussion and conclusion.

2 Related Work

The diagnosis and treatment of retinal diseases are critical in preventing blindness, especially in diabetic patients. With the increasing amount of research being conducted, it is crucial to conduct a comprehensive examination of the prevalent techniques for implementing supervised learning ML, transfer learning, and DL in the diagnosis of diabetes mellitus. Specifically, the categorization of the retinal blood vessels, prediction and identification, classification and recognition, and analysis procedures are differentiated. Modern evidence in scholarly literature may generally be split into two categories: classical learning approaches and deep learning approaches. For the automated grading of DR, various traditional approaches are based on the lesions shown in fundus photographs. Textural and transform-based qualities are used to categorize most traditional approaches. There has been a widespread application of deep learning techniques to enhance the precision of detecting and categorizing different retinal diseases, using ML-based identification of eye diseases.

Traditional ML techniques have been widely employed in the study of diabetic retinopathy. One such technique involves manually crafting features to analyze fundus images. These features are then fed into a classifier for disease classification. Various studies have used this approach to diagnose a range of eye conditions, such as cataracts. Other conventional techniques have also been utilized for DR diagnosis based on features such as textural and transform-based properties. Despite their success, traditional methods have limitations in extracting complex features from images, which may lead to inaccuracies in diagnosis. Consequently, DL methods have gained popularity in recent years for their ability to automatically extract more intricate features. In this regard, several studies have employed DL methods for the diagnosis of diabetic retinopathy, utilizing CNN and transfer learning for feature extraction, followed by classification using machine learning algorithms. These approaches have demonstrated promising results in achieving higher accuracy in DR diagnosis.

Junjun He et al. employed support vector machines and genetic algorithms to classify images as cataract or non-cataract. They segmented fundus images into 16 blocks and extracted texture features by applying a gray-level co-occurrence matrix and frequency response analysis with a Haar wavelet. Subsequently, GA was utilized to weigh the features, and SVM was used to classify the images [14]. Omar et al. developed an ML model for the classification of diabetic retinopathy using fundus images. The researchers identified multiple features in the images, such as vessels, hematoma, capillaries, exudates, and the optic disc, which were utilized to categorize the pictures into mild, moderate, and severe stages of non-proliferative diabetic retinopathy or proinflammatory diabetic retinopathy [15]. Burlina et al. proposed a model for the classification of diabetic

retinopathy, which extracts both segmented and non-segmented visual attributes from fundus images. The non-segmented attributes include Contrast, Association, Homogeneity, Vitality, and Volatility, while the segmented attributes include Exudates, Veins, and Optic Disc. The extracted features are then fed into an SVM model, which employs 10-fold cross-validation using three different kernels: radial bias function, polynomial, and linear [16]. Ting DSW et al. created a DL model to diagnose diabetic retinopathy and other diabetic-related eye diseases using retinal images from a multiethnic population with diabetes [17]. S. Aslani et al. conducted a study on classifying diabetic retinopathy using fundus images of the retina [18]. Wejdan Alyoubi et al. developed a technique for the automatic diagnosis of glaucoma that employs SVM and Adaboost classifiers [19].

2.1 Ensemble Learning-Based Identification of Eye Diseases

Ensemble learning is a technique used in ML to improve the performance of predictive models [20]. By combining the strengths of different models, ensemble learning can achieve better accuracy and robustness than any single model [21]. Deep Ensemble is a technique in which multiple deep neural networks are trained on different subsets of the data and the outputs are combined through averaging or voting [22]. In their investigation, Costa et al. employed ensemble techniques to categorize retinal disorders [23]. To extract high-level attributes from the fundus images, these algorithms were trained on a sizable dataset of fundus images. An ensemble classifier then processed the combined results from these models to produce the final forecast. The accuracy of disease diagnosis in their study was increased by the integration of different models utilizing ensemble learning. Deshmukh et al. exceeded each of the separate models, reporting an accuracy of 95.71% [24]. Wang et al. used the CNN approach to extract features. Efficient-NetB3 served as the feature extractor for the model [25]. Bulut et al. provided a 21-disease classification approach. Here, Xception is used with Dropout layers, Global Average Pooling, 128-batch size, and 0.001 learning rate. As the collection comprises many photos, it was serialized and loaded in 100–200 MB pieces. The 9565-image dataset is skewed. The imbalanced dataset affects training and testing performance [26].

2.2 Deep Learning-Based Identification of Eye Diseases

Ophthalmology is a field of study that can benefit from the use of classification methods such as convolutional neural networks, particularly about the widespread problem of glaucoma and retinopathy. Ahmad et al. developed a framework for eye disease classification using ANNs. In pre-processing, color histogram-based texture-based feature extraction is performed [27]. Berrimi and Moussaoui et al. suggested a deep learning model that performs significantly better than pre-trained transfer learning approaches. The proposed architecture consists of three CNN layers. The Dropout levels and CNN layers were incorporated into this architecture so that it might be improved even further [28]. Yao et al. proposed a correlation module with a DC network-based model, in which the DC network extracts features, the spatial correlation module examines the relationships between the attributes, and the classification layer categorizes multi-label eye disorders [29].

There has been a significant increase in the amount of research dedicated to improving DL-based technologies in the field of e-healthcare [30]. This research has the potential to revolutionize the way we approach healthcare and can make significant improvements in patient care. This has been driven by the ready availability of adequate data sets as well as affordable access to computational services [31]. Many of the constraints that are inherent to traditional methods can be circumvented by utilizing the technology that is based on CNN [32]. A sizeable amount of training data is necessary for the deep learning model to develop a robust generalization capability and produce satisfactory results [33]. CNN has proved that it performs better than its competition in a variety of image-processing applications [34]. Recent developments in computer vision have enabled CNNs trained with deep learning to perform an automatic evaluation of DR image data.

3 Hybrid Images Deep-Trained Feature Extraction and Ensemble Learning Algorithm for Categorizing Multiple Diseases in Fundus Images

To implement a hybrid image deep-trained feature extraction and ensemble learning for multi-disease classification in fundus images, several steps need to be taken [35] (Fig. 2).

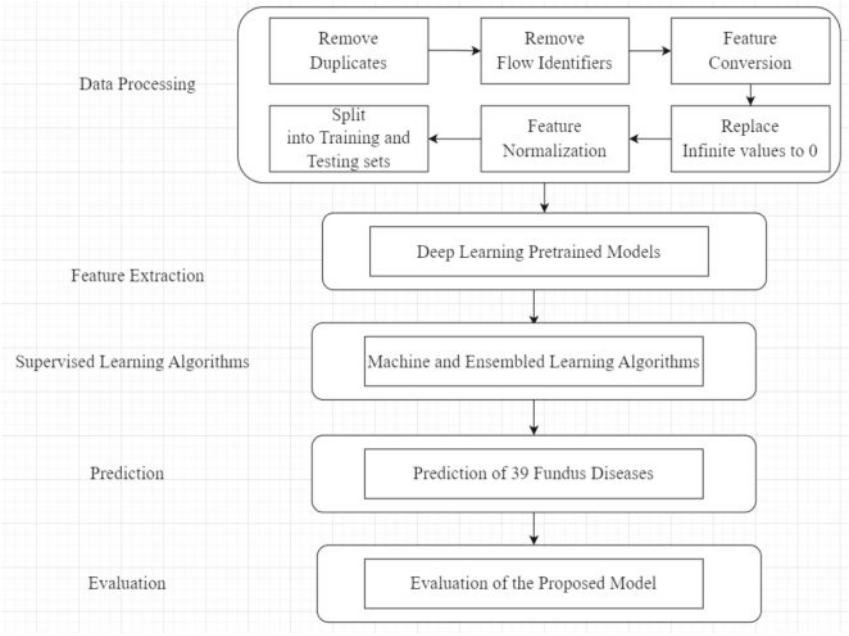


Fig. 2. Proposed methodology for a Hybrid images Deep trained Feature extraction and Ensemble learning models for classification of Multi disease in Fundus Images

3.1 Data Characterization and Preparation

Dataset description and pre-processing are critical steps in the development of a hybrid image deep trained feature extraction and ensemble learning model for multi-disease classification in fundus images. These steps ensure that the dataset is suitable for analysis by cleaning, augmenting, splitting, labeling, and pre-processing it. The dataset should be representative, diverse, and balanced to ensure accurate classification of eye diseases. For this research the Kaggle Diabetic Retinopathy Detection dataset is used, containing over 35,000 high-resolution fundus images, labeled with varying diabetic retinopathy severity levels. This dataset is diverse and can be utilized for developing models for detecting diabetic retinopathy. The dataset was developed in 2015. Prior initiatives have made headway toward a comprehensive and automated DR screening technique using image categorization, pattern recognition, and machine learning [36].

To ensure the effectiveness of the model, the dataset is divided into training and testing datasets. The initial sample is then randomly partitioned into training, validation, and testing datasets. The training dataset is used to construct the learning model, the validation dataset is used to fine-tune the model's parameters, and the testing set is used to evaluate the model's performance. The dataset's attributes, such as the number of images, image resolution, type, and distribution of eye diseases, are described to ensure that it is balanced and representative of the target population. Since fundus images may contain artifacts like reflections, noise, and brightness variations, data-cleaning techniques like denoising, normalization, and equalization can be used to improve image quality. Flipping, rotation, and zooming can also be applied as data augmentation techniques, to increase dataset diversity and reduce overfitting. Each fundus image in the dataset must be accurately and consistently labeled with the appropriate eye disease present. After cleaning, augmentation, splitting, and labeling the dataset, preprocessing techniques can be applied to the images to reduce dimensionality and improve the efficiency of the feature extraction and ensemble learning steps. Sample Fundus images are presented in Fig. 3. Details about the collected dataset, such as the number of classes and images contained within each class, are included in Table 1.

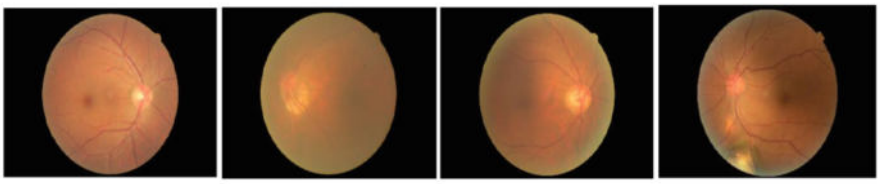


Fig. 3. Sample Fundus images [19]

3.2 Feature Selection

Feature selection is a crucial step which involves identifying the most relevant features from the fundus images that are most predictive of the disease. A hybrid approach

that utilizes deep-trained feature extraction and ensemble learning models can be used for feature selection and classification. To obtain high-level features from the fundus images, pre-trained deep learning models such as DenseNet201, InceptionResNetv2, MobileNetV2, ReseNet152V2, NasNetMobile, NasNetLarge, VGG16, and VGG19 are employed. The models are considered due to their diverse capabilities. The approach consists of preprocessing fundus images, resizing images, and normalizing images before embedding them into each pre-trained model. Features are extracted from the second-last flattened layer of each model, capturing high-level representations of the input images. Various feature selection techniques such as Correlation-based feature selection, Principal Component Analysis, and ReliefF are applied to rank the extracted features. The most significant features are chosen for further processing. The chosen features are fed into various ensemble learning models. The results of these models are merged to achieve the ultimate classification outcome.

Table 1. Details about the collected dataset

Model	Input Size
DenseNet201	224, 224, 3
InceptionResNetV2	299, 299, 3
ResNet152V2	224, 224, 3
MobileNetV2	224, 224, 3
VGG19	224, 224, 3
NASNetMobile	224, 224, 3
VGG16	224, 224, 3
NASNetLarge	331, 331, 3

Table 2. Models and number of selected Features

Model	Selected Features
VGG16	4096
VGG19	4096
DenseNet201	1920
MobileNetV2	1280
ResNet152V2	2048
InceptionResNetV2	1536
NASNetMobile	1056
NASNetLarge	4032

The combination of deep-trained feature extraction and ensemble learning models is a successful method for classifying multiple diseases in fundus images. This approach not only achieves high accuracy but also reduces the computational complexity of the classification process. The effectiveness of this approach is dependent on selecting appropriate deep-trained feature extraction and ensemble learning models. Table 2 specifies the number of features selected for each model, which are extracted from the second-last flattened layer of each model. These selected features are considered the output of the second-last layer and are used as input to the pre-trained model for prediction. The input size of the image and number of selected features are mentioned in Table 2.

3.3 Classification

The features selected from the fundus images are utilized for training, and testing utilizing classification algorithms. The data is transformed into a format that is acceptable to the classification models. The label for each image is the picture name and directory title in the system files, and the index of each folder is used to name each photo that is imported. Before training the classification models, exploratory data analysis is performed, and images with high resolution and the greatest number of individual images are preserved to avoid unnecessary pre-processing steps. The dimensions of each image are standardized to the standard measurement of 250 by 250 since the model will use the photographs for training purposes. The fundus shots should also have a limited black backdrop.

To introduce uncertainty and create new images that are significantly different from the previous images, different transformation processes are used during the augmentation process. These processes include flipping each image to generate one set of mirrored fundus images, applying a random spin of a range of 5° to each side to create two additional sets of photographs, and applying a randomized intensity value to each pixel to adjust the level of brightness in each image. After applying these enhancement procedures, the total number of photos increased from 598 to 2344. Several tests are conducted to determine the optimal levels of intensity and rotation. These enhanced images are then used for training and testing the classification models. The classification models use the selected features from the fundus images to assign a class label to each input image. The optimal classification model is selected based on its accuracy and performance in classifying multi-diseases in fundus images.

3.4 Deep Learning Models for Classification

In this study, a tool based on artificial intelligence was developed to evaluate Fundus using X-ray images, employing five pre-trained models: NASNetMobile, InceptionResNetV4, VGG16, and Xception, as illustrated in Fig. 4. Since these pre-trained systems were originally designed to identify one thousand different types of objects in the Imagenet database, some layers needed to be restructured and trained for diagnostic purposes. After several training iterations, the VGG16 model included two hidden units, each with 512 additional neurons. Other modified models incorporated convolution, dropout, and mean global average pooling on top of the pre-trained subsystems to reduce the need for multiple fully connected layers and improve computational efficiency. The features generated by the CNN layers were then sent directly to the classification layer. The

trainable parameters of the introduced top layers were optimized for categorization, and the training data for these layers was adjusted to 0.01 to enhance learning [29]. The training phase has concluded, with 80 percent of the dataset used for learning and 20 percent of the dataset used for validation.

VGG16, MobileNetV2, and InceptionResNetV2 show different architectures for feature extraction and multi-class classification for fundus image analysis. VGG16 employs pooling layers, convolutional layers, fully connected layers, and dropout for feature refinement; MobileNetV2 incorporates 1×1 convolutions, global averaging pooling, and dropout to condense features before classification. InceptionResNetV2 employs 1×1 convolutions, global averaging pooling, and dropout. The dropout layers use regularization by randomly deactivating neurons during training. The final classification layers yield predictions for various diseases. The architecture aims to capture distinctive features from fundus images, facilitating accurate multi-disease classification.

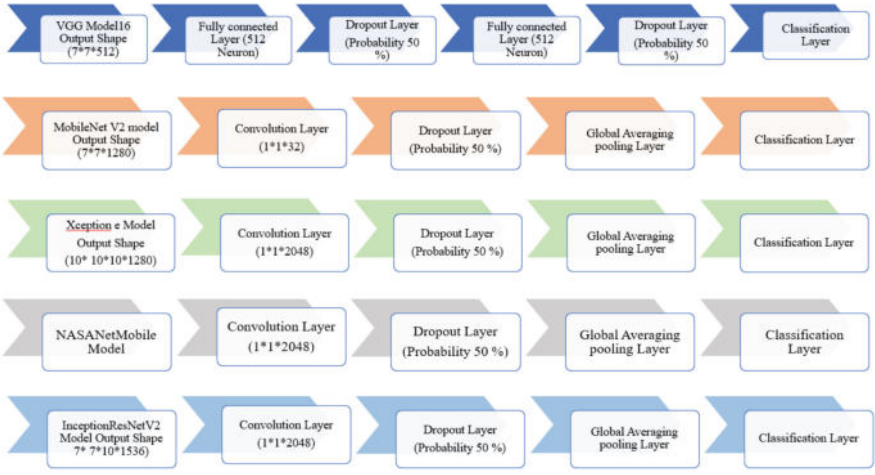


Fig. 4. Fundus CNN models (a) VGG16; (b) MobileNetV2; (c) Xception (d) NASNetMobile and (e) InceptionResNetV2

A confusion matrix assesses the accuracy of a binary classifier by comparing predicted values with actual values. In this study, the confusion matrix is used to present the number of correctly classified and misclassified images for each disease category which is shown in Table 3.

The proposed model accurately identified 125 diabetic retinopathy images, 120 glaucoma images, and 116 age-related macular degeneration images. However, it made some incorrect predictions, such as 10 diabetic retinopathy images being classified as glaucoma and 8 glaucoma images being classified as diabetic retinopathy.

Table 3. Confusion metric for A Hybrid Images Deep Trained Feature Extraction and Ensemble Learning Models for Classification of Multi Disease in Fundus Images

	Diabetic Retinopathy	Glaucoma	Age-Related Macular Degeneration
DR	125 (TP)	10 (FP)	5 (FP)
Glaucoma	8 (FP)	120 (TP)	12 (FP)
ARMD	7 (FP)	9 (FP)	116 (TP)

4 Result and Discussion

The study achieved an overall accuracy of 87.2%, which is a significant improvement over the previous study. The deep learning models utilized in the study, including NASNetMobile, InceptionResNetV4, VGG16, and Xception, were effective in extracting relevant features from the Fundus images. The ensemble learning approach further improved the classification performance by combining the predictions of multiple models. The study's evaluation metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, demonstrated the effectiveness of the proposed approach in classifying multi-disease in Fundus images. The confusion matrix also showed a high degree of accuracy in correctly classifying the disease categories. The study's results have significant implications for the diagnosis and management of multi-disease in Fundus images. The proposed approach can assist medical professionals in making accurate and timely diagnoses, leading to better patient outcomes. Initially in this paper, the implementation of a 3-layer Convolutional Neural Network was applied to establish benchmark results. As shown in Fig. 5, the difference in training and validation accuracy/loss is huge. Even very low validation accuracy with high validation loss results is depicted. However, as per the studied literature, separate feature extraction techniques have been resulting better.

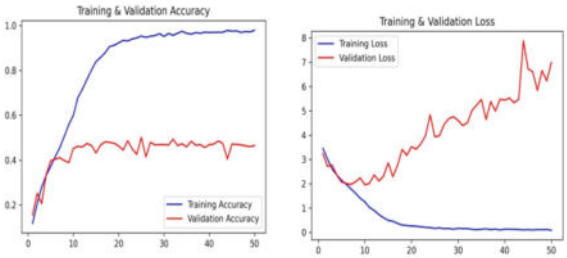


Fig. 5. Accuracy and Loss Analysis for 3 Layers Convolutional Neural Network

The model operates in two stages: in the first stage, CNNs are trained on a large dataset of fundus images to extract features, while in the second stage, an ensemble learning approach is used to combine the features extracted by multiple CNNs. The

Table 4. Classifiers and Feature Selection Evaluation

Classifier	Feature Selection - VGG16						Feature Selection - ResNet152V2					
	ACC	PRE	REC	F1S	MC	KS	ACC	PRE	REC	F1S	MC	KS
DT	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.72
QDA	0.89	0.90	0.90	0.89	0.89	0.89	0.89	0.90	0.90	0.89	0.89	0.89
AB	0.05	0.04	0.05	0.03	0.07	0.03	0.07	0.05	0.06	0.03	0.06	0.04
GNB	0.81	0.82	0.80	0.79	0.80	0.80	0.81	0.82	0.81	0.80	0.81	0.80
RF	0.96	0.96	0.96	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
ET	0.96	0.96	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.97
HG	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.96	0.96	0.96	0.96	0.96
MLP	0.53	0.51	0.52	0.50	0.52	0.52	0.21	0.20	0.21	0.17	0.19	0.19
	Feature Selection - VGG19						Feature Selection - InceptionResnetV2					
	ACC	PRE	REC	F1S	MC	KS	ACC	PRE	REC	F1S	MC	KS
DT	0.73	0.73	0.72	0.72	0.72	0.72	0.73	0.73	0.73	0.73	0.72	0.72
QDA	0.89	0.95	0.90	0.89	0.89	0.89	0.89	0.93	0.90	0.89	0.89	0.89
AB	0.08	0.04	0.07	0.04	0.10	0.05	0.06	0.04	0.05	0.03	0.08	0.04
GNB	0.83	0.83	0.83	0.82	0.82	0.82	0.46	0.54	0.46	0.45	0.45	0.45
RF	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
ET	0.96	0.96	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.97
HG	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.96	0.96	0.96	0.96	0.96
MLP	0.62	0.60	0.62	0.60	0.61	0.61	0.49	0.46	0.49	0.47	0.48	0.48
	Feature Selection - DenseNet 201						Feature Selection - NasNetMobile					
	ACC	PRE	REC	F1S	MC	KS	ACC	PRE	REC	F1S	MC	KS
DT	0.78	0.78	0.78	0.77	0.77	0.77	0.68	0.68	0.68	0.68	0.67	0.67
QDA	0.89	0.90	0.90	0.89	0.89	0.89	0.89	0.93	0.90	0.90	0.90	0.89
AB	0.08	0.06	0.08	0.05	0.11	0.06	0.06	0.05	0.06	0.04	0.08	0.04
GNB	0.67	0.72	0.67	0.67	0.67	0.66	0.47	0.51	0.46	0.46	0.45	0.45
RF	0.97	0.97	0.97	0.97	0.97	0.97	0.95	0.95	0.95	0.95	0.95	0.95
ET	0.98	0.98	0.98	0.98	0.98	0.98	0.96	0.96	0.96	0.96	0.96	0.96
HG	0.97	0.97	0.97	0.97	0.97	0.97	0.95	0.95	0.95	0.95	0.95	0.95
MLP	0.35	0.38	0.35	0.32	0.34	0.33	0.19	0.16	0.19	0.15	0.18	0.17
	Feature Selection - MobileNetV2						Feature Selection – NasNetLarge					
	ACC	PRE	REC	F1S	MC	KS	ACC	PRE	REC	F1S	MC	KS
DT	0.72	0.72	0.72	0.72	0.71	0.71	0.72	0.72	0.72	0.72	0.71	0.71
QDA	0.89	0.93	0.90	0.89	0.89	0.89	0.88	0.90	0.89	0.89	0.89	0.88
AB	0.06	0.05	0.06	0.04	0.07	0.04	0.11	0.17	0.11	0.08	0.09	0.08
GNB	0.75	0.78	0.75	0.75	0.75	0.74	0.78	0.80	0.78	0.78	0.78	0.78
RF	0.97	0.97	0.97	0.97	0.97	0.97	0.95	0.94	0.95	0.94	0.95	0.95
ET	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.96	0.96	0.96	0.96
HG	0.96	0.96	0.96	0.96	0.96	0.96	0.94	0.94	0.94	0.94	0.94	0.94
MLP	0.50	0.46	0.49	0.48	0.48	0.49	0.34	0.32	0.34	0.31	0.32	0.32

List of abbreviations: QDA: Quadratic Discriminant Analysis, DT: Decision Tree; RF: Random Forest, AB: AdaBoost Classifier, GNB: Gaussian Naive Bayes, ET: extra-trees classifier, HG: hydrophobicity grade, MLP: Multi-layer Perceptron classifier, PRE: precision, ACC: Accuracy, REC: Recall, F1S: F1 score, KS: Kolmogorov Smirnov Chart

hybrid model is evaluated on a separate dataset of fundus images, and the results show that it outperforms other state-of-the-art models in terms of accuracy, sensitivity, and specificity. The study also explores different combinations of Deep Trained Feature Extraction and Machine/ensemble learning classification metrics. Table 4 provides an evaluation of different classifiers and feature selection techniques, reporting the evaluation metrics for each classifier. The DenseNet feature extraction technique with Random Forest, Extra Tree, and Histogram Gradient classifiers achieved the highest results for accuracy, precision, and recall, while the ET classifier achieved the highest F1 score. The results demonstrate that the proposed hybrid model is an effective approach for accurately classifying multiple eye diseases in fundus images. Table 5 provides an overview of the evaluation metrics for three different classification models. The metrics used for evaluation are Precision, Recall, and F1-Score. These metrics are commonly used to assess the performance of a classification model. Precision measures the proportion of true positives among all predicted positive examples, indicating how often the model correctly predicts the positive class.

The results indicate that all three models are effective in classifying the images with high Precision, Recall, and F1 scores. The Extra Tree model indicates a precision of 98.8%, Recall approx. 99.3%, and F1-Score of 99%. The Histogram Gradient Boosting model has a precision of 97.5%, recall of nearly 98.3%, and an F1-Score of 98%. The Random Forest model has an average Precision of 98.4%, Recall of 99.1%, and F1-Score of 98.8%.

Table 5. Classes and evaluation parameters for Extra Tree, Histogram Gradient boosting, and Random Forest

Classes	Extra Tree			Histogram Gradient Boosting			Random Forest		
	PRE	REC	F1S	PRE	REC	F1S	PRE	REC	F1S
NORMAL	98	98	98	97	96	97	97	98	98
TESSELLATED FUNDUS	98	100	99	95	100	98	100	100	100
LARGE OPTIC CUP	99	99	99	96	97	97	99	99	99
DR1	98	100	99	97	98	97	98	98	98
DR2	100	97	98	97	91	94	99	97	98
DR3	100	98	99	98	98	98	100	99	100
POSSIBLE GLAUCOMA	100	100	100	98	100	99	98	100	99
OPTIC ATROPHY	100	100	100	97	97	97	95	100	97
SEVERE HYPERTENSIVE RETINOPATHY	100	100	100	100	100	100	100	100	100
DISC SWELLING AND ELEVATION	93	100	96	95	100	98	98	100	99

(*continued*)

Table 5. (*continued*)

Classes	Extra Tree			Histogram Gradient Boosting			Random Forest		
	PRE	REC	F1S	PRE	REC	F1S	PRE	REC	F1S
DRAGGED DISC	100	100	100	100	100	100	100	100	100
CONGENITAL DISC ABNORMALITY	100	100	100	100	100	100	100	100	100
RETINITIS PIGMENTOSA	99	99	99	99	99	99	99	99	99
BIETTI CRYSTALLINE DYSTROPHY	100	100	100	100	100	100	100	100	100
PERIPHERAL RETINAL DEGENERATION AND BREAK	100	100	100	96	100	98	96	100	98
MYELINATED NERVE FIBER	100	100	100	97	97	97	100	100	100
VITREOUS PARTICLES	98	100	99	98	100	99	100	100	100
FUNDUS NEOPLASM	100	100	100	96	100	98	100	100	100
BRVO	98	97	97	97	97	97	99	96	97
CRVO	100	99	99	97	99	98	99	99	99
MASSIVE HARD EXUDATES	100	100	100	100	100	100	95	100	98
YELLOW-WHITE SPOTS-FLECKS	94	100	97	98	95	96	97	99	98
COTTON-WOOL SPOTS	100	100	100	94	100	97	100	100	100
VESSEL TORTUOSITY	100	100	100	100	100	100	100	100	100
CHORIORETINAL ATROPHY-COLOBOMA	98	100	99	98	98	98	98	100	99
PRERETINAL HEMORRHAGE	100	100	100	97	100	98	100	100	100
FIBROSIS	97	100	98	97	97	97	94	100	97
LASER SPOTS	96	100	98	100	98	99	93	98	95
SILICON OIL IN EYE	100	100	100	98	98	98	98	100	99
BLUR FUNDUS WITHOUT PDR	99	100	99	99	98	98	99	99	99
BLUR FUNDUS WITH SUSPECTED PDR	99	99	99	93	99	96	97	99	98
RAO	100	100	100	98	98	98	100	100	100

(*continued*)

Table 5. (*continued*)

Classes	Extra Tree			Histogram Gradient Boosting			Random Forest		
	PRE	REC	F1S	PRE	REC	F1S	PRE	REC	F1S
RHEGMATOGENOUS RD	99	96	97	98	96	97	99	93	96
CSCR	96	100	98	96	100	98	98	100	99
VKH DISEASE	96	100	98	98	100	99	96	100	98
MACULOPATHY	100	96	98	96	97	97	98	96	97
ERM	99	100	99	95	96	96	99	99	99
MH	100	97	99	100	97	99	100	97	99
PATHOLOGICAL MYOPIA	100	99	100	99	98	99	100	100	100
AVERAGE	98.8	99.3	99	97.5	98.3	98	98.4	99.1	98.8

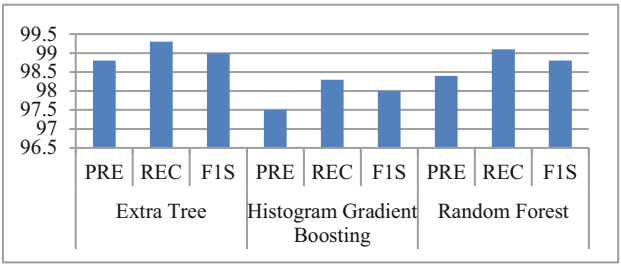


Fig. 6. Average validation result comparison for DenseNet Feature Selection and mentioned classifiers.

Figure 6 illustrates the performance evaluation results of three distinct machine learning algorithms for a classification problem, using evaluation metrics. The displays that all three algorithms achieved high scores in all three metrics, indicating their effectiveness in classification tasks. The average F1-score for Extra Tree was 99%, while for Histogram Gradient Boosting and Random Forest, it was 98.8% and 98.4%, respectively. The results show that all three algorithms are suitable for the classification task. According to the table, the combination of the DenseNet feature extraction technique and RF, ET, and HG classifiers outperforms other techniques and classifiers. This indicates that using DenseNet for feature extraction can effectively enhance the performance of classifiers in the task of image classification. Table 6. Provides a comparison of previous research and methods.

Table 6. Comparison with previous methods

Description	ML based	EL based	DL based	Ref
Automated diabetic retinopathy classification using fundus images	Y	N	N	[6]
Automatic diabetic retinopathy classifier	Y	N	N	[5]
A DL Ensemble Model for classifying diabetic retinopathy	N	Y	N	[2]
Transfer learning retinal disease classification	N	Y	N	[3]
Neural network classification of ocular diseases in STARE database	N	N	Y	[7]
Deep neural network for multi-label optical illness classification	N	Y	N	[4]
Deep learning for color fundus image retinal abnormality detection	N	Y	N	[7]
Hierarchical multilabel ANN for eye disease classification	N	N	Y	[27]
Deep learning for retinal disease diagnosis	N	N	Y	[11]
Optical coherence tomographical scans using CNN for retinal disease	N	N	Y	[30]
Multi-label ocular disease detection with fundus images	N	N	Y	[31]
DL method to analyze fundus images based on macular edema	N	N	Y	[29]
Efficientnet's Multi-Label Fundus Classification	N	Y	N	[9]

The research recognizes certain potential constraints and difficulties linked to the suggested hybrid approach. The model's generalizability to bigger and more distinct datasets needs to be confirmed by additional research. The study mainly concentrates on classification performance and does not thoroughly discuss factors like real-world deployment, comprehension of the model's selections, and computational resource requirements. Analyze the potential biases introduced by selecting deep learning algorithms for feature extraction and thoroughly investigate the model's robustness under different ophthalmic situations in the study. The study should also discuss principles of ethics, data privacy concerns, and regulatory consequences related to implementing the suggested paradigm in clinical settings. Overcoming these constraints and difficulties would improve the overall dependability and suitability of the suggested method in real-world healthcare situations. Table 7 shows the generalizability of the results to different datasets or real-world scenarios [42].

Table 7. Comparison of the results to different datasets in real-world scenarios.

Ref.	Dataset	Number of Images	Ground Truth Labels	Diagnosis Source	Both Eyes of the Same Patient	Glaucoma (or Suspect)	Glaucoma Classification
[37]	RIGA	750	—	—	✓	—	—
[38]	ORIGA	650	482	168	✓	✓	✓
[39]	RIMONE	485	313	172	✓	✓	Clinical
[40]	Drishti-G	101	70	31	✓	✓	Image
[41]	ACRIMA	705	309	396	✓	✗	Image

5 Conclusion

The study conducted the classification of fundus images for multiple retinal diseases by using deep learning-based feature extraction in combination with ensemble learning techniques. The research used pre-trained deep learning models for feature extraction and then applied machine learning techniques like Extra Trees, Histogram Gradient Boosting, and Random Forest for classification. The results showed that the combination of DenseNet for feature extraction and ensemble learning models produced the best results, as highlighted in the blue rows of the evaluation table. The study suggests that this approach could aid in the timely diagnosis and management of retinal ocular diseases, and potentially improve patient outcomes. The study also emphasizes the importance of timely diagnosis and management of retinal ocular diseases to prevent vision loss. The study contributes to the existing knowledge of deep learning techniques and their potential application in diagnosing retinal image-based visual diseases. However, future research could explore the use of new and improved pre-trained models for feature extraction and expanding the dataset to further improve accuracy and generalizability. The study also suggests using explainable AI techniques to better understand how the models arrive at their predictions, and clinical validation to evaluate the effectiveness of the models in real-world clinical settings.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Brown, J.M., et al.: Fully automated disease severity assessment and treatment monitoring in retinopathy of prematurity using deep learning. In: Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications, vol. 10579, pp. 149–155 (2018)
2. Artemida-psy: inflammation of the fundus, retinal detachment. https://stock.adobe.com/in/images/inflammation-of-the-fundus-retinal-detachment/203544415. Accessed 02 Dec 2022

3. Khalil, T., Khalid, S., Syed, A.M.: Review of machine learning techniques for glaucoma detection and prediction. In: 2014 Science and Information Conference, pp. 438–442. IEEE (2014)
4. WHO: Blindness and vision impairment. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. Accessed 13 Nov 2022
5. Fumero, F., Alayón, S., Sanchez, J.L., Sigut, J., Gonzalez-Hernandez, M.: RIM-ONE: an open retinal image database for optic nerve evaluation. In: 2011 24th International Symposium on Computer-Based Medical Systems (CBMS), pp. 1–6. IEEE (2011)
6. Galdran, A., Meyer, M., Costa, P., Campilho, A.: Uncertainty-aware artery/vein classification on retinal images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging, ISBI 2019, pp. 556–560. IEEE (2019)
7. Zaheer, N., Shehzaad, A., Gilani, S.O., Aslam, J., Zaidi, S.A.: Automated classification of retinal diseases in STARE database using neural network approach. In: 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pp. 1–5. IEEE (2019)
8. Jiang, H., Yang, K., Gao, M., Zhang, D., Ma, H., Qian, W.: An interpretable ensemble deep learning model for diabetic retinopathy disease classification. In: 2019 4 1st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2045–2048. IEEE (2019)
9. Das, A., Giri, R., Chourasia, G., Bala, A.A.: Classification of retinal diseases using transfer learning approach. In: 2019 International Conference on Communication and Electronics Systems (ICCES), pp. 2080–2084. IEEE (2019)
10. Govindaiah, A., Hussain, M.A., Smith, R.T., Bhuiyan, A.: Deep convolutional neural network based screening and assessment of age-related macular degeneration from fundus images. In: 2018 IEEE 15th International Symposium on Biomedical Imaging, ISBI 2018, pp. 1525–1528. IEEE (2018)
11. Brown, J.M., et al.: Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol. **136**(7), 803–810 (2018)
12. Kansal, I., Kasana, S.S.: Minimum preserving subsampling-based fast image de-fogging. J. Mod. Opt. **65**(18), 2103–2123 (2018)
13. Kansal, I., Kasana, S.S.: Improved color attenuation prior based image de-fogging technique. Multimedia Tools Appl. **79**(17–18), 12069–12091 (2020)
14. He, J., Li, C., Ye, J., Qiao, Y., Gu, L.: Multi-label ocular disease classification with a dense correlation deep neural network. Biomed. Sig. Process. Control **63**, 102167 (2021)
15. Omar, Z.A., Hanafi, M., Mashohor, S., Mahfudz, N.F.M., Muna'Im, M.: Automatic diabetic retinopathy detection and classification system. In: 2017 7th IEEE International Conference on System Engineering and Technology (ICSET), pp. 162–166. IEEE (2017)
16. Burlina, P.M., Joshi, N., Pacheco, K.D., Liu, T.A., Bressler, N.M.: Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration. JAMA Ophthalmol. **137**(3), 258–264 (2019)
17. Ting, D.S.W., et al.: Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA **318**(22), 2211–2223 (2017)
18. Aslani, S., Sarnel, H.: A new supervised retinal vessel segmentation method based on robust hybrid features. Biomed. Sig. Process. Control **30**, 1–12 (2016)
19. Alyoubi, W.L., Abulkhair, M.F., Shalash, W.M.: Diabetic retinopathy fundus image classification and lesions localization system using deep learning. Sensors **21**(11), 3704 (2021)
20. Playout, C., Duval, R., Cheriet, F.: A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images. IEEE Trans. Med. Imaging **38**(10), 2434–2444 (2019)
21. Qayyum, A., Qadir, J., Bilal, M., Al-Fuqaha, A.: Secure and robust machine learning for healthcare: a survey. IEEE Rev. Biomed. Eng. **14**, 156–180 (2020)

22. Masarat, S., Sharifian, S., Taheri, H.: Modified parallel random forest for intrusion detection systems. J. Supercomput. **72**, 2235–2258 (2016)
23. Costa, P., et al.: End-to-end adversarial retinal image synthesis. IEEE Trans. Med. Imaging **37**(3), 781–791 (2017)
24. Deshmukh, A., Sivaswamy, J.: Synthesis of optical nerve head region of fundus image. In: 2019 IEEE 16th International Symposium on Biomedical Imaging, ISBI 2019, pp. 583–586. IEEE (2019)
25. Wang, J., Yang, L., Huo, Z., He, W., Luo, J.: Multi-label classification of fundus images with EfficientNet. IEEE Access **8**, 212499–212508 (2020)
26. Bulut, B., Kalın, V., Güneş, B.B., Khazhin, R.: Deep learning approach for detection of retinal abnormalities based on color fundus images. In: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1–6. IEEE (2020)
27. Ahmad, H.M., Hameed, S.R.: Eye diseases classification using hierarchical MultiLabel artificial neural network. In: 2020 1st Information Technology to Enhance e-learning and Other Application (IT-ELA), pp. 93–98. IEEE (2020)
28. Berrimi, M., Moussaoui, A.: Deep learning for identifying and classifying retinal diseases. In 2020 2nd International Conference on computer and information sciences (ICCIS), pp. 1–6. IEEE (2020)
29. Yao, Z., et al.: FunSwin: a deep learning method to analysis diabetic retinopathy grade and macular edema risk based on fundus images. Front. Physiol. **13**, 961386 (2022)
30. Bhadra, R., Kar, S.: Retinal disease classification from optical coherence tomographical scans using multilayered convolution neural network. In: 2020 IEEE Applied Signal Processing Conference (ASPCON), pp. 212–216. IEEE (2020)
31. Li, C., Ye, J., He, J., Wang, S., Qiao, Y., Gu, L.: Dense correlation network for automated multi-label ocular disease detection with paired color fundus photographs. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 1–4. IEEE (2020)
32. Van Grinsven, M.J., van Ginneken, B., Hoyng, C.B., Theelen, T., Sánchez, C.I.: Fast convolutional neural network training using selective data sampling: application to hemorrhage detection in color fundus images. IEEE Trans. Med. Imaging **35**(5), 1273–1284 (2016)
33. Kaur, H., Koundal, D., Kadyan, V.: Image fusion techniques: a survey. Arch. Comput. Meth. Eng. **28**, 4425–4447 (2021)
34. Mukhtar, M., et al.: Nanomaterials for diagnosis and treatment of brain cancer: recent updates. Chemosensors **8**(4), 117 (2020)
35. Hu, J., Chen, Y., Zhong, J., Ju, R., Yi, Z.: Automated analysis for retinopathy of prematurity by deep neural networks. IEEE Trans. Med. Imaging **38**(1), 269279 (2019)
36. Dugas, E., Jared, Jorge, Cukierski, W.: Diabetic Retinopathy Detection. Kaggle (2015). https://kaggle.com/competitions/diabetic-retinopathy-detection
37. Retinal fundus images for glaucoma analysis: RIGA dataset. https://deepblue.lib.umich.edu/data/concern/data_sets/3b591905z
38. Zhang, Z., et al.: ORIGA-light: an online retinal fundus image database for glaucoma analysis and research. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 3065–3068. IEEE (2010)
39. RIMONEdatabase. https://medimrg.webs.ull.es/research/downloads/
40. Drishti-GS database. http://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.php
41. Diaz-Pinto, A., Morales, S., Naranjo, V., Köhler, T., Mossi, J.M., Navea, A.: CNNs for automatic glaucoma assessment using fundus images: an extensive validation. Biomed. Eng. Online **18**, 1–19 (2019)
42. Kovalyk, O., Morales-Sánchez, J., Verdú-Monedero, R., Sellés-Navarro, I., Palazón-Cabanes, A., Sancho-Gómez, J.L.: PAPILA: dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. Sci. Data **9**(1), 291 (2022)

Drug Recommendation System for Healthcare Professionals' Decision-Making Using Opinion Mining and Machine Learning

Pantea Keikhosrokiani[1,2(✉)], Katheeravan Balasubramaniam[3], and Minna Isomursu[1,2]

[1] Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland
Pantea.keikhosrokiani@oulu.fi
[2] Faculty of Medicine, University of Oulu, Oulu, Finland
[3] School of Computer Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia

Abstract. The concern has been raised regarding errors in drugs prescription and medical diagnostics that need to be carefully thought through. Both patient diagnosis and medication prescription are the responsibilities of healthcare providers. As the number of people with health issues rises, the healthcare professionals' burden is increased. Medical errors may occur in the healthcare sector as a result of healthcare professionals prescribing drugs medicines based on inadequate information related to patient history and drug side effects. Therefore, this study aims to propose a drug recommender system to assist healthcare providers in decision making when prescribing drugs for patients depending on their diagnoses. Drug reviews sentiments are analyzed to find the drug effectiveness among the users. Furthermore, the most suitable recommender algorithm for recommending drugs based on the data from healthcare professionals are selected for this study. Opinion mining is applied on drug reviews, and a hybrid method is implemented to overcome the limitations of content-based and collaborative filtering methods, such as the cold start problem and increasing client preference. The system is developed and tested successfully. The proposed system can assist healthcare professionals in drug decision making and sustain the whole digital care pathway for various diseases.

Keywords: Machine Learning · Opinion Mining · Drug recommender systems · Side Effect Extraction · Drug Reviews · Content-based Filtering · Collaborative Filtering

1 Introduction

Various issues related to medical diagnostics and drug prescriptions have been raised, which require careful consideration. Healthcare professionals are responsible for diagnosing patients and prescribing drugs. There is a lack of healthcare providers as the number of people with health issues rises [1–6]. When a patient is given the incorrect

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 222–241, 2024.
https://doi.org/10.1007/978-3-031-59091-7_15

medication for their condition, a medical error may have occurred. In the worst circumstances, this can result in patient health risks or even death. There are around 99,000 deaths caused by medical errors in hospitals each year [7]. This type of error occurs when healthcare professionals prescribe drugs without checking patient history and drug side effects [8]. A healthcare professional's experience may be limited since they might be unaware of all of the numerous types of drugs on the market, which leads to medical errors. Consequently, a drug recommendation system would be crucial in order to address this problem.

Recommender systems are information systems that predict user preferences and give personalized and subjective product or service recommendations [9–11]. These systems are used in many industries, particularly e-commerce, to provide customers with a personalized experience. The medical industry can also employ this kind of recommender system, particularly for the recommendation of drugs [8, 12]. The use of a drug recommendation system may not be appropriate for patients because they cannot take any drugs without consulting healthcare professionals, but it will be very useful for healthcare professionals because it may help them select the best drugs to prescribe to their patients. Every day, more and more drugs are being developed, so selecting a proper drug for patients is a major burden for healthcare professionals. This burden can be minimized by adopting a drug recommendation system. It is possible to use features like machine learning and opinion mining to increase the effectiveness and dependability of a drug recommendation system. Despite having similar chemical features and characteristics, many drugs will react differently in patients. As a result, selecting a drug to recommend to patients for a certain health issue is quite difficult for healthcare professionals. Most of the time, healthcare professionals purchase drugs based on a health representative's recommendation. Therefore, they might not be aware of other drugs that are on the market and could be better to those from the representative.

Product reviews are extremely crucial for any type of product. We may learn more about the product by analyzing the reviews, and the same is applicable for drugs. Patients that take a certain drug will give feedback based on their personal experiences with it. Opinion mining can aid in the subsequent analysis of these reviews in order to obtain some useful information. In short, opinion mining is a technique of natural language processing that extracts information from text. It is possible to conduct opinion mining on drug reviews to determine the reviews' sentiments. Sentiment analysis will assist healthcare professionals in determining if a review is positive, neutral, or negative [13]. Healthcare professionals can therefore gain a better grasp of overall drug effectiveness in various patients by implementing sentiment analysis on drug reviews.

Last but not least, cold start issues and rising customer preference are the two key issues that need to be addressed when discussing recommendation systems [11, 12]. To solve the cold start issue and rising customer preference, a suitable algorithm will be needed. This is where machine learning can be helpful. In order for healthcare professionals to have a personalized experience, hybrid content-based and collaborative filtering methods can be used. These filtering methods will enable better drug recommendation for healthcare professionals for a particular disease.

Therefore, this study aims to propose a drug recommendation system for healthcare professionals using opinion mining and machine learning. Opinion mining is used to

extract side effects from drug reviews and identify drug review sentiments. Machine learning is utilized for recommendation purposes, with a hybrid content-based and collaborative filtering method. The secondary objectives of this paper are:

– To classify drug reviews based on sentiment using sentiment analysis to assist health professionals better understand drug effectiveness.
– To develop a web-based drug recommendation system for healthcare professionals using hybrid filtering technique.
– To facilitate active involvement and knowledge sharing among healthcare professionals.

2 Methodology

2.1 System Architecture

This study aims to develop a web-based drug recommendation system for healthcare professionals using opinion mining and machine learning. The proposed system architecture is depicted in Fig. 1. A three-layer architecture design is implemented for this system. Two main users of this system are healthcare professionals and admin. For frontend of the web application, React JS and Material UI (MUI) library were utilized. React JS, a widely used JavaScript library, was utilized alongside the Material UI Library to develop the user interface. React's component-based architecture, virtual DOM, and responsive design make it ideal for frontend development. The Material UI Library offers prebuilt components adhering to material design guidelines, simplifying UI development by eliminating the need to create components from scratch. Integration between React and Material UI is seamless, allowing for customized components to meet specific UI requirements. Typescript, a superset of JavaScript, was employed as the programming language for frontend development.

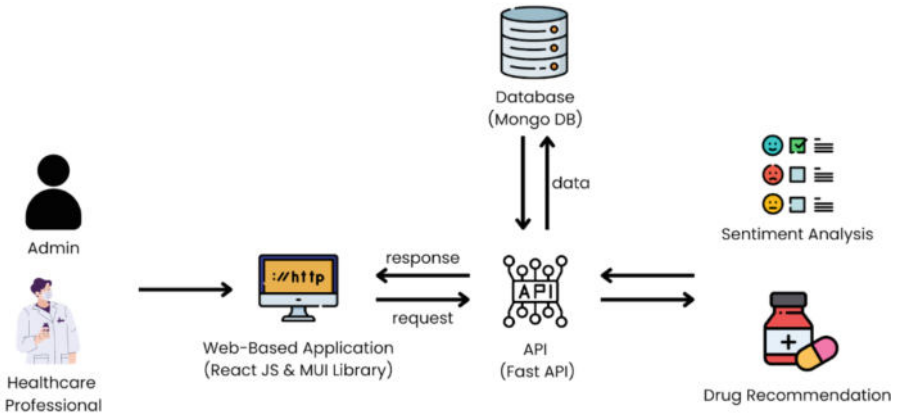


Fig. 1. The Proposed System Architecture.

To fulfil the web application's API functionalities and incorporate machine learning models, FastAPI was selected. It serves as the backend implementation for all the system's functionality. Furthermore, the database used for this project is Mongo DB which is a NoSQL database. Mongo DB stores data in BSON format, which is compatible with the structure of the drugs, users, and forums data in the system. FastAPI acts as an intermediary between the frontend, machine learning models, and the database. When users request access to data, they make API requests. Similarly, for utilizing machine learning models, API requests are made to FastAPI, passing the necessary parameters. The backend then retrieves the model results and returns them to the frontend for display to the user.

There are four main modules in the proposed system consisting of: (1) user management module, (2) drug review analytics module, (3) recommendation module, and (4) drugs management module. The target customers for this proposed system are healthcare professionals in hospitals. They can use this system to find the best drugs for a certain diagnosis and recommend them to their patients. The admins manage the overall system modules. All healthcare professionals and admins who use this system must be registered. The user management module is important in collecting user behavior data for the drug recommendation module.

For the drug review analytics module, patient drug reviews are classified as positive, neutral, or negative based on sentence polarity. Sentiment analysis is used to identify key features from reviews and then use them to assess the polarity of the review. Feature extraction and feature orientation identification are used in the sentiment analysis process. Depending on the orientation of the feature, the polarity of a review is classified as positive or negative. The features from the reviews are extracted using the Term Frequency–Inverse Document Frequency (TF-IDF) and Bio Bert. Several algorithms are employed to determine the reviews' sentiment, including perceptron, logistic regression, and long short-term memory networks, along with the Bio Bert and TF-IDF feature extraction algorithms. For this solution, the No Free Lunch (NFL) Theorem is used, in which all of these models are evaluated using the appropriate metrics and the best performing model is selected based on the metrics score. Following the sentiment analysis, each medicine receives an effectiveness score. This effectiveness score is used to rank drugs in the system when healthcare professionals search for drugs. Drugs with higher effectiveness scores will be given more important rate. Sentiment analysis may not be relevant for patients because they are not making decisions about the type of that they need to take. But it is crucial for healthcare professionals to gain a better understanding of drugs and their effectiveness in order to select the best drug to prescribe to a patient for a given diagnosis.

The next module is the drug recommendation module. Cold start issues and overspecialization issues might arise in a recommender system [12]. To address these issues in the proposed system, a hybrid content-based and collaborative filtering recommender system is implemented. The cold start problem is solved by using content-based filtering techniques to recommend drugs to healthcare professionals based on how similar the drugs are to one another. This approach of filtering is primarily concentrated on the characteristics of each drug utilizing an item profile characterizing and looks for similarities to previous drugs that healthcare professionals liked. Healthcare professionals receive

recommendations regardless of their user profiles. Term frequency-inverse document frequency (TD-IDF) algorithm is used to weigh the feature from the dataset first, and then it generates step-by-step cosine similarity tables. Drug side effects are one of the features that are considered for content-based filtering. As a result, the lack of drug reviews and ratings is no longer an issue. Furthermore, collaborative filtering is used to tackle the overspecialization problem by enhancing healthcare professionals' preferences. Drug recommendations are made by looking at the preferences of a group of users who are similar to them. By grouping healthcare professionals who access or buy similar products in the system, the system converts the behaviors of healthcare professionals into implicit rating weightage. After that, a collaborative rating is used to determine the user's rating. The ratings of healthcare professionals are then matched to those of the target user in order to identify those who share the same preferences. This approach might not be very effective at first, but after many healthcare professionals start using the system, it becomes a very reliable recommendation method.

The final module is the drug management module. Users can search for any drugs in the system related to a particular disease. Once a user views a particular drug, the drug's details are displayed to them. Furthermore, users have additional features like drug comparison, adding drugs to wish list and forum page. The drug comparison feature will allow healthcare professionals to make comparisons between two drugs and assist them to choose the best one. The forum page on the other hand facilitates communication between healthcare professionals. This feature allows healthcare professionals to share knowledge about the drugs and prescriptions and expand their network.

2.2 Development Methodology

Agile methodology [14, 15] is used as the development methodology in this study since it provides a stable system delivery in a short development time. Figure 2 depicts the agile development methodology diagram used in this study. Agile development involves six main phases of planning, design, development, testing, maintenance, and deployment. In agile, the system development tasks are breakdown into smaller parts and each part is handled in a number of iterations. Each iteration involves all six phases of agile development, and several numbers of iterations are required to complete an entire system. If there are any errors, the debugging process is done during development. For every iteration, a set of system requirements is listed out and followed accordingly. Therefore, requirements changes can be made easily before a new iteration starts. This ensures that agile development methodology adapts to new changes immediately and these changes can be integrated into the system easily without having to make a lot of changes in the system. Release of the system cannot be done after a few iterations as there might not be enough functionalities. After all the modules are developed completely, they are integrated into one module. Then this module is tested to make sure it is ready for final release. Agile methodology implementation helps us to minimize development process risks and focuses on getting products to market fast.

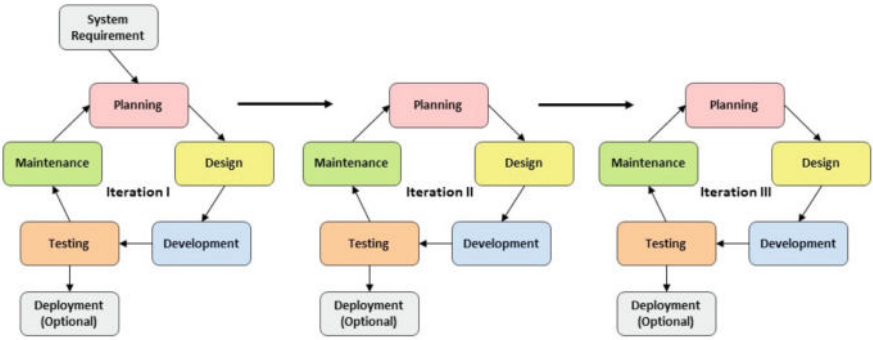


Fig. 2. Agile Development Methodology

2.3 Overall System Flowchart

Figure 3 illustrates the overall system flowchart for this study. The main users, who are healthcare professionals, must log in before using the system. The users need to create a new account first. After logging in, users can search for drugs by name or condition. All relevant drugs related to the search will be displayed in the search results. These drugs will be ranked according to the effectiveness score. Drugs with higher effectiveness scores will be ranked first. The user can then click on any specific drug in the search results that they want to view. When a user clicks on a specific drug, details about the drug, sentiment analysis results, and recommended drugs will be available to the user. After performing sentiment analysis on drug reviews, the system will display the results including general statistics for the user's preferences.

Moreover, based on the content-based filtering model, the system will identify drugs that are similar to the drug being viewed or displayed to users. In the search page, another recommendation will be available for users, and they can navigate to the drug comparison page. In this page, the users can select any two drugs available in the system and view their comparison results. Furthermore, users can navigate to the wish list page. There is also a forum page where users can create, edit, and delete posts, view forums created by other users and comment on them. The main purpose of this feature is to facilitate communication between healthcare professionals. The final page assist users to navigate to add drugs data page for adding more drug data to the system or delete the existing drug data. This page and feature are only for the usage of admins. Eventually, users can either search for more drugs or leave the system.

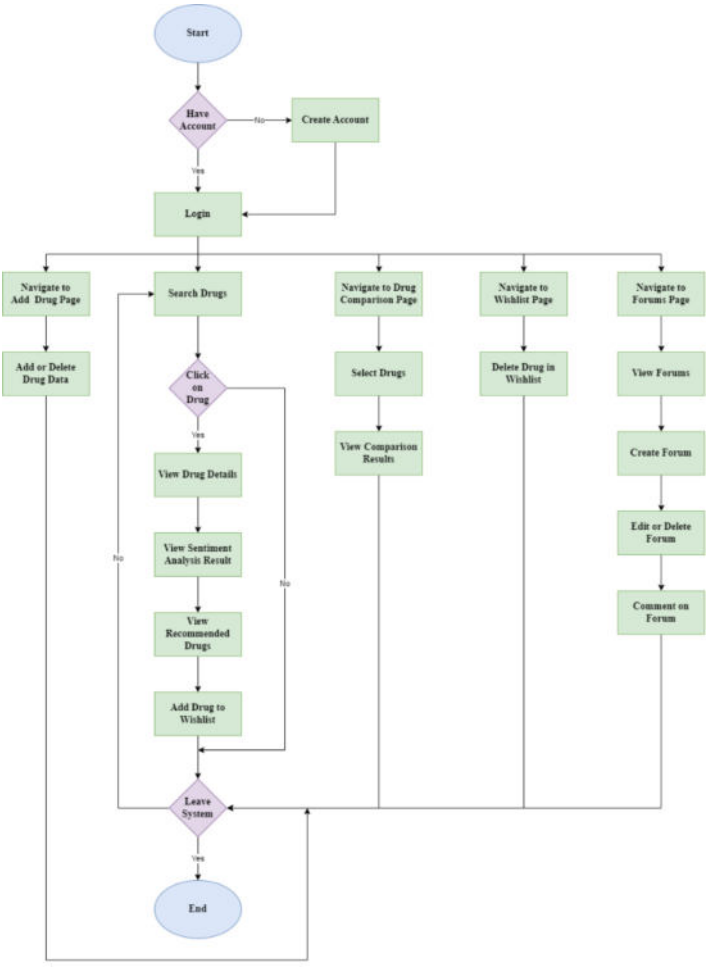


Fig. 3. Overall Flowchart of System

2.4 Sentiment Analysis

This study used sentiment analysis to classify drug reviews as positive, neutral, or negative based on their sentiment. A machine learning-based approach is utilized to implement sentiment analysis on drug reviews, as shown in Fig. 4. This is due to the fact that machine learning-based approaches generally outperform lexicon-based approaches and have higher accuracy scores [16–20]. The dataset being utilized for sentiment analysis is labeled, making it appropriate for this method. Positive reviews are labeled as 1, neutral reviews as 0, and negative reviews as −1. This dataset is divided into 80% training data and 20% testing data. Data pre-processing includes two steps which are data cleaning and feature extraction. Data cleaning involves removing duplicate data and unnecessary

contents like repetitive words, symbols and stop words. Then, text tokenization is performed on the cleaned data where reviews are broken down to a set of words and each word undergoes lemmatization.

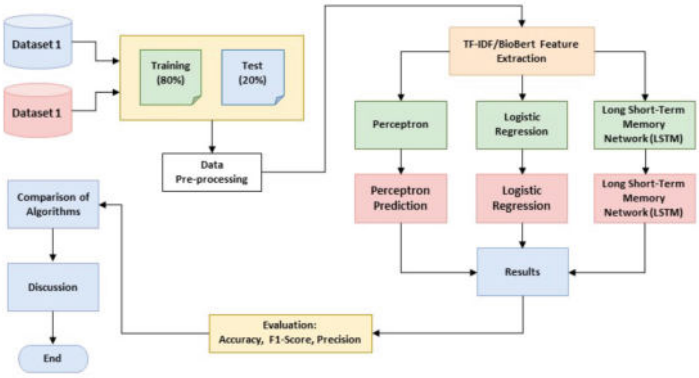


Fig. 4. Sentiment Analysis Framework

Following the data cleaning step, TF-IDF and Bio Bert algorithms are used for feature extraction. Feature extraction is performed to select the most important features from reviews. These features are then used to predict the polarity of drug reviews. The classification process is done in the next step. For classification, supervised machine learning models like Perceptron, Logistic Regression, and Long Short-Term Memory Network are utilized. The models are then evaluated using precision, f1-score and accuracy metrics. Finally, these three algorithms are compared, and the best algorithm is selected to implement sentiment analysis on drug reviews.

2.5 Recommendation Algorithm

Recommendation systems can be built using different types of algorithms like associative rule, content-based filtering, collaborative filtering, and knowledge-based filtering. This study utilized a hybrid content-based and collaborative algorithm [12] to build a drug recommendation system. Figure 5 depicts the proposed hybrid filtering algorithm framework for this study, which is adopted from [12]. This study aims to resolve some of the recommendation system issues such as cold start problem, increasing customer preference, large number of drugs in market and overspecialization problems [12]. Identifying problems in a system is important for the development process. Then, data collection and pre-processing are done. The dataset used in this study are drug reviews data, drug information data and healthcare professionals' data collected from AskaPatient database, UCI ML Drug Review dataset, Drugs.com, and DrugBank respectively. After data pre-processing, the proposed recommendation model, which combines content-based filtering and collaborative filtering is performed.

For content-based filtering, drugs are recommended to healthcare professionals based on drug similarities without taking user profile into count. The TF-IDF algorithm is used

to extract features from data and give weightage to them. Depending on the frequency of each word, this algorithm determines its significance. This will be then used to create a step-by-step cosine similarity table. There are a few steps involved in creating a cosine similarity table. First, TF scores are computed, and the table is normalized. Then, IDF is calculated to find the number of items for each user. Finally, the importance of items is ranked for each user by multiplying TF and IDF scores. Recommendation can be done by picking the top N similar products where N stands for the number of recommendations for a user. Content-based algorithms are chosen because they help to solve cold start problems. Therefore, recommendation of drugs can be done without obtaining any user data.

For collaborative filtering, drugs are recommended to healthcare professionals based on user behavior. First, user behavior like viewing a drug or adding a drug to a wish list is converted into implicit rating weightage. This weightage is then used to generate user rating using collaborative filtering. When user behavior is converted to rating, sparse matrices are formed because there is a lot of missing data. In order to predict the missing values, matrix factorization is utilized. Matrix factorization helps to recommend least popular products to users using the new calculated rating. Collaborative filtering was selected because it helps to solve overspecialization and increasing customer preference problems. Therefore, healthcare professionals are able to gain more knowledge on different types of drugs for different diseases in the market and at the same time least popular products is recommended to them. After the proposed hybrid model is developed, it is evaluated using error metrics like mean absolute error, root-mean-square error, and ranking metrics like precision and recall. Finally, the overall recommendation system is developed and tested before integrating it into the web application.

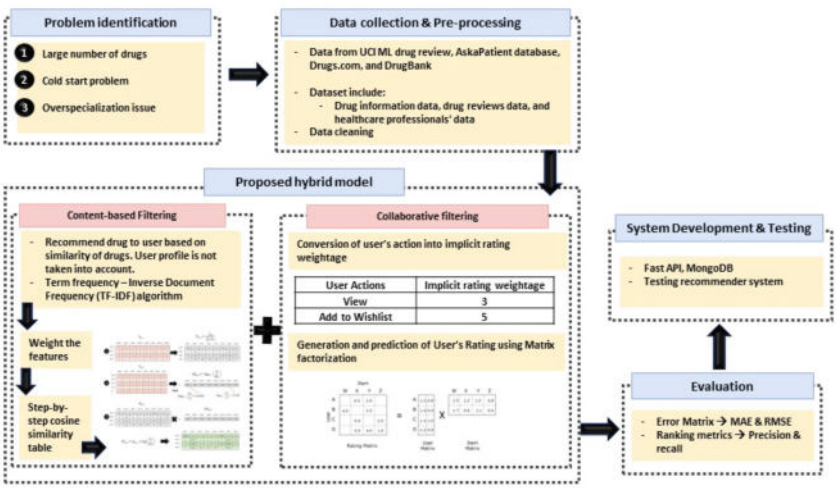


Fig. 5. Hybrid Filtering Algorithm Framework

2.6 Description of Data Source

In this study, two sets of data are required, which are drug reviews data for developing sentiment analysis models and drug details data for the system. Drug review data are obtained from mainly two sources which are AskaPatient database and UCI ML Drug Review dataset. The dataset for UCL ML Drug Review has a total of 232000 reviews for different drugs for different conditions. This dataset has attributes like drug name, health condition, drug review, date of review, drug rating and useful count. The dataset from AskaPatient contains drug review as well, but the attributes are rating, reason, side effects, comments, sex, age, dosage, and date. This dataset is very huge as it contains thousands of drug reviews for many types of drugs in the market. Both datasets do not have positive or negative labels for the reviews. Therefore, the rating of drugs is used to label the drug review data either positive, neutral, or negative for training the model. Positive reviews are labelled as 1, neutral review as 0, and negative review as −1.

Furthermore, another important dataset for this study is the drug details data from Drugs.com and Drug Bank Online. Both datasets include all the necessary information about drugs, which is extracted and shown to healthcare professionals for their reference. Data from both of these sources are combined together into a single drug dataset to be used for Dr. Drugs website. Examples of the data that can be retrieved from these datasets are generic names, dosing information, drug characteristics, associated conditions, and adverse effects.

3 Results and Analysis

3.1 Sentiment Analysis Model

The sentiment analysis model is developed to classify drug reviews of patients into positive, neutral, and negative classes. There are several steps involved in developing this model. The first step is data retrieval from the selected databases. For the sentiment analysis model, only review and rating attributes were used. The preprocessing tasks that were performed on the reviews are stop words removal, special characters removal, removing white spaces, transforming text to lower case, and stemming. A machine learning approach is employed to create this model, necessitating a labeled dataset. However, the existing dataset lacked labels for every review, although it did contain rating information for each entry. The ratings data were transformed into labels as follows: Ratings higher than 7 were classified as positive and given a label of 1. Ratings lower than 4 were classified as negative and assigned a label of −1. Ratings falling between 4 to 7 were considered as neutral and labeled as 0. Table 1 shows ratings data with their respective labels.

To handle the large dataset, it was downsized to contain 60,000 rows of data only. Out of 60,000 rows, 20,000 rows were labelled as Positive, another 20,000 as Neutral, and the remaining 20,000 as Negative. This selection ensured a balanced representation of different sentiment categories. Subsequently, only this reduced dataset was utilized for the subsequent modelling tasks. Then, the dataset was split into two parts: 80% for training data and 20% for testing data. This division allowed for the model to be trained on a majority of the data while preserving a separate subset for evaluating its

Table 1. Drug Ratings and Labels

Rating	Label
8–10	Positive
4–7	Neutral
1–3	Negative

performance. After the completion of data splitting, the feature extraction process was applied to both the training and testing datasets. Two distinct algorithms were employed for this purpose: TF-IDF and Bio BERT. Bio BERT is an adapted version of the BERT algorithm designed specifically for biomedical text analysis. It undergoes pre-training on a vast collection of biomedical literature and clinical text to acquire a deep understanding of word and sentence contexts. By learning from this extensive corpus, Bio BERT gains the ability to generate contextualized representations of biomedical terms and phrases. Once the dataset has been transformed into vectors, the next step involves training the sentiment analysis models. For this task, three distinct algorithms were employed: Logistic Regression, Perceptron, and LSTM (Long Short-Term Memory). A total of six sets of models were developed for sentiment analysis. These models were created by combining two feature extraction algorithms with the three aforementioned training algorithms. This approach allowed for a comprehensive exploration of various combinations, enabling the comparison and evaluation of the different feature extraction techniques and training algorithms for sentiment analysis. Grid search was applied to determine the optimal parameters for training each model across all the training algorithms. The resulting models were then evaluated, and their accuracy, macro-average f1-score, and macro-average precision scores were recorded. Table 2 summarizes and compares the performance metrics of the trained models.

Table 2. Performance Comparison for Different Sentiment Analysis Algorithms

Algorithm	Bio Bert			TF-IDF		
	Accuracy	F1-Score	Precision	Accuracy	F1-Score	Precision
Logistic Regression	0.5719	0.5701	0.5708	0.7441	0.7449	0.7444
Perceptron	0.5452	0.5499	0.5252	0.7274	0.7261	0.7264
LSTM	0.5651	0.5659	0.5677	0.7319	0.7310	0.7314

Based on the results shown in Table 2, it is evident that the models utilizing the TF-IDF vectorizer for feature extraction generally outperformed the Bio BERT-based models. Among the TF-IDF models, the Logistic Regression algorithm demonstrated the highest performance compared to Perceptron and LSTM. It achieved an accuracy of 0.7441, an F1-score of 0.7449, and a Precision of 0.7444 respectively. Based on these findings, to develop an effective sentiment analysis model, the TF-IDF algorithm

is chosen for feature extraction, while Logistic Regression is selected as the preferred model training algorithm.

The best trained sentiment analysis model and vectorizer are converted into.pkl files and incorporated into the backend, which utilizes FastAPI. This model serves two purposes within the system. Firstly, when an administrator adds a new drug to the system, all the accompanying reviews uploaded for that drug are automatically classified according to their sentiment using the model. This allows for the categorization of reviews as positive, neutral, or negative. Secondly, healthcare professionals are given the option to contribute new reviews for any specific drug, which are also subjected to sentiment classification using the trained model. Before prediction, each review undergoes preprocessing similar to the data cleaning process applied to the dataset. By employing this model, the system ensures that all newly added reviews, whether uploaded by administrators or healthcare professionals, undergo sentiment analysis to provide valuable insights into the sentiment associated with specific drugs. This enables the assignment of a sentiment rating to each drug. A sample interface for the sentiment analysis is illustrated in Fig. 6.

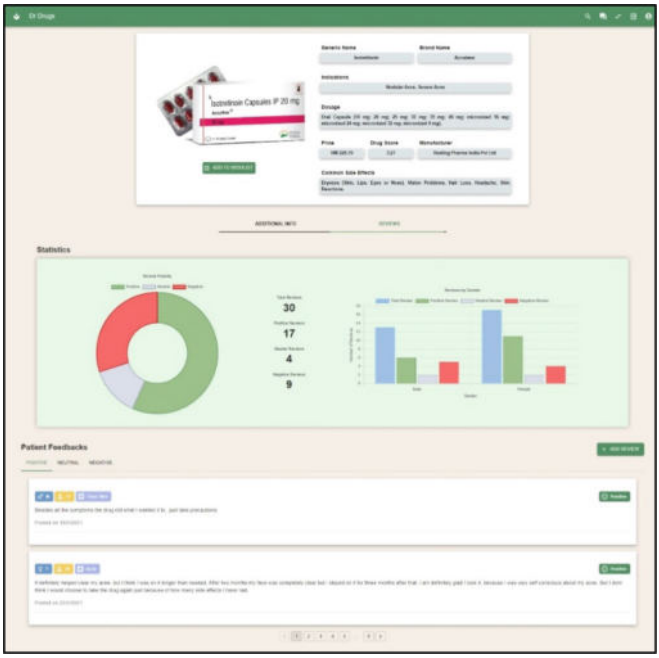


Fig. 6. Drug Details Page (Review Section)

3.2 Content-Based Filtering Model

To provide drug recommendations based on similarity, the system incorporates a content-based filtering model. This model analyses the characteristics of drugs to identify similarities and make recommendations. When a user views a specific drug, the model suggests other drugs that are similar to the one being viewed. The development of this model involves several steps, starting with the acquisition of dataset containing drug information and details. The drugs data used for this model includes the following specific attributes: Generic Name, Brand Name, Common Side Effects, Description, Dosage, Food Interaction, Indications, Ingredients, Manufacturer, Price, Rating, and Things to Avoid. All these attributes are important to determine the similarity between drugs. After reading the data, the data undergoes several text pre-processing steps. Since all data is in string format, the text pre-processing steps include conversion to lower case, removal of stop words, removal of special characters and lemmatization. After these pre-processing steps, the individual attributes of each drug are joined together to create a corpus.

Each drug in the database has its own corpus consisting of its attributes. Then the TF-IDF algorithm is used to perform feature extraction on these corpuses. All corpora are converted to vectors. In order to find similar drugs for a given drug, attributes of that drug undergo text preprocessing, and are then combined to create a corpus as well. This corpus is subsequently vectorized using the TF-IDF algorithm, transforming it into a numerical representation. After completing these steps, the similarity between drugs is determined. The Cosine Similarity algorithm is employed for this purpose. Both the vectors of drugs in the database and the vector of the drug being viewed are fitted into the Cosine Similarity algorithm. The corpus of the drug being viewed is compared with the corpuses of all other drugs in the database and the similarity scores of each of the drugs is computed. These drugs are then sorted based on ranking using the argsort() function depending on the similarity score. The indices of the top 5 drugs are retrieved based on the similarity scores. Using these indices, the corresponding drugs are retrieved from the database. Essentially, this approach constitutes a content-based recommender system, where drugs being viewed by users are compared with the rest of the drugs in the database. The top five most similar drugs are recommended to the user based on this comparison.

During the implementation of the model into the system, an initial step involves generating TF-IDF vectors for all the drugs beforehand. The resulting TF-IDF vectors for all drugs are then stored in a.pkl file for easy retrieval. When a user interacts with the system's user interface and selects a specific drug, the TF-IDF vectors of all drugs in the database are loaded from the.pkl file. These vectors, along with the vector representation of the drug being viewed, are then fitted to the cosine similarity algorithm. Finally, the recommended drugs are retrieved from the model and displayed to the user. Whenever a new drug is added to the database, the TF-IDF vectors associated with the drug data are regenerated. Subsequently, the existing.pkl file is updated to include these new vectors. Figure 7 and 8 depict the user interface of the content-based filtering model.

As shown in Fig. 7, the focus is on the drug 'Eletriptan', commonly used for migraines. The top three recommended drugs are 'Rizatriptan,' 'Sumatriptan,' and 'Venlafaxine,' which are also used for migraine treatment. The remaining two drugs are used for different purposes but are the most similar options available to 'Eletriptan' within

the limited database of only four drugs for migraine treatment. Another example is demonstrated in Fig. 8, where the drug under consideration is 'Cephalexin,' utilized for treating 'bladder infection.' The top two recommended drugs are 'Nitrofurantoin' and 'Ciprofloxacin,' which are also commonly used for 'bladder infection.' As there are only three drugs in the database for 'bladder infection,' the remaining displayed drugs represent the closest alternatives compared to the others in the database. This successful outcome illustrates the effective functionality of the content-based filtering model.

Fig. 7. Content Based Filtering Model Recommended Drugs UI for "Eletriptan"

Fig. 8. Content Based Filtering Model Recommended Drugs UI for "Cephalexin"

3.3 Collaborative Filtering Model

The collaborative filtering model has been developed to recommend drugs to users based on their similarity to other users. Data for this model is collected directly from the system, capturing user behavior such as drug views, reviews, and additions to the wish list. These interactions, which can be considered as user ratings, are stored in the database and used to construct the collaborative filtering model. The initial steps involve loading the user behavior data and drug data, followed by preprocessing, and encoding. To ensure a comprehensive set of drugs is considered for recommendations, the unique drugs from both datasets are combined. A Label Encoder is then utilized to assign numeric IDs to each drug, facilitating efficient processing of categorical drug names. The core element of the collaborative filtering approach is the user-item matrix. An empty matrix is created to store user ratings for drugs. The user interactions with drugs are assigned different weights: adding drugs to the wish list is given a weight of 2.0, viewing a drug is given a weight of 1.0, and reviewing a drug is given a weight of 3.0. These weightings reflect the importance placed on each user's behavior. Using the drug encodings and assigned weights, the user ratings are populated in the user item matrix. Some drugs may have missing ratings, indicating that no users have rated those specific drugs. Handling missing ratings is crucial to address the sparsity of the user-item matrix. If a drug has not been rated by any user, its rating is replaced with the average rating of all drugs. This step ensures that the model can provide reasonable recommendations for unrated drugs based on collective user behavior.

Once the user-item matrix is prepared, it undergoes transformation using Singular Value Decomposition (SVD), which is a dimensionality reduction technique. The TruncatedSVD class from scikit-learn is employed to perform SVD on the user-item matrix, allowing for lower-dimensional representations. After transforming the matrix, the next step involves calculating the similarity between users. Cosine similarity is used to measure the similarity between the transformed user-item matrix representations of different users. This similarity matrix captures the extent of similarity between users' behaviors in drug ratings. Moving on to the recommendation process, similarity scores between the user in need of recommendations and other users in the system are calculated. These scores are utilized to identify similar users based on their similarity to the target user. Drug ratings from similar users are collected, forming the basis for generating recommendations for the target user. Recommendations are based on drugs that similar users have rated, but the target user has not yet rated. Finally, the inverse transform function is applied to the encoder to retrieve the recommended drug names based on their IDs. The code snippet for the recommendation process is shown below.

The collaborative filtering model is seamlessly integrated into the system, streamlining the data loading and user item matrix transformation steps. The user item matrix is transformed and stored in a.pkl file for efficient retrieval. When a recommendation is requested for a particular user, only the user index is needed to retrieve their similarity scores. Using these similarity scores, the model identifies users who exhibit similar preferences and behaviors. It then collects drug ratings from these similar users, which are aggregated to determine the overall ratings for each drug. Based on these aggregated ratings, the drugs are sorted to prioritize the most highly recommended ones. The updated recommendations are displayed to the user in the main page of the system, specifically

in a dedicated section titled "Recommended Drugs for You," as depicted in Fig. 9. Each user in the system will receive a personalized set of recommended drugs based on their individual behaviors. For newly registered users, recommendations would not be available initially. However, once the model is trained again and they start to interact with the system, they start receiving personalized recommendations. It is important to note that recommended drugs for similar users dynamically adapt over time, reflecting their own behaviors as well as those of other users within the system.

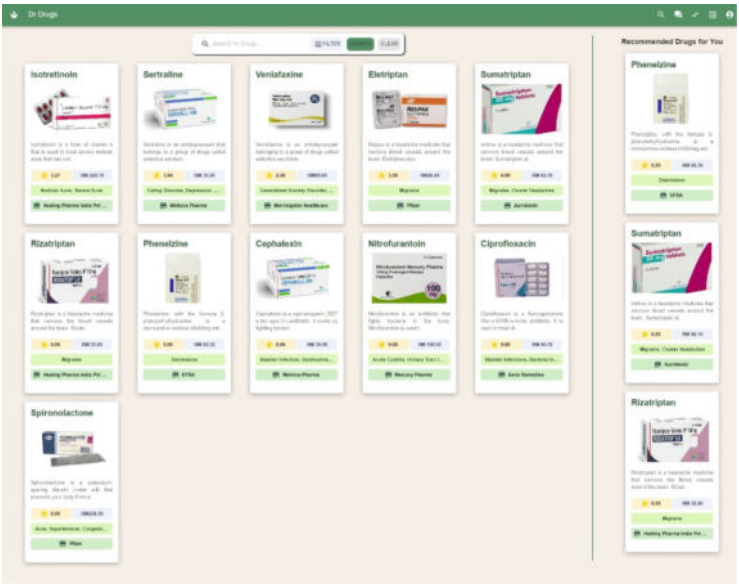


Fig. 9. Collaborative Filtering Model Recommended Drugs UI

4 Discussion and Conclusion

The goal of this study was to develop a drug recommendation system for healthcare professionals to aid them in selecting the best drug for their patients. For this reason, sentiment analysis and hybrid content-based and collaborative filtering algorithms were implemented. Currently, there is no other recommender system developed for the usage of healthcare professionals, which makes the system unique and very important. All three objectives of the study are fulfilled, which are to classify drug reviews based on sentiment using sentiment analysis, to assist health professionals better understand drug effectiveness, and to develop a web-based drug recommendation system for healthcare professionals using hybrid filtering technique and to facilitate active involvement and knowledge sharing among healthcare professionals. The system was successfully implemented and tested.

The implementation strategy adopted for this study follows a bottom-up approach. This means that the entire system is divided into smaller sub-parts, which are built and

tested independently before being integrated into a complete system. Each module represents a distinct feature of the system and developing them separately offers several benefits. By developing the modules independently, the possibility of errors is decreased, and the complexity of the entire system is effectively managed. This system's development was carried out methodically, with a focus on finishing, testing, and debugging every component before integrating it into the main application. The first step was the development of the main user interface (UI).

Three machine learning models were then developed and integrated into the system, with thorough testing carried out at each step. This methodical integration process made certain that each model's usability and functionality were carefully evaluated before continuing. In addition, additional features were developed individually so they could be tested and improved before being integrated into the overall system. The main benefit of this bottom-up approach was that it reduced the risks involved in developing a large, complex system. Any concerns or issues that arose may be readily addressed and handled by breaking it down into smaller, manageable components. Throughout the development process, we ensure that each module matched the required criteria for performance and usability by frequent testing and debugging. This approach not only enhanced the overall functionality of the system but also allowed for better flexibility in adjusting and refining individual components before integrating them into the overall system.

The first part of the system development includes the development of three different intelligent computing models. The first model is a sentiment analysis model which is developed using Bio Bert, TF-IDF, Logistic Regression, Perceptron and LSTM algorithms. This model was used to classify each drug review according to polarity positive, neural, and negative. The second model is content-based filtering model which was developed using TF-IDF algorithm and cosine similarity algorithm. This model is used to recommend similar drugs to users based on the current drug they are viewing. The final model is collaborative filtering algorithm, which was developed using cosine similarity algorithm and Singular Value Decomposition (SVD) algorithm. This model was used to recommend drugs to users based on similar users which was determined using user behavior in the system. All the three developed models were working well without any issues, which meant that the first and second objectives were met.

The second part of system development focused on a web application called Dr.Drugs. The web application was developed using FARM (FastAPI, React, MongoDB) stack and Material UI library. A three-layer architecture was used to develop the system. Overall, the developed system has achieved the requirements planned during the system requirement and analysis phase. The development methodology used to develop this system is a bottom-up approach where all the small components of modules were developed first and then integrated into a bigger module. The entire system was tested in terms of unit testing, integration testing and system testing, and it can be said that the system passed all the tests without any issues and the expected results were achieved. This indicates that the developed web application is a fully functional website. In the web application, additional features like drug comparison, forums and adding drugs to wish list were added. These features are added to give more assistance for users when selecting a drug. The forums feature helps to achieve the third objective. This feature

helps healthcare professionals to communicate with each other, expand their knowledge and expand drug knowledge.

Unit testing, integration testing, and system testing were done to ensure the proposed system is free of bugs and functions efficiently as intended. Based on the test results, it can be concluded that the proposed system has met the expected requirements that were established during the system requirement and analysis phase. The system is able to function under different input conditions without any major errors. This is due to the fact that unit testing and integration testing was done very carefully for every module. Thus, Dr.Drugs web application is a fully functionally web application that meets its objectives and requirements.

In terms of strengths, the system's unique features of content-based and collaborative filtering algorithms recommendation are its primary strength. There are limited recommender systems focused for healthcare professionals on recommending drugs [12]. The collaborative model is able to recommend personalized drugs to users based on their user behavior while the content-based filtering model is able to recommend similar drugs based on drugs being viewed accurately. Moreover, sentiment analysis is another strength of the system. Drug reviews are usually not utilized to the fullest potential. By implementing sentiment analysis on this system, healthcare professionals are able to understand drug effectiveness in various patients. This is another unique feature of the proposed system. Besides that, the system also includes a forum page feature. This is implemented to facilitate communication between healthcare professionals. By doing so, healthcare professionals will be able to share drug knowledge with each other.

In terms of limitations, the system does not provide specific information based on the healthcare professional's specialization. Thus, personalized recommendations based on healthcare professionals' specialization could not be made. Moreover, another limitation is drug data added to the system could not be edited. Once admin have added data to the system, they are only able to delete the drug data do not edit it. Additionally, when adding drug data, admin can upload reviews, but the file must be in csv file in a specific format. If not, review data would not be added correctly.

As future work, there are some suggestions that can be made to the system to improve it even further. The first suggestion is implementing a personalized, patient-centric, or disease specific drug recommendation. Currently, the system recommends drugs based on similar drugs and similar user behavior. This recommendation can be improved to include the patient or disease-based drug recommendation as well. This will assist healthcare professionals to select drugs that are specific to a patient as different patient will have different reaction or effectiveness with different drug.

Moreover, another feature that can be included is the drug availability feature. This feature will allow healthcare professionals to know the location at which they can purchase the drugs they need. The system needs to show all the pharmacy locations that are currently selling a particular drug.

Disclosure of Interests. There is no conflict of interest.

References

1. Keikhosrokiani, P.: Perspectives in the Development of Mobile Medical Information Systems: Life Cycle, Management, Methodological Approach and Application (2019)
2. Keikhosrokiani, P., Kamaruddin, N.S.A.B.: IoT-based in-hospital-in-home heart disease remote monitoring system with machine learning features for decision making. In: Mishra, S., González-Briones, A., Bhoi, A.K., Mallick, P.K., Corchado, J.M. (eds.) Connected e-Health. Studies in Computational Intelligence, vol. 1021, pp. 349–369. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-97929-4_16
3. Augustine, C.A., Keikhosrokiani, P.: A hospital information management system with habit-change features and medial analytical support for decision making. Int. J. Inf. Technol. Syst. Approach (IJITSA) **15**, 1–24 (2022). https://doi.org/10.4018/IJITSA.307019
4. Keikhosrokiani, P., Mustaffa, N., Zakaria, N.: Success factors in developing iHeart as a patient-centric healthcare system: a multi-group analysis. Telematics Inform. **35**, 753–775 (2018). https://doi.org/10.1016/j.tele.2017.11.006
5. Keikhosrokiani, P., Mustaffa, N., Zakaria, N., Abdullah, R.: Assessment of a medical information system: the mediating role of use and user satisfaction on the success of human interaction with the mobile healthcare system (iHeart). Cogn. Technol. Work **22**, 281–305 (2020). https://doi.org/10.1007/s10111-019-00565-4
6. Jinjri, W.M., Keikhosrokiani, P., Abdullah, N.L.: Machine learning algorithms for the classification of cardiovascular disease- a comparative study. In: Proceedings of the 2021 International Conference on Information Technology, ICIT 2021 (2021)
7. Silvén, H., Savukoski, S.M., Pesonen, P., et al.: Association of genetic disorders and congenital malformations with premature ovarian insufficiency: a nationwide register-based study. Hum. Reprod. **38**, 1224–1230 (2023). https://doi.org/10.1093/humrep/dead066
8. Garg, S.: Drug recommendation system based on sentiment analysis of drug reviews using machine learning. In: 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 175–181 (2021)
9. Zhao, X., Keikhosrokiani, P.: Sales prediction and product recommendation model through user behavior analytics. Comput. Mater. Continua **70**, 3855–3874 (2022). https://doi.org/10.32604/cmc.2022.019750
10. Xian, Z., Keikhosrokiani, P., XinYing, C., Li, Z.: An RFM model using k-means clustering to improve customer segmentation and product recommendation. In: Keikhosrokiani, P. (ed.) Handbook of Research on Consumer Behavior Change and Data Analytics in the Socio-Digital Era, pp. 124–145. IGI Global, Hershey, PA, USA (2022)
11. Bhat, S., Aishwarya, K.: Item-based hybrid recommender system for newly marketed pharmaceutical drugs. In: 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp 2107–2111 (2013)
12. Keikhosrokiani, P., Fye, G.M.: A hybrid recommender system for health supplement e-commerce based on customer data implicit ratings. Multimed. Tools Appl. (2023). https://doi.org/10.1007/s11042-023-17321-6
13. Cavalcanti, D., Prudêncio, R.: Aspect-based opinion mining in drug reviews. In: Oliveira, E., Gama, J., Vale, Z., Cardoso, H.L. (eds.) Progress in Artificial Intelligence, pp. 815–827. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-65340-2_66
14. Chugh, M., Chugh, N.: A deep drive into software development agile methodologies for software quality assurance. In: Agile Software Development, pp. 235–255 (2023)
15. Kayanda, A., Busagala, L., Oyelere, S., Tedre, M.: The use of design science and agile methodologies for improved information systems in the Tanzanian Higher Education context. Electron. J. Inf. Syst. Dev. Countries **89**, e12241 (2023). https://doi.org/10.1002/isd2.12241

16. Keikhosrokiani, P.: Big Data Analytics for Healthcare: Datasets, Techniques, Life Cycles, Management, and Applications (2022)
17. Keikhosrokiani, P.: Handbook of research on consumer behavior change and data analytics in the socio-digital era (2022)
18. Keikhosrokiani, P., Asl, M.P.: Handbook of Research on Artificial Intelligence Applications in Literary Works and Social Media (2022)
19. Suhendra, N.H.B., Keikhosrokiani, P., Asl, M.P., Zhao, X.: Opinion mining and text analytics of literary reader responses: a case study of reader responses to KL noir volumes in goodreads using sentiment analysis and topic. In: Keikhosrokiani, P., Pourya, A.M. (eds.) Handbook of Research on Opinion Mining and Text Analytics on Literary Works and Social Media, pp. 191–239. IGI Global, Hershey, PA, USA (2022)
20. Keikhosrokiani, P., Asl, M.P.: Handbook of Research on Opinion Mining and Text Analytics on Literary Works and Social Media. IGI Global (2022)

Enhancing Arrhythmia Diagnosis with Data-Driven Methods: A 12-Lead ECG-Based Explainable AI Model

Emmanuel C. Chukwu[1,2] and Pedro A. Moreno-Sánchez[1(✉)]

[1] Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
pedro.morenosanchez@tuni.fi
[2] Eindhoven University of Technology, Eindhoven, Netherlands

Abstract. Accurate and early prediction of arrhythmias using Electrocardiograms (ECG) presents significant challenges due to the non-stationary nature of ECG signals and inter-patient variability, posing difficulties even for seasoned cardiologists. Deep Learning (DL) methods offer precision in identifying diagnostic ECG patterns for arrhythmias, yet they often lack the transparency needed for clinical application, thus hindering their broader adoption in healthcare. This study introduces an explainable DL-based prediction model using ECG signals to classify nine distinct arrhythmia categories. We evaluated various DL architectures, including ResNet, DenseNet, and VGG16, using raw ECG data. The ResNet34 model emerged as the most effective, achieving an Area Under the Receiver Operating Characteristic (AUROC) of 0.98 and an F1-score of 0.826. Additionally, we explored a hybrid approach that combines raw ECG signals with Heart Rate Variability (HRV) features. Our explainability analysis, utilizing the SHAP technique, identifies the most influential ECG leads for each arrhythmia type and pinpoints critical signal segments for individual disease prediction. This study emphasizes the importance of explainability in arrhythmia prediction models, a critical aspect often overlooked in current research, and highlights its potential to enhance model acceptance and utility in clinical settings.

Keywords: Arrhythmia · ECG · prediction model · deep learning · heart rate variability · explainable AI

1 Introduction

Cardiovascular disease (CVD) is the leading cause of death in Europe and the EEUU, causing 3.9 million and 1.8 million deaths annually [1]. Traditional CVD diagnosis relies on rule-based evaluation of patient history and clinical examinations. This approach struggles with the volume and diversity of data and depends heavily on medical expertise, leading to challenges in resource-limited settings like developing countries.

The electrocardiogram (ECG) is a key, non-invasive tool for diagnosing cardiac conditions, utilizing a 12-lead setup to capture heart's electrical activity through distinct P, Q, R, S, and T waves [2]. While ECG, especially in identifying cardiac arrhythmias,

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 242–259, 2024.
https://doi.org/10.1007/978-3-031-59091-7_16

is straightforward, interpreting these signals, particularly in complex cases, remains challenging and prone to errors with serious implications [3]. Additionally, Heart Rate Variability (HRV) analysis, which examines variations in consecutive heartbeats, has emerged as a crucial technique in cardiac assessment. It assesses the autonomic nervous system's impact on the heart by analyzing the R-R interval, the time between successive R wave peaks, and the N-N interval, the duration between consecutive QRS complexes. These measures help in understanding the cardiac system's dynamic state [4].

Arrhythmias, a common and varied group of CVDs diagnosed using ECG, are characterized by irregular heartbeats due to improper electrical signaling, leading to abnormally fast, slow, or inconsistent heart rhythms. This work focuses on several arrhythmia classes including Atrial fibrillation (AF), Right and Left Bundle Branch Blocks (RBBB and LBBB), First-degree atrioventricular block (IAVB), and Premature Atrial and Ventricular Contractions (PAC and PVC), along with Myocardial Infarction (MI) [5]. Diagnosing arrhythmias is challenging due to: i) absence of symptoms during ECG recording; ii) high inter-patient ECG signal variability; iii) non-stationary signal morphology affected by physical state, noise, and artifacts; and iv) the need for large data volumes to avoid false diagnoses [5].

Computer-Aided Diagnosis Systems (CADS) address arrhythmia diagnosis challenges by leveraging digital technologies for the analysis of physiological and clinical data, aiding clinicians in making more informed decisions. Traditional ECG analysis techniques in CADS rely on automated detection of ECG components and classifying them based on fixed rules, but they often fall short due to outdated rules and sensitivity to imperfect ECG recordings. In the medical field, Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL), has significantly enhanced CADS. AI combines mathematical and computer science theories to create systems capable of intelligent actions, with DL being notable for its ability to process large volumes of data through artificial neural networks. These networks perform sequential transformations to highlight crucial input features for classification and regression tasks. Modern arrhythmia diagnosis models increasingly use DL, credited for its precision in identifying ECG waveforms like QRS complexes, and P and T waves, facilitating the calculation of vital clinical measures including heart rate and axis deviation [6, 7].

AI's potential in various fields, including healthcare, is often hindered by its 'black box' nature, leading to trust issues due to a lack of transparency [8]. Healthcare professionals need to understand the reasoning behind AI-recommended treatments. Without this level of explainability, AI's adoption in healthcare can be negatively impacted. Explainable AI (XAI) addresses this challenge by providing insights into the decision-making processes of AI systems. XAI aims to make the logic behind AI algorithms clear, thereby aligning advanced AI capabilities with the healthcare sector's need for transparent decision-making.

In recent literature, advancements in arrhythmia classification using DL have been notable, especially with the application of Convolutional Neural Networks (CNNs). A diverse range of studies has utilized various DL architectures, showing significant progress in ECG signal analysis. Chen et al. [9] combined a CNN with RestNet-34 layers and bi-directional LSTM, achieving an accuracy of 0.81 on 12-lead ECG samples. Their study, while promising, highlighted the need for balanced datasets. Cheng

et al. [10] used a modified 1-D CNN on the MIT-BIH arrhythmia database, focusing on compressed ECG signals suitable for wearable devices. Gao et al. [11]'s approach involved a 4-layer LSTM model, achieving an accuracy of 0.992, demonstrating robustness against noise and normal ECG beat dominance. Niu et al. [12] introduced a novel DL method based on adversarial domain adaptation, while Romdhane et al. [13] focused on enhancing minority class classification accuracy. Wang et al. [14, 15] employed 1-D CNNs and continuous wavelet transform techniques, demonstrating effectiveness against noise. Yildirim et al. [16] developed a 1D-CNN suitable for mobile and cloud computing applications due to its efficiency. Zhang [17] and Zhang et al. [18] used 1D-CNN networks, showing the superiority of 12-lead over single-lead ECGs. Rai et al. [19] tested a CNN + LSTM ensemble approach, improving minority class accuracy. Finally, Toma et al. [20] presented a parallel approach combining RNN and 2D CNN, effectively capturing temporal and spatial ECG signal characteristics. These studies illustrate the advancements and diversity in DL applications for arrhythmia detection, with a trend towards optimizing network architectures and inputs for enhanced classification accuracy. However, the integration of explainable AI (XAI) remains a largely unexplored area, crucial for the clinical applicability of these models.

In our study, we developed an explainable model for detecting cardiac arrhythmias using 12-lead ECG signals. We explored two approaches: one using raw ECG signals as input for various DL architectures like ResNet, VGG, and DenseNet, and another combining raw ECG signals with HRV features in a hybrid model. We thoroughly evaluated the performance of these classifiers. Furthermore, we assessed the explainability of the most effective model by analyzing the significance of different leads in arrhythmia classification and presenting case examples that illustrate the ECG signal segments influencing the predictions.

The remainder of this paper is organized as follows: Section 2 outlines the dataset, the DL algorithms, training and testing methodologies, and the explainability technique used in building the arrhythmia detector. Section 3 details the evaluation results, including the predictive performance of the DL approaches (using raw ECG and the hybrid method with HRV features) and the explainability analysis of the most effective model. Section 4 discusses these results, and Sect. 5 concludes the paper with key findings.

2 Material and Methods

2.1 Dataset Description

In our research, we utilized the 'China Physiological Signal Challenge 2018 (CPSC 2018)' dataset to investigate arrhythmia predictions [21]. CPSC 2018 is an extensive dataset that was collected and curated to facilitate research within the domain of physiological signal processing to encourage the development of algorithms for the detection of morphological abnormalities. This dataset also comprises 12 lead ECG recordings and was sourced through collaboration with 11 hospitals in China. CPSC 2018 includes a total of 6,877 individual data samples, with a gender distribution of 3,178 females and 3,699 males. The ECG recordings are sampled at a frequency of 500 Hz, and they vary in length, ranging from 6 to 60 s. Within the CPSC 2018 dataset, researchers have access to ECG recordings representing nine distinct cardiac states, including atrial fibrillation

(AF) with 1098 recordings, intrinsic paroxysmal atrioventricular block (I-AVB) with 704 recordings, left bundle branch block (LBBB) with 207 recordings, normal heartbeat (SNR) with 918 recordings, premature atrial contraction (PAC) with 574, premature ventricular contraction (PVC) with 653 recordings, right bundle branch block (RBBB) with 1695 recordings, ST-segment depression (STD) with 826 recordings, and ST-segment elevation (STE) with 202 recordings. It's worth noting that among the 6,877 recordings, 476 of them have two or three different labels, indicating their complexity.

2.2 HRV Features Extraction

In this study, we computed 33 Heart Rate Variability (HRV) features from each subject's entire ECG using the pyHRV Python library [22] and BioSPPy toolbox for biosignal processing [23]. BioSPPy's ECG processing and R-peak detection algorithms enabled us to calculate the Normal-to-Normal Interval (NNI) series, from which we extracted HRV features covering time-domain, frequency-domain, and non-linear parameters.

The HRV features included: maximum and minimum NNI, standard deviation (SD) of heart rate (HR), maximum and minimum HR, mean HR, root mean square of NNI difference, number of NN intervals differing by more than 20 ms and 50 ms, ratios of NN20 and NN50 to total NNI, SD1 and SD2 (standard deviations of the major and minor axes), ratio of SD1 to SD2, maximum and minimum NNI difference, mean NNI difference, sample entropy, area S of the fitted ellipse, fast Fourier transform (FFT) metrics, number of NNI, TINN (baseline width of the interpolated triangle) computation values, triangular index, and AR (autoregression) metrics.

2.3 Deep Learning Algorithms

Deep learning (DL) algorithms have revolutionized AI, particularly in healthcare, by outperforming traditional machine learning methods in complex tasks. In this subsection, we explore the DL algorithms applied in our study, detailing their architecture, training strategies, and their specific use in arrhythmia prediction. For a comprehensive comparative analysis, we utilized models such as ResNet34, ResNet50, VGG16, and DenseNet, each chosen for their distinct strengths in deep learning applications. These experiments utilize a learning rate of 0.0001, the Adam optimizer, and Binary Cross Entropy (BCE) with Logits Loss as loss function.

ResNet34 and ResNet50, part of the Residual Neural Network family by He et al. [24], are designed primarily for image recognition tasks. ResNet50 has a deeper structure with 50 layers, compared to the 34 layers of ResNet34. Both architectures follow a similar design, featuring convolutional, pooling, activation, and fully connected layers to extract detailed features. A key innovation in ResNet is the introduction of residual learning to tackle the vanishing or exploding gradient problem in deep networks. This is achieved through skip connections, enabling the network to learn residual functions and maintain performance in deeper layers.

In our study, we implemented both ResNet34 and ResNet50 for arrhythmia prediction. These architectures, chosen for their proven efficacy across various tasks, are particularly well-suited for this purpose. Their ability to handle intricate features makes them ideal for our ECG analysis, which uses 1D CNNs to process raw ECG data (12 leads, 30-s recordings at 500 Hz). This approach, leveraging 1D CNNs' effectiveness in time-series data [25], allows for detailed feature extraction and classification across nine diagnostic categories. To mitigate overfitting, dropout regularization with a probability of 0.2 was implemented after the initial convolutional layer's activation function, stochastically zeroing activations. By using these architectures, we aim to enhance our understanding of model performance and specialization in ECG classification.

The VGG16 model, a 16-layer convolutional neural network developed by the Visual Graphics Group at the University of Oxford [26], is also used in our study. Renowned for its simplicity and effectiveness in image recognition, we adapted VGG16 as a feature extractor for ECG signals. Its architecture, composed predominantly of convolutional and max-pooling layers, is adept at learning complex hierarchical features from raw ECG data. In our research, VGG16 excelled in identifying detailed features in the ECG signals, crucial for high-level cardiac state classification. By processing 12-lead ECG signals through its 1D CNN layers, VGG16 transformed them into structured representations for downstream classification tasks. This allowed us to detect intricate patterns and categorize data effectively.

DenseNet, or Densely Connected Convolutional Networks, is a cutting-edge CNN architecture employed in our ECG classification framework. It addresses deep neural network challenges like information propagation and feature reuse by introducing dense connections between layers for improved information flow and gradient propagation [27]. In contrast to traditional CNNs with sequential layer connections, DenseNet layers receive inputs from all previous layers and pass their feature maps to all subsequent layers in a dense block, enhancing feature reuse and efficient learning of discriminative features. In our study, DenseNet efficiently extracts hierarchical representations from raw ECG signals. Notably, DenseNet is recognized for its parameter efficiency, delivering competitive performance with fewer parameters than other architectures. This efficiency was crucial in our research, allowing for effective feature extraction with reduced computational complexity.

2.4 Training, Testing and Performance Metrics

For consistent and reliable model development in our deep learning training process, we adopted a standardized data split approach using a custom function. Our preprocessed ECG dataset was divided into three subsets: training, validation, and testing, with proportions of 80%, 10%, and 10% respectively. The majority (80%) of the data was allocated for training, ensuring the model's exposure to a wide array of ECG signals and patterns. The validation set, constituting 10%, was used during training to monitor and fine-tune the model's performance, helping to prevent overfitting. The remaining 10% comprised the test set, which was completely unseen during the training and validation phases. This setup allowed for an unbiased assessment of the model's performance on new, unseen ECG samples.

In assessing our models' performance in predicting nine distinct arrhythmia classes, we employed six diverse evaluation metrics for a comprehensive analysis. These metrics included accuracy, recall, precision, F1-score, Area Under the Receiver Operating Curve (AUROC), and the confusion matrix. Together, they provide a holistic view of model effectiveness, measuring not only the accuracy in classifying arrhythmias but also the ability to differentiate between various arrhythmia types and other class instances.

2.5 Explainability AI Techniques

SHAP (SHapley Additive exPlanations), introduced by Lundberg et al. [28], is a key interpretability framework used in our study to elucidate the predictions of complex machine learning models. Utilizing tools referenced in [17], we applied SHAP values to determine the importance of each feature in our 12-lead ECG input data, identifying the most influential leads in our predictive models. SHAP values offer an in-depth analysis of how each feature contributes to individual predictions, providing insights into the model's decision-making process. Additionally, they enable a broader examination of the model's behavior by summarizing feature impacts across all predictions, revealing general patterns and trends in the data. SHAP also facilitates model comparison and selection by evaluating different models based on their feature contributions, aligning with principles from game theory for a fair and mathematically sound attribution of feature importance.

In our study, we utilized the SHAP DeepExplainer class, a specialized tool for interpreting deep learning models with an efficient computational approach. The SHAP DeepExplainer approximates conditional expectations of SHAP values, integrating multiple background samples to summarize the difference between expected model outputs (based on these samples) and the actual model outputs. This method provides a practical way to understand the model's reasoning by comparing its predictions with a baseline derived from the background data.

3 Results

This section details the classification outcomes of two methods examined in our study: DL with raw ECG signal, and a hybrid method combining DL with HRV and raw ECG signals. For each method, we highlight the algorithm with the highest performance, based on AUROC scores, and present its confusion matrix and ROC curve.

Moreover, we delve into the approach that yielded the most accurate classification, emphasizing explainability. This involves analyzing the significance of different ECG leads in identifying arrhythmia classes and providing explanations for individual instances, specifically highlighting the ECG segments that contributed to arrhythmia classification.

3.1 Deep Learning Classification with ECG Raw Signal

The raw ECG signal is used to fit the several deep learning architecture proposed, i.e. ResNet34, ResNet50, DenseNet, VGG16. Their performances are shown, respectively, in Table 1, Table 2, Table 3 and Table 4, where the metrics denote the model's performances for each arrhythmia following the one-versus-the-rest (OVR) approach. The average for each metric is also shown.

Table 1. ResNet34 classifier performance

Arrhythmia category	Accuracy	Precision	Recall	F1-Score	AUC
SNR	0.96	0.77	0.84	0.80	0.98
AF	0.98	0.94	0.93	0.94	0.99
IAVB	0.98	0.96	0.90	0.93	0.99
LBBB	0.99	0.95	0.83	0.88	1.00
RBBB	0.96	0.93	0.94	0.94	0.99
PAC	0.95	0.66	0.71	0.68	0.96
PVC	0.97	0.87	0.86	0.87	0.99
STD	0.96	0.84	0.85	0.85	0.97
STE	0.97	0.63	0.50	0.56	0.95
Average	*0.97*	*0.84*	*0.82*	*0.83*	*0.98*

Table 2. ResNet50 classifier performance

Arrhythmia category	Accuracy	Precision	Recall	F1-Score	AUC
SNR	0.95	0.74	0.82	0.78	0.97
AF	0.98	0.95	0.93	0.94	0.99
IAVB	0.99	0.95	0.92	0.93	1.00
LBBB	0.99	0.78	0.91	0.84	1.00
RBBB	0.96	0.90	0.96	0.93	0.99
PAC	0.95	0.70	0.66	0.68	0.94
PVC	0.97	0.90	0.85	0.87	0.99
STD	0.94	0.69	0.84	0.76	0.98
STE	0.97	0.53	0.67	0.59	0.97
Average	*0.97*	*0.79*	*0.84*	*0.81*	*0.98*

Table 3. DenseNet classifier performance

Arrhythmia category	Accuracy	Precision	Recall	F1-Score	AUC
SNR	0.94	0.73	0.71	0.72	0.97
AF	0.97	0.87	0.97	0.92	0.99
IAVB	0.99	0.96	0.95	0.95	0.99
LBBB	0.99	0.86	0.83	0.84	0.96
RBBB	0.96	0.91	0.96	0.93	0.99
PAC	0.89	0.42	0.67	0.52	0.88
PVC	0.94	0.71	0.67	0.69	0.95
STD	0.96	0.88	0.76	0.82	0.97
STE	0.97	0.54	0.58	0.56	0.95
Average	*0.96*	*0.76*	*0.79*	*0.77*	*0.96*

Table 4. VGG16 classifier performance

Arrhythmia category	Accuracy	Precision	Recall	F1-Score	AUC
SNR	0.95	0.71	0.82	0.76	0.97
AF	0.97	0.90	0.93	0.92	0.99
IAVB	0.97	0.92	0.80	0.86	0.98
LBBB	0.99	0.95	0.78	0.86	0.95
RBBB	0.95	0.86	0.97	0.92	0.98
PAC	0.90	0.42	0.41	0.42	0.82
PVC	0.96	0.90	0.74	0.81	0.96
STD	0.95	0.86	0.66	0.75	0.95
STE	0.97	0.52	0.63	0.57	0.96
Average	*0.96*	*0.78*	*0.75*	*0.76*	*0.95*

Assessing the average AUROC and subsequently the F1-score of the four DL classifiers, ResNet34 emerges as the best performer with an AUROC of 0.98. Thus, the ROC curve and confusion matrix of ResNet34 are shown in Fig. 1 and Fig. 2, respectively.

3.2 Hybrid Approach: Optimal Classifier with ECG Raw Signal and HRV Features

While DL algorithms effectively process raw ECG signals for arrhythmia classification, we propose a hybrid approach that combines these with HRV features to, in theory, enhance their performance, especially in misclassification-prone categories. Illustrated in Fig. 3, our model integrates HRV features into the fully connected layer of the ResNet34

network, aiming to combine raw signal processing with feature-based analysis for better classification accuracy. The fully connected layer is dimensioned to accommodate the flattened convolutional feature representations as well as the auxiliary inputs (HRV features). The integration of auxiliary features into the model's decision-making process is an attempt to enhance its adaptability and performance.

Similar to the approaches using pure DL networks, the classification performance of the hybrid approach is documented in Table 5. Consequently, the ROC curve and confusion matrix of the hybrid approach are shown in Fig. 4 and Fig. 5, respectively. Figure 6 shows a comprehensive comparison of all classifiers' AUROC performance for each of the arrhythmias considered in the diagnosis. By using Kruskal-Wallis test, we confirm that there is a static statistically significant difference in performance among the classifiers ($p = 0.0008 << 0.05$).

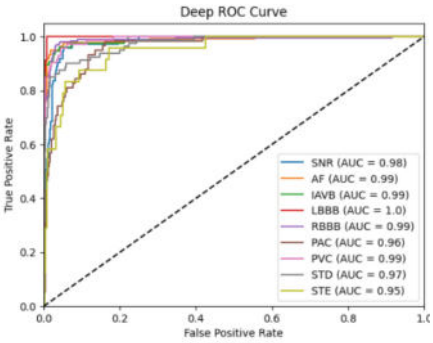


Fig. 1. ROC curve of ResNet34

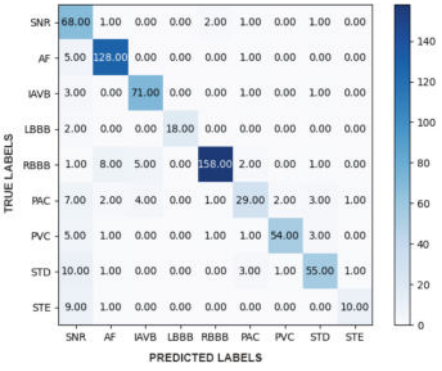


Fig. 2. Confusion Matrix of ResNet34

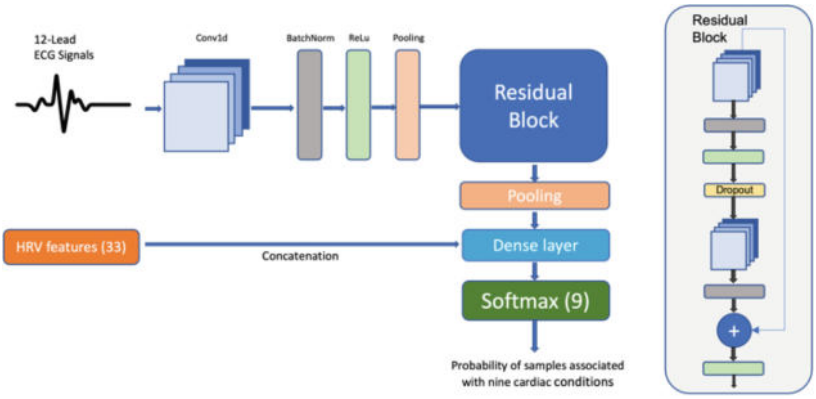


Fig. 3. Hybrid approach architecture

Table 5. Hybrid approach performance

Arrhythmia category	Accuracy	Precision	Recall	F1-Score	AUC
SNR	0.95	0.76	0.75	0.76	0.95
AF	0.19	0.19	1.00	0.33	0.95
IAVB	0.97	0.93	0.82	0.86	0.94
LBBB	0.98	0.68	0.83	0.75	0.90
RBBB	0.28	0.28	1.00	0.44	0.96
PAC	0.93	0.59	0.68	0.63	0.87
PVC	0.96	0.77	0.82	0.79	0.92
STD	0.95	0.81	0.73	0.76	0.92
STE	0.97	0.60	0.50	0.54	0.85
Average	0.79	0.623	0.791	0.65	0.92

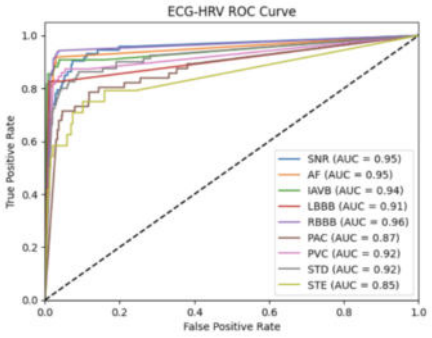


Fig. 4. ROC curve of the hybrid approach

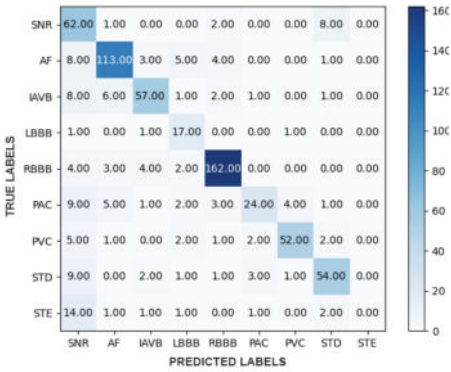


Fig. 5. Confusion Matrix of hybrid approach

3.3 Explainability Analysis of the Optimal Model

According to the objectives of our study, we have conducted an explainability analysis of the best-performing model, specifically ResNet34. For that purpose, the XAI technique SHAP will be used to offer a global and local explainability analysis.

For global explainability, our focus is on determining the significance of each of the 12 ECG leads in predicting different arrhythmia categories. The results of this global explainability analysis are depicted in Fig. 7.

By aggregating the SHAP values for each lead across all predicted instances, we can ascertain the positive or negative influence of each lead on the prediction of a specific arrhythmia. It is important to note, given the multiclass nature of our classification problem, that the negative values observed for each category may not be intuitively interpretable since they could be affected by any of the other category's predictions.

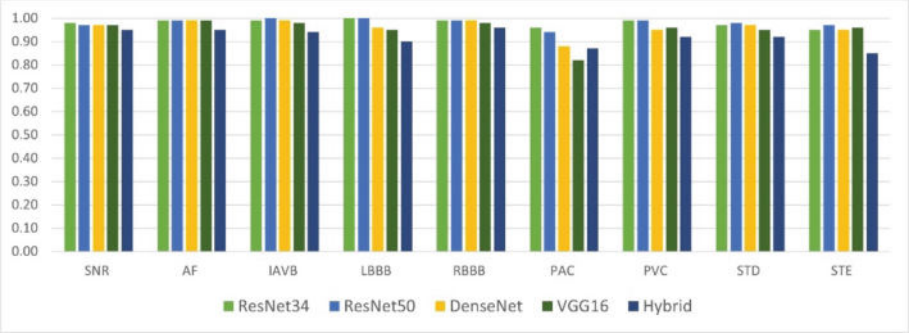


Fig. 6. AUROC comparison of classifiers' performance for each arrhythmias classes

Therefore, by observing the bars with a positive influence we can propose the following Table 6 that indicates the most relevant features for each of the arrhythmia categories. This relevant feature ranking allows us to propose the presentation of an individual explainability approach since SHAP can indicate which region of the ECG signal contributes to the prediction. Therefore, we show in Fig. 8 one instance predicted for each of the arrhythmia categories and ECG segment more relevant for such prediction on the relevant features identified in the global explainability approach.

Table 6. Prominent lead for each cardiac state detection

Arrhythmia category	SNR	AF	IAVB	LBBB	RBBB	PAC	PVC	STD	STE
Relevant ECG leads	II, avR	V2, V3	V2, V5	avR, avF[1]	V2, V4	V2, V3	V4	V1, V2	V1, V2

4 Discussion

ML and DL are increasingly important in CVD detection, leveraging ECGs as key data sources. ECGs are crucial for improving diagnostic accuracy in data-driven predictive models for CVD [29]. These technologies not only facilitate early CVD detection, leading to better health outcomes, but also help address the demand for skilled cardiologists in ECG data analysis. Recent research in clinical cardiology suggests that ML and DL, especially in combination with other methods, provide superior predictive power for cardiovascular or overall mortality compared to traditional clinical or imaging techniques alone [7]. DL's strength lies in its ability to capture temporal signal variations and autonomously learn from complex inputs like ECG signals, provided there's enough high-quality training data. This learning bypasses the need for predefined feature processing, offering an end-to-end solution that minimizes errors in feature calculation, thereby enhancing model accuracy [30–34].

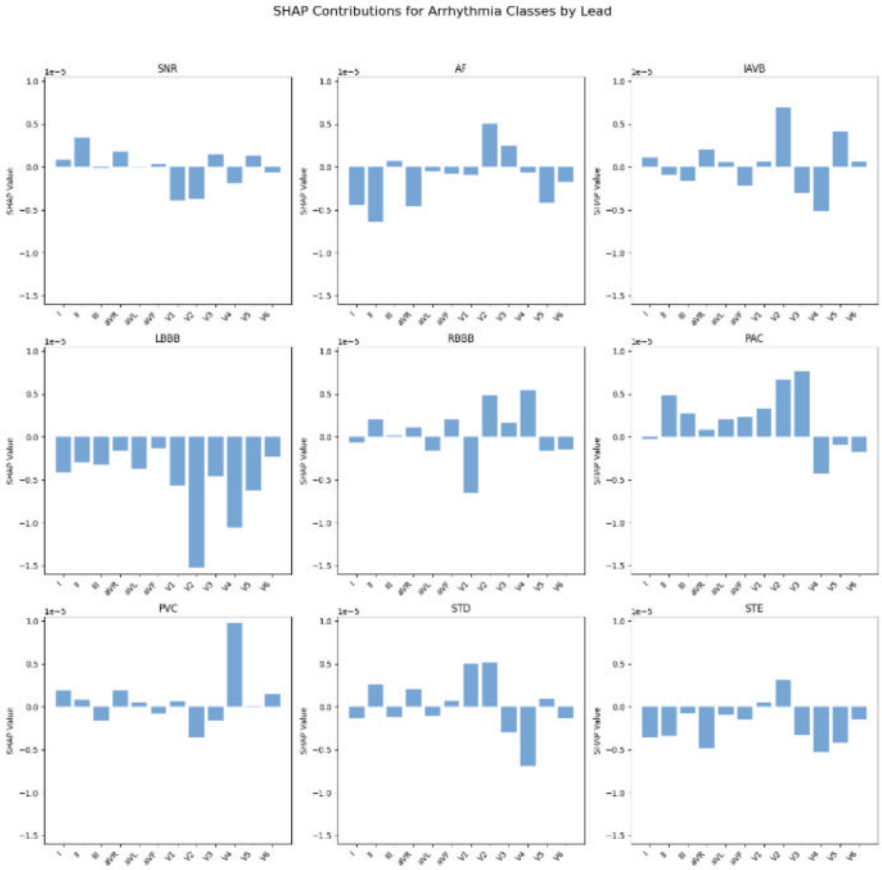


Fig. 7. SHAP contributions for arrhythmia categories per ECG lead.

In our study, we developed an explainable model for detecting cardiac arrhythmias using 12-lead ECG signals by exploring two different approaches: one using raw ECG signals as input for various DL architectures, and another combining raw ECG signals with HRV features in a hybrid model. We identify the best-performing model in the multiclass arrhythmia prediction by assessing the AUROC for each arrhythmia category and their average. If classifiers exhibit equal AUROC values, we also evaluate the F1-score, an important metric in multiclass classification scenarios.

In our evaluation of DL models for arrhythmia detection, both ResNet models outperformed DenseNet and VGG16, with ResNet34 and ResNet50 showing similar effectiveness. However, ResNet34 marginally leads as the optimal model with an AUROC of 0.98 and an F1-score of 0.83, compared to ResNet50 (AUROC: 0.98, F1-score: 0.81), DenseNet (AUROC: 0.96, F1-score: 0.77), and VGG16 (AUROC: 0.95, F1-score: 0.76).

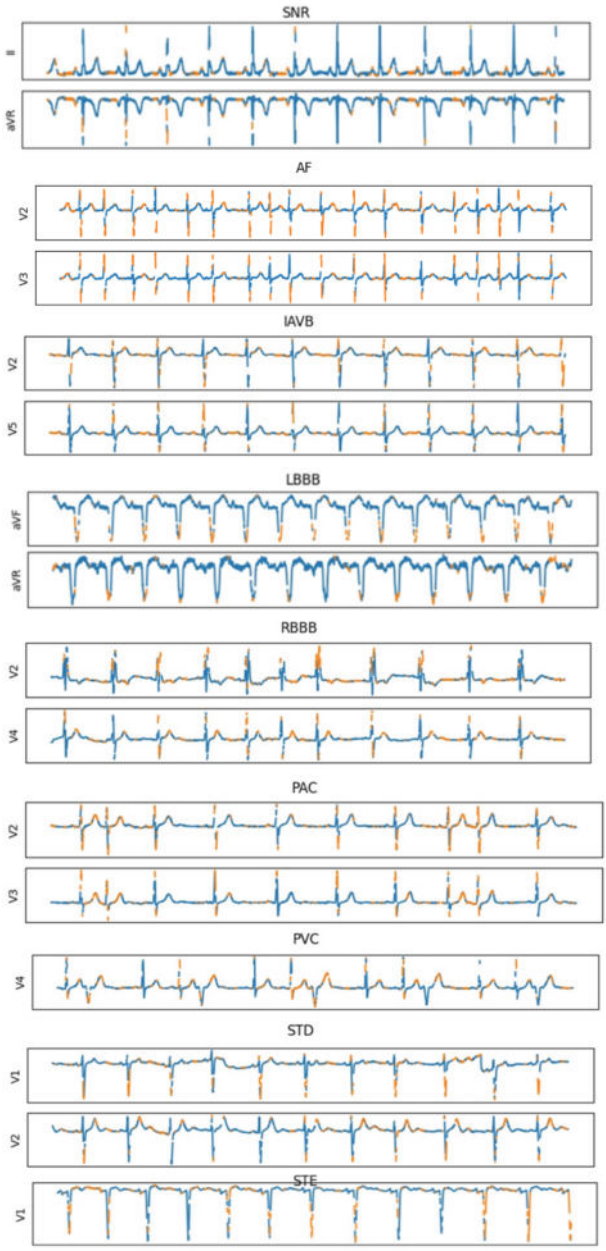


Fig. 8. SHAP contributions in the ECG signal per the relevant lead for each arrhythmia category

By inspecting the different ROC curves and the confusion matrix for individual arrhythmia classifications, ResNet34 excels in detecting SNR, LBBB, RBBB, PAC, and STD, while ResNet50 is superior in classifying AF, IAVB, PVC, and STE.

The hybrid approach, combining ResNet34 with Heart Rate Variability (HRV) features, did not enhance performance in any arrhythmia category, including overall average. Notably, it showed reduced accuracy and precision, especially in AF and RBBB categories. This suggests that instead of providing ResNet34 with additional informative features, the HRV integration might have impeded the network's learning, leading to a decline in classification performance. Thus, the integration of HRV features appears to compromise rather than improve the model's efficacy in arrhythmia classification.

This study's results indicate that the employed DL algorithms have the potential to aid clinicians and healthcare professionals in detecting cardiac conditions that might otherwise be missed or diagnosed later through specialist evaluations or echocardiography. Leveraging both short-term and long-term learning, these methods enable early detection of cardiovascular diseases (CVD), facilitating timely treatment initiation. This early intervention can lead to improved health outcomes, while delays or missed diagnoses could exacerbate health conditions.

ML and DL tools offer precise predictions but their 'black box' nature poses significant interpretability challenges, hindering clinician acceptance due to difficulties in understanding decision-making processes. This lack of transparency is especially problematic in clinical practice, where clinicians need clarity for effective decision-making. The interpretability and application of these advanced models, particularly in identifying critical features, remain complex, potentially limiting their utility in settings without computer assistance. This opacity significantly impedes the adoption of AI models by healthcare professionals who require comprehensible explanations of AI-derived results. In response, the emergence of XAI is a critical development, aiming to demystify AI models and outcomes to enhance accessibility and user trust. XAI facilitates the identification of key features influencing predictions and explores causal relationships between features and clinical outcomes. Despite its importance, research focusing on the understandability and trust in ML models, particularly those aiding in the diagnosis or prognosis of CVD, is still scarce, highlighting a vital area for future exploration.

In our study, we employed the SHAP technique to analyze explainability at both global and individual levels using the 12-lead ECG data. Globally, SHAP enabled us to perform a lead-specific relevance analysis for each arrhythmia category, guiding clinical experts to focus on the most influential leads identified. For example, the V2 lead is highlighted as the most significant feature in six out of nine arrhythmia categories, with avR, V1, V3, and V4 also being important. This information is valuable for researchers focusing on specific arrhythmia predictions, as it suggests prioritizing these leads in ECG monitoring. Interestingly, this contrasts with the common use of ECG lead II in arrhythmia detection, which our model found to be less relevant, highlighting a divergence from established practices noted in related literature.

The significant advantage of our explainable model lies in the synergy created by merging global and individual explainability approaches. Clinicians can use features deemed important globally to scrutinize individual arrhythmia predictions. This process allows them to verify if the ECG segments identified by SHAP as influential align with established clinical knowledge. Essentially, the model facilitates a comprehensive cross-verification, enabling clinicians to confirm the clinical validity of globally relevant

features by examining their contribution to specific arrhythmia classifications on a case-by-case basis.

This study's primary limitation is its reliance on a single database for constructing the prediction model, potentially affecting both the model's generalizability and the validity of explainability analysis. Future research should include additional databases like MIT-BIH to validate and benchmark our findings. Additionally, our study focused exclusively on ECG signal characteristics, excluding other potentially informative patient data such as demographic details, medical history, and laboratory results, due to database constraints. Expanding future models to datasets with comprehensive patient information is recommended. Finally, our study lacks clinical validation of the explainability results, a common challenge in CVD prediction models using XAI. Future efforts should aim to correlate XAI findings with established clinical knowledge to enhance their medical relevance.

5 Conclusions

This study developed a deep learning (DL) based prediction model for nine arrhythmia classes using 12-lead ECG signals and tested various DL architectures including ResNet, DenseNet, and VGG16. ResNet34 was identified as the most effective model, achieving an AUROC of 0.98 and an F1-score of 0.826. A hybrid approach combining raw ECG signals with Heart Rate Variability (HRV) features was also tested, but it did not outperform the raw signal model. Additionally, we conducted an explainability analysis using SHAP within the XAI framework. This analysis pinpointed key ECG leads and signal segments critical for arrhythmia prediction, enhancing the model's transparency and potential usability for clinical professionals. The study highlights the significance of explainable models in cardiovascular disease (CVD) prediction, promoting their acceptance in clinical settings due to the clarity of model decisions.

Acknowledgement. This work is funded by the project PerCard (Personalised Prognostics and Diagnostics for Improved Decision Support in Cardiovascular Diseases) in ERA PerMed supported by the Research Council of Finland (decision number 351846), under the frame of ERA PerMed; by the Federal Ministry of Education and Research, Germany [i.e. "Bundesministerium für Bildung und Forschung (BMBF)"], grant number 01KU2211; and by the Fondazione Regionale della Ricerca Biomedica (FRRB) under the frame of ERA PerMed.

References

1. Heart Failure and Cardiovascular Diseases. https://ehnheart.org/publications-and-papers/publications/1202:heart-failure-and-cardiovascular-diseases.html. Accessed 25 Sept 2023
2. Agrawal, A., Chauhan, A., Shetty, M.K., Gupta, M.D., Gupta, A.: ECG-iCOVIDNet: interpretable AI model to identify changes in the ECG signals of post-COVID subjects. Comput. Biol. Med. **146**, 105540 (2022). https://doi.org/10.1016/j.compbiomed.2022.105540
3. Bogun, F., et al.: Misdiagnosis of atrial fibrillation and its clinical consequences. Am. J. Med. **117**, 636–642 (2004). https://doi.org/10.1016/j.amjmed.2004.06.024

4. Nenna, A., et al.: Heart rate variability: a new tool to predict complications in adult cardiac surgery. J. Geriatr. Cardiol. **14**, 662–668 (2017). https://doi.org/10.11909/j.issn.1671-5411.2017.11.005
5. Ebrahimi, Z., Loni, M., Daneshtalab, M., Gharehbaghi, A.: A review on deep learning methods for ECG arrhythmia classification. Exp. Syst. Appl. X. **7**, 100033 (2020). https://doi.org/10.1016/j.eswax.2020.100033
6. Moreno-Sánchez, P.A.: Data-driven early diagnosis of chronic kidney disease: development and evaluation of an explainable AI model. IEEE Access. **11**, 38359–38369 (2023). https://doi.org/10.1109/ACCESS.2023.3264270
7. Al'Aref, S., et al.: Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. Eur. Heart J. **40**, 1975 (2019). https://doi.org/10.1093/eurheartj/ehy404
8. Moreno-Sánchez, P.A.: Improvement of a prediction model for heart failure survival through explainable artificial intelligence. Front. Cardiovasc. Med. **10** (2023). https://doi.org/10.3389/fcvm.2023.1219586
9. Chen, Y.-J., Liu, C.-L., Tseng, V.S., Hu, Y.-F., Chen, S.-A.: Large-scale classification of 12-lead ECG with deep learning. In: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pp. 1–4 (2019). https://doi.org/10.1109/BHI.2019.8834468
10. Cheng, Y., Ye, Y., Hou, M., He, W., Pan, T.: Multi-label arrhythmia classification from fixed-length compressed ECG segments in real-time wearable ECG monitoring. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference 2020, pp. 580–583 (2020). https://doi.org/10.1109/EMBC44109.2020.9176188
11. Gao, J., Zhang, H., Lu, P., Wang, Z.: An effective LSTM recurrent network to detect arrhythmia on imbalanced ECG dataset. J. Healthcare Eng. **2019**, 6320651 (2019). https://doi.org/10.1155/2019/6320651
12. Niu, L., Chen, C., Liu, H., Zhou, S., Shu, M.: A deep-learning approach to ECG classification based on adversarial domain adaptation. Healthcare **8** (2020). https://doi.org/10.3390/healthcare8040437
13. Romdhane, T., Alhichri, H., Ouni, R., Atri, M.: Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss. Comput. Biol. Med. **123** (2020). https://doi.org/10.1016/j.compbiomed.2020.103866
14. Wang, H., Shi, H., Chen, X., Zhao, L., Huang, Y., Liu, C.: An improved convolutional neural network based approach for automated heartbeat classification. J. Med. Syst. **44** (2019). https://doi.org/10.1007/s10916-019-1511-2
15. Wang, T., Lu, C., Sun, Y., Yang, M., Liu, C., Ou, C.: Automatic ECG classification using continuous wavelet transform and convolutional neural network. Entropy **23** (2021). https://doi.org/10.3390/e23010119
16. Yildirim, O., Plawiak, P., Tan, R., Acharya, U.: Arrhythmia detection using deep convolutional neural network with long duration ECG signals. Comput. Biol. Med. **102**, 411–420 (2018). https://doi.org/10.1016/j.compbiomed.2018.09.009
17. Zhang, D., Yang, S., Yuan, X., Zhang, P.: Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram. iScience **24**, 102373–102373 (2021). https://doi.org/10.1016/j.isci.2021.102373
18. Zhang, Y., Yu, J., Zhang, Y., Liu, C., Li, H.: A convolutional neural network for identifying premature ventricular contraction beat and right bundle branch block beat. Presented at the 2018 International Conference on Sensor Networks and Signal Processing (SNSP 2018) (2018). https://doi.org/10.1109/SNSP.2018.00037

19. Rai, H., Chatterjee, K.: Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data. Appl. Intell. **52**, 5366–5384 (2022). https://doi.org/10.1007/s10489-021-02696-6
20. Toma, T.I., Choi, S.: A parallel cross convolutional recurrent neural network for automatic imbalanced ECG arrhythmia detection with continuous wavelet transform. Sensors (Basel) **22** (2022). https://doi.org/10.3390/s22197396
21. Liu, F., et al.: An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. J. Med. Imaging Hlth. Inform. **8**, 1368–1373 (2018). https://doi.org/10.1166/jmihi.2018.2442
22. Highlights — pyHRV - OpenSource Python Toolbox for Heart Rate Variability 0.4 documentation. https://pyhrv.readthedocs.io/en/latest/. Accessed 19 Jan 2024
23. Welcome to BioSPPy — BioSPPy 0.6.1 documentation. https://biosppy.readthedocs.io/en/stable/. Accessed 19 Jan 2024
24. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. IEEE, Las Vegas, NV, USA (2016). https://doi.org/10.1109/CVPR.2016.90
25. Guessoum, S., et al.: The short-term prediction of length of day using 1D convolutional neural networks (1D CNN). Sensors **22**, 9517 (2022). https://doi.org/10.3390/s22239517
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. http://arxiv.org/abs/1409.1556 (2015)
27. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. http://arxiv.org/abs/1608.06993 (2018). https://doi.org/10.48550/arXiv.1608.06993
28. Lundberg, S., Lee, S.-I.: A unified approach to interpreting model predictions. http://arxiv.org/abs/1705.07874 (2017). https://doi.org/10.48550/arXiv.1705.07874
29. Castelyn, G., Laranjo, L., Schreier, G., Gallego, B.: Predictive performance and impact of algorithms in remote monitoring of chronic conditions: a systematic review and meta-analysis. Int. J. Med. Inf. **156** (2021). https://doi.org/10.1016/j.ijmedinf.2021.104620
30. Baek, S., Jang, J., Yoon, S.: End-to-end blood pressure prediction via fully convolutional networks. IEEE Access **7**, 185458–185468 (2019). https://doi.org/10.1109/ACCESS.2019.2960844
31. Jangra, M., Dhull, S., Singh, K.: ECG arrhythmia classification using modified visual geometry group network (mVGGNet). J. Intell. Fuzzy Syst. **38**, 3151–3165 (2020). https://doi.org/10.3233/JIFS-191135
32. Tadesse, G., et al.: DeepMI: deep multi-lead ECG fusion for identifying myocardial infarction and its occurrence-time. Artif. Intell. Med. **121** (2021). https://doi.org/10.1016/j.artmed.2021.102192
33. Butun, E., Yildirim, O., Talo, M., Tan, R., Acharya, U.: 1D-CADCapsNet: one dimensional deep capsule networks for coronary artery disease detection using ECG signals. Phys. Med-Eur. J. Med. Phys. **70**, 39–48 (2020). https://doi.org/10.1016/j.ejmp.2020.01.007
34. Dai, H., Hwang, H.-G., Tseng, V.S.: Convolutional neural network based automatic screening tool for cardiovascular diseases using different intervals of ECG signals. Comput. Methods Programs Biomed. **203**, 106035 (2021). https://doi.org/10.1016/j.cmpb.2021.106035

Real-Time Gait Anomaly Detection Using 1D-CNN and LSTM

Jakob Rostovski[1(✉)], Mohammad Hasan Ahmadilivani[2], Andrei Krivošei[1], Alar Kuusik[1], and Muhammad Mahtab Alam[1]

[1] TJS Department of Electronics, Tallinn University of Technology, Tallinn, Estonia
{jakob.rostovski,andrei.krivosei,alar.kuusik,muhammad.alam}@taltech.ee
[2] Computer Systems Department, Tallinn University of Technology, Tallinn, Estonia
mohammad.ahmadilivani@taltech.ee

Abstract. Anomaly detection and fall prevention represent one of the key research areas within gait analysis for patients suffering from neurological disorders. Deep Learning has penetrated into healthcare applications, encompassing disease diagnosis and anomaly prediction. Connected wearable medical sensors are emerging due to computationally expensive machine learning tasks, which traditionally require use of remote PC or cloud computing. However, to reduce needs for wireless communication channel throughput, for data processing latency, and increase service reliability and safety, on device machine learning is gaining attention. This paper presents an innovative approach that leverages one dimensional convolutional neural network (1D-CNN) and long-short term memory (LSTM) neural network for the real-time detection of abnormal gait patterns during the step. Real-time anomaly detection pertains to the algorithm's ability to promptly detect true gait abnormality occurrence during the swing phase of an ongoing step.

For the experiments, we have collected eight different common gait anomalies, simulated by 22 persons, using motion sensors containing multidimensional inertial measurement units (IMUs).

Results have demonstrated that the proposed 1D-CNN-AD algorithm achieves an average accuracy of 95% and an average F1-score of 88% for all gait types and can run in true real-time. Average earliness for 1D-CNN-AD algorithm was 0.6 s, which is mid-swing phase of the step. Proposed LSTM-AD algorithm achieved average accuracy of 87% and average F1-score of 70% for all gait types.

Keywords: Human gait · Anomaly detection · Gait analysis · Machine learning · Real-time · 1D-CNN · LSTM · Wearable sensors

This work has been supported by Estonian Research Council, via research grant No PRG424, by the Center of Excellence (TK) project TAR16013 (EXCITE) and Estonian IT Academy project "Sustainable Artificial Internet of Things (SAIoT)".

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 260–278, 2024.
https://doi.org/10.1007/978-3-031-59091-7_17

1 Introduction

According to the World Health Organisation (WHO) report about one billion persons are affected by neurological disorders worldwide [3]. Neurological diseases ranging from migraine to stroke, and Alzheimer are the leading causes of Disability Adjusted Life Years (DALY) loss [7]. For instance, there is a substantial risk of falling for patients with gait impairments from neurological diseases [23]. It is especially true for patients suffering from neuromuscular diseases, because high variability and deviations from the optimal gait pattern can be seen in their gait [13]. Therefore, it is challenging to analyze patients' gait patterns in real-time. The gait of a person can be described by a set of parameters such as: step length, duration of individual step phases, muscle force, etc. [19]. Wearable motion sensors, containing multidimensional Inertial Measurement Units (IMUs), are the most widely used gait assessment devices in recent years for supporting daily activities [25]. For example, motion sensors are used to detect initial and final contact events of the gait cycle for different persons - healthy, with stroke, and with other neurological disorders, and select the best algorithms and sensor placements for correct classification between them [10]. Motion sensors can be employed to detect activities of daily life, fall events and their directions [9], to determine rehabilitation progress and analyze gait normalcy index [2,36]. Also such devices can be used to discover environment dependent differences in gait, which will help with context-aware decisions [29]. Finally, in combination with Neural Networks (NNs), identify if person has balance disorder [20], to track rehabilitation progress for broken limbs [4] etc.

It is shown that Functional Electrical Stimulation (FES) can be used to assist walking and help with fall prevention [12] as well as for generic gait improvements [17]. Long-term gait deviation analysis and efficient run-time control of FES devices require automated real-time recognition of gait deviations. Average swing phase of a step is 300–400 ms long [8], and the time of full contraction of the muscle using electrical stimulation is 100–200 ms long [5], thus the detection time of step pattern deviations should be under 100 ms. Considering that the incoming signal must be processed, a correct decision made, and stimulation actuation started, a detection time of 50 ms is required since the gait abnormality has started.

Connected wearable medical sensors are emerging due to computationally expensive machine learning tasks, which traditionally require use of remote PC or cloud computing [14]. Nowadays, it is common to offload such data analysis from wearable sensors to wirelessly connected smartphones [11]. For example, data processing unit, sensors and muscle stimulator shall be wireless for gait correction system, i.e. based on Bluetooth or SmartBAN standard. However, to reduce needs for wireless communication channel throughput, for data processing latency, and increase service reliability and safety, on device machine learning is gaining attention [31]. Existing real-time algorithms are used in gait analysis for identification by gait [15]; detecting of gait events like heel-strike and toe-off for elderly healthy subjects; stroke patients and patients with Parkinson disease [35], as well as with other impairments [24,37]; haptic biofeedback devices are

implemented using inertial measurement units (IMUs), to correct toe-in or toe-out during walking in real-time [32].

Notably, there are not found state-of-the-art solutions in gait analysis for real-time anomaly detection of realistic gait deviations during the ongoing step, caused by neurological diseases.

In our prior research work [27,28] we proposed a base method for real-time anomaly detection in gait during the ongoing step, with an algorithm based on Support Vector Machines (SVM), which is one of the most popular algorithms used in gait analysis. On the other hand, NNs are widely adopted in gait analysis [30]. They are capable of solving complex tasks in time-series data. Nonetheless, to the best of our knowledge, there is no research exploiting NNs for real-time anomaly detection during the ongoing step in gait analysis. In this paper, for the first time, we leverage Convolutional Neural Network (CNN) and Long Short-Term Memory NNs for real-time anomaly detection during the ongoing step in human gait.

The contributions of this work are:

- Estimation of the performance of One Dimensional-Convolutonal Neural Network-Anomaly Detection algorithm (1D-CNN-AD) and Long Short-Term Memory Neural Network-Anomaly Detection algorithm (LSTM-AD) on the collected simulated gait deviation dataset in comparison to the Real-time tsSVM Anomaly Detection algorithm (RTtsSVM-AD).
- Exploiting hyperparameters for the neural networks to optimize performance on simulated gait dataset for real-time in-step anomaly detection.

This paper consists of six sections: after the introduction, in Sect. 2 data acquisition and gait types are described, as well as metrics used for analysis in addition to presenting the proposed 1D-CNN-AD and LSTM-AD algorithms, then in Sect. 3 we briefly describe evaluation metrics and the SVM-based algorithm – RTtsSVM-AD, which is continued with experimental setup in Sect. 4; this is followed by the results and discussion in the Sect. 5 and the paper is concluded in Sect. 6.

2 Methodology

2.1 Dataset

Data Acquisition. The dataset in our experiments is collected from twenty-two healthy persons of different genders, ages, heights and weights (Table 1), while walking in a straight line and simulating abnormalities. Simulations are recreating actual patients' video recordings of gait deviations in collaboration and guidance from a professional physiotherapist of Tallinn East Central Hospital. We have included the most frequent human gait abnormalities, regarding reference [1]: Ataxic, Diplegic, Hemiplegic, Hyperkinetic, Parkinonian, Slap, Steppage, and Trendelenburg (lurch). Table 2 shows eight under-study gait types and the number of collected gait recordings per gait type. Collected data is labeled

Table 1. Persons' Information Used in This Study (Mean ± *Standard Deviation)*

No. of subjects	Age (years)	Height (cm)	Mass (kg)
15 (Male)	32.1 ± 11.1	177.7 ± 5.5	76.8 ± 15.1
7 (Female)	26.3 ± 5.5	169.5 ± 6.2	62.7 ± 8.9

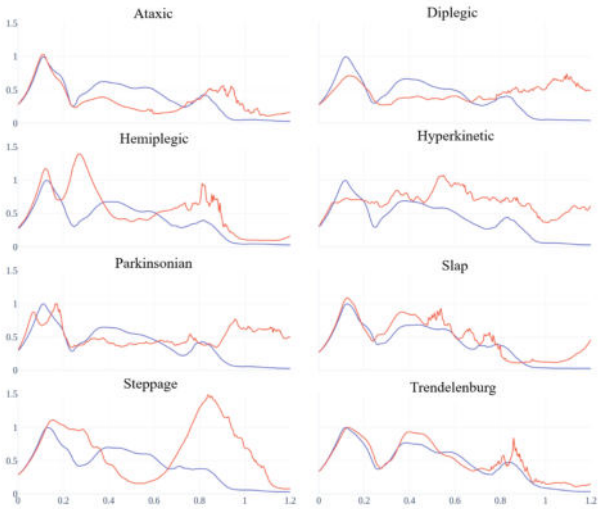


Fig. 1. Example of the typical shape of simulated step of studied gait types in comparison to normal step shape, from the data used in this study. Blue line is normal step shape and red line is corresponding typical shape for this gait type. On X-axis is time in seconds and on Y-axis is normalized magnitude of angular velocities of gyroscope. (Color figure online)

Table 2. Labeled data collected for this study.

Gait type	Total number of recordings for all persons
Ataxic	32
Diplegic	25
Hemiplegic	17
Hyperkinetic	6
Parkinsonian	29
Slap	8
Steppage	32
Trendelenburg	6

step-wise, thus all steps are annotated as *normal* or *abnormal*. Figure 1 illustrates the patterns of each gait type in comparison with a normal step.

Such dataset to the best of authors knowledge is first to have combination of normal and abnormal steps in one dataset. Other datasets are focusing on normal gait patterns; have only abnormal steps in the dataset; compare separate normal gait datasets and abnormal gait datasets, etc. [6,18,26,33].

Data Preprocessing. The collected data is in a form of time-series including a three-axis gyroscope and their calculated magnitude (1).

$$\mathbf{Mag}(X, Y, Z) = \sqrt{\mathbf{X}^2 + \mathbf{Y}^2 + \mathbf{Z}^2}, \tag{1}$$

where $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ are gyroscope axes data vectors, $\mathbf{X} = [x_0, x_1, \ldots, x_i]^T$, $\mathbf{Y} = [y_0, y_1, \ldots, y_i]^T$ and $\mathbf{Z} = [z_0, z_1, \ldots, z_i]^T$, sample index $i \in \mathbb{Z}$. And the $\mathbf{Mag}(X, Y, Z)$ is the magnitude vector of these axes.

To address future works with embedded devices in regard to data transmission and data gathering, data is collected into chunks. One chunk contains M samples for each gyroscope axis. The collected data sample rate is $256 Samples/s$ in the current study. Collected data is labeled stepwise as *"normal"* step or *"abnormal"* step.

Data Preparation for Real-Time Anomaly Detection. For 1D-CNN-AD and LSTM-AD algorithms each person's data is assessed separately. Data for one gait type is prepared by separating training and validation datasets. One gait recording is used as a validation dataset in real-time step anomaly detection estimation, and all other recordings are combined into one training dataset. The ratio between the training and validation datasets can change depending on the person, gait type and available gait recordings for particular gait type.

To enable real-time abnormality detection in the swing phase of the ongoing step, training dataset is divided into overlapping sliding windows. Figure 2 depicts how the windowing of the dataset is designed. As it is shown, each window contains P chunks (i.e., window factor), and each chunk includes M samples and the overlap is N chunks.

Labeling of the windows is conducted according to the labels of the steps. In edge cases, where one step is ending and new step is begging, label is assigned by the proportion of samples of abnormal steps in the window. If this proportion is less than *abnormality proportion threshold* then the window is labeled as *normal*, if more, then it is labeled as *abnormal*.

One of the key advantages of the sliding windows for this study is independence of the anomaly detection algorithms from gait phases.

As a part of hyperparameters optimization, hyperparameters, which affect sizes, overlaps and labels of the sliding windows are investigated. These hyperparameters are a) chunk duration – time in milliseconds, where number of samples M in one chunk is calculated from chunk duration as $M = round(Chunk\ duration * Sample\ rate)$; b) window factor P – determines window size and is proportional to P chunks; c) Abnormality proportion

threshold – fraction of the window, which should contain abnormal samples, to consider the label of the window to be abnormal.

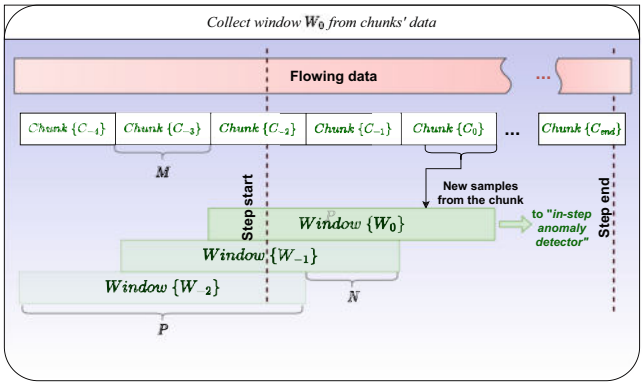


Fig. 2. Windowing of the data for training and for real-time anomaly detection performance estimation. Ongoing gait data is incoming as flowing data, which is split into chunks. From these chunks sliding windows are collected and used in real-time in-step anomaly detector. Step start can be misaligned with sliding window. Chunks are aligned with the sliding windows If abnormality is detected during the chunk C_0, then earliness is time between step start and end of the chunk C_0.

2.2 Proposed Neural Networks

One Dimensional-Convolutonal Neural Network-Anomaly Detection Algorithm. The hypothesis of the 1D-CNN-AD algorithm is following: if real-time gait data could be collected in the form of sliding windows, and neural network could be trained on the dataset using same form of sliding windows with known labels, then it is possible to detect abnormalities in gait during the ongoing step.

The CNN in this study consists of two 1D convolutional layers, max pooling layer, and two fully-connected (dense) layers to provide a binary classification. The 1D-CNN-AD algorithm has the following hyperparameters: i) number of filters; ii) kernel size; iii) batch size; iv) and number of epochs. These hyperparameters would be optimized in this study to achieve the best performance for 1D-CNN-AD algorithm.

In the explorations, the CNN is initialized with a fixed seed of parameters (i.e., weights and bias). The neural network is trained on training dataset with Adam optimizer and cross-entropy loss function. Moreover, a 20% dropout is also considered between the convolutional layer and dense layers.

Long Short-Term Memory Neural Network-Anomaly Detection Algorithm. Hypothesis of the LSTM-AD algorithm is identical to the hypothesis of the 1D-CNN-AD algorithm.

The LSTM-AD algorithm in this work consists of one layer of LSTM followed by two fully-connected (dense) layers to provide a classification probability. The number of cells in the LSTM layer is equal to the number of neurons in the first dense layer. The LSTM-AD algorithm has the following hyperparameters: i) number of LSTM cells; ii) batch size; iii) and number of epochs. These hyperparameters would be optimized in this study to achieve the best performance for LSTM-AD algorithm.

In the explorations, the LSTM is initialized with a fixed seed of parameters (i.e., weights and bias). The LSTM-AD algorithm is trained on training dataset with Adam optimizer and cross-entropy loss function.

2.3 Anomaly Detection

To estimate performance of the real-time anomaly detection of the algorithms, validation dataset is processed in online-fashion. It means, that data is arriving sample by sample. Each sample is collected into chunks. Chunks are collected into windows, as was described in the Sect. 2.1. Algorithms return anomalous class probability for each window, which is collected to the buffer. After the real-time estimation, collected probabilities are analyzed. Different thresholds for anomalous class probability are estimated to achieve best results. This results in the binary classification. These classification results are compared to the labels of the validation dataset, resulting in confusion matrix. Accuracy and F1 score are calculated from confusion matrix.

3 Baseline and Evaluation

3.1 Real-time tsSVM Anomaly Detection Algorithm

RTtsSVM-AD algorithm is based on a *tslearn* [34] Python library. Optimization of hyperparameters is done by dividing training data into two datasets: training and testing with ratio of 70%:30%. Trained classifier with best results for test dataset is used in real-time performance estimation. Model step is calculated as normal step ensemble average from test dataset, which consist of normal steps that have been classified correctly.

The hypothesis of the algorithm is following: if full time-series step pattern could be collected in real-time by combining the average normal step from training phase with the ongoing step data, then anomaly could be detected during the swing phase of the ongoing step by the RTtsSVM-AD algorithm.

We have adopted the RTtsSVM-AD algorithm in our prior work [28] as a baseline for comparing the results of the proposed 1D-CNN-AD and LSTM-AD algorithms in this paper.

Brief overview of the algorithm. Data is collected chunk wise, when step start is detected. Step start and end events are detected if the step detection

threshold crosses 20% of the gyroscope magnitude range. Hyperparameter γ is optimized on training and testing datasets. This hyperparameter is used by the global alignment kernel (GAK), where γ is the hyperparameter controlling soft dynamic time warping (softDTW) smoothness [34]. Multiple classifiers with different values of γ could have same performance. Average normal step is created from correctly classified normal steps from training dataset. In real-time in-step gait anomaly detection performance estimation, if step start is detected, data is collected into a chunk. This chunk is replacing corresponding chunk in the model step. Such chunkwise replacement converts regular time-series SVM into the real-time anomaly detection algorithm.

3.2 Evaluation Metrics

For evaluation, several metrics are exploited: Accuracy, F1-score, *earliness*, and real-time factor (RTF). *Earliness* in this paper is defined as – time between the beginning of a step and the moment in time when anomaly is detected in this step. The minimal achievable earliness naturally depends on the gait deviation type. Such a measure has been introduced, because the concrete moment when anomaly starts to occur can fluctuate, depending on a gait type.

3.3 Score and Alarm

For estimation of the performance of the algorithms, anomalous class probability is collected from the classifier. Binary decision is performed later in post-processing of the results. *Score* is the resulting anomalous class probability. For RTtsSVM-AD algorithm *Score* is average score from used classifiers in estimation, because multiple classifiers could be used simultaneously. *Score* is compared with the selected *threshold*, giving alarm signal in (2), finalizing the anomaly detection.

$$Alarm = \begin{cases} 1, & \text{if } S > threshold \\ 0, & \text{if } S \leq threshold \end{cases} \tag{2}$$

If *Alarm* is triggered, then earliness is the time duration from the beginning of the step to the current moment in time.

4 Experimental Setup

For 1D-CNN-AD and LSTM-AD algorithms, the considered hyperparameters are presented in the Table 3 and Table 4.

For the RTtsSVM-AD algorithm parameters used in this work are as follows: a) one chunk is $M = 12$ samples; b) predefined γ values are in the range from 100 to 1000 with an increase of 100 and in the range from 5 to 100 with an increase of 10; c) the step detection threshold is $200°/s$.

All training and validation experiments are implemented in Python 3.10.13, tslearn 0.6.2, and TensorFlow 2.9.1 and performed on a prebuilt HP computer

Table 3. Global hyperparameters for 1D-CNN-AD and LSTM-AD algorithms

Hyperparameter	Values
Window factor (P)	6 to 10. Default 8
Chunk size	25 ms to 100 ms. Default 50 ms
Samples in a chunk (M)	6 to 25. Default 12
Sliding window overlap (N)	1
Abnormality proportion threshold	50% to 90%. Default 70%
Batch size	2^n where n is from 3 to 8. Default n is 5
Number of epochs in training	1 to 30. Default 20

Table 4. Algorithm-Specific Hyperparameters

Algorithm	Specific Hyperparameters
LSTM-AD	Number of LSTM cells: 20, 25, 30. Default: 25
1D-CNN-AD	Number of filters in convolutional layer: 2^n where n is from 3 to 8. Default n is 6 Kernel size in convolutional layer: 2, 3, 5, 7, 9, 11. Default 5 Dense layer with 100 neurons

with Intel Core i7 and 16Gb of DDR4 memory. We conducted CPU experiments to model the execution on the embedded devices in future works.

5 Experimental Results and Discussion

Results for the 1D-CNN-AD, LSTM-AD and RTtsSVM-AD algorithms are presented in this section.

5.1 Optimization of 1D-CNN-AD and LSTM-AD Algorithms Hyperparameters

In this paper, optimization is performed by one parameter at a time, while the other parameters are set to their default values.

Chunk Length. The first hyperparameter to consider is the length of the chunk. Table 5 shows the best mean F1 scores with corresponding chunk sizes. It is observed that the best results are achieved with chunk sizes of 75 and 100 ms for all gait types for LSTM-AD and most of the gait types for 1D-CNN-AD. Chunk size of 40 and 50 ms performed better for Steppage, and Trendelenburg gait types for 1D-CNN-AD algorithm. Despite the better performance with longer chunks for some gait types, chunk size is set to 50 ms, with consideration of fast anomaly detection. Larger chunk sizes would lead to slow anomaly detection.

Table 5. Best mean F1 scores for different chunk sizes (CS)

Gait type	LSTM-AD		1D-CNN-AD	
	F1	CS, ms	F1	CS, ms
Ataxic	62.63%	100	79.28%	75
Diplegic	72.36%	100	87.49%	100
Hemiplegic	81.03%	100	83.52%	75
Hyperkinetic	75.95%	100	96.3%	75
Parkinsonian	75.24%	100	84.65%	100
Slap	57.45%	75	78.7%	75
Steppage	75.09%	100	84.17%	40
Trendelenburg	59.65%	75	81.3%	50

Window Factor and Abnormality Proportion. These hyperparameters should be considered in correlation with each other because both of them change the number of samples in the window, which can change the final label of the window. Table 6 presents the best mean F1 scores for combination of window factor and abnormality proportion. It could be seen, that 1D-CNN-AD algorithm is performing best with shorter windows for most of the gait types, whereas LSTM-AD algorithm is performing best with longer windows for most of the gait types. In terms of abnormality proportion threshold, for most of the gait types for both 1D-CNN-AD and LSTM-AD algorithms higher threshold is needed. Only for Hyperkinetic and Steppage gait types it was 70% for LSTM-AD and 60% for 1D-CNN-AD algorithms respectively. It means, that for Hyperkinetic and Steppage gait types edge cases are important for correct anomaly detection. Thus, in general, most of the windows should contain mostly abnormal samples to be labeled abnormal for best performance. With the default settings for other parameters, 1D-CNN-AD algorithm achieves mean F1 scores of 96.3% for Hyperkinetic gait type. On the other hand, LSTM-AD algorithm achieves best mean F1 score of 73.78% for Hemiplegic gait type.

Diplegic and Hyperkinetic gait types have anomalies in the middle and end of the step, thus short windows should be best suited for them to detect abnormality early, as can be seen in 1D-CNN-AD algorithm results. Both Ataxic and Parkinsonian gait types have multiple abnormal steps in a row, which can be similar to normal steps, thus requiring well defined long abnormal windows during the training phase. Slap gait is usually characterized by the sharp short peak at the end of the step, whereas the rest of the step can be similar to normal, thus making it more critical to have a correct classification in edge cases. Steppage gait type have different amplitudes from the normal step for its peaks when the knee is raised up to compensate for lack of movement in the forefoot. Hemiplegic gait type can be similar to a normal gait, which makes it more difficult to differentiate from normal steps which require well-defined shorter windows.

Table 6. Best mean F1 scores for different window factor (WF) and abnormality proportion threshold (AP)

Gait type	LSTM-AD			1D-CNN-AD		
	F1	WF	AP	F1	WF	AP
Ataxic	58.16%	9	90%	78%	10	90%
Diplegic	58.31%	8	90%	82.81%	7	80%
Hemiplegic	73.78%	10	90%	84.03%	6	90%
Hyperkinetic	69.05%	10	70%	95.15%	7	90%
Parkinsonian	63.25%	9	90%	84.38%	10	80%
Slap	62.55%	8	80%	86.9%	6	80%
Steppage	64.83%	10	80%	88.39%	6	60%
Trendelenburg	64.75%	10	80%	83.75%	6	90%

Number of Filters and Kernel Size in the Convolutional Layer for 1D-CNN-AD Algorithm and Number of LSTM Cells for LSTM-AD Algorithm. As presented in Table 7, the best scores for 1D-CNN-AD algorithm are generally achieved with a higher number of filters of 128 and 256, except for Diplegic gait type with 32 filters. This means that extracting more features from the data improves the performance of the 1D-CNN-AD algorithm demonstrating the complexity of the human gait. For Diplegic gait type a smaller network is best suited, meaning that extracting too many features can confuse the 1D-CNN-AD algorithm, because the shapes of the abnormal steps for them are more defined than the ones in other gait types.

Best performance is achieved for 1D-CNN-AD algorithm with medium kernel size of 7 except for Hyperkinetic and Steppage gait types with a kernel size of 11 and for Parkinsonian and Trendelenburg gait types with kernel size of 9. For Hyperkinetic, Steppage, Parkinsonian and Trendelenburg gait types bigger kernel size is needed to neglect the variance between individual abnormal steps in the data.

For LSTM-AD algorithm larger number of LSTM cells results in a better performance, due to the complexity of the gait signal. For Ataxic, Diplegic and Slap gait types algorithm performs best with 25 cells showing, that they have simpler shapes, compared to other gait types. For Parkinsonian gait type the best performance was with 20 cells, meaning, that this gait type, has more pronoun shape, compared to other gait types.

Batch Size and Number of Epochs in Training. As presented in Table 8, the best scores are generally achieved with a bigger batch size of 128 and 256, except for Trendelenburg gait type with a size of 32 for 1D-CNN-AD algorithm, and Slap and Trendelenburg gait types with size of 16 and 64 respectively for LSTM-AD algorithm. This means, that a more accurate training gradient of the neural network is needed for these gait types.

Table 7. Best mean F1 scores for different numbers of LSTM cells (#C) and 1D-CNN kernel size (KS) and number of filters (#F)

Gait type	LSTM-AD		1D-CNN-AD			
	F1	#C	F1	KS	F1	#F
Ataxic	52.4%	25	75.31%	7	75.04%	128
Diplegic	56%	25	81.67%	7	80.3%	32
Hemiplegic	61.37%	30	82.3%	7	87.87%	256
Hyperkinetic	69.5%	30	92.45%	11	86.7%	128
Parkinsonian	57.49%	20	85.45%	9	87.74%	128
Slap	56.95%	25	87.1%	7	81.8%	256
Steppage	60.17%	30	85.23%	11	87.43%	256
Trendelenburg	59.85%	30	83.05%	9	81.75%	128

Table 8. Best mean F1 scores for different batch size (B) and number of epochs (#E)

Gait type	LSTM-AD				1D-CNN-AD			
	F1	B	F1	#E	F1	B	F1	#E
Ataxic	56.11%	128	55.95%	5	77.52%	256	78.95%	3
Diplegic	68.77%	256	71.06%	5	85.92%	256	89.33%	5
Hemiplegic	77.08%	256	81.32%	2	85.08%	256	85.45%	4
Hyperkinetic	60.7%	256	73.2%	2	92.45%	256	98.1%	5
Parkinsonian	68.43%	256	66.38%	4	86.81%	256	88.36%	2
Slap	59.8%	16	55.8%	10	83.9%	256	90.8%	4
Steppage	64.89%	128	67.8%	5	87.7%	256	88.1%	2
Trendelenburg	55.4%	64	57.05%	10	81.3%	32	81.3%	20

In terms of the amount of training required by the algorithms, it is clear, that more than 5 epochs could lead to overfitting, thus reducing classification quality in this study. Only LSTM-AD algorithm performed better with 10 epochs for Slap and Trendelenburg gait types, and 1D-CNN-AD algorithm performed better with 20 epochs for Trendelenburg gait type. This could be due to similarities between normal step and typical step shape for Trendelenburg gait type, thus needing more time to properly fit the network. Considering the overall performance of 55.8% for Slap gait type for LSTM-AD algorithm in epoch optimization, algorithm struggled with this gait type. Training dataset usually contains around 5000 windows, thus every epoch has around 39 iterations with batch size of 128. Normal and abnormal steps have mostly consistent shapes in one gait type. Thus, smaller number of epochs can fit such data better. Larger number of epochs could lead to lower performance due to overfitting of the training data and would trigger anomaly detection while classifying unknown data. Therefore, better results are generally achieved with 2 to 5 epochs.

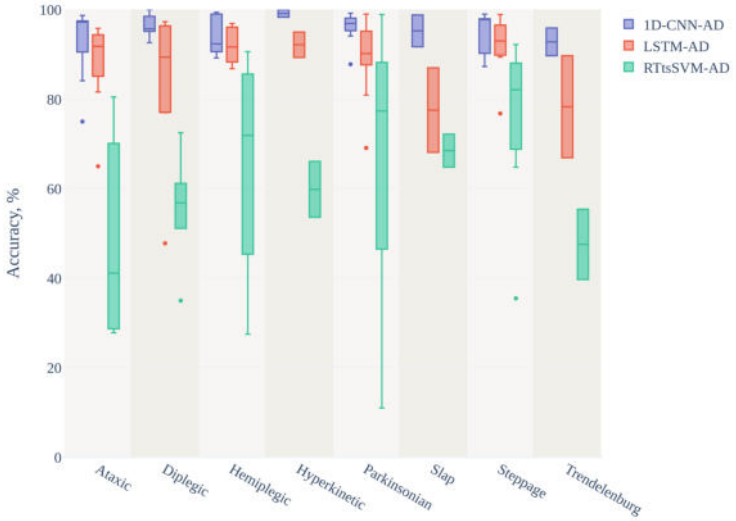


Fig. 3. Distribution of accuracy across different algorithms for all persons for different gait types. On y-axis is accuracy in percents or time in seconds, on x-axis are different gait types.

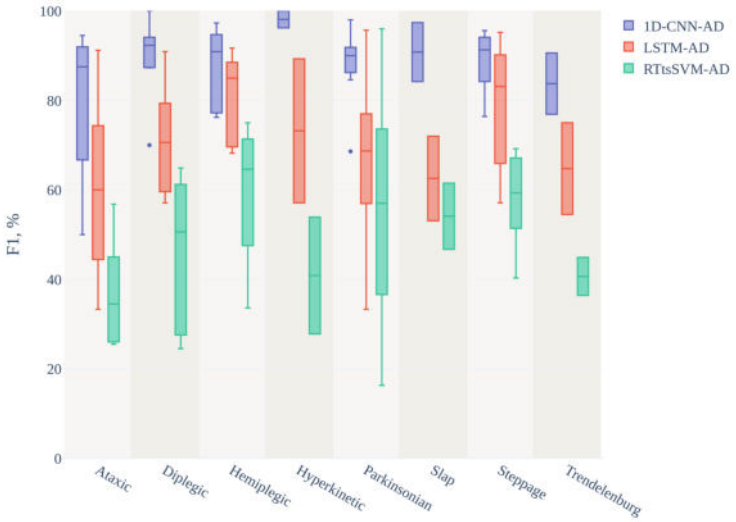


Fig. 4. Distribution of F1 scores across different algorithms for all persons for different gait types. On y-axis is F1 score in percents, on x-axis are different gait types.

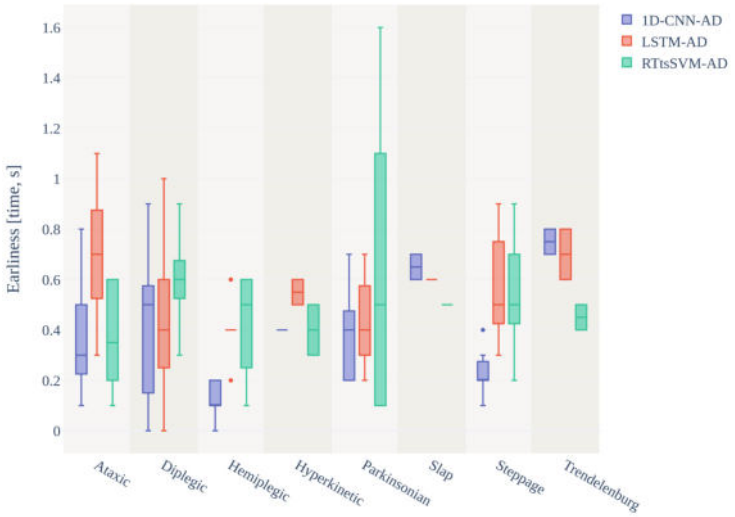


Fig. 5. Distribution of Earliness across different algorithms for all persons for different gait types. On y-axis is time in seconds, on x-axis are different gait types.

Comparison of Algorithms. In Fig. 3 and Fig. 4 could be seen, that both 1D-CNN-AD and LSTM-AD algorithms are outperforming the RTtsSVM-AD base comparison algorithm. The best scores for all gait types are achieved by 1D-CNN-AD algorithm with an average accuracy of 95% and average F1-score of 88%. LSTM-AD algorithm achieved an average accuracy of 87% and average F1-score of 70%. Best results for 1D-CNN-AD algorithm are for Hyperkinetic and Slap gait types with F1 scores of $98.1 \pm 2.7\%$ and $90.8 \pm 9.3\%$ respectively. It could be observed that for Ataxic, Hemiplegic, Slap, Steppage, and Trendelenburg gait types there are some deviations in results from person to person, that could be improved with additional optimization. Best result for LSTM-AD algorithm is achieved for Hemiplegic gait type with average F1 score of $81.32 \pm 9.96\%$. 1D-CNN-AD algorithm is achieving accuracies over 92.6% for all gait types and F1 scores of over 83% for all gait types, except for Ataxic gait type with F1 score of $78.95 \pm 15.43\%$. LSTM-AD algorithm achieved accuracies over 78.3% for all gait types with F1 scores of $71.06 \pm 12.47\%$, $73.2 \pm 22.77\%$ and $79.61 \pm 14.92\%$ for Diplegic, Hyperkinetic and Steppage gait types respectively. Lowest F1 score of $60.24 \pm 18.65\%$ is achieved for Ataxic gait type.

Time of detection is relevant, when classification accuracy is high. Typical normal step length in this study is ranging from 1 to 1.2 s depending on the person, whereas abnormal step duration ranges from 1 to 1.7 s, depending on the person and gait type. Mid-swing phase of the step is starting at around 0.2–0.4 s from the step beginning. Therefore, for the earliness metric depicted in Fig. 5, it

could be observed that for most gait types the earliness is less than one second. For Steppage gait type the most common earliness measure is around 0.6 s for RTtsSVM-AD and LSTM-AD algorithms and 0.2 s for 1D-CNN-AD algorithm which is in the middle or at the beginning of a step. For other gait types it could be observed that detection was mainly in the middle of a step, which shows, that algorithms can detect anomalies early, during the mid-swing phase of a step. For some gait types RTtsSVM-AD and LSTM-AD are detecting abnormality earlier than 1D-CNN-AD, but in combination with quality of prediction, 1D-CNN-AD is outperforming other presented algorithms.

Table 9. Average real-time factor for all algorithms

Algorithm	RTF
1D-CNN-AD	0.09 ± 0.03
LSTM-AD	1 ± 0.07
RTtsSVM-AD	9.13 ± 6.54

In Table 9, it can be observed that the main issue of RTtsSVM-AD algorithm is *computational real-time factor*. It means that for every second of incoming data, it takes 9.13 ± 6.54 s to classify it, which is 3 to 15 times longer than the amount of collected data in real-time. The main reason for this is the usage of prediction probability in *tslearn* classifier, which uses an expensive 5-fold cross-validation method to calculate probability. Using regular class prediction is not possible due to inaccurate results from the classifier, as it outputs only zero or one as class identification, drastically reducing classification quality. LSTM-AD algorithm is performing classification in near real-time but not faster than it, because recurrent operations of the algorithm are computationally expensive. Thus, 1D-CNN-AD algorithm is most suitable for real-time applications, for example, to operate in real-time on a real gait assistive device.

This work have several limitations: a) Simulated gait deviation could differ from the real patient's gait with neurological disorders. However, the main goal in this study is to classify the step as normal or abnormal during the mid-swing phase of the step. If patient's normal step pattern after the rehabilitation is sufficiently different from the patient's abnormal step pattern (i.e. because of fatigue or other reasons), then algorithms will be able to detect gait abnormalities during the mid-swing phase of the step as they are able to detect them in this study with simulated gait. Also, as it was stated in the Sect. 2.1: simulations are recreating actual patients' video recordings of gait deviations in collaboration and guidance from a professional physiotherapist of Tallinn East Central Hospital. Thus, such simulated gait types are representing real gait types as close as possible. b) Neural networks in this study are not aware of the gait phase, thus multiple alarms could be triggered during one abnormal step, thus they would be optimized further. Cross-correlation of different hyperparameters could improve classification performance and would be studied in future work.

6 Conclusion

Proposed in this study real-time in-step anomaly detection algorithms are at the very beginning of the research towards context aware assistive devices, which will help to improve gait quality and reduce falling risk for patients suffering from neurological disorders.

Results of this study shows that 1D-CNN-AD algorithm is suitable for real-time anomaly detection in realistic gait deviations during the ongoing step with average earliness of 0.4 s. An average accuracy of 95% and average F1 score of 88% across different studied gait types is achieved for 1D-CNN-AD algorithm, with best F1 score of $98.1 \pm 2.7\%$ for Hyperkinetic gait type. Benefits of this algorithm are, that it is not dependent on gait phases, resistant to the non-optimal hyperparameters and can run in real-time. Second proposed LSTM-AD algorithm achieved average accuracy of 87% and average F1-score of 70% across different studied gait types and best result is achieved for Hemiplegic gait type with F1 score of $81.3 \pm 9.96\%$.

Future gait correction systems and assistive devices will benefit from context awareness in a form of real-time anomaly detection algorithms, leading to more tailored approach for patients suffering from neurological disorders. This will help them to maintain better gait quality, which they obtained after rehabilitation, giving higher chance to continue daily living activities without major restrictions. Main benefit of context aware assistive devices compared to regular assistive devices would be less muscle fatigue from using it. Considering, that FES is used in current assistive devices [16,21,22], where electrical stimulation is given every step, context aware FES would be used only, when step deviation is detected and stimulation is necessary.

Future work will be focusing on further optimization of the presented algorithms, in-step abnormality estimation with more persons and real-time in-step abnormality detection tests with embedded devices running proposed in this study algorithms.

References

1. Stanford Medicine 25: Gait abnormalities. https://stanfordmedicine25.stanford.edu/the25/gait.html
2. Anwary, A.R., Arifoglu, D., Jones, M., Vassallo, M., Bouchachia, H.: Insole-based real-time gait analysis: feature extraction and classification. In: 2021 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL), pp. 1–4 (2021). https://doi.org/10.1109/INERTIAL51137.2021.9430482
3. Bertolote, J.M.: Neurological disorders affect millions globally: WHO report. World Neurol. **22**(1), 1 (2007). https://worldneurologyonline.com/wp-content/uploads/2013/03/WFN-March-2007-Issue.pdf
4. Boompelli, S.A., Bhattacharya, S.: Design of a telemetric gait analysis insole and 1-D convolutional neural network to track postoperative fracture rehabilitation. In: 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech), pp. 484–488 (2021). https://doi.org/10.1109/LifeTech52111.2021.9391975

5. Cameron, M.H.: Physical Agents in Rehabilitation: From Research to Practice, 4 edn. Elsevier/Saunders, St. Louis, Mo (2013)
6. Chang, C.W., Yan, J.L., Chang, C.N., Wen, K.A.: IMU-based real time four type gait analysis and classification and circuit implementation. In: 2022 IEEE Sensors, pp. 1–4 (2022). https://doi.org/10.1109/SENSORS52175.2022.9967269
7. Feigin, V.L., et al.: Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the global burden of disease study 2016. Lancet Neurol. **18**(5), 459–480 (2019). https://doi.org/10.1016/S1474-4422(18)30499-X
8. Hollman, J.H., McDade, E.M., Petersen, R.C.: Normative spatiotemporal gait parameters in older adults. Gait & Posture **34**(1), 111–118 (2011). https://doi.org/10.1016/j.gaitpost.2011.03.024
9. Hsieh, C., Shi, W., Huang, H., Liu, K., Hsu, S.J., Chan, C.: Machine learning-based fall characteristics monitoring system for strategic plan of falls prevention. In: 2018 IEEE International Conference on Applied System Invention (ICASI), pp. 818–821 (2018)
10. Hsu, W.C., et al.: Multiple-wearable-sensor-based gait classification and analysis in patients with neurological disorders. Sensors **18**(10), 3397 (2018)
11. Huan, J., et al.: A wearable skin temperature monitoring system for early detection of infections. IEEE Sens. J. **22**(2), 1670–1679 (2022). https://doi.org/10.1109/JSEN.2021.3131500
12. Kluding, P.M., et al.: Foot drop stimulation versus ankle foot orthosis after stroke: 30-week outcomes. Stroke **44**(6), 1660–1669 (2013)
13. Kuusik, A., Gross-Paju, K., Maamägi, H., Reilent, E.: Comparative study of four instrumented mobility analysis tests on neurological disease patients. In: 2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops, pp. 33–37. IEEE (2014)
14. Lavado, D.M., Vela, E.A.: A wearable device based on IMU and EMG sensors for remote monitoring of elbow rehabilitation. In: 2022 E-Health and Bioengineering Conference (EHB), pp. 1–4 (2022). https://doi.org/10.1109/EHB55594.2022.9991526
15. Li, R., Song, C., Wang, D., Meng, F., Wang, Y., Tang, Q.: A novel approach for gait recognition based on CC-LSTM-CNN method. In: 2021 13th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), pp. 25–28. IEEE, Hangzhou, China, August 2021. https://doi.org/10.1109/IHMSC52134.2021.00014
16. Matsumoto, S., et al.: Effect of functional electrical stimulation in convalescent stroke patients: a multicenter, randomized controlled trial. J. Clin. Med. **12**(7) (2023). https://doi.org/10.3390/jcm12072638. https://www.mdpi.com/2077-0383/12/7/2638
17. Miller, L., et al.: Functional electrical stimulation for foot drop in multiple sclerosis: a systematic review and meta-analysis of the effect on gait speed. Arch. Phys. Med. Rehabil. **98**(7), 1435–1452 (2017)
18. Moura Coelho, R., Gouveia, J., Botto, M.A., Krebs, H.I., Martins, J.: Real-time walking gait terrain classification from foot-mounted inertial measurement unit using convolutional long short-term memory neural network. Expert Syst. Appl. **203**, 117306 (2022). https://doi.org/10.1016/j.eswa.2022.117306
19. Murray, M.: Gait as a total pattern of movement. Am. J. Phys. Med. **46**(1), 290–333 (1967)
20. Napieralski, J.A., et al.: Classification of subjects with balance disorders using 1D-CNN and inertial sensors. IEEE Access **10**, 127610–127619 (2022). https://doi.org/10.1109/ACCESS.2022.3225521

21. O'Dell, M.W., et al.: Response and prediction of improvement in gait speed from functional electrical stimulation in persons with poststroke drop foot. PM&R **6**(7), 587–601 (2014). https://doi.org/10.1016/j.pmrj.2014.01.001. https://onlinelibrary.wiley.com/doi/abs/10.1016/j.pmrj.2014.01.001
22. Peishun, C., Haiwang, Z., Taotao, L., Hongli, G., Yu, M., Wanrong, Z.: Changes in gait characteristics of stroke patients with foot drop after the combination treatment of foot drop stimulator and moving treadmill training. Neural Plast. **2021**, 1–5 (2021). https://doi.org/10.1155/2021/9480957
23. Pirker, W., Katzenschlager, R.: Gait disorders in adults and the elderly. Wien. Klin. Wochenschr. **129**(3), 81–95 (2017)
24. Pérez-Ibarra, J.C., Siqueira, A.A.G., Krebs, H.I.: Real-time identification of gait events in impaired subjects using a single-IMU foot-mounted device. IEEE Sens. J. **20**(5), 2616–2624 (2020). https://doi.org/10.1109/JSEN.2019.2951923
25. Ramdhani, R.A., Khojandi, A., Shylo, O., Kopell, B.H.: Optimizing clinical assessments in Parkinson's disease through the use of wearable sensors and data driven modeling. Front. Comput. Neurosci. **12**, 72 (2018)
26. Robles, D., et al.: Real-time gait pattern classification using artificial neural networks. In: 2022 IEEE International Workshop on Metrology for Living Environment (MetroLivEn), pp. 76–80 (2022). https://doi.org/10.1109/MetroLivEnv54405.2022.9826927
27. Rostovski, J., Krivošei, A., Kuusik, A., Ahmadov, U., Alam, M.M.: SVM time series classification of selected gait abnormalities. In: Ur Rehman, M., Zoha, A. (eds.) BODYNETS 2021. LNICS, vol. 420, pp. 195–209. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-95593-9_16
28. Rostovski, J., Krivošei, A., Kuusik, A., Alam, M.M., Ahmadov, U.: Real-time gait anomaly detection using SVM time series classification. In: 2023 International Wireless Communications and Mobile Computing (IWCMC), pp. 1389–1394 (2023). https://doi.org/10.1109/IWCMC58020.2023.10182666
29. Roth, N., et al.: Do we walk differently at home? A context-aware gait analysis system in continuous real-world environments. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1932–1935 (2021). https://doi.org/10.1109/EMBC46164.2021.9630378
30. Saboor, A., et al.: Latest research trends in gait analysis using wearable sensors and machine learning: a systematic review. IEEE Access **8**, 167830–167864 (2020)
31. Sayeed, M.A., Nasrin, F.: An edge-computing platform for low-latency and low-power wearable medical devices for epilepsy. In: 2023 IEEE Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS), pp. 1–4 (2023). https://doi.org/10.1109/WMCS58822.2023.10194265
32. Shull, P.B., Xia, H., Charlton, J.M., Hunt, M.A.: Wearable real-time haptic biofeedback foot progression angle gait modification to assess short-term retention and cognitive demand. IEEE Trans. Neural Syst. Rehabil. Eng. **29**, 1858–1865 (2021). https://doi.org/10.1109/TNSRE.2021.3110202
33. Singh, Y., Vashista, V.: Gait classification with gait inherent attribute identification from Ankle's kinematics. IEEE Trans. Neural Syst. Rehabil. Eng. **30**, 833–842 (2022). https://doi.org/10.1109/TNSRE.2022.3162035
34. Tavenard, R., et al.: Tslearn, a machine learning toolkit for time series data. J. Mach. Learn. Res. **21**(118), 1–6 (2020)
35. Wang, F.C., Li, Y.C., Kuo, T.Y., Chen, S.F., Lin, C.H.: Real-time detection of gait events by recurrent neural networks. IEEE Access **9**, 134849–134857 (2021). https://doi.org/10.1109/ACCESS.2021.3116047

36. Wang, L., Sun, Y., Li, Q., Liu, T., Yi, J.: IMU-based gait normalcy index calculation for clinical evaluation of impaired gait. IEEE J. Biomed. Health Inform. **25**(1), 3–12 (2021). https://doi.org/10.1109/JBHI.2020.2982978
37. Zhang, M., Wang, Q., Liu, D., Zhao, B., Tang, J., Sun, J.: Real-time gait phase recognition based on time domain features of multi-MEMS inertial sensors. IEEE Trans. Instrum. Meas. **70**, 1–12 (2021). https://doi.org/10.1109/TIM.2021.3108174

Research for JYU: An AI-Driven, Fully Remote Mobile Application for Functional Exercise Testing

Neil Cronin[1(✉)], Ari Lehtiö[2], and Jussi Talaskivi[2]

[1] Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland
neil.j.cronin@jyu.fi
[2] Digital Services, University of Jyväskylä, Jyväskylä, Finland

Abstract. As people live longer, the incidence and severity of health problems increases, placing strain on healthcare systems. There is an urgent need for resource-wise approaches to healthcare. We present a system built using open-source tools that allows health and functional capacity data to be collected remotely. The app records performance on functional tests using the phone's built-in camera and provides users with immediate feedback. Pose estimation is used to detect the user in the video. The x, y coordinates of key body landmarks are then used to compute further metrics such as joint angles and repetition durations. In a proof-of-concept study, we collected data from 13 patients who had recently undergone knee ligament or knee replacement surgery. Patients performed the sit-to-stand test twice, with an average difference in test duration of 1.12 s (range: 1.16–3.2 s). Y-coordinate locations allowed us to automatically identify repetition start and end times, while x, y coordinates were used to compute joint angles, a common rehabilitation outcome variable. Mean difference in repetition duration was 0.1 s (range: −0.4–0.4 s) between trials 1 and 2. Bland-Altman plots confirmed general test-retest consistency within participants. We present a mobile app that enables functional tests to be performed remotely and without supervision. We also demonstrate real-world feasibility, including the ability to automate the entire process, from testing to analysis and the provision of real-time feedback. This approach is scalable, and could form part of national health strategies, allowing healthcare providers to minimise the need for in-person appointments whilst yielding cost savings.

Keywords: Computer Vision · Remote Rehabilitation · Mobile Health App

1 Introduction

The global population continues to grow, and with advances in living standards, life expectancy has gradually increased in most nations over recent decades. As people live longer, the incidence and severity of health problems increases, placing strain on national health systems. There is an urgent need for resource-wise approaches to healthcare that free up medical staff to focus on life-threatening cases, whilst also minimising wait

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 279–287, 2024.
https://doi.org/10.1007/978-3-031-59091-7_18

times for patients with less critical needs. The recent COVID pandemic also highlights the need for remote solutions that reduce the need to get in-person access to a healthcare professional [1].

Recent advances in technology, particularly in the field of AI, have made the prospect of remote healthcare solutions feasible. For example, it is now possible to monitor heart rate dynamics, blood pressure and sleep behaviour via smartphone applications [2, 3]. However, these applications tend to be fragmentary and narrow in scope, focusing on a single variable or function, and thus only giving a limited window into a person's health status. Moreover, existing solutions are almost always proprietary, making it difficult to scale up their use or add new functionality.

Open-source tools would give healthcare providers more opportunities to monitor patient function, whilst also giving patients more freedom and flexibility, by allowing new tools and functionality to be developed based on patient needs. Moreover, in line with the recent rise in citizen science applications [4], mobile apps enable healthcare interventions to reach more people, including those in poorer or more remote regions. This in turn allows data to be collected from larger and more diverse populations. The aims of this paper are: 1) to present a system built using open-source tools that allows health and functional capacity data to be collected remotely with minimal user input, and that also provides patients with immediate test feedback. 2) To demonstrate a practical use case of this approach in a hospital setting.

State-of-the-Art. Several studies have presented methods for remote testing of functional performance in different clinical groups. For example, Brooks [5] developed a self-administered 6-min walking test mobile application (SA-6MWTapp) for independent use at home, and Hwang [6] did similar work using video conferencing to supervise the tests. Boswell [7] created a smartphone app to examine sit-to-stand test performance remotely. As well as the sit-to-stand test, the timed up and go test and step tests can be performed at home by patients with chronic respiratory disease [8–10]. Netz [11] used smartphone accelerometer data to remotely examine balance, strength and flexibility. Hellsten [12] summarised the potential of markerless AI algorithms for remote monitoring, and these applications have proliferated in recent years [13–15]. However, existing applications for remote testing often focus on a single task, limiting their practical value as monitoring tools. Moreover, data are typically post-processed, so participants may not receive performance feedback, and interactive applications (e.g. gamification) are not possible.

2 Methods

We have developed a smartphone app that records individuals performing various functional tests using the phone's built-in camera and provides users with immediate feedback. The app currently offers the following functionality: individual sign-in using a QR code; access to different research projects via the app home screen; the ability to perform functional exercise tests fully remotely, and to receive instant feedback about performance. Here we first describe the technical implementation of the app, and then report results from a study performed in a clinical setting.

2.1 Approach

The first key step is to detect a person from a video automatically and in real-time. This can be achieved using a number of existing open-source pose estimation algorithms such as OpenPose [16]. Since we require real-time tracking on current smartphones, we instead use MoveNet [17], which runs with reasonable accuracy in real-time (3–6 Hz; see below). The algorithm detects x, y coordinates of key body landmarks in an image, and the coordinates are used to compute further metrics such as joint angles, distances, rep times/durations etc.

2.2 Technical Details

A schematic of how the app works is shown in Fig. 1. The mobile app is interfaced with open-source software developed at the University of Jyväskylä called Vasara, which is a Hyperautomation platform. From the user's perspective, a test session begins when the user scans a QR code using the phone's camera (QR codes can be supplied to study participants by email, for example). After scanning the code, the user enters some basic demographic data such as their age and body mass, and can then proceed to performing a functional test. In the schematic shown in Fig. 1, there are 2 different tests, but in theory we can add as many as desired. When a test is selected, the user is first presented with an instructional video demonstrating how the test is done. After this, they are given audible instructions via the app, for example "take a seat within view of the camera". Pose estimation is used to check that all body keypoints that need to be visualized are in view before the test can start. If, for example, part of the participant's body is obscured or outside of the camera view, the user is instructed to move forward/backward etc. Once the test is initiated, pose estimation detects the key body landmarks (see right side of Fig. 1), and the x, y coordinates of these points are saved in JSON format. Once the test is finished, the video is deleted immediately, thereby minimising the potential for applicants to be identifiable from the data, as well as minimising the data footprint of the app.

2.3 Use Case – Functional Testing of Patients After Knee Surgery

In this proof-of-concept study, we collected data from patients who had recently undergone knee ligament reconstruction or knee replacement surgery ($n = 13$). As part of their regular appointment with a physiotherapist, patients were offered the opportunity to participate in this study, which involved performing brief functional tests, and repeating each test after a short rest period.

In this study we used the sit-to-stand test as an example. The sit-to-stand test involves the functional movement of rising from a seated position and then sitting again, which is repeated several times. The test is often used as a proxy measure of lower limb muscle strength [18] and is suitable for a wide range of populations, including hip and knee osteoarthritis, and adults of different ages [19]. The test is also well suited to remote applications because it is quick and easy to administer and interpret, whilst requiring minimal equipment and space. Importantly, there is also no need to calibrate the camera since output variables are based on angles rather than distances.

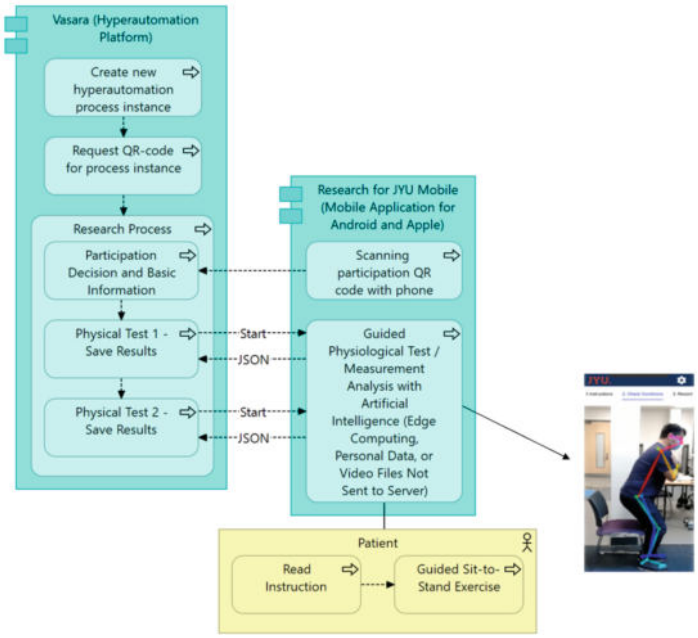


Fig. 1. Schematic of the mobile application. The screenshot to the right is taken from the actual app and shows the detected body landmarks overlaid on the participant in real-time.

Patients who agreed to participate provided written informed consent before testing. For the sit-to-stand test, participants were advised (via the app) to start the test by sitting in the chair. Once they were seated within view of the camera, the following instruction was: "when you are ready to start, raise your hand". This motion was detected with pose estimation, and a countdown (5 to 1) was initiated. Participants were instructed to perform 5 sit-to-stand repetitions to complete the test. Upon completion, the maximum knee joint angle and test time were displayed on the screen. In this study, each participant repeated the test approximately 2 min after the first test, but in theory, tests can be repeated at any interval, for example as part of a long-term intervention protocol performed over several months or even years.

3 Results and Interpretation

The average time difference between the duration of the first and second trials- i.e. the time taken to perform 5 sit-to-stand repetitions- was 1.12 s (range: 1.16–3.2 s). Figure 2A shows examples of y-coordinate trajectories for the right hip during the test for 4 different participants. Most participants performed the movement with high repeatability, although some performed one test clearly faster than the other, likely due to the learning effect associated with an unfamiliar task. The y-coordinates of the key body parts can be used for simple analyses such as identifying the start and end of each rep, or the maximum/minimum height of a body part (e.g. using peak finding or gradient-based techniques), as shown in Fig. 2B. By combining x, y coordinate data with simple

mathematics, we can also compute joint angles (Fig. 2C), which are a common outcome variable in rehabilitation and sports movement analysis. Similarly, by differentiating over time, limb velocities can be computed, which could be useful for biofeedback applications where patients are given a target movement profile to follow.

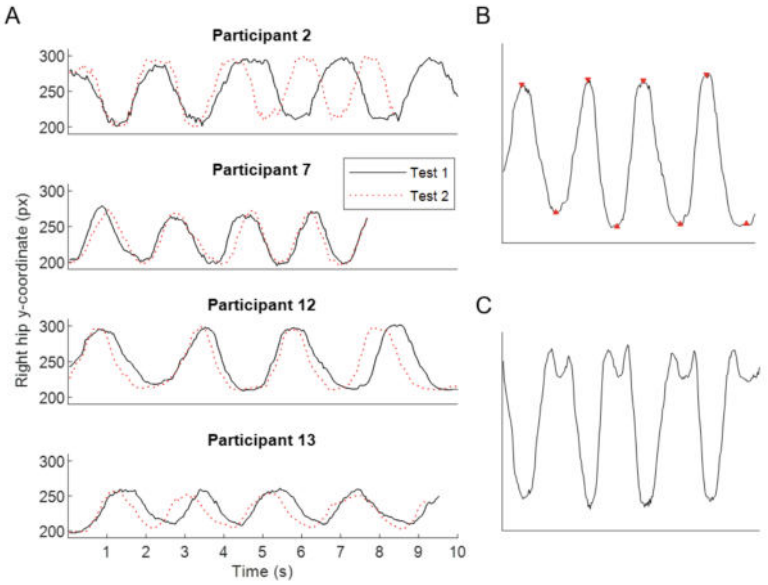


Fig. 2. A: Example of test-retest results from four different participants. Each panel shows the y-coordinate trajectories over time for the right hip, one panel per participant. B: Using peak detection to identify the transitions between ascending and descending motion. C: The corresponding hip joint angle based on the data in B.

As the segregation of individual repetitions within a trial can be automated, it is also easily possible to perform repetition-level analyses. For example, the mean difference in repetition duration- i.e. the time taken to stand up fully and return to the seated position once- was 0.1 s (range: −0.4–0.4 s) between trials 1 and 2. Data for all repetitions by all participants are compared in the Bland-Altman plot [20] in Fig. 3.

From this figure it is clear that the majority of datapoints fall within the limits of agreement, demonstrating general consistency within participants, i.e. between trials 1 and 2. Although not performed here, the analysis can be further enhanced by comparing the duration of the ascending and descending phases of a repetition, which can give important information about muscle function.

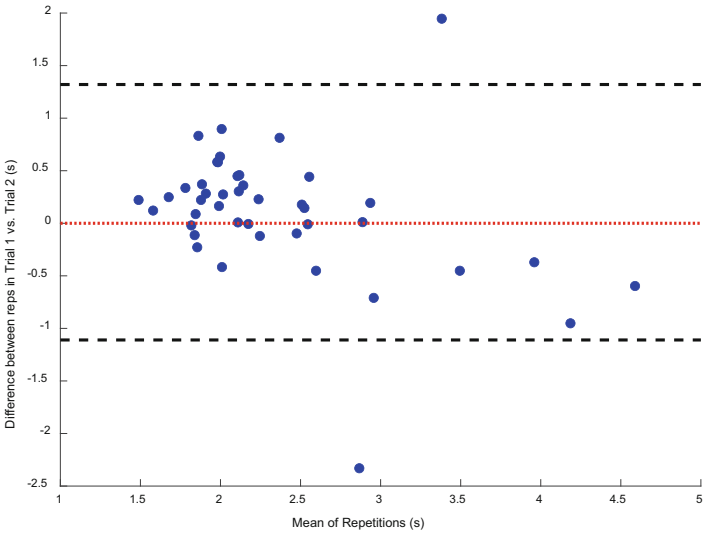


Fig. 3. Bland-Altman plot comparing the individual repetition durations between trials 1 and 2 for all participants. The dotted red line denotes the mean difference between trials 1 and 2 (bias), and the dashed black lines indicate the limits of agreement.

4 Discussion and Future Perspectives

Around 85% of the global population owns a smartphone, which are typically equipped with a camera and movement sensors. Phone literacy is also extremely high and improving continually, including among older adults [21] and in poorer countries [22]. Thus, smartphones have huge potential as health monitoring tools in large populations, enabling the remote completion of functional tests from anywhere. We present a mobile app that enables functional tests to be performed remotely and without supervision. We also present real-world evidence of the feasibility of this approach, including the ability to automate the entire process, from testing to analysis and the provision of real-time feedback. Using computer vision to detect people in images, we can quantify movement metrics such as the time taken to complete a task, joint angles, or limb velocities, with levels of accuracy that generally match those of a human physiotherapist [13, 14]. Our mobile app is also sufficiently user-friendly to be used in a real clinical environment by patients and/or medical staff.

A limitation of our approach is that only angle and repetition-related metrics can easily be extracted. In clinical settings it is often also desirable to compute distances during a test (e.g. maximum distance between left and right ankles). Such metrics require camera calibration, introducing possible scaling errors [14]. Nonetheless, our approach offers several advantages. We have tested our app on 4 different smartphones and as the underlying pose estimation algorithms are quite robust to image quality and resolution, the main performance difference between phones is inference speed, which ranged between 10–30 frames per second, which is satisfactory for clinical applications. We also minimise the risk of GDPR issues as our app performs real-time, on-device analysis and

then discards the video, thus also minimising data storage needs. In future work we will expand the app's functionality in several ways. Most importantly, we will develop strong protocols for handling sensitive data. We will also add more functional tests (e.g. range of motion, balance) and tailor the specific feedback that is given to users immediately after each test. Moreover, we will include the ability to administer e-questionnaires, with the option of completing the forms in writing or by recording spoken responses, which will then be transcribed using large language models.

The approach presented here could be used as a tool to implement follow-up research protocols, such as rehabilitation, training interventions, or monitoring of at-risk groups. Taking a citizen science approach will help to grow our datasets, and in turn could enable new applications in the future such as predictive modelling. By allowing patients to give consent dynamically, we can also facilitate biobank type applications, which could allow longitudinal profiles of individuals to be built up. Finally, this approach is scalable, and could form part of national health strategies, allowing healthcare providers to minimise the need for in-person appointments, and freeing up medical staff for other tasks. This could in turn lead to significant cost savings for healthcare providers.

The infrastructure on which our app is built is highly customisable. We recently used the core components of the app and added new functionality to produce an entirely different application that allows users to record the sound of birdsong outdoors. The soundtrack is then fed to an AI algorithm trained to detect different species. To date, this app has attracted over 140,000 users, who have collectively submitted over 3 million recordings. In ongoing work funded by the Jane and Aatos Erkko Foundation, we are developing several more mobile citizen science applications that span numerous domains, including human-nature interaction and learning difficulties in children. By developing open-source tools, we aim to increase the broad participation of regular citizens in scientific research.

Acknowledgments. The authors thanks Dr. Juhani Multanen for assistance with study logistics and the staff physiotherapists at Nova hospital for helping with data collection. This study did not receive any separate funding.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Uimonen, M., Kuitunen, I., Paloneva, J., et al.: The impact of the COVID-19 pandemic on waiting times for elective surgery patients: a multicenter study. PLoS One **16** (2021). https://doi.org/10.1371/JOURNAL.PONE.0253875
2. Christien Li, K.H., White, F.A., Tipoe, T., et al.: The current state of mobile phone apps for monitoring heart rate, heart rate variability, and atrial fibrillation: narrative review. JMIR Mhealth Uhealth **7** (2019). https://doi.org/10.2196/11606
3. Holz, C., Research, M., Edward Wang, R.J.: Glabella: continuously sensing blood pressure behavior using an unobtrusive wearable device. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **1**, 1–23 (2017). https://doi.org/10.1145/3132024

4. Marks, L., Laird, Y., Trevena, H., et al.: A scoping review of citizen science approaches in chronic disease prevention. Front. Public Health **10**, 743348 (2022). https://doi.org/10.3389/FPUBH.2022.743348/BIBTEX
5. Brooks, G.C., Vittinghoff, E., Iyer, S., et al.: Accuracy and usability of a self-administered 6-minute walk test smartphone application. Circ. Heart Fail. **8**, 905–913 (2015). https://doi.org/10.1161/CIRCHEARTFAILURE.115.002062
6. Hwang, R., Bruning, J., Morris, N.R., et al.: Home-based telerehabilitation is not inferior to a centre-based program in patients with chronic heart failure: a randomised trial. J. Physiother. **63**, 101–107 (2017). https://doi.org/10.1016/J.JPHYS.2017.02.017
7. Boswell, M.A., Kidziński, Ł., Hicks, J.L., et al.: Smartphone videos of the sit-to-stand test predict osteoarthritis and health outcomes in a nationwide study. npj Digit. Med. **2023**, 6:1–7 (2023). https://doi.org/10.1038/s41746-023-00775-1
8. Holland, A.E., Malaguti, C., Hoffman, M., et al.: Home-based or remote exercise testing in chronic respiratory disease, during the COVID-19 pandemic and beyond: a rapid review. Chron. Respir. Dis. **17** (2020). https://doi.org/10.1177/1479973120952418
9. Vilarinho, R., Caneiras, C., Montes, A.M.: Measurement properties of step tests for exercise capacity in COPD: a systematic review. Clin. Rehabil. **35**, 578–588 (2021). https://doi.org/10.1177/0269215520968054/ASSET/IMAGES/LARGE/10.1177_0269215520968054-FIG1.JPEG
10. Vaidya, T., Chambellan, A., de Bisschop, C.: Sit-to-stand tests for COPD: a literature review. Respir. Med. **128**, 70–77 (2017). https://doi.org/10.1016/j.rmed.2017.05.003
11. Netz, Y., Yekutieli, Z., Arnon, M., et al.: Personalized exercise programs based upon remote assessment of motor fitness: a pilot study among healthy people aged 65 years and older. Gerontology **68**, 465–479 (2022). https://doi.org/10.1159/000517918
12. Hellsten, T., Karlsson, J., Shamsuzzaman, M., Pulkkis, G.: The potential of computer vision-based marker-less human motion analysis for rehabilitation. Rehabil. Process Outcome **10**, 1–12 (2021). https://doi.org/10.1177/11795727211022330
13. Hannink, E., Mansoubi, M., Cronin, N., et al.: Computer-vision aided functional movement measurement in people with and without axial spondyloarthritis – validation and feasibility study protocol. OSF Preprints (2021). https://doi.org/10.31219/OSF.IO/HSV7P
14. Cronin, N., Mansoubi, M., Hannink, E., et al.: Accuracy of a computer vision system for estimating biomechanical measures in axial spondyloarthropathy patients and healthy subjects. Clin. Rehabil. **37**, 1087–1098 (2022)
15. Mazéas, A., Blond, M., Chalabaev, A., Duclos, M.: Validity and reliability of an app-based medical device to empower individuals in evaluating their physical capacities. PLoS One **18** (2023). https://doi.org/10.1371/JOURNAL.PONE.0289874
16. Cao, Z., Hidalgo Martinez, G., Simon, T., et al.: OpenPose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans. Pattern Anal. Mach. Intell. 1–1 (2019). https://doi.org/10.1109/tpami.2019.2929257
17. The TensorFlow Hub Authors: MoveNet (2021). https://github.com/tensorflow/docs/blob/master/site/en/hub/tutorials/movenet.ipynb
18. Csuka, M., McCarty, D.J.: Simple method for measurement of lower extremity muscle strength. Am. J. Med. **78**, 77–81 (1985). https://doi.org/10.1016/0002-9343(85)90465-6
19. Bennell, K., Dobson, F., Hinman, R.: Measures of physical performance assessments: Self-Paced Walk Test (SPWT), Stair Climb Test (SCT), Six-Minute Walk Test (6MWT), Chair Stand Test (CST), Timed Up & Go (TUG), Sock Test, Lift and Carry Test (LCT), and Car Task. Arthritis Care Res (Hoboken) **63**(Suppl 11) (2011). https://doi.org/10.1002/ACR.20538
20. Bland, J.M., Altman, D.G.: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet **1**, 307–310 (1986)
21. Zickuhr, K., Madden, M.: Main Report | Pew Research Center (2012)

22. Rotondi, V., Kashyap, R., Pesando, L.M., et al.: Leveraging mobile phones to attain sustainable development. Proc. Natl. Acad. Sci. U S A **117**, 13413–13420 (2020). https://doi.org/10.1073/PNAS.1909326117/

Exploring and Extending Human-Centered Design to Develop AI-Enabled Wellbeing Technology in Healthcare

Laura Tahvanainen[1,2(✉)], Birgitta Tetri[1,3], and Outi Ahonen[1]

[1] Laurea University of Applied Sciences, Vantaa, Finland
laura.tahvanainen@laurea.fi
[2] University of Lapland, Rovaniemi, Finland
[3] University of Helsinki, Helsinki, Finland

Abstract. Digital transformation and digitalisation are rapidly affecting the society. The gradually increasing applications of different types of AI into solutions and services are welcome, but there are associated risks. These include, for example, within human aspects of care undermining fundamental rights, ethical considerations, sustainability, and policies and regulations. This change permeates every societal level, but it is especially evident in the healthcare sector due to the ageing population and shortage of professionals. This situation also places pressure on the development of competencies among healthcare professionals. A human-centered approach in design and design methods can promote the development of AI-based solutions in transdisciplinary and cross-disciplinary processes encompassing numerous stakeholders, scientific orientations, and perspectives. There is a need for research and evaluation of Human-Centered Design (HCD) processes and design methods to develop and gain more insights for future development.

This study was conducted as research through design. It aimed to elucidate the application and insights gained from the adopted Service design process for AI-enabled services and HCD approach while developing AI-empowered solution, Voima-chatbot. One of this research's main conclusions and realization is the shift from purely HCD towards Life-Centered design of AI-enabled solutions with a human-in-the-loop. In addition, this project increased the understanding of the deep importance of having a transdisciplinary dialogue with developers during the process of developing digital well-being devices and combining different professional competencies to achieve the best working solutions.

Keywords: Human-centered design · AI-enabled solution · Transdisciplinary · Design process · Healthcare · eHealth

1 Introduction

In Europe, the working conditions in healthcare are undergoing major changes as digital transformation extends more and more widely to different job descriptions. The use of mobile technologies, telemedicine and other digital tools intended to support clinical

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 288–306, 2024.
https://doi.org/10.1007/978-3-031-59091-7_19

decisions has improved health workers' performance and mental health, as well as their competencies [1]. Digitalisation also requires new ethical reflection skills from healthcare professionals to understand factors of guiding and promoting ethical approaches [2]. There is an urgent need for effective digital tools and technologies and an unprecedented rush to implement eHealth services, including telemedicine consultation and digital contact tracing, in countries across the WHO Region. Strategic alignments are made to support this change [1, 3]. The need for digitalisation is due to the ageing of the population in Europe [4], the shortage of healthcare professionals in all occupational groups [5, 6] and the aftermath of the COVID-19 pandemic [1]. Different kinds of AI applications are predicted to have a growing role in healthcare and wellbeing devices and services [7]. The recent European Artificial Intelligence Act is going to change the use of AI within the EU region in the upcoming years [8].

The maturity of healthcare information management varies from country to country. Also, citizens' skills in the use of digital technology vary in different countries. In Finland, almost 80% of adult age of citizens have at least a low level of digital skills [7].

There is also variation in digitalisation and informatics skills in different countries and between professions. The introduction and smooth use of new technology require expertise from the individual, but the maturity level of digitalisation in society and organizations is also important [9]. In healthcare, devices must meet the regulation for a medical device [10, 11], but wellbeing technology is also widely used by citizens, and they utilize the information they produce to maintain their own well-being [12].

Digitalisation is most welcome, but there are risks involved, for example, in terms of human aspects of care and undermining of fundamental rights [7]. Multiple health-related institutions and stakeholders, including World Health Organization (WHO), are promoting the adoption and scale-up of digital health technologies (DHT) innovations worldwide. These promotional initiatives aim to translate scientific research into action and enhance knowledge through scientific engagement, assessing and linking geographical needs with innovation pipelines, and implementing practical approaches that balance the benefits and risks of DHTs [13].

Studies have shown that mobile applications (apps) can effectively support lifestyle-related health. Demographic and personal factors of the target group should be considered when developing health apps. The inclusion of appropriate functionalities and their personalization can ensure a high uptake of health apps in workplaces [14].

Conversational AI and chatbots have been used in the last decade to improve access to mental health services [15–21]. Chatbots are automated systems which replicate users' behavior on one side of the chat communication. They are mimic systems which imitate the conversations between two individuals [22]. Chatbots can facilitate interactions with those who are reluctant to seek mental health advice due to stigmatization and allow more conversational flexibility [15, 16]. Threats to the chatbot include the cost of cloud services, the still-developing field of AI, and the unethical over-imitation of a human therapist or its replacement [16, 21].

When designing medical devices and wellness technology, it is crucial to consider evidence-based design and experience-based approaches in service design [23]. The ethical perspective is also essential [2], particularly during the design process, which involves transdisciplinary cooperation with professionals from various sectors [24].

Service design can offer a method to research and develop AI-enabled solutions in the complex healthcare sector. These approaches can inspire and support individuals to participate in the development process [25–27]. HCD is the design approach that centers people and their needs, motivations, emotions, behavior, and perspective in the development of a design. Both users and service provider stakeholders are involved in design activities during and potentially after the service design process, supporting the change that co-design brings [27]. Multiprofessional, cross-sectoral healthcare involving multiple care system levels is a design context that requires context-specific knowledge, such as evidence-based care and specific design competencies, to include the perspectives of diverse actors in design processes [26].

Service design can promote transdisciplinary and cross-disciplinary processes encompassing numerous stakeholders, scientific orientations, and perspectives. It can clarify how to work together to ensure all aspects are considered when innovating new or developing existing technologies [24].

This study aims to elucidate the application and insights gained from the Service design process with HCD approach while developing an AI-enabled empowering solution, Voima-chatbot. The paper provides a detailed account of the development phases, using the case of Voima-chatbot as an exemplar. The objective is to enhance understanding of the feasibility of the service design process with HCD approach in developing AI-enabled technological solutions and point out future research insights.

The research question is:

What are the implications and results of applying Human-Centric design in developing an AI-enabled technological solution?

2 Methods of HCD in Developing AI-Enabled Technical Solutions

This study was conducted as research through design, meaning that design was an integral part of the process, providing both the data for the research and the practical artefacts from the workshops and other interventions, such as the ideation questionnaire. By using research through design [28, 29], a dual benefit was obtained: the design process with HCD approach, along with the co-designed artefacts, helps to better understand both the factors affecting the development of an AI-enabled chatbot and gaining understanding of the healthcare workers occupational well-being.

2.1 Human-Centered Design (HCD) and Service Design

The HCD is the design approach that centers people and their needs, motivations, emotions, behavior, and perspective in the development of a design. HCD is a shift of viewing humans not as a part of the system but central in every aspect of the design. HCD has a long history, and it can play an essential role in dealing with today's complex care challenges [30, 31].

In HCD, as in all design disciplines using HCD principles, designers rely heavily on the tools, methods and insights from the Human Factor discipline, as illustrated by the definition of HCD by the International Standards Organization (ISO): 'Human-Centered Design is an approach to interactive systems development that aims to make

systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors/ergonomics, usability knowledge, and techniques [32]. HCD includes many methods but is essentially a frame of reference and a value system to be considered and applied by the designer. HCD begins with a deep respect for the user, and a realization that the user is the most important partner in design [33].

Human-Centered Artificial Intelligence (HCAI) is based on the concept of human-centered technology development and combines HCD, artificial intelligence, and machine learning. This includes the fundamental starting points of understanding user needs and the contextual and sociotechnical factors of system design, as well as introducing new ones specific to AI as a technology. Designing AI with a human focus is crucial for end-users' well-being and for addressing ethical issues that may lead to unwanted societal-level consequences [34–36]. HCAI advocates for the development of AI applications that are trustworthy, usable, and based on human needs. Many Human-Centered AI principles include explainable, transparent, ethical, fair, trustworthy, responsible, and sustainable AI [34, 36].

Awad et al. [37] state that these rationalistic guidelines provide advice on the development (process) and application (product) of trustworthy, ethical, and robust AI. However, such general guidelines do not represent real-world complexity when laws and policies often evolve slower than technological development. Ethical principles and moral choices are not universal as surveyed and identified. Robustness does not represent real-world complexity as the social impact of AI is hard to predict or foresee. The humanistic design perspective may provide a more suitable approach to examining the societal impact of AI as laws may not be up-to-date, universal principles cannot answer context-specific ethical questions, and robustness does not prevent unintended consequences.

More research is needed to develop the design processes, methods, tools and HCD approach when dealing with today's challenges with digital AI-enabled services in complex healthcare contexts.

It is also posited that HCD is prone to sampling bias by using methods that often rely on studying a relatively small sample in-depth. By default, not everyone can participate in the sessions, resulting in under- or overrepresentation of certain groups (a selection bias). End-user input might be biased and limited, leading to an overreliance on fresh end-user input. End-users are only a subset of the people who should be heard during eHealth design, and HCD tends to overlook ethical, societal, and political aspects [38, 39]. Developing diverse design teams can prevent machine biases in the design of AI systems. Building AI systems that overcome biases is not only a matter of having more diverse, and diversity-minded design teams as AI systems themselves can help identify for example gender and racial biases [31]. These are all significant aspects and objectives for developing the design process.

'Design' is a broadly defined term used for both the process of designing and the outcome of that process. Service design employs the double diamond model and design thinking and methods for designing new services [40]. The development project adapted and utilized the current service design process for AI-enabled services from the original version created by Jylkäs et al. [41, 42] seen in the Fig. 1. This process model which is based on this original double-diamond model [40].

This process has three layers in the 10-phase service design process: business, design, and technology. The case example of "Voima-chatbot" development is used to describe how Human-Centric design approach was utilized in the service design process for AI-enabled services and what was learned during the development process.

Jylkäs et al. [41, 42] observed that the 10 process phases— 'discover', 'define', 'ideate', 'design', 'prototype', 'test', 'develop', 'implement', 'operate', and 'scale'—are more sufficient when communicating the main activities in designing AI assistants.

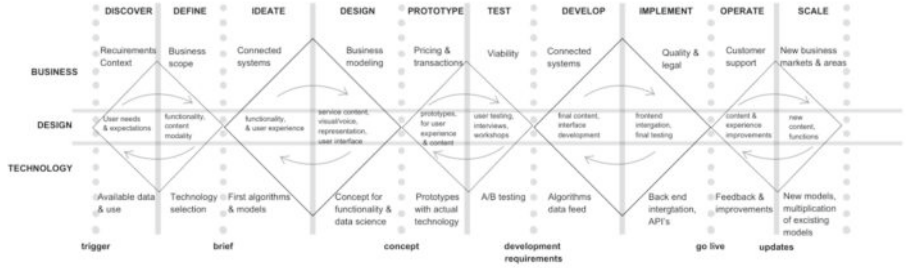


Fig. 1. Service design process for AI-enabled services. (Picture: Laura Tahvanainen, 2024, adapted from the original version created by Jylkäs et al., 2019 under the CC BY-NC-SA 4.0 DEED license [41, 42].

Designers need to recognize their role, ideology, and socio-economical processes in which they are embedded to design AI systems beneficial for society [43]. It is found that in many of the companies there is a clear separation between the AI and UX (User experience) teams. UX practitioners are not considered to be a part of the AI team, nor are they involved in the early-phase development [44].

The phases of the Service design process with the Human-Centric design approach and the technological development process will be presented later in the intervention chapter of this article. For the transdisciplinary design perspective and process it is important to notice that there are different ways and methods to measure and follow development and realization of things in different disciplines and approaches. For example, Technology Readiness Levels (TRL) are a type of measurement system used to assess the maturity level of a particular technology. There are nine technology readiness levels, with TRL 1 being the lowest and TRL 9 the highest. The TRL scale was developed at the National Aeronautics and Space Administration (NASA) in the 1970s as a standardized technology maturity assessment tool for complex system development [45].

By incorporating the Human-centered approach in the design of AI systems, the focus of the design challenges moves away from purely technical problems to the enhancement and support of human capabilities through the AI system [31].

3 Case Intervention: Applying HCD and Technological Development Process to Develop Voima-Chatbot

The development process of Voima-chatbot was part of a wider project that focused on the well-being of healthcare workers with pre-existing system stresses relating to resource constraints, crisis management, growing demand, recruitment, and retention. The project involved several phases from 2021 to 2023, with different methods, participants, and results (see Table 2).

Voima-chatbot was developed to build a scalable digital well-being service (non-medical device) that supports the well-being of healthcare professionals with empowering and solution-focused methods. Voima-chatbot utilizes an asset-based approach, which means that it works within the individual's own world of meanings, supporting a functional interpersonal relationship and pursuits towards activating one's own abilities, skills, strengths, and assets to enable a positive change [46, 47].

Voima-chatbot is an AI-enabled technological solution that uses a conversational AI platform to operate. Conversational AI is based on several advanced technological components, such as Natural Language Processing (NLP), Machine Learning, intent recognition, entity extraction and speech-to-text converters [48]. Voima-chatbot is not a medical device or therapy [11]. Its area of use is in the early prevention level as a well-being device.

Table 1 provides a comprehensive overview of the development process in the project, detailing the various phases, methods used, participants involved, and the deliverables at each stage. This chapter highlights the current Service design process for AI-enabled services adhered to an HCD approach, but also the problems that were resolved or not resolved with HCD approach and what are the findings for future development of service design process for AI-enabled services with HCD approach in healthcare. This model is illustrated earlier in Fig. 1.

Next, this article presents and evaluates the phases seen in Table 1. Based on the Service Design process with Human-Centered approaches and HCAI principles.

3.1 Discover and Define Phases of the Development Process

During the discover and define phases HCD methods like workshops and questionnaires were conducted with healthcare professionals and students to identify factors affecting occupational wellbeing and to gather content for the chatbot's intent tree. A stakeholder map was created, and existing applications and chatbots were benchmarked and tested. Using qualitative analysis helped design an investigation of a phenomenon of interest and helped construct the intent tree for the chatbot. The stakeholders recognized were healthcare professional, healthcare organization, project group, technical development group, the server holder, regulation, and legislation of the AI-enabled chatbots and data security and safety regulation.

The qualitative desktop study revealed that prior research [16, 21] has gathered perceptions and viewpoints about chatbots. This helped the project group to build understanding about the chatbots and helped with the ideation questionnaire. The key discoveries encompass positive and negative aspects for developing chatbot and opportunities for healthcare sector.

Table 1. Development phases of the chatbot with Technological readiness level (TRL)

Design Phase (timing) + TRL	Method (Level)	Participants	Results
Discover (2021–2022)	Workshops in healthcare organizations (Design)	Healthcare professionals n = ca. 100	Enhancing and hindering things affecting the occupational well-being of the healthcare workers, Content for the intent tree in the chatbot
Discover (2021)	Virtual workshops (Design)	Healthcare professionals n = 30	Enhancing and hindering things affecting the occupational well-being of the healthcare workers. Content for the intent tree in the chatbot
Discover (2021)	Qualitative questionnaires together with project partner (Design)	Students n = 437 Healthcare professionals n = ca. 4000	Enhancing and hindering things affecting the occupational well-being of the healthcare workers. Content for the intent tree in the chatbot
Discover (2021)	Qualitative Desk-top study on previous research (Design)	Project group and student's thesis work done in the project n = ca. 20	Enhancing and hindering things affecting the occupational well-being of the healthcare workers
Discover (2022) TRL 1	Stakeholder map (Design)	Project group n = ca. 6	Stakeholders for developing the chatbot
Discover (2022) TRL 1	Desk-top study on research and Benchmarking and testing of the existing applications and chatbots (Business)	Project group n = ca. 6	Information and experiences on using chatbot for supporting mental well-being, existing applications and chatbots

(*continued*)

Table 1. (*continued*)

Design Phase (timing) + TRL	Method (Level)	Participants	Results
Define (2022)	Ideation questionnaire of the chatbot (Design, business)	Healthcare professionals and students n = 77 (n = 64 for chatbot persona)	Information and feedback from the focus group about the idea of the chatbot, user insights for the use of the chatbot and the chatbot persona
Define (2022) TRL 1	Mockup (Design)	Project group n = ca. 6	The basic idea of the chatbot, including basic visuals and conversation flow
Define (2022)	Ethical considerations (Technology, design, business)	Project group and organization n = ca. 8	Anonymous service, service level of the chatbot (not a medical device)
Define (2022)	Meaning and use area of the chatbot (Business)	Project group n = 10	Identifying the use area of the chatbot
Ideate (2022)	Ideation questionnaire of the chatbot (Design, business)	Healthcare professionals and students n = 70	Information and feedback from the focus group about the idea of the chatbot, user insights for the use of the chatbot and for the chatbot persona
Design (2022)	Chatbot character, name, fonts, colors, way of talking (Technology, design, business)	Project group and Bachelor students n = 15	Chatbot persona
Design (2022) TRL 2	Empowering conversation flow (design, technology)	Project group and technical professionals n = 15	Basic/ground idea for the empowering conversation flow
Design (2022)	Ideating and designing digital service paths for the Chatbot (Design, business)	Master's students n = 30	Integration ideas to other services, data management plans

(*continued*)

Table 1. (*continued*)

Design Phase (timing) + TRL	Method (Level)	Participants	Results
Prototype (2022) TRL 3	Mockup (Design, technology)	Project group n = ca. 6	The basic idea of the bot, including basic visuals and conversation flow
Prototype (2022) TRL 1	Testing empowering conversations online person-to-person (Design, technology)	Healthcare professionals and students n = 10	Information for the empowering conversation flow
Prototype (2022) TRL 5	Empowering conversations in Slack platform person-to-person (Design, technology)	Project group n = ca. 4, Healthcare professionals n = 10 and students n = 20	Information for the empowering conversation flow, test- and training data for the intent tree in the chatbot
Develop (2023) TRL 2 &TRL 6	Technical development of empowering conversation flow (Technology)	Project groups & technical professionals n = ca. 15	Technical scope for the empowering conversation in the chatbot platform
Develop (2022–2023) TRL 6	Intent tree in chatbot platform (Design, technology)	Project group & technical professionals n = ca. 15	Intents in the intent tree in the chatbot platform
Develop (2023) TRL 6	Test & training data for the chatbot platform (Technology)	Project group & technical professionals n = ca. 15	Only some of the intents were tested during the first user testing
Implement (2023)	Information safety and security of the chatbot (Technology, design, business)	Safety and security professionals from the technical side and project group n = ca. 15	Data collection and storage, privacy statement, anonymity
Implement (2023) TRL 7	User testing (Technology, design, business)	Healthcare professionals and students n = ca. 90 n = 11 online testing healthcare units and n = 2 onsite testing healthcare units	User experience, usability, feedback on empowering conversation, improvements, test-/training data

(*continued*)

Table 1. (*continued*)

Design Phase (timing) + TRL	Method (Level)	Participants	Results
Implement (2023)	Qualitative feedback questionnaire, open questions, and Likert 1–5 (Technology, design, business)	Healthcare professionals and students n = 24	User experience, usability, feedback on empowering conversation, improvements, test-/training data
Implement (2023)	Implementation plan and future insights (Technology, design, business)	Project group and organization	Future development plans and projects, improvements, in.egration

The benchmarking resulted in the identification of several applications such as Woebot, an AI-powered chatbot that uses Cognitive Behavioral Principles, Dialectical Behavior Therapy, Mindfulness, Symptom Tracking/Self-Monitoring, Psychoeducation & Information [49]. Wysa, an AI-enabled mental health app, uses Cognitive-behavioral therapy (CBT) techniques, meditation, breathing and mindfulness exercises, and micro-actions to help users build mental resilience skills through its conversational interface [50]. ChatPal, a chatbot application, uses Positive psychology to support users' mental health and wellbeing. The scripts used in the ChatPal chatbot are freely available as an output from the ChatPal project [51].

Based on the benchmarking data, it was found that existing applications and services employ various methodologies and techniques for conversation flow. All applications utilized evidence-based and clinically validated methods such as Cognitive-Behavioral Therapy (CBT) and positive psychology (PP), albeit with different emphases. No applications were found that exclusively used Solution-Focused Brief Therapy (SFBT) or Empowering methods. After testing the existing applications and chatbots, it was observed that conversations quickly led to predetermined answers and conclusions. CBT tends to be more problem-solving oriented, whereas SFBT sessions had significantly higher positive content than CBT sessions [21, 52]. Research and practices around occupational well-being have traditionally been focusing on the problems causing the issues [53, 54]. The asset-based approach is an evidence-based, human-centered approach formed from different theoretical and practical elements connected to each other [23].

These tasks provided general information on chatbots and confirmed the solution-focused and empowering approach in chatbots. This also meant innovation on the chatbot's main conversation flow when the chatbot is asking the questions instead of the person. In healthcare, the importance of evidence-based approaches is paramount.

When comparing these methods and process to Jylkäs et al. [41, 42] model and HCAI principles [36] at this phase it was recognized that methods that were used helped to understand healthcare worker and the context where the chatbot is going to take place. It also provided understanding on what has already been done with supportive

conversational chatbots and AI. Ethical, data regulation and legislation were recognized, but these are currently missing from Jylkäs et al. [41, 42] model. It was also clear that the development process had already began without the actual technology provider. Business level in Jylkäs et al. [41, 42] model in this case is healthcare sector. Our research pointed out the need for an evidence-based approach and early recognition on whether the solution will be a medical device. Stakeholder mapping conducted during this research revealed a demand for new tiers for the model. Some aspects were discerned later during this research, such as sustainability. This indicates a necessity for novel methods for the mapping during the process that can be applied and revisited as needed. Recognizing the product's lifecycle impact at a global or national level is important.

3.2 Ideate and Design Phases of the Development (TRL Level 1)

The ideation questionnaire for a chatbot was designed to gather insights from healthcare professionals and students in diverse age groups (N = 77). This facilitated the acquisition of user feedback regarding the suitability of this solution for the intended objective. No technological innovation can have an impact if it is not adopted [14].

The questionnaire was divided into two parts: the idea of the chatbot and the chatbot persona. The respondents indicated a willingness to use the chatbot across various contexts and preferred platforms for using the chatbot. Feedback on the chatbot mockup was predominantly positive, with the chatbot seen as a tool for reflection, aiding those with difficulty speaking to others, enhancing occupational well-being, and providing quick help. The chatbot persona questions helped shape the Voima-chatbot character and persona and provided insights into prejudices towards the chatbot idea.

During this phase, the project team critically evaluated the questionnaire from a human-centered design viewpoint. They questioned the adequacy of the information gathered and the depth of user understanding achieved. They also scrutinized whether the methods employed sufficiently captured the necessary user needs. This introspection served as a valuable insight into the process.

One of the main findings from phase and feedback was that the chatbot is not a person and the expected lack of empathy. The design of artificial empathy is one of the most essential issues in social robotics. Based on views from developmental robotics, empathic behaviors are expected to be learned through social interactions with humans [55]. This also contributes to the HCAI transparency of AI since it is important for the person to know that there is AI talking and not a real human. The results from the ideation questionnaire were similar to previous research.

3.3 Prototype and Test Phases of the Development Process (TRL Levels 4–6)

The prototyping and testing phases were divided into two parts. The first part involved empowering, asset-based conversations online conducted by master's degree students. Students offered these conversations to healthcare professionals as part of their relevant methodological studies. The second part of the prototyping was in the Slack application utilized for training and prototyping written person-to-person empowering conversations between the master's students and healthcare professionals. The aim was to simulate chatbot conversation and gather insights for the chatbot's conversation flow.

Feedback from the participants in the first prototype was positive, indicating a need for such an approach. The second prototype yielded various results for development and valuable test and training data for the chatbot and helped build an understanding of the interaction and "tone of voice". Challenges arose due to the lack of verbal cues and visible expressions in the online interaction. It was difficult for participants to let go of the idea of playing a "chatbot" and focus solely on the conversation.

The design perspective utilized interventions through prototyping to also examine the emerging ethical dilemmas in the interactions between people and AI systems. The implication of these choices indicates that design researchers need to consider various aspects of human implication in the design experiment beyond merely paying close attention to human and social factors [31].

Prototypes of the solution are efficacious and cost-effective means to evaluate in a realistic context. In this research, these elicited emotions and challenges towards the solution that were advantageous in the development phase. Prototypes also aided in building confidence that this solution is feasible.

3.4 Develop and Implement Phases of the Development (TRL Levels 2–6 and TRL 7)

The technological development and kick-off phase marked the commencement of the AI-powered chatbot platform. Training Conversational AI, not an IT project, involves providing example questions or requests for the neural network to analyze and understand semantics. Modern NLP-optimized networks require only a few questions per intent, an NLP term referring to a user's area of interest or request [48].

The first model of the asset-based foundational conversation in the chatbot platform was collaboratively created and evaluated with the platform provider. This new approach in the conversation flow required development by the platform provider. Existing knowledge from empowering conversation was integrated to build the empowering conversation flow, which typically consists of five stages or phases [47, 56]. A novel approach requiring technical resolution was the chatbot asking questions instead of the person. The intent tree, which can be described as a classification or "catalogue" built on the intents of a user, was structured by combining these five stages [47, 56] to a classification model with two levels of components (Table 2). This intent tree, necessary for the proper functioning of the AI, helped the technical team determine the required technical properties and coding.

User testing was conducted with a closed webpage set up for online testing. Various healthcare units were approached for user testing. The testing period yielded 91 conversations, about half of which were complete. Post-testing, participants were instructed to fill out a questionnaire, including scaled and open questions addressing various aspects of the chatbot. A modest number of participants responded to the questionnaire after testing. The aim was to understand the performance of the empowering conversation and whether users found it meaningful and would recommend the chatbot to others. Feedback indicated that the chatbot's repetitive questioning caused frustration, and some of its questions were difficult to understand.

Questions regarding the chatbot persona received a slightly higher average score. The overall rating of the chatbot was slightly above the midpoint on Likert scale 1–5. Written

Table 2. Empowering foundational conversation flow in the chatbot

Basic Empowering conversation flow
Baseline survey Likert scale 1–5 1. Building a connection Joint phase questions 1.1 I still wish to clarify my challenge 1.2. I wish to continue to goal clarification
2. Desire for change/Goal clarification Joint phase questions 2.1. I still wish to clarify my desire for change 2.2. I wish to continue mapping my assets next 3. Mapping and promoting of the assets and strengths Joint phase questions: 3.1 I still wish to continue mapping my assets 3.2 I wish to continue planning the start of the journey 4. Start of the journey Joint phase questions: 4.1.1. I still wish to think about the change 4.2 I wish to continue 5. Conclusion and review Feedback survey: Likert scale 1–5

feedback revealed that participants noticed the mechanical nature of the conversation. Positive feedback highlighted the chatbot's ease of use, low threshold, clarity, immediate availability, and ability to help recognize targets and positives. However, there was consensus on the need for further development, with negative feedback focusing mainly on the lack of empathy and personal approach. This was also boosted by a notable incident when the chatbot incorrectly predicted self-harming intentions, highlighting the need for quality test data. Following this, all intents were turned off, and only the empowering conversation flow was tested.

This phase underscored the importance of effectively understanding and measuring user emotions in human-centric design. It was recognized that developing the chatbot's intent tree and responses required more authentic user conversations. Finding a test group and the need for more active feedback collection methods beyond a simple questionnaire was challenging. It is imperative to employ diverse methods to comprehensively understand the target group, for instance, from varying age groups, geographical regions, and backgrounds. This phase also pointed out the importance of openness from the HCAI point of view. This phase also highlighted the meaning of transdisciplinary work on data information safety and security aspects and the chatbot user testing on the actual Voima-chatbot platform. The design process's previous work significantly contributed to the technological development. Feedback from the technological platform provider was positive.

This phase encompassed inventive dialogues on how to construct an empowering AI-enabled solution. The technical development team grasped the empowering conversations process [46, 47, 57]. The project team had knowledge of AI-enabled solutions [36, 41], eHealth and informatics [7], data security and human-centered design [34, 41, 58].

3.5 Operate and Scale Phases of the Development (TRL Levels 8–9)

In the context of typical technical development processes, the development period of Voima-chatbot (2022–2023) has been relatively brief. Currently, Voima-chatbot is still in the nascent stages of technical development and will require ongoing enhancements in the future. While plans for new applications and integrations have been conceptualized, they have not yet been actualized.

4 Discussion and Future Insights

The objective of this research was to augment the comprehension of the feasibility of the HCD approach in creating AI-enabled technological solutions. This was accomplished during the development process of an AI-enabled solution, Voima-chatbot. The application and insights derived from the service design process, which utilized HCD design approach for AI-enabled services with 10 phases [41, 42] is presented in Fig. 1. The model incorporated three levels: technology, business, and design.

In addressing the research question - 'What are the implications and results of applying Human-Centric design in the development of an AI-enabled technological solution?' - a methodology of research through design was employed.

These development process phases produced information analyzed through HCD and HCAI point of views. Results emphasized the importance of observing transdisciplinary and cross-disciplinary processes, which involve numerous stakeholders and scientific orientations when working in the healthcare context.

This study has identified various strategies that merit integration into the model delineated by Jylkäs et al. [41, 42]. The healthcare context [23, 26], where understanding the various factors at play is crucial, but also research and evidence-based research when applied in healthcare [23]. It is mandatory when there is a development process for medical devices [10, 11].

Ethics and sustainability [2, 35, 38] factors guiding and promoting ethical activities contribute to the realization of the reflective process of ethical and sustainable activities. In ethical problem-solving, professionals base their judgement on legislation and ethical guidelines as well as on the ethical basis of social and healthcare. Ethical activities are promoted and facilitated, for example, through ethical management, organizational structures, and operational culture [2]. Sustainability also refers to Future insights into AI-enabled services. AI-enabled technological solutions are constantly developing [59]. Information safety and security, regulations, laws, EU data interoperability and policies [7, 8, 11]. It is one of the fundamental elements to decide if the application is a medical device or not. The categorization as a medical device brings a plethora of laws and regulations that necessitate careful consideration throughout the process [11]. Throughout this process, deliberations were held regarding the scope of the device and the possibilities of technical solutions. At the inception of this process, a decision was made to engineer a device aimed at promoting well-being.

Enhanced understanding of the development process, user needs, and expectations during the development process was cultivated through the application of Human-Centered design methodologies employed in the developmental stages (refer to Table 2). This research posits that the design process can foster and elucidate collaborative efforts

when stakeholders are acknowledged during the process. This ensures comprehensive consideration of all facets when innovating new technologies or refining existing ones [25, 26], for example, Technology Readiness Levels (TRL) [45].

User-testing technique yielded data indicating a need for more understanding of human behavior, interactions, and the human-machine relationship. It was also acknowledged that conventional approaches, such as questionnaires, might be too simplistic, with the small number of participants skewed for obtaining this information. It would be beneficial to enhance understanding of different facets of human-AI interaction, such as emotions, cognition, assets, mutual learning, or failure/success. Interfaces such as chatbots are shared boundaries between the sociotechnical systems of computers, connecting hardware, software, and human users. Ethnographical research can incorporate technical walkthroughs and interfaces to complement participant observations in local settings where interfaces are accessed. Interface ethnography can be utilized with multi-sited fieldwork designs since interfaces are components of transnational networks and mediate between different actors [60].

Pervasive themes for all stages of the development process were comprehending the context where the user and the intended technological solution are situated. Collecting user insights and genuine material for testing from diverse range of people. Humans can provide training data for machine learning applications and directly perform tasks that are challenging for computers in the pipeline with the assistance of machine-based approaches. This is a way to avoid biases [38, 61]. Developing the AI solution with transdisciplinary team is essential. It was learned that when healthcare professionals are involved in the process understanding on AI-enabled services was built among the participants [9, 13, 26].

Accompanying the findings of this research to the existing Jylkäs et al. model, it was contemplated that a broader perspective and implications of implementing Life-Centered design ought to be addressed incorporating Human-Centered design [35, 62]. This research could inform the development of similar AI-enabled health technologies by presenting a need for updating a new model and methods for the design process with a Life-centered design approach of AI-enabled solutions that could be prototyped and tested in the following research. Life-centered design, for example, expands human-centered design to include consideration for nature and vulnerable humans by merging practices such as circular design, biomimicry, systems thinking, and futuring, and aligning designers with global goals, such as the United Nation's Sustainable Development Goals [62–64]. As life-centered design is still emerging, its practices vary and is practiced by only a few. More research is needed to evaluate Life-centered design in digital services that utilize AI.

Acknowledgements. The empowering Voima-chatbot was developed as part of the ESF-funded project during 2021–2023 in Laurea University of Applied Sciences in association with Finnish Nurses Association. The project focused on the healthcare worker's well-being in Finland. We want to thank all the project group professionals, healthcare professionals, students and partners contributing to this research.

Disclosure of Interests. The authors have no competing interests to declare relevant to this article's content.

References

1. World Health organization: Empowerment through Digital Health. https://www.who.int/europe/initiatives/empowerment-through-digital-health
2. Koskinen, R., Helminen, K., Koski, A., Malkavaara, M., Sanerma, P., Sihvo, P.: Modus operandi for ethical action in social and health care in the era of digitalisation. Fin. J. eHealth eWelfare **14** (2022). https://doi.org/10.23996/FJHW.113414
3. Ministry of Social Affairs and Health: Strategy for digitalisation and information management in healthcare and social welfare emphasises flexibility. https://stm.fi/-/sosiaali-ja-terveyden huollon-digitalisaation-ja-tiedonhallinnan-strategia-painottaa-joustavuutta
4. Eurostat: Population structure and ageing - Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_structure_and_ageing
5. International Labour Organization: Occupational Health (Occupational Safety and Health). https://www.ilo.org/safework/areasofwork/occupational-health/lang--en/index.htm
6. European Commission: Employment and Social Developments in Europe (ESDE) (2023). https://op.europa.eu/webpub/empl/esde-2023/index.html
7. European Commission: 2023 Report on the state of the Digital Decade | Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/library/2023-report-state-digital-decade
8. Artificial Intelligence Act: Artificial Intelligence Act. https://artificialintelligenceact.com/
9. Fergusson, L.C., Brömdal, A., Gough, M., Mears, S.: Competency, capability and professional identity: the case for advanced practice. Work Based Learning e-Journal **9** (2020)
10. Fimea: Medical devices. https://fimea.fi/en/medical-devices
11. Eur Lex: Medical Device Regulation 2017/745. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=uriserv:OJ.L_.2017.117.01.0001.01.ENG
12. Ordish, J., Murfet, H., Hall, A.: Algorithms as Medical Devices (2019)
13. Azzopardi-Muscat, N., et al.: The global effect of digital health technologies on health workers' competencies and health workplace: an umbrella review of systematic reviews and lexical-based and sentence-based meta-analysis. Rev. Lancet Digit. Health **5**, 534–578 (2023)
14. Billmann, M., Böhm, M., Krcmar, H.: Use of workplace health promotion apps: analysis of employee log data. Health Policy Technol. **9**, 285–293 (2020). https://doi.org/10.1016/J.HLPT.2020.06.003
15. Abd-Alrazaq, A.A., Rababeh, A., Alajlani, M., Bewick, B.M., Househ, M.: Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J. Med. Internet Res. **22** (2020). https://doi.org/10.2196/16021
16. Abd-Alrazaq, A.A., et al.: Perceptions and opinions of patients about mental health chatbots: scoping review. J. Med. Internet Res. **23** (2021). https://doi.org/10.2196/17828
17. Cameron, G., et al.: Best Practices for Designing Chatbots in Mental Healthcare – A Case Study on iHelpr (2018). https://doi.org/10.14236/EWIC/HCI2018.129
18. Cameron, G., et al.: Assessing the usability of a chatbot for mental health care. In: International Conference on Internet Science (2018)
19. Vaidyam, A.N., Wisniewski, H., Halamka, J.D., Kashavan, M.S., Torous, J.B.: Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can. J. Psychiat. **64**, 456–464 (2019). https://doi.org/10.1177/0706743719828977
20. Følstad, A., Brandtzaeg, P.B.: Users' experiences with chatbots: findings from a questionnaire study **5**, 3 (2020). https://doi.org/10.1007/s41233-020-00033-2

21. Haque, M.D.R., Rubya, S.: An overview of chatbot-based mobile mental health apps: insights from app description and user reviews. JMIR Mhealth Uhealth **11** (2023). https://doi.org/10.2196/44838
22. Bhirud, N., Tataale, S., Randive, S., Nahar, S.: A literature review on chatbots in healthcare domain. Int. J. Sci. Technol. Res. **8**, 7 (2019)
23. Carr, V.L., Sangiorgi, D., Büscher, M., Junginger, S., Cooper, R.: Integrating evidence-based design and experience-based approaches in healthcare service design. HERD **4**, 12–33 (2011). https://doi.org/10.1177/193758671100400403
24. Choi, B.C.K., Pak, A.W.P.: Multidisciplinarity, interdisciplinarity, and transdisciplinarity in health research, services, education and policy: 2. Promotors, barriers, and strategies of enhancement. Clin. Investig. Med. **30** (2007). https://doi.org/10.25011/cim.v30i6.2950
25. Salmi, A., Ahonen, O., Pöyry-Lassila, P.: Crossing asymmetries in multistakeholder service design in integrated care. Serv. Des. Pract. Healthc. Innov., 133–156 (2022). https://doi.org/10.1007/978-3-030-87273-1_7
26. Alhonsuo, M.: Early Phase of Healthcare-Related Service Design (2021). http://urn.fi/URN:ISBN:978-952-337-296-2
27. Rönnholm, R.: Co-design of change: why changing what people do should be the key ingredient in service design. In: An Introduction to Industrial Service Design. Routledge, Taylor & Francis Group, London , New York (2017)
28. Frayling, C.: Research in Art and Design (Royal College of Art Research Papers, vol. 1, no. 1, 1993/4) (1994)
29. Zimmerman, J., Forlizzi, J., Evenson, S.: Research through design as a method for interaction design research in HCI. In: Conference on Human Factors in Computing Systems – Proceedings, pp. 493–502 (2007). https://doi.org/10.1145/1240624.1240704
30. Melles, M., Albayrak, A., Goossens, R.: Innovating health care: key characteristics of human-centered design. Int. J. Qual. Health Care, 37–44 (2021). https://doi.org/10.1093/intqhc/mzaa127
31. Auernhammer, J.: Human-centered AI: the role of Human-centered Design Research in the development of AI. DRS2020: Synergy **1** (2020). https://doi.org/10.21606/DRS.2020.282
32. National Institution of Standards and Technology: Human Centered Design (HCD). https://www.nist.gov/itl/iad/visualization-and-usability-group/human-factors-human-centered-design
33. Burns, C.: Human-centred design. eHealth Res. Theory Dev. Multidiscip. Approach, 208–227 (2018). https://doi.org/10.4324/9781315385907-10
34. Hartikainen, M., Väänänen, K., Lehtiö, A., Ala-Luopa, S., Olsson, T.: Human-centered AI design in reality: a study of developer companies' practices a study of developer companies' practices. In: NordiCHI 2022: Nordic Human-Computer Interaction Conference, pp. 1–11 (2022)
35. Borthwick, M., Tomitsch, M., Gaughwin, M.: From human-centred to life-centred design: considering environmental and ethical concerns in the design of interactive products. J. Responsible Technol. **10**, 100032 (2022). https://doi.org/10.1016/J.JRT.2022.100032
36. Shneiderman, B.: Human-centered artificial intelligence: reliable, safe & trustworthy. Int. J. Hum. Comput. Interact. **36**, 495–504 (2020). https://doi.org/10.1080/10447318.2020.1741118
37. Awad, E., et al.: The moral machine experiment. Nature **563**, 59–64 (2018). https://doi.org/10.1038/S41586-018-0637-6
38. Van Velsen, L., Ludden, G., Grünloh, C.: The limitations of user-and human-centered design in an ehealth context and how to move beyond them. J. Med. Internet Res. **24**, e37341 (2022). https://doi.org/10.2196/37341

39. Oppermann, L., Boden, A., Hofmann, B., Prinz, W., Decker, S.: Beyond HCI and CSCW: Challenges and Useful Practices Towards a Human-Centred Vision of AI and IA (2019). https://doi.org/10.1145/3363384.3363481
40. Design Council: Framework for Innovation: Design Council's evolved Double Diamond - Design Council. https://www.designcouncil.org.uk/our-work/skills-learning/tools-frameworks/framework-for-innovation-design-councils-evolved-double-diamond/
41. Jylkäs, T.: Service Design and Artificial Intelligence in Designing Human-Centred Digital Services. http://urn.fi/URN:ISBN:978-952-337-227-6
42. Jylkäs, T., Augsten, A., Miettinen, S.: From hype to practice: revealing the effects of AI in service design. In: Conference: Academy for Design Innovation Management Conference 2019 (2019)
43. Böckle, M., Kouris, I.: Design thinking and AI : a new frontier for designing human-centered AI solutions. Des. Manage. J. **18**, 20–31 (2023). https://doi.org/10.1111/DMJ.12085
44. Hartikainen, M., Väänänen, K., Lehtiö, A., Ala-Luopa, S., Olsson, T.: Human-centered AI design in reality: a study of developer companies' practices, 22. https://doi.org/10.1145/3546155.3546677
45. Manning, C.: Technology Readiness Levels – NASA. https://www.nasa.gov/directorates/somd/space-communications-navigation-program/technology-readiness-levels/
46. Franklin, C., Zhang, A., Froerer, A., Johnson, S.: Solution focused brief therapy: a systematic review and meta-summary of process research. J. Marital Fam. Ther. **43**, 16–30 (2017). https://doi.org/10.1111/JMFT.12193
47. Teater, B.: Solution focused brief therapy. In: Davies, M. (ed.) The Blackwell Companion to Social Work. Wiley-Blackwel, Hoboken (2018)
48. FrontAI: What is Conversational AI? | Text or Voice-based Virtual Assistants. https://front.ai/conversational-ai/
49. Woebot Health: Relational Agent for Mental Health. https://woebothealth.com/
50. Wysa: Wysa - Everyday Mental Health. https://www.wysa.com/
51. ChatPAL Digital Wellbeing Conversations: ChatPAL. https://chatpal.interreg-npa.eu/
52. Jordan, S.S., Froerer, A.S., Bavelas, J.B.: Microanalysis of positive and negative content in solution-focused brief therapy and cognitive behavioral therapy expert sessions **32**, 46–59 (2013). https://doi.org/10.1521/JSYT.2013.32.3.46
53. De Lange, A.H., Løvseth, L.T., Rui, K., Teoh, H., Christensen, M.: Editorial: healthy healthcare: empirical occupational health research and evidence-based practice. Front. Psychol. **11** (2020). https://doi.org/10.3389/fpsyg.2020.02236
54. Hassard, J., Teoh, K., Thomson, L., Blake, H.: Understanding the cost of mental health at work: an integrative framework. In: The Sage Handbook of Organizational Wellbeing, pp. 9–25. SAGE Inc, London (2021)
55. Asada, M.: Towards artificial empathy how can artificial empathy follow the developmental pathway of natural empathy? Int. J. Soc. Robotics **7**, 19–33 (2015). https://doi.org/10.1007/s12369-014-0253-z
56. Ruutu, S.: Coachin työkalupakki. Alma Talent, Helsinki (2020)
57. Adams, R.: Concept of empowerment. In: Social Work and Empowerment, pp. 1–24 (1996). https://doi.org/10.1007/978-1-349-14033-6_1
58. Chow, B.E., Pilarski, A., Schmitt, J., Decker, M.C., Ark, T., Davis, C.S.: Using human-centered design to improve a surgery resident well-being program. J. Surg. Res. **277**, 157–162 (2022). https://doi.org/10.1016/J.JSS.2022.02.043
59. Hines, A., Bishop, P.C.: Framework foresight: exploring futures the Houston way. Futures **51**, 31–49 (2013). https://doi.org/10.1016/J.FUTURES.2013.05.002
60. Ritter, C.S.: Rethinking digital ethnography: a qualitative approach to understanding interfaces. Qual. Res. **22**, 916–932 (2022). https://doi.org/10.1177/14687941211000540/ASSET/IMAGES/LARGE/10.1177_14687941211000540-FIG2.JPEG

61. Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., He, L.: A survey of human-in-the-loop for machine learning. Futur. Gener. Comput. Syst. **135**, 364–381 (2022). https://doi.org/10.1016/j.future.2022.05.014
62. Lutz, D.: The Life-Centred Design Compass. https://uxdesign.cc/the-life-centred-design-compass-25a98f129c96
63. United Nations, D. of E. and S.A.S.D.: THE 17 GOALS | Sustainable Development. https://sdgs.un.org/goals
64. Rossi, E., Attaianese, E.: Research synergies between sustainability and human-centered design: a systematic literature review. Sustainability **15**, 12884 (2023). https://doi.org/10.3390/SU151712884

Health Technology Assessment and Impact Evaluation

Finnish Digi-HTA Assessment Model for Digital Health and an International Comparison

Jari Haverinen[1,2(✉)], Jarno Suominen[1], Rauli Kaksonen[3], Paula Veikkolainen[2], Merja Voutilainen[1], Jarmo Reponen[2,4], Juha Röning[3], and Petra Falkenbach[1]

[1] Finnish Coordinating Center for Health Technology Assessment (FinCCHTA), Oulu University Hospital, Oulu, Finland
jari.haverinen@oulu.fi
[2] FinnTelemedicum, Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland
[3] Biomimetics and Intelligent Systems Group, University of Oulu, Oulu, Finland
[4] Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland

Abstract. New health technology assessment (HTA) models for digital health are continuously being developed and are already in use. In Finland, the HTA model for digital health, named Digi-HTA, has been employed since 2020. Internationally and also in Finland, the need for harmonization of these HTA models has been recognized. In order to harmonize the models, it is necessary to first identify the key features and requirements of existing models. In this study, three key assessment models for digital health identified as central in the Finnish context were analyzed. After the analysis, the results were compared to the Finnish Digi-HTA assessment model, and a final synthesis was created regarding the similarities and differences between the assessment models. The comparison includes German DiGA model, the global CEN-ISO/TS 82304-2:2021 technical specification, and the Nordic-designed NordDEC assessment model. There was a great deal of similarity in the evaluated models, although certain differences in emphasis were found. The key differences relate to reimbursement process, maturity of the assessment process and supported product categories as well as cost and effectiveness evaluation. The results of this study can be utilized in harmonizing assessment models for digital health.

Keywords: Health technology assessment · digital health · artificial intelligence · robotics · digital therapeutics

1 Introduction

Health Technology Assessment (HTA) involves the systematic evaluation of the properties, effects, and/or impacts of health technology. Its main purpose is to inform decision-makers to better support the introduction of new health technologies [1]. New digital health solutions, such as digital health applications, AgeTech, Digital Therapeutics (DTx), artificial intelligence (AI), and robotics, enable further development of healthcare

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 309–321, 2024.
https://doi.org/10.1007/978-3-031-59091-7_20

services, but their introduction should follow the same criteria as other healthcare methods. They must provide evidence-based benefits and be safe to use, and their impacts on patients and organizations need to be clarified [2]. In case of digital health, the data security and privacy of the products must also be ensured in all situations, and they should be user-friendly for all assumed user groups [2, 3].

The new and innovative digital health products also set new demands on HTA models as well [2]. In Finland, the need for new models to support the HTA work of digital health was identified. Therefore, in 2018, the Finnish Ministry of Social Affairs and Health commissioned the development of a new HTA model for digital health [4]. A new HTA model, named Digi-HTA, that supports a wide range of digital health products such as digital health applications, AgeTech, AI, and robotic solutions, was published in 2019 [2–4]. The Digi-HTA model utilizes the Digi-HTA assessment framework as well as criteria developed in the Kyber-Terveys project, which are used for assessing data security and protection aspects [2–4]. Since 2020, Digi-HTA has been part of the daily HTA activities of the Finnish Coordinating Center for Health Technology Assessment (FinCCHTA), and Digi-HTA assessments have been published on various digital health products, such as digital health applications, medicine dispensing, and rehabilitation robotics, as well as digital platform solutions [5].

HTA for digital health is still a growing trend globally, not only in Finland, and new models are constantly being developed, with some of them already in use [6, 7]. Some of these models are national, such as German Fast-Track process for digital health applications, while others are developed for international use, such as the CEN-ISO/TS 82304-2:2021 Health software – Part 2: Health and wellness apps – Quality and reliability technical specification (hereinafter referred to as "the CEN-ISO/TS 82304-2:2021") [8, 9]. Some models aim to address the assessment needs of a specific region, such as the Nordic Digital Health Evaluation Criteria (NordDEC) model developed for the Nordic countries [10].

In 2019, Germany enacted the Digital Healthcare Act (Digitale-Versorgung-Gesetz), which defines the so-called Fast-Track procedure for the assessment and reimbursement of digital health applications. Digital health applications covered by the German Fast-Track process are referred to as "DiGAs" ("Digitale Gesund-heitsAnwendungen"). [8] The details of the requirements for the DiGA are regulated in the Digital Health Applications Ordinance (Digitale Gesundheitsanwendungen-Verordnung, DiGAV) (Bundesministerium für Gesundheit, 2022) [11]. The German Federal Institute for Drugs and Medical Devices (BfArM) is the body that carries out assessments and approvals for DiGA [8].

The CEN-ISO/TS 82304-2:2021 was published in July 2021. The development was motivated by the fact that the number of digital health applications had already exceeded 300.000, yet there was no standard in place for assessing their quality. The background of the development was a commission from the European Commission, and the development was carried out in collaboration with the International Organization for Standardization (ISO). [9] The adoption of CEN-ISO/TS 82304-2:2021 is being promoted in the Label2Enable project funded by the European Union (EU) [12].

The goal of the NordDEC assessment model is to support the assessment of digital health applications in the Nordic countries and enable cross-border assessment work

[10]. The requirements of the NordDEC assessment process for digital health applications are defined in the Nordic Digital Health Evaluation Criteria [13]. The development of NordDEC is managed by the Nordic Interoperability Project, jointly funded by Nordic Innovation and the Nordic health tech industry. The assessment model is developed and operated by the Organisation for the Review of Care and Health Apps (ORCHA) [10].

Since many assessment models have been developed from national or regional perspectives, such as DiGA or NordDEC, there may be significant differences or emphasis variations in the requirements of different models [8, 10]. It has been recognized among EU member states that voluntary cooperation is needed to harmonize these models, and one example of this collaboration is the European Taskforce for Harmonised Evaluations of Digital Medical Devices (DMDs) [14]. In Finland as well, it has been recognized in the EU-funded Finnish Recovery and Resilience Plan program that the existing Digi-HTA model should be further developed. For that reason, understanding the key features and requirements of available HTA models for digital health is crucial. In this study, the models selected for evaluation were considered relevant in the context of Finland. DiGA was chosen because it has already become a benchmark for assessing DTx applications and integrating assessments into reimbursement processes since 2020. As the purpose of CEN-ISO/TS 82304-2:2021 is to serve as a global criterion for digital health applications, it provides a valuable point of comparison in a global context. The NordDEC assessment model is based on the long-standing ORCHA assessment model and is designed to meet the needs of the Nordic countries, making it a good point of comparison from a Nordic perspective. Through this study, it is possible to develop and harmonize the Finnish HTA model at the national level, as well as utilize these results in international harmonization efforts.

2 Aim of the Study

1. To evaluate the features, domains, and aspects that are included in the DiGA, CEN-ISO/TS 82304-2:2021, and NordDEC assessment models.
2. To identify the similarities and differences between the evaluated assessment models and the Digi-HTA model.

3 Materials and Methods

Information about the key features of different assessment models was gathered from the websites of organizations conducting assessments, guidelines, and scientific articles [8–10, 12]. The information about the assessment frameworks of DiGA and NordDEC models was collected from information available on their websites [11, 13]. The DiGAV criteria, which were available on the website as of January 23, 2023, were included in the comparative work [11]. The comparative work included the version of the Nordic Digital Health Evaluation Criteria that was last updated on June 15, 2022 [13]. Information about the requirements included in CEN-ISO/TS 82304-2:2021 was obtained directly from the technical specification, which was published on August 20, 2021 [15]. The comparative work included the Digi-HTA assessment model criteria that was in use in the process between May 2022 and April 2023 [2, 3]. During this period, there were no changes in the criteria.

In the first phase, the key features of each assessment model and its associated process were listed. This included, for example, what product categories the assessment process supported and whether the assessment were linked to reimbursement processes. In the next phase, the assessment frameworks and the included domains were compared. Each assessment framework was reviewed at the level of individual questions. After that, the questions were grouped into key identified domains. However, it should be noted that different naming practices were in use for the domains that mainly addressed the same issues, such as technical stability in Digi-HTA and NordDEC, and robustness in DiGA. Therefore, these were attempted to be consolidated under the same domain. With regard to data security and protection, the comparison was conducted based on the product requirement categories and category groups presented in the article 'Common cybersecurity requirements in IoT standards, best practices, and guidelines' [16]. Subsequently, a comparison was made between each individual assessment model and the Digi-HTA assessment model. The individual comparative works were carried out between May 2022 and April 2023. This study includes the final top-level synthesis between different assessment models based on three individual comparative reports [17–19].

4 Results

Figure 1 illustrates the key elements included in typical assessment processes for digital health products. In all evaluated models, the technology company initiates the process. Afterward, the product is assessed using an assessment framework that includes detailed questions about the product being assessed. These questions are divided into different HTA domains. In addition to the questions, other documentation, such as research studies, is required, providing sufficient evidence to support the claims. Assessment is carried out by the entity responsible for assessments in each model. The assessment team may include various types of expertise, such as HTA experts and cybersecurity experts. Completed assessments are published on the web portal. There are two different scenarios for utilizing the assessments. Completed assessments can lead to a formal product reimbursement process, or assessments can be used more freely as part of procurement or product introductions.

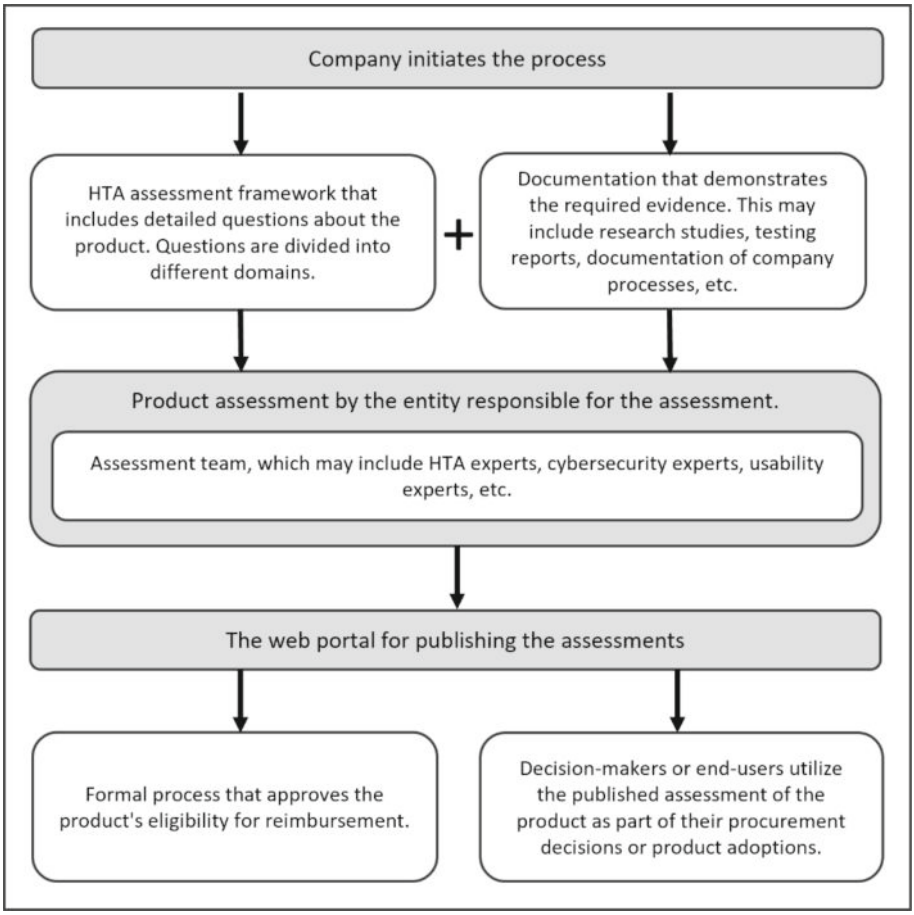


Fig. 1. Typical assessment process for digital health products

The key features of evaluated assessment models are presented in Table 1. In all evaluated models, the assessment frameworks have been published and are available. The assessment model is still under development for the CEN-ISO/TS 82304-2:2021 in Label2Enable project. Only the DiGA assessment model includes a clear reimbursement model, while in others, it is still under development. There were differences in the supported product categories among the assessment models.

The domains included in different assessment frameworks are presented in Table 2. In the examined assessment frameworks, there were a lot of similarities in terms of the key assessment domains they included. The main differences were related to effectiveness, costs, robotics, AI, ethics and consumer protection.

Table 1. Features of evaluated models

	Assessment model			
Features	Digi-HTA	Fast-Track procedure for DiGA	CEN-ISO/TS 82304-2:2021	NordDEC
Region	Finland	Germany	Global	Nordic countries
Coordinating body	FinCCHTA	BfArM	ISO. Assessment and business models are under development	The Nordic Interoperability Project
Readiness level	In production since 2020	In production since 2020	Framework published 2021. Assessment and business models are under development	Published 2022
Assessment process fee for the company	Free of charge	BfArM charges fees in accordance with regulations	Assessment and business models under development	Not publicly defined
The duration of the assessment	2 months	3 months	Not specified	Not specified
The publication portal for assessments exists	Yes	Yes	Under development	Yes
Suitable for medical devices	Yes, all classes	Yes, classes I, IIa	Yes, all classes	Yes, all classes
Suitable for non-medical devices	Yes	No	Yes	Yes
Suitable for digital health products in the form of native apps, web apps or websites	Yes	Yes	Yes	Yes

(*continued*)

Table 1. (*continued*)

	Assessment model			
Features	Digi-HTA	Fast-Track procedure for DiGA	CEN-ISO/TS 82304-2:2021	NordDEC
Suitable for digital health technologies in addition of digital health products in the form of native apps, web apps or websites	Yes, the main functionality should be digital. For example, digital platforms, AgeTech, AI, and robotic solutions can be assessed	Yes, hardware components can be included but the main functionality should be digital	No	No
Country of origin of evidence	Not specified	Studies performed in the German healthcare context preferred	Not specified	Not specified
Link to reimbursement process	Not at the moment. Can be linked to regional decisions. Reimbursement process under investigation	Yes, Fast-Track process	Not at the moment, but assessments can be linked to national reimbursement models	Not at the moment, but assessments can be linked to national reimbursement models

Table 2. Domains included in evaluated assessment frameworks

	Assessment framework			
Domain	Digi-HTA	DiGAV	CEN-ISO/TS 82304-2:2021	NordDEC
Information about the product and its functionalities	Yes	Yes	Yes	Yes
Effectiveness/Clinical evidence patient and end-user point of view	Yes, randomized controlled trials (RCTs) preferred	Yes, RCTs preferred	Yes, requirements are based on Evidence Standard Framework (ESF) Tier levels	Yes, requirements are based on ESF Tier levels

(*continued*)

Table 2. (*continued*)

	Assessment framework			
Domain	Digi-HTA	DiGAV	CEN-ISO/TS 82304-2:2021	NordDEC
Effectiveness/benefits from organizational point of view	Yes	Partly, improvements of structure and processes in healthcare should be patient relevant	Yes, requirements are based on ESF Tier levels	Yes, requirements are based on ESF Tier levels
Cost evaluation	Yes, economic evidence will be assessed	No, economic evidence will not be assessed	No, economic evidence will not be assessed	No, economic evidence will not be assessed
Safety	Yes	Yes, by default, the CE marking ensures safety	Yes	Yes
Usability	Yes, evidence about end-user testing is required	Yes, evidence about end-user testing is required	Yes, evidence about end-user testing is required. Evaluate if the app is age-appropriate	Yes, evidence about end-user testing is required
Accessibility	Yes, accessibility statement required and WCAG 2.1. AA guidelines should be followed	Yes, the product should be accessible for people with disabilities. Accessibility statement not required	Yes, accessibil-ity statement required and WCAG 2.1. AA and AAA guidelines should be fol-lowed	Yes, accessibility statement required and WCAG 2.1. AA and AAA guidelines should be followed
Technical stability/robustness	Yes	Yes	Yes	Yes
Interoperability	Yes, integrations within the Finnish healthcare context	Yes, integrations within the German healthcare context	Yes	Yes

(*continued*)

Table 2. (*continued*)

	Assessment framework			
Domain	Digi-HTA	DiGAV	CEN-ISO/TS 82304-2:2021	NordDEC
Data security and protection	Yes, a total of 108 different categories and 23 category groups are covered	Yes, a total of 77 different categories and 22 category groups are covered	Yes, a total of 84 different categories and 21 category groups are covered	Yes, a total of 63 different categories and 18 category groups are covered
Robotics	Yes, own domain for robotics aspects	No	No	No
Artificial intelligence	Yes, own domain for AI aspects	No	No	No
Ethics	No	No	Yes	No
Consumer protection	No	Yes, own domain for consumer protection issues	Partly	Partly

5 Discussion

The aim of this study was to investigate the key features and requirements of existing well-known HTA models for digital health. This study synthesized the similarities and differences between the models. The results of the study are intended to facilitate the further development of the Finnish Digi-HTA model. The goal is to ensure that Digi-HTA covers as many perspectives of existing well-known assessment models as possible and to identify key aspects that ensure regulated market access in different countries. The results of this study can also be utilized as part of international harmonization efforts. The comparison included the DiGA, CEN-ISO/TS 82304-2:2021, and NordDEC assessment models, which were found to be the most relevant in the Finnish context.

According to this study, the published assessment framework was available for all models, but the assessment process was still under development for CEN-ISO/TS 82304-2:2021 in the Label2Enable project. Only DiGA assessments were linked to a formal reimbursement process, while in others, this was still under development. DiGA focused solely on Class I and IIa medical devices, whereas others covered both medical and non-medical devices. The Digi-HTA model covers the widest range of different digital health products, such as digital health applications, AI, robotics, and various digital platform solutions, while others primarily focus on native applications, web-based applications, or websites. However, in the DiGA process, it is stated that products may include hardware

components, but the primary functionality must be digital. After completed assessments, it is crucial that information about the conducted assessments is also publicly disseminated to all those who need assessment information. All models, except for CEN-ISO/TS 82304-2:2021, had an existing publication portal where completed assessments could be viewed. The goal of the development of CEN-ISO/TS 82304-2 is that the quality label obtained through assessments would become a part of app stores or libraries, or it would be incorporated into trusted websites used by patients or clinicians [9].

Traditionally, key domains of HTA have included effectiveness, costs, and safety [20]. However, digital health products introduce new key aspects that should also be considered in addition to these traditional domains [2]. The key observation of this study was that there was a great deal of similarity in the key domains of all assessment models, although there were differences in emphasis within these domains. This may indicate that the entities developing the models have each identified the essential domains that should be considered in the adoption of digital health products. For example, all models assess the usability of digital health products, and according to research, the ease of use of digital health products has been identified as a factor that promotes their use [2, 21]. Since the DiGA process evaluates only products classified as medical devices, the safety and functionality of the products are assumed to be demonstrated by the CE marking. However, additional evidence from the product manufacturer may be required if necessary. In other models, there were more detailed requirements for product safety or safety-related company processes. Only the Digi-HTA model included the evaluation of costs as part of the assessment process. In other models, there was a requirement that the costs associated with using the product should be communicated transparently to end users. Even though only DiGA included its own domain on consumer protection, in other models as well, except in Digi-HTA, these perspectives had been partially addressed. For instance, CEN/ISO TS 82304-2:2021 required that age restrictions for applications should be clearly communicated to consumers.

All models assessed data security and privacy issues, which should fundamentally be in order for all digital health products to ensure user trust and prevent the leakage of sensitive information to unauthorized parties [21, 22]. In the domain of data security and privacy, the Digi-HTA model had the broadest coverage. For example, CEN/ISO TS 82304-2:2021 focuses on digital health applications, while Digi-HTA covers the entire IT system. The former has very few requirements beyond applications. Digi-HTA model's data security and protection requirements covered 108 categories and 23 different category groups, while NordDEC's requirements were the most limited, encompassing 63 categories and 18 category groups.

Digital health products or services have the potential to offer benefits to patients, but also to healthcare service providers, for example, through improved efficiency in care processes [23]. Each evaluated model assesses effectiveness/clinical benefit from the patients' perspective. The DiGA process emphasizes, in all aspects of product benefits, that the achieved benefits must be relevant to the patients. Benefits solely from the perspective of healthcare organizations are not sufficient evidence of effectiveness in the DiGA procedure. However, in other models, benefits obtained solely from the organization's perspective, such as improvements in care processes, are also considered.

CEN-ISO/TS 82304-2:2021 and NordDEC assessment models define the required evidence of product benefits based on the Evidence Standard Framework (ESF) developed by the National Institute for Health and Care Excellence (NICE). According to the ESF, products are classified into three different categories (Tier A, B and C) based on the potential risk they may pose. The higher the risk classification, the more compelling evidence is required. Digital health products that do not have direct outcomes related to patient health or care, but instead provide system services aimed at saving time or cost, are included in Tier A. [24] The DiGA process emphasizes that studies should be conducted in Germany or companies must demonstrate that research results from other countries can be transferred to the context of German healthcare. In other models, the origin of research results is not precisely defined. However, in the Finnish Digi-HTA process, it is always assessed on a case-by-case basis whether the results can be transferred to the context of Finnish healthcare.

The three key separate comparative works, on which the synthesis of this study is based, were conducted between May 2022 and April 2023 [17–19]. At the time of the study, these three assessment models, namely DiGA, CEN-ISO/TS 82304-2:2021, and NordDEC, were considered the most relevant for conducting the comparative work in the Finnish context. However, since then, new assessment models have been published, with one of the most significant being the French Early Access to Reimbursement for Digital Devices (PECAN) assessment and reimbursement model released in the spring of 2023. The process defines assessment and reimbursement models for products that can be included in the categories of DTx and remote monitoring. The PECAN process is designed for products classified as medical devices. However, unlike the German DiGA process, products from all risk classes can be included in the process [7, 25]. The ongoing development of models emphasizes the need to continue comparative work to identify all key perspectives that should be included in HTA models assessing digital health products.

6 Conclusion

In this study, the key features and requirements of four different assessment models for digital health were analyzed. The study included the Digi-HTA, DiGA, CEN-ISO/TS 82304-2:2021, and NordDEC assessment models. There was a great deal of similarity in the evaluated models, although certain differences in emphasis were found. The key differences relate to reimbursement process, maturity of the assessment process and supported product categories as well as cost and effectiveness evaluation. The information from this study can be utilized in the harmonization efforts of HTA models for digital health.

Acknowledgments. This study part of Finnish Recovery and Resilience Plan which is funded by the European Union - NextGenerationEU funding.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Panteli, D., Busse, R.: Health technology assessment at age 25 - squaring the circle of strong methodology and context-dependency. Health Policy **123**(2), 115–117 (2019)
2. Haverinen, J., Keränen, N., Falkenbach, P., Maijala, A., Kolehmainen, T., Reponen, J.: Digi-HTA: health technology assessment framework for digital healthcare services. Finnish J. EHealth EWelfare **11**(4), 326–341 (2019)
3. Jääskelä, J., et al.: Digi-HTA, assessment framework for digital healthcare services: information security and data protection in health technology – initial experiences. Finnish J. EHealth EWelfare **14**(1), 19–30 (2022)
4. Haverinen, J., Turpeinen, M., Falkenbach, P., Reponen, J.: Implementation of a new Digi-HTA process for digital health technologies in Finland. Int. J. Technol. Assess Health Care **38**(1) (2022)
5. Digi-HTA. https://oys.fi/fincchta/digi-hta/about-digi-hta/. Accessed 23 Jan 2024
6. Mezei, F., Horváth, K., Pálfi, M., Lovas, K., Ádám, I., Túri, G.: International practices in health technology assessment and public financing of digital health technologies: recommendations for Hungary. Front Public Health **11** (2023)
7. van Kessel, R., et al.: Digital health reimbursement strategies of 8 European countries and israel: scoping review and policy mapping. JMIR Mhealth Uhealth **11**, e49003 (2023)
8. Federal Institute for Drugs and Medical Devices. The Fast-Track Process for Digital Health Applications (DiGA) according to Section 139e SGB V A - Guide for Manufacturers, Service Providers and Users. https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.pdf?__blob=publicationFile. Accessed 20 Jan 2024
9. Hoogendoorn, P., et al.: What makes a quality health app-developing a global research-based health app quality assessment framework for CEN-ISO/TS 82304-2: Delphi study. JMIR Format. Res. **7**, e43905 (2023)
10. ORCHA. Nordic Digital Health Evaluation Criteria. https://norddec.org/. Accessed 24 Jan 2024
11. Verordnung über das Verfahren und die Anforderungen zur Prüfung der Erstattungsfähigkeit digitaler Gesundheitsanwendungen in der gesetzlichen Krankenversicherung (Digitale Gesundheitsanwendungen-Verordnung - DiGAV). https://www.gesetze-im-internet.de/digav/BJNR076800020.html. Accessed 23 Jan 2023
12. Label2Enable. https://label2enable.eu/. Accessed 20 Jan 2024
13. Nordic Digital Health Evaluation Criteria – NordDEC. https://confluence.external-share.com/content/7a7e9bac-472b-49ada8e4-d1fd13b95199. Accessed 13 Jan 2023
14. European Taskforce for Harmonised Evaluations of Digital Medical Devices (DMDs). https://eithealth.eu/european-taskforce-for-harmonised-evaluations-of-digital-medical-devices-dmds/. Accessed 22 Jan 2024
15. International Organization for Standardization. Health software -part 2: health and wellness apps—quality and reliability (ISO/TS 82304-2) (2021). https://www.iso.org/standard/78182.html. Accessed 30 Sept 2022
16. Kaksonen, R., Halunen, K., Röning, J.: Common cybersecurity requirements in IoT standards, best practices, and guidelines. In: Proceedings of the 7th International Conference on Internet of Things, Big Data and Security - IoTBDS, pp. 149–156. SciTePress (2022)
17. Suominen, J., Veikkolainen, P., Kaksonen, R., Voutilainen, M., Haverinen, J.: Comparison report of Digi-HTA and CEN/ISO TS 82304-2:2021. https://oys.fi/fincchta/wp-content/uploads/sites/21/2023/02/comparison-report-of-digi-hta-and-cen-iso-ts-82304-2-2021.pdf. Accessed 19 Jan 2024
18. Voutilainen, M., Kaksonen, R., Haverinen, J.: Comparison report of Digi-HTA and DiGAV ordinance. https://oys.fi/fincchta/wp-content/uploads/sites/21/2023/04/comparison-report-of-digi-hta-and-digav-ordinance.pdf. Accessed 19 Jan 2024

19. Suominen, J., Veikkolainen, P., Kaksonen, R., Haverinen, J.: Comparison report of Digi-HTA and NordDEC, https://oys.fi/fincchta/wp-content/uploads/sites/21/2023/04/comparison-report-of-digi-hta-and-norddec.pdf. Accessed 19 Jan 2024
20. Sihvo, S., Ikonen, T., Mäkelä, M.: Implementing health technology assessment-based recommendations in Finland: managed uptake of medical methods. Int. J. Technol. Assess. Health Care **33**(4), 430–433 (2017)
21. Deng, Z., Hong, Z., Ren, C., Zhang, W., Xiang, F.: What predicts patients' adoption intention toward mhealth services in China: empirical study. JMIR Mhealth Uhealth **6**(8), e172 (2018)
22. Kyytsönen, M., Vehko, T., Jylhä, V., Kinnunen, U.M.: Privacy concerns among the users of a national patient portal: a cross-sectional population survey study. Int. J. Med. Inform. **183** (2024)
23. Haverinen, J., Harju, T., Mikkonen, H., Liljamo, P., Turpeinen, M., Reponen, J.: Digital care pathway for sleep apnea patients in specialized care: mixed methods study. JMIR Hum. Factors **11**, e47809 (2024)
24. The National Institute for Health and Care Excellence (NICE). Evidence standards framework for digital health technologies. https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies. Accessed 22 Jan 2024
25. Early access to reimbursement for digital devices (PECAN). http://gnius.esante.gouv.fr/en/financing/reimbursement-profiles/early-access-reimbursement-digital-devices-pecan. Accessed 22 Jan 2024

Influencing Factors in Digital Health Intervention Uptake: The Interplay of Education, Lifestyle, and Digital Literacy

Ilona Ruotsalainen[1(✉)], Mikko Valtanen[2,3], Riikka Kärsämä[2], Adil Umer[1], Suvi Parikka[2], Annamari Lundqvist[2], and Jaana Lindström[2,4]

[1] VTT Technical Research Centre of Finland Ltd., Espoo, Finland
ilona.ruotsalainen@vtt.fi
[2] Population Health Unit, Finnish Institute for Health and Welfare, Helsinki, Finland
[3] Department of Mathematics and Statistics, University of Turku, Turku, Finland
[4] Institute of Public Health and Clinical Nutrition, School of Medicine, University of Eastern Finland, Kuopio, Finland

Abstract. Chronic diseases strain global healthcare economically, and integrating digital solutions are proposed to help in meeting the rising demand. Digital health interventions (DHIs) offer promise for personalized, and cost-effective health services, however, factors influencing their uptake remain unclear. We examined whether the probability of lifestyle DHI uptake varies among individuals with different educational levels and lifestyles, based on their attitudes and usage of e-services. We also examined the effect of sex and age, and the association between DHI uptake and both educational attainment and overall lifestyle. A possibility to start using a web-based lifestyle DHI was offered to a subgroup (n = 6978) of Healthy Finland survey participants and adjusted logistic regression models were used to investigate the factors affecting uptake. We found that higher education and healthier lifestyle, as indicated by lifestyle score, were related to higher odds of DHI uptake. However, the effects of age, sex, independence of e-service use, and competence to use online services varied across lifestyle score groups. No significant interactions were observed related to educational attainment. These results imply that lifestyle DHIs are less likely to reach individuals with less-healthy lifestyle habits and lower educational attainment. In addition, some predictors affected the uptake differently across lifestyle score groups, suggesting that implementations of DHIs might attempt strategies to optimize the participation rates in especially targeted subgroups.

Keywords: Digital health intervention · Lifestyle · Education · Uptake

1 Introduction

Chronic diseases pose a significant global health challenge, imposing a growing economic burden on healthcare services [1, 2]. Digital solutions integrated into service pathways have been proposed to offer a solution to increasing service demand.

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 322–332, 2024.
https://doi.org/10.1007/978-3-031-59091-7_21

Lifestyle habits play an important role in individual's health, with factors like smoking, unhealthy diets, and physical inactivity contributing to several chronic diseases [3]. Adopting a healthy lifestyle can reduce the risk of such diseases by up to 80% [3]. Digital health interventions (DHIs) are interventions utilizing digital technologies (such as apps, digital platforms, and wearables) to enhance individual's health [4]. DHIs hold the potential to improve personalization, accessibility, and effectiveness of health-promoting services, while concurrently reducing scaling-up costs [5]. Despite the enormous potential of DHIs, our understanding of the factors influencing their uptake remains limited. Understanding these factors can help researchers and healthcare providers develop targeted interventions, thereby enhancing effectiveness.

In addition to lifestyle-related factors, educational level has also been shown to be associated with health [6]. Individuals with higher educational attainment generally experience better health and longer lives compared to those with lower educational attainment [6, 7]. When aiming to large-scale implementation of DHIs, it is essential to understand whether the intended population is reached. Particularly, the uptake of the DHI among individuals with less-healthy lifestyle habits and lower educational levels is crucial, as these groups are at a heightened risk of experiencing poorer health outcomes.

Previous research has focused on examining the lifestyle habits of users of health apps, but this approach introduces a potential bias, as the usage of the app itself may already impact lifestyle habits. The studies regarding health app users suggest a connection between smoking and health app usage [8], but findings on diet quality are controversial [9–11]. Regarding education, existing evidence suggests that individuals with higher educational levels are more inclined to use health apps [9, 12, 13]. Nonetheless, not all studies have found this association [14]. Other critical factors that could impact the uptake and adoption of DHIs are the previous use of electronic services and individuals' attitudes towards such services. Although the influences of these aspects remain largely understudied, there is some evidence suggesting that privacy concerns might play a significant role [15].

The main aim of the current study was to examine whether probability of DHI uptake in individuals with different educational levels and lifestyle vary based on their attitudes and usage of e-services. We also examined the effect of sex and age, and the association between DHI uptake and both educational attainment and overall lifestyle.

2 Methods

2.1 Participants and Study Design

This cross-sectional study was a sub-study of a Healthy Finland Survey [16], where a questionnaire on health, well-being and service use was sent to randomly selected persons over the age of 20, representing the entire adult population of Finland. Finnish speaking individuals aged 20 to 74 years who had answered the Healthy Finland questionnaire by the end of 2022 and who were not invited to participate to a health examination part of the main study, were considered eligible for this current study. An SMS invitation was sent in February 2023 to all eligible individuals with a known phone number (n = 4978 [71%]). Additionally, 2000 (29%) individuals from the rest of the eligible population were sampled based on 5-year age-groups to receive a letter invitation in February

2023. The invitation letter included web address and SMS message a direct link to the project's web page where information regarding the study, a link to the lifestyle DHI app (BitHabit), and brief information about the app and instructions on how to begin the use were given. Three participants asked their data to be removed resulting in a final sample size of 6975 individuals. The ethical approval for the study was obtained from the research ethics committee of the Finnish Institute for Health and Welfare (THL/5335/6.02.01/2022).

2.2 Digital Health Intervention App (BitHabit)

The BitHabit web-based app was developed to support the formation of healthy lifestyle habits in adults at an increased risk of type 2 diabetes [17, 18]. It is based on habit formation and self-determination theories and it aims to help the app user to try small healthy habits in their everyday life, gradually building a healthier, permanent lifestyle. Invitees had approximately three weeks to start using the app, and those who began using it within this timeframe were considered as having started using the app. To log into the app, the users had to provide a phone number and a user id that was given in the invitation. The uptake of the app was defined as accepting the invitation to participate, agreeing to the terms of the BitHabit app, and registration to the app with a phone number and user id.

2.3 Study Variables

Age and sex were obtained from Finnish National Population Register and all other study variables were obtained from the Healthy Finland survey. Age was categorized into four classes (20 – 34, 35 – 49, 50 – 64, 65 – 74 years).

The participants' educational levels were assessed by asking about the number of years of education. We categorized the participants within each age group into three education level groups (low, middle, and high) by dividing them into tertiles based on the length of their education. Overall lifestyle was evaluated with a summary score on questions about diet quality, amount of sleep (do you get enough sleep: yes, almost always or often, rarely or hardly ever, or not sure), smoking (daily, occasionally, not at all, or have never smoked) and amount of physical activity (whether or not, the participants achieve the Finnish physical activity recommendations). Diet quality included questions related frequency of consumption of different foods and drinks, and quality score was created following a method by Lindström et al. [19]. A higher lifestyle score indicated healthier lifestyle. The participants where then assigned into three groups based on their score (low, middle, high).

Use of electronic services was assessed in 6 categories: independence of e-service use, competence to use e-services, non-accessibility of e-services, concerns about data security, poor internet connections, and perceived benefits of e-services. Individual answers were used as categories for all questions except for perceived benefits of e-services. For the latter, a summary score was calculated based on participants' level of agreement with six claims, and then assigning the participants into three groups based on tertiles. Questions and the answer options are presented in Table1.

Table 1. Questions used to assess the use of electronic services

Category	Question(s)	Answer options
Use of e-services	Do you use the Internet to access e-services (e.g. My Kanta, MyTax, OmaKela)?	1) I use it independently, 2) I use it with another person's help or someone else uses it on my behalf, 3) I don't use it
Competence to use e-services	How would you rate your competence to use online services (on a computer or smartphone)?	1) No competence or low competence, 2) Moderate competence, 3) High or very high competence
Non-accessibility of e-services	How do you feel about the following statement: the electronic services are not accessible to me e.g. due to my visual impairment	1) Completely agree or somewhat agree, 2) Neither agree nor disagree, 3) Somewhat disagree or strongly disagree
Data security	How do you feel about the following statement: I am concerned about data security when it comes to my personal details	1) Completely agree, 2) Somewhat agree, 3) Neither agree nor disagree, 4) Somewhat disagree, 5) Strongly disagree
Data connections	How do you feel about the following statement: data connections are poor in my area	1) Completely agree or somewhat agree, 2) Neither agree nor disagree, 3) Somewhat disagree or strongly disagree
Benefits of digital services	Electronic services… 1) Help me to assess the need for services, 2) Support me in finding and choosing the most suitable services, 3) Make it easier for me to use services regardless of where I am and when, 4) Make it easier for me to collaborate with professionals 5) Help me to take an active role in looking after my own health and welfare, 6) Help me to take care of the health, welfare and functional capacity of family or friends	1) Completely agree, 2) Somewhat agree, 3) Neither agree nor disagree, 4) Somewhat disagree, 5) Strongly disagree

2.4 Statistical Methods

Logistic regression models were used to assess the associations of education level and lifestyle with the uptake of the DHI as well as the interaction of either lifestyle or education level and age, sex, and use of e-services. The results are presented as adjusted

odds ratios (aORs) with 95% confidence intervals (CIs). The models were adjusted for contact method (SMS or letter), age, sex, and annual household income. Model implied probabilities of uptake, shown in Fig. 1, 2, and 3, were calculated for a hypothetical population distributed uniformly over the adjusted covariates. The confidence intervals for the probabilities were obtained by non-parametric bootstrap methods, using 1000 bootstrapped samples of the study population.

The overall significance of the associations between the categorical predictors and the outcome were assessed with likelihood ratio tests (LRT), for which p-values less than 0.05 were determined to indicate significant associations. The R software, version 4.3.2. Was used to perform all statistical analyses [20].

3 Results

Of the 6975 (57% females) invitees, 1282 (67% females) started using the application. The distribution of sex, age, education, and lifestyle score is presented in Table 2 for those who started using the app and those who did not. Information regarding educational attainment was missing from 95 participants and lifestyle score was missing from 886 persons due to a missing answer in one or more of the sub questions.

Table 2. Comparison of sex, age, lifestyle score, and education among those who started using the DHI app, those who did not, and total population. Data presented as N (%).

	All	Started using DHI	Did not start using the DHI
Sex			
Female	3975 (57%)	868 (67%)	3107 (55%)
Male	3000 (43%)	419 (33%)	2581 (45%)
Age (years)			
20 – 34	1229 (18%)	250 (19%)	979 (17%)
35 – 49	1254 (18%)	303 (24%)	951 (17%)
50 – 64	2066 (30%)	361 (28%)	1705 (30%)
65 – 74	2426 (35%)	373 (29%)	2053 (36%)
Lifestyle score			
Low	1150 (19%)	172 (15%)	978 (20%)
Middle	2350 (39%)	422 (36%)	1928 (39%)
High	2588 (43%)	576 (49%)	2013 (41%)
Education			
Low	1791 (26%)	206 (16%)	1585 (28%)
Middle	3036 (44%)	549 (43%)	2487 (44%)
High	2053 (30%)	522 (41%)	1531 (27%)

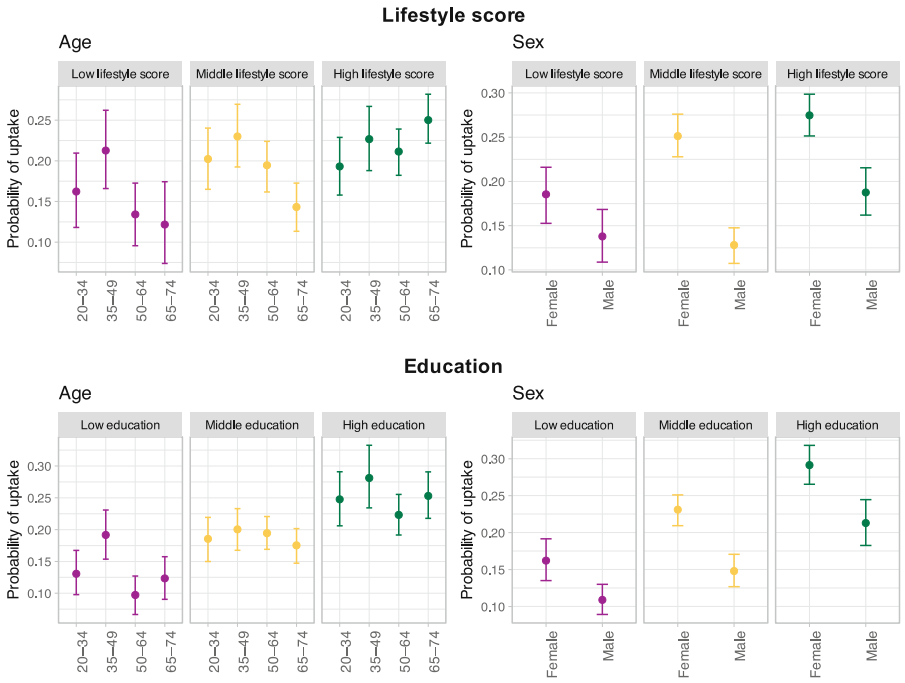


Fig. 1. Probabilities (with 95% CI) of DHI uptake in different lifestyle score and educational groups by age and sex.

Overall lifestyle and educational level were significantly associated with DHI uptake. Higher odds of DHI uptake were found in those with higher educational level (middle vs. low: aOR: 1.52, 95% CI: 1.27–1.82; high vs. low: 2.19, 1.82–2.63). Those with healthier overall lifestyle score had also higher odds of the DHI uptake (middle vs. low: 1.25, 1.03–1.53; high vs. low: 1.58, 1.30–1.93).

The interaction analyses did not reveal statistically significant interaction between education and either age or sex, but the interaction between lifestyle and age ($p = 0.001$) and sex ($p = 0.029$) was significant. In all lifestyle score groups men had significantly lower odds for DHI uptake than women but this difference was smallest in the low lifestyle score group (low aOR: 0.70, 0.50–0.98; middle 0.43, 0.34–0.54; high 0.60, 0.49–0.74). When comparing the uptakes in different lifestyle and age groups, in high lifestyle score group oldest participants (60–74-years) had significantly higher odds for uptake compared to youngest group (1.40, 1.03–1.90). On the contrary, in the middle lifestyle score group oldest age group had significantly lower odds for uptake than the reference youngest group (0.65, 0.46–0.92). In the low lifestyle score group, no age group differed significantly from the reference group. The probabilities for each category are presented in Fig. 1.

The interaction analyses regarding lifestyle and use of e-services revealed a significant interaction between lifestyle and independence of e-service use ($p = 0.042$) and competence to use online services ($p = 0.039$). Within-group comparisons showed that

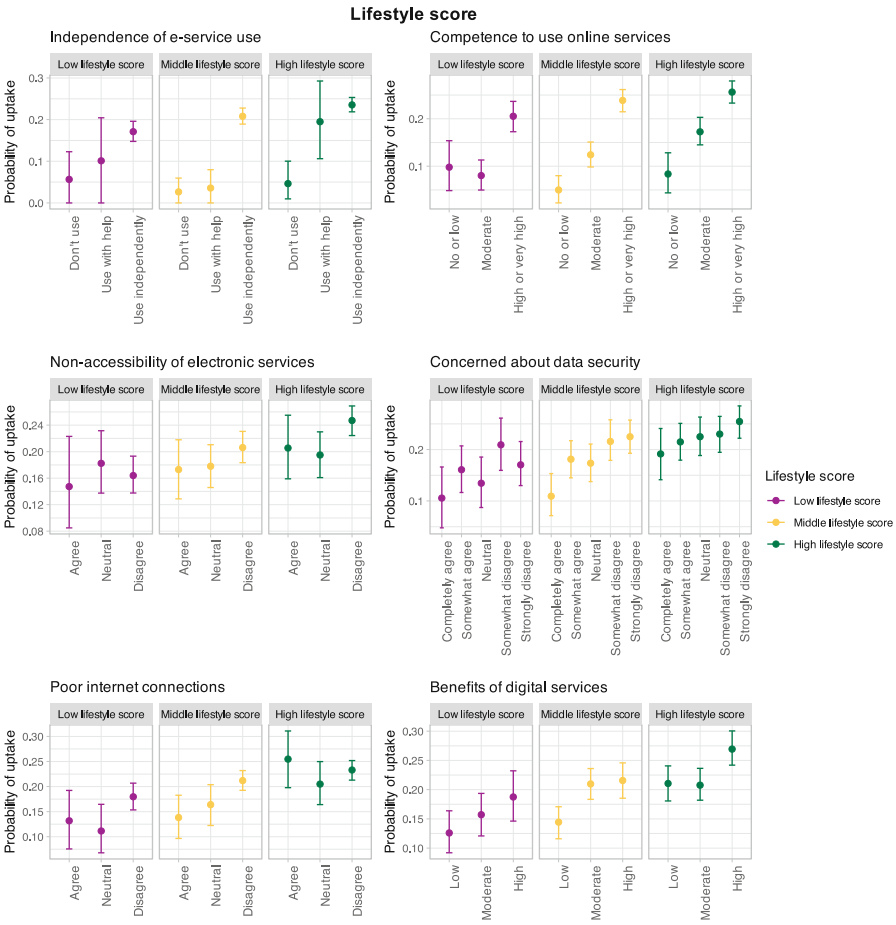


Fig. 2. Probabilities (with 95% CI) of DHI uptake in different lifestyle groups with use of e-services.

in low and middle lifestyle score groups those reporting to use e-services independently had significantly higher odds of DHI uptake than those who do not use e-services (low: 3.47, 1.07–11.2; middle 9.74, 3.07–30.86). In high lifestyle score group, those reporting to use e-services with help (5.01, 1.54–16.31) or independently (6.43, 2.35–17.62) had significantly higher odds to start using the DHI than those who did not report to use e-services. Regarding competence to use e-services, in low lifestyle score group those with high or very high competence had significantly higher odds for DHI uptake than those with no or low competence (2.44, 1.30–4.6). In middle and high lifestyle score groups those with moderate competence (middle: 2.72, 1.42–5.22; high: 2.33, 1.31–4.17) or high or very high competence (middle: 6.20, 3.30–11.64; high: 3.93, 2.22–6.95) had

higher odds of uptake than those with no or low competence. Other interactions with lifestyle were non-significant. The probabilities for each category are presented in Fig. 2. There were no significant interactions between educational attainment and any of the use of e-services variable. Probabilities for different educational groups are presented in Fig. 3.

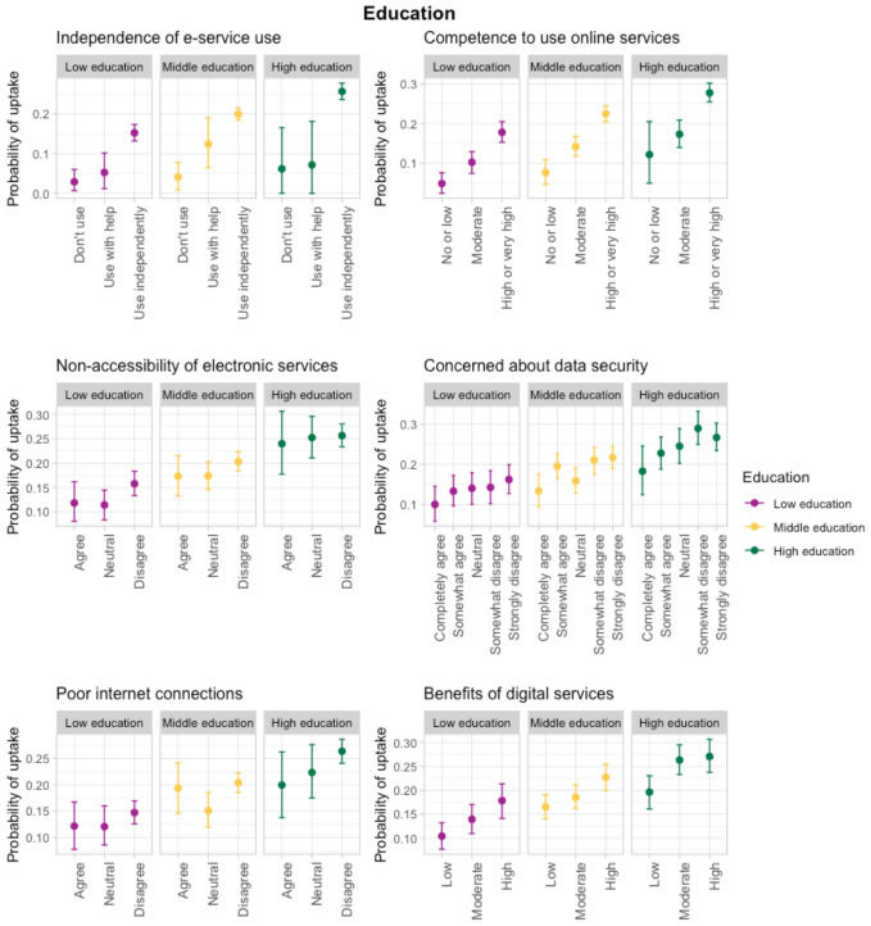


Fig. 3. Probabilities (with 95% CI) of DHI uptake in different educational groups with use of e-services.

4 Discussion

Our study showed that higher education and healthier lifestyle as indicated by higher lifestyle score are related to higher odds of lifestyle DHI uptake. However, we found that the effects of age, sex, independence of e-service use, and competence to use online services varied across lifestyle score groups. No significant interactions were observed related to educational attainment.

Lifestyle interventions, including DHIs, would be most effective in decreasing the burden of chronic diseases if they would reach individuals with less-healthy lifestyle. Reaching these individuals is crucial for preventive healthcare as these individuals are at higher risk for chronic diseases [3]. While analyzing overall barriers and facilitators of DHI uptake is crucial, a deeper understanding of these factors in various subgroups is equally essential. Firstly, adjusted logistic regression revealed that those with healthier lifestyle have higher odds of the DHI uptake. These results imply that when offering DHIs, we might not efficiently get individuals with less-healthy lifestyle habits to start using DHIs. Secondly, we found significant interactions between lifestyle groups and variables related to skills on e-service use. These findings suggest a nuanced relationship between DHI uptake, lifestyle, and competence or independence of e-service usage highlighting the importance of digital proficiency across lifestyle strata. Interestingly, significant interactions also between lifestyle groups and both age and sex were observed. In the group with the least beneficial lifestyle the probability of uptake was lowest in the oldest age group but in group with best lifestyle the probability of uptake was highest in the oldest age group. Thus, when targeting older adults with less-healthy lifestyle habits there is a need for tailored recruitment strategies and interventions.

While higher educational attainment was related to higher probability of DHI uptake the effects of age, sex and use of e-services did not seem to vary across educational levels. The uptake probability was consistently higher in those with more favorable attitudes and better skills on using e-services. The results regarding the association between education and DHI uptake in the whole study population align with earlier evidence on health app usage [9, 12, 13]. While these results show the importance of education on DHI uptake, the differences in DHI uptake among individuals with different educational attainment does not vary based on their digital literacy. These results regarding education imply a potential existence of a digital divide, wherein individuals with lower education levels may be at a disadvantage when it comes to using digital tools for managing their health. This may have a potential to worsen health inequalities as individuals with lower educations are at a heightened risk of experiencing poorer health outcomes.

Strengths of this study include large sample size and the available knowledge of the background characteristics of the whole approached population, instead of only studying health app users that is a common approach in prior research. This study was conducted in Finland and the caution should be exercised in generalizing these findings to populations in other countries with potentially different cultural, social, and economic contexts.

Acknowledgments. This study was funded by the Academy of Finland (grant numbers 332464 and 332466).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Mathers, C.D., Loncar, D.: Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. **3**, e442 (2006). https://doi.org/10.1371/journal.pmed.0030442
2. World Health Organization: World Health Organization global status report on noncommunicable diseases 2014. World Health Organization, Geneva (2014)
3. Ford, E., Bergmann, M.M., Kröger, J., Schienkiewitz, A., Weikert, C., Boeing, H.: Healthy living is the best revenge: findings from the european prospective investigation into cancer and nutrition–Potsdam study. Arch. Intern. Med. **169**, 1355 (2009). https://doi.org/10.1001/archinternmed.2009.237
4. Murray, E., et al.: Evaluating digital health interventions. Am. J. Prev. Med. **51**, 843–851 (2016). https://doi.org/10.1016/j.amepre.2016.06.008
5. Gentili, A., et al.: The cost-effectiveness of digital health interventions: a systematic review of the literature. Front. Public Health **10**, 787135 (2022). https://doi.org/10.3389/fpubh.2022.787135
6. Raghupathi, V., Raghupathi, W.: The influence of education on health: an empirical assessment of OECD countries for the period 1995–2015. Arch. Public Health. **78**, 20 (2020). https://doi.org/10.1186/s13690-020-00402-5
7. Zajacova, A., Lawrence, E.M.: The Relationship between education and health: reducing disparities through a contextual approach. Annu. Rev. Public Health **39**, 273–289 (2018). https://doi.org/10.1146/annurev-publhealth-031816-044628
8. Stühmann, L.M., et al.: Health app use and its correlates among individuals with and without type 2 diabetes: nationwide population-based survey. JMIR Diabetes. **5**, e14396 (2020). https://doi.org/10.2196/14396
9. Carroll, J.K., Moorhead, A., Bond, R., LeBlanc, W.G., Petrella, R.J., Fiscella, K.: Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J. Med. Internet Res. **19**, e125 (2017). https://doi.org/10.2196/jmir.5604
10. Coughlin, J.W., et al.: Electronic health record-based recruitment and retention and mobile health app usage: multisite cohort study. J. Med. Internet Res. **24**, e34191 (2022). https://doi.org/10.2196/34191
11. Ernsting, C., et al.: Using smartphones and health apps to change and manage health behaviors: a population-based survey. J. Med. Internet Res. **19**, e101 (2017). https://doi.org/10.2196/jmir.6838
12. Reiners, F., Sturm, J., Bouw, L.J.W., Wouters, E.J.M.: Sociodemographic factors influencing the use of ehealth in people with chronic diseases. Int. J. Environ. Res. Public Health **16**, 645 (2019). https://doi.org/10.3390/ijerph16040645
13. Szinay, D., Jones, A., Chadborn, T., Brown, J., Naughton, F.: Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J. Med. Internet Res. **22**, e17572 (2020). https://doi.org/10.2196/17572
14. Xie, Z., Nacioglu, A., Or, C.: Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR MHealth UHealth. **6**, e103 (2018). https://doi.org/10.2196/mhealth.9002
15. Jokisch, M.R., Schmidt, L.I., Doh, M.: Acceptance of digital health services among older adults: Findings on perceived usefulness, self-efficacy, privacy concerns, ICT knowledge, and support seeking. Front. Public Health **10**, 1073756 (2022). https://doi.org/10.3389/fpubh.2022.1073756

16. Lundqvist, A., Tolonen, H.: Healthy Finland survey as the latest uptake for 50 years of population health monitoring in Finland. Eur. J. Public Health. **33**, ckad160.720 (2023). https://doi.org/10.1093/eurpub/ckad160.720
17. Harjumaa, M., et al.: Internet-based lifestyle intervention to prevent type 2 diabetes through healthy habits: design and 6-month usage results of randomized controlled trial. JMIR Diabetes **5**, e15219 (2020). https://doi.org/10.2196/15219
18. Lavikainen, P., et al.: Digitally supported lifestyle intervention to prevent type 2 diabetes through healthy habits: secondary analysis of long-term user engagement trajectories in a randomized controlled trial. J. Med. Internet Res. **24**, e31530 (2022). https://doi.org/10.2196/31530
19. Lindström, J., et al.: Formation and validation of the healthy diet index (HDI) for evaluation of diet quality in healthcare. Int. J. Environ. Res. Public Health **18**, 2362 (2021). https://doi.org/10.3390/ijerph18052362
20. R Core Team: A language and environment for statistical computing. R Foundation for Statistical Computing (2023). https://www.R-project.org/

A Mobile Application Can Be Used as an Alternative to the Traditional Preparation Method for Parents in Pediatric Day Surgery: A Randomized Controlled Trial

Heli Kerimaa[1,3(✉)], Marianne Haapea[3,4], Mervi Hakala[1,2,3], Willy Serlo[2], and Tarja Pölkki[1,3]

[1] Research Unit of Health Sciences and Technology, University of Oulu, Oulu, Finland
Heli.kerimaa@oulu.fi
[2] Oulu University Hospital, Oulu, Finland
[3] Medical Research Center Oulu, University Hospital and University of Oulu, Oulu, Finland
[4] Research Service Unit, Oulu University Hospital, Oulu, Finland

Abstract. Background: Digital preparation programs for day surgery are now available through smartphones; however, research on the effectiveness of digital interventions among parents is lacking.

Aim: This study aimed to assess the effectiveness of a mobile application intervention in preparing parents for pediatric day surgery and to describe the correlations between parents' anxiety, stress, and satisfaction.

Methods: A total of 70 parents of preschool children who were scheduled for elective day surgery were randomly divided into two groups: the intervention group (IG; n = 36) and the control group (CG; n = 34). The study took place in the pediatric day surgical department of a university hospital in Finland. The IG used a mobile application, while the CG used routine methods. Parents' anxiety, stress and satisfaction were measured using validated instruments.

Results: There was no significant difference in parental anxiety levels between the two groups, both before and after the surgery. After the surgery, both groups of parents reported feeling less anxious while at home. Pre-surgery, most parents experienced no/mild stress at home. However, post-surgery, intervention group parents reported significantly less stress at home than control group parents. The mean VAS score for parents' satisfaction in both groups was high: 8.8 for the intervention group (SD 1.9) and 8.6 for the control group (SD 0.9). These mean scores did not significantly differ. Anxiety, stress, and satisfaction showed a significant correlation in most cases at both T1 and T4.

Conclusions: A mobile application can serve as an alternative to the traditional method of preparing parents for pediatric day surgery.

Keywords: Anxiety · Pediatric day surgery · Mobile application · Preparation · Parent · Randomized controlled trial · Satisfaction · Stress

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 333–348, 2024.
https://doi.org/10.1007/978-3-031-59091-7_22

1 Introduction

Globally, there is an increase in the number of day surgery procedures being performed on children [1–4]. Children can return home and resume normal activities the same day after day surgery. Parents must be educated about the different stages involved in day surgeries [5, 6]. Parents often face anxiety and stress when preparing their child for surgery, as they must bear responsibility and provide support [7–9]. Digital preparation programs for day surgery are now available through smartphones, thanks to advancements in information and communication technologies [10, 11].

1.1 Background

In day surgery, parents should be informed about the procedure to improve cooperation between the child and healthcare staff. Research indicates that parental anxiety and stress can be transmitted to children, leading to feelings of pain and fear [7–9, 12, 13]. Parents often feel anxious during their child's surgery due to a lack of control in a new environment [14]. This can result in feelings of guilt, lack of awareness, separation anxiety, and loss of control. Preparing children for day surgery can be a challenging task for parents who need to address their concerns and prevent negative experiences such as anxiety and stress [14]. Parents should support their child experiencing fear and anxiety, as these dynamics can negatively impact day surgery procedure and recovery [5, 6, 15, 16]. The situation can be challenging for children due to their active imaginations, which are stronger at younger ages and prevent them from employing abstract logical thinking [17].

Insufficient time for preparation has been associated with increased levels of anxiety and stress among parents [11]. While there is considerable knowledge on preparing parents for day surgery [5, 7, 18], there is lack on knowledge about the effectiveness of digital a interventions in surgical preparation [19]. According to Liu et al. [19] the mobile application was effective in preparing parents for their child's hernia surgery. According to the findings, the application resulted in higher parental awareness and decreased the number of surgery cancellations [19].

Smartphones and their applications are integral to daily life. The WHO [20] has reported that digitalization supports human health, improves access to quality health services and enhances the efficiency of health systems. Integrating multiple methods before, during, and after day surgery may increase effectiveness. To optimize solutions that benefit both parents and children, it's crucial to identify the key factors that lead to the best outcomes [21]. This study was conducted to determine if a mobile application intervention, which included audio-visual content, instructions, images, and timelines, was effective to decrease parental anxiety and stress. Mobile application interventions for pediatric day surgery preparation in parents have not been widely studied. Previous research has mainly concentrated on studying different age groups and children with long-term illnesses [22]. Research on digital interventions to support families and children is still lacking [23].

2 Methods

2.1 Aim

This study aimed to assess the effectiveness of a mobile application intervention in preparing parents for pediatric day surgery. Our hypothesis was as follows: parents in the intervention group have lower levels of anxiety (primary outcome) and stress during the day surgery process, as well as be more satisfied with the preparation they received (secondary outcomes), than parents in the control group. The study also aimed to describe the correlations between anxiety, stress and satisfaction experienced by parents.

2.2 Design

The study was designed as a randomized controlled trial (RCT) with two groups.

2.3 Participants

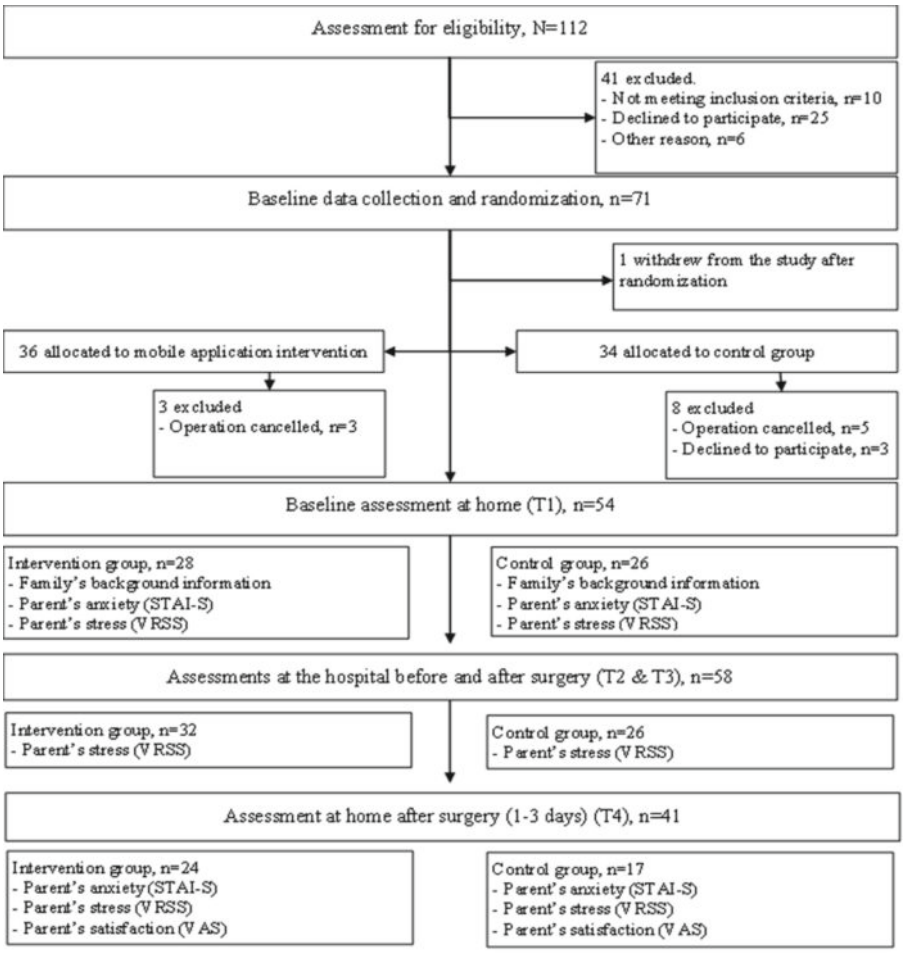
The study was carried out between 2018 and 2019 at the pediatric day surgical department of one university hospital in Finland. The study included parents of children ages 2–6 who had elective day surgery under general anesthesia. (Table 1). The mobile application allowed parents to prepare their children up to 3–4 weeks prior to the operation. The sample size calculation was based on the study of Kain et al. [24]. They found that 46% of parents experience anxiety before their child's surgery, according to the State-Trait Anxiety Inventory (STAI) as the primary outcome. For this study, we used an independent-sample t-test with an alpha value of 0.05 and 80% power. Based on this, we estimated that we would need 50 participants in total, with 25 participants in the intervention group and 25 participants in the control group. Sample size adjusted for 30% dropout rate, resulting in 70 parents: 51% (36 parents) in the intervention group and 49% (34 parents) in the control group.

Table 1. Study inclusion criteria.

Context	Inclusion criteria
Child surgery	Hernias, foreskin stenosis, testicular repairs, skin and subcutaneous tissues, orthopaedics
Risk classification	ASA 1–2
Pain management	Pre-medicated analgesic and local anesthesia after surgery
Pre-medication	If needed
Parents	Parents who have an Android or iOS phone, iPad, or internet browser can have access
Other criteria	Families with Finnish-speaking

Participants Randomization

The eligible participants were divided into five age groups (2, 3, 4, 5, and 6 years) and then randomly assigned to each group with a 1:1 ratio using stratified simple randomization [25]. The researcher prepared two envelopes beforehand; one for 2-year-olds and the other for children up to six years old. Each age group envelope contained 10 notes: 5 for the intervention group and 5 for the control group. The researcher randomly selected one note from the envelope after a phone call with the parents to determine the family's group assignment. Both the participants and the researcher did not know the group allocation beforehand, and ethical rules were followed. A flow chart is shown in Fig. 1.



STAI-S = State Anxiety Inventory, VAS = Parent's satisfaction, VRSS = The Verbal Rating Scale for Stress.

Fig. 1. The study CONSORT diagram.

2.4 Participants Randomization

Development of the Mobile Application
A mobile application (BuddyCare) was developed in collaboration with the pediatric hospital unit and the commercial company to support parents before and after surgery. The application offered clear and easy-to-follow instructions for parents in various formats, including videos and photos from the pediatric day surgery unit, guidance on surgery and pain care, directions to the hospital and the ward, notifications, required forms, as well as written instructions. The application lacked a chat feature through which users could communicate with healthcare providers. Users shared this information with the hospital, allowing healthcare professionals to stay on track with the parents' preparation for the pediatric surgery.

Mobile Application for Parents
The intervention was tested on five families in the pediatric day surgical department between December 2016 and January 2017 using a mobile application. The hospital staff distributed materials to parents via an automated system upon surgery confirmation. The application allowed for the completion of pre-information forms and reminders about preparing the child and provided necessary surgery information. Parents could access the application conveniently and contact the hospital if needed.

Parents in the intervention group were given access to the BuddyCare application or web portal three to four weeks prior to pediatric day surgery. The application provided color-coded reminders before the surgery, using a timeline with spacers for easy understanding. All parents were provided with identical information, but they had the freedom to select the materials they preferred and the amount they wanted to use. The forms were then sent to the hospital staff in day surgery unit. The nurses kept a close eye on the application's usage, and no problems were reported.

Traditional Preparation for Parents
For pediatric day surgery, parents in the control group were provided with written instructions and a video and were given a possibility to contact the hospital if needed. One day prior to the day surgery, a nurse reached out to the parents and provided them with the date and time of the surgery. During the call, the nurse checked the child's health and filled out a form with information provided by the parents. The form contained questions about the child's allergies, underlying medical conditions, and current health status. The nurse provided instructions to parents on when it was allowed to drink and eat before the surgery, and how to prepare for arrival at the hospital. Additionally, parents were given the opportunity to ask any questions they may have had. This call was similar to the intervention provided through the mobile application.

2.5 Data Collection

The researcher contacted each parent to inquire about participation. Before that a pediatric surgeon had evaluated the need for surgery and secretaries had been trained by the researcher identified participants based on inclusion criteria. The intervention group

received instructions for a mobile application to help prepare for surgery, while the control group received conventional instructions.

Parents brought a consent form and first measurements including demographic data and self-reported stress and anxiety, to the hospital (T1). Before taking measurements, the nurses were not informed whether the families belonged to the intervention or control group, and the researcher did not participate in data collection. The data on parents' stress levels before (T2) and after (T3) their child's surgery in the hospital was collected by the nurse. Afterwards (T4), the parents assessed their anxiety, stress, and satisfaction with their childs' preparation (Fig. 1).

Measurements

For this study, the State-Trait Anxiety Inventory (STAI) S-Anxiety scale was used. This scale comprised of 20 items that were rated on a four-point Likert scale to measure the levels of anxiety experienced by parents. The responses ranged from "rarely" to "almost constantly." The total score for the S-Anxiety scale ranged from mild anxiety (20 to 39) to moderate anxiety (40 to 59) and finally to intense anxiety (60 to 80) [26]. According to the study of Gustafson, the internal consistency of STAI has demonstrated good reliability [27].

The Verbal Rating Scale for Stress (VRSS) is a tool that was used to evaluate the parents' stress levels. The scale ranged from 0 to 5, where 0 indicated no stress at all and 5 indicated the highest possible level of stress. The VRSS has demonstrated good reliability and validity in Alven's research [28].

The satisfaction levels of parents were measured using the Visual Analogue Scale (VAS). This scale consisted of a 100 mm line, with "I was not satisfied" marked at one end and "I was delighted" marked at the other. Parents indicated their level of satisfaction by marking the line at a point between 0 and 10. The VAS is a reliable and valid measurement tool [29].

2.6 Ethical Issues

The study conducted followed the Helsinki declaration and was approved by the Northern Ostrobothnia Regional Ethics Committee Board (EETTMK: 53/2017). At all stages of the research, ethical considerations were considered, including privacy, data protection, and participants' right to information, respect, and honesty. Parents were informed about the study on preparing their child for day surgery and written consent was obtained.

2.7 Data Analysis

All statistical analyses were conducted using the IBM SPSS statistical software for Windows (version 27; SPSS Inc., Chicago, IL). The t-test was used to assess the significance of between-group differences in the measured variables, while the Chi-square test was employed to compare the changes within and between pre-operative (T1, T2) and post-operative (T3, T4) measurements. Repeated Measures Analysis of Variance revealed differences between the intervention group and the control group in the STAI pre-surgery (T1) and post-surgery measurements (T4). The significance level for statistics was set at $p < 0.05$. Parental anxiety was assessed before and after the day surgery,

and the scores of all parents in each group were summed up to calculate the average anxiety. Additionally, Spearman's rho coefficient (T1, T4) was used to examine the correlations between anxiety, satisfaction, and stress.

2.8 Validity and Reliability

The nurses who took part in data collection were given training by a researcher (HK). Buddy Healthcare employees trained the hospital staff on how to use the mobile application intervention. The study results were reported in accordance with the CONSORT Statement [30] and registered at ClinicalTrials.gov (NCT03774303). The data was collected by parents using validated measurements.

3 Results

3.1 Demographic Data

In this study, 70 parents were involved, with 36 in the intervention group and 34 in the control group. Before randomization, six families refused to participate in the study, citing reasons such as lack of time (n = 3), fear of surgery (n = 1), language barriers (n = 1), and recent participation in another study (n = 1). The loss rate for the intervention group was 22%, while the control group had a 24% loss rate. Apart from gender distribution, no significant differences were found in the demographic data (Table 2).

3.2 Anxiety

The study found no significant difference in anxiety levels between the intervention and control groups before and after surgery. Pre-operative anxiety score was 36.7 (SD 9.9; 78%) in the intervention group and 36.9 (SD 12.3; 76%) in the control group ($p = 0.95$). Post-operative anxiety score was 28.1 (SD 6.9; 67%) in the intervention group and 30.2 (SD 7.06; 50%) in the control group ($p = 0.34$).

Most of the parents in both groups experienced mild anxiety before (in the intervention group 68% and in the control group 69%; $p = 0.77$) and after the surgery (in the intervention group 88% and in the control group 94%; $p = 0.63$).

The results indicated a significant decrease in anxiety levels for both the intervention and control groups after surgery (IG: $p = 0.003$; CG: $p = 0.002$). There was no significant difference in anxiety levels between the two groups ($p = 0.13$). Parental anxiety decreased overall from pre-surgery (mean 36.3, SD 10.3) to post-surgery assessment (mean 29.4, SD 6.9).

3.3 Stress

The stress levels experienced by parents in both groups at home were similar before surgery. The majority of parents in both groups did not experience any stress or only experienced mild stress. Specifically, 61% of parents in the intervention group and 50% in the control group reported no or mild stress, with no significant difference between

Table 2. Demographic data for the participants.

	Intervention group n (%)	Control group n (%)	Total n (%)	p-value
Participants	28 (51.8%)	26 (41.9%)	54 (100%)	
Parents age category (years)				
25–30 31–35 36–40 41-over 50	8 (28.6) 6 (21.4) 10 (35.7) 4 (11.3)	6 (23,.1) 6 (23.1) 10 (38.4) 4 (15.4)	14 (25.9) 12 (22.2) 20 (37.0) 8 (14.9)	ns
Gender				
Female Male	28 (100.0) 0	19 (73.1) 7 (26.9)	47 (87.0) 7 (13.0)	0.004
Marital status				
Married Cohabitation Single parent-other	21 (75.0) 5 (17.8) 2 (7.2)	20 (76.9) 6 (23.1) 0 (0.0)	41 (75.9) 11 (20.3) 3 (3.8)	ns
Educational level				
No education				
Vocational education or A college- or polytechnic education University education	1 (3.6) 22 (78,6) 5 (17.8)	2 (7.7) 18 (69.2) 6 (23.1)	3 (5.6) 40 (74,0) 11 (20.4)	ns
Child's age (years)				
2–4 5–6	13 (46,4) 15 (53.6)	11 (42.3) 15 (57.7)	24 (44.5) 30 (55.5)	ns
Previous hospital experience				
No Yes, once or many times	10 (35.7) 18 (64.3)	13 (50.0) 13 (50.0)	23 (42.6) 31 (57.4)	ns

ns = non-significant

the groups ($p = 0.61$). However, before surgery at the hospital, most parents in the intervention group felt mild stress (77%) or moderate to intense stress (23%), while in the control group, 23% of parents felt no stress. There was a statistically significant difference between the groups, with a p-value of 0.02. After surgery at the hospital, most parents in both groups experienced no stress, with 47% in the intervention group and 50% in the control group reporting no stress ($p > .99$). After surgery at home, none of the parents in the intervention group, but 18% of parents in the control group experienced moderate to intense stress. The groups showed a statistically significant difference ($p = 0.05$). The stress levels decreased significantly after surgery in both groups, with a significant decrease observed for both the intervention group ($p = 0.003$) and the control group ($p = 0.004$) from T1 to T4.

3.4 Satisfaction

The mean VAS score for parents in both groups was high: 8.8 for the intervention group (SD 1.9) and 8.6 for the control group (SD 0.9). These mean scores did not significantly differ ($p = 0.794$). Of the parents who participated in the research, 10 (24%) gave the maximum score of 10 when asked about preparation, 11 (27%) scored the preparation as 9–9.9, and 11 (27%) scored their satisfaction with the preparation as 8–8.9. The results revealed that 20% of parents were not completely satisfied.

3.5 Correlations Between Anxiety, Stress and Satisfaction

Before the surgery (T1) there was a significant correlation between the anxiety experienced by parents and the perceived satisfaction in the intervention group (-0.624; $p = 0.002$) but not in the control group (-0.449; $p = 0.071$). After the surgery (T4) there was a significant correlation between the anxiety experienced by parents and the perceived satisfaction in both groups (IG; -0.565; $p = 0.05$) and (CG; -0.640; $p = 0.006$) (Fig. 2).

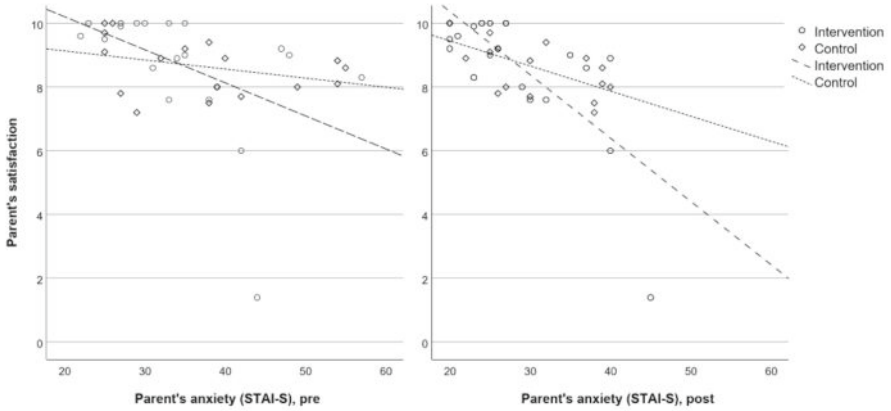


Fig. 2. Correlation between anxiety and satisfaction levels of parents in the intervention and control groups, before (T1) and after (T4) pediatric day surgery at home.

Before the surgery (T1) there was a significant correlation between the stress experienced by parents and the perceived anxiety in both groups (IG; 0.527; $p = 0.004$ and CG; 0.823; $p = 0.000$). After the surgery (T4) there was a significant correlation between the stress experienced by parents and the perceived anxiety in the CG (0.725; $p = 0.001$), but not in the IG (0.159; $p = 0.457$) (Fig. 3).

Before the surgery (T1) there was not a significant correlation between the stress experienced by parents and the perceived satisfaction in both groups (IG; 0.028; $p = 0.903$ and (CG; -0.229; $p = 0.260$). After the surgery (T4) there was also not a significant correlation between the stress experienced by parents and the perceived satisfaction in both groups (IG; 0.204; $p = 0.351$ and CG; -0.476; $p = 0.053$) (Fig. 4).

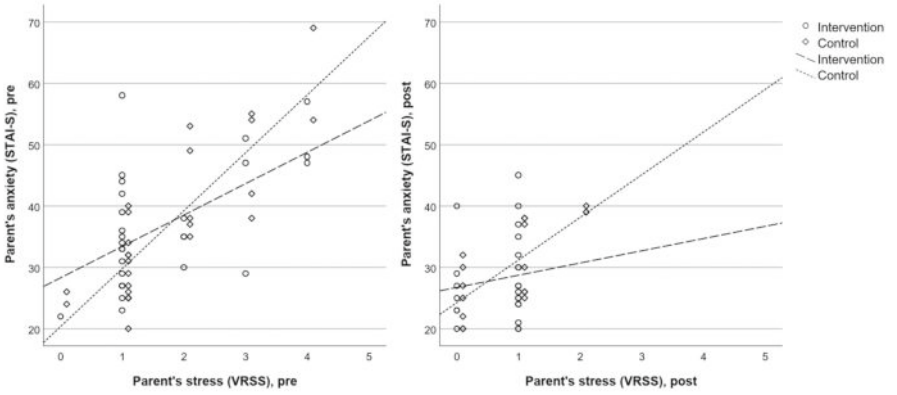


Fig. 3. Correlation between stress and anxiety levels of parents in the intervention and control groups, before (T1) and after (T4) pediatric day surgery at home.

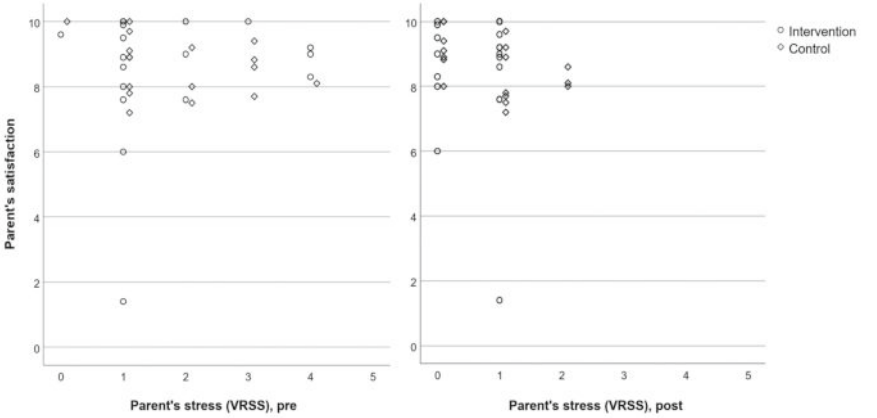


Fig. 4. Correlation between stress and satisfaction of parents in intervention and control groups before (T1) and after (T4) pediatric day surgery at home.

4 Discussion

This study evaluated the effectiveness of a mobile application intervention for parents of preschool children who are preparing for day surgery. The intervention had no effect on reducing parents' anxiety levels, but parents in both groups experienced a significant decrease in anxiety levels when comparing before to after the surgery. Before surgery, only a small percentage of parents in the intervention group experienced mild stress, compared to the control group where 23% of parents experienced mild stress. Parents in the intervention group experienced less stress after surgery, and both groups showed a

decrease in stress levels. Anxiety, stress, and satisfaction were significantly correlated, highlighting the importance of considering parental anxiety. The mobile application offers affordable parental support, encouraging future usage.

Based on the findings, it is common for parents to experience mild anxiety and stress prior to their child's day surgery; a similar finding was reported by Justus [15]. It was noteworthy that parents in the control group experienced at home more post-surgery stress than parents in the intervention group. This finding may be explained by parents' concerns about the child's recovery, pain management, and possible nausea, among others. Previous studies have found that parents want to take responsibility for supporting their child in the best possible way [5, 6]. This means that parents still have a significant burden to bear at home after the procedure and, as such, they will need support, help, and adequate follow-up instructions from healthcare providers during this period. In addition, parental anxiety and stress transmission to a child is known to cause fear and pain among children, which can affect their post-surgery recovery [16]. To counteract this, healthcare providers should pay attention to the continuity of post-surgery care. This is an important research topic for the future, as the number of day surgeries among preschool children is increasing on a global level.

The families who participated in the study expressed overall satisfaction with the preparation they received for day surgery, though it is important to note that the procedure can be a significant event for both the child and their family [31, 32]. Therefore, preparation for day surgery should consider the whole family's needs and be flexible and supportive. Hospitals should prioritize the development of tools that are family-oriented and take into account the needs of each individual when it comes to preparation [33, 34]. This can help strengthen parents' coping. Information presented in an age-appropriate way is more effective for children's understanding [32].

Based on the results presented, we can conclude that using a mobile application is equally effective to the traditional preparation method. Parents often experience stress and anxiety while their child is hospitalized, as reported by many studies [12, 13, 35]. The results of a longitudinal pilot study by Wray [35] demonstrated high levels of anxiety and stress among parents shortly after their child's admission, with these levels remaining elevated at discharge. According to our research, high-quality and adaptable preparation can help alleviate the anxiety and stress experienced by parents during hospital stays. Providing accurate information before and during the stay is crucial in reducing parental anxiety [36]. A mobile application intervention can achieve the same goal, designed to support the needs of both healthcare providers and families.

An mHealth app delivers information and preparation material instantly to parents regardless of their location or environment through videos and images [37]. Well-designed and versatile mobile applications are also suitable for families who need individual guidance and travel a lot, which is becoming more common in contemporary society. It is crucial to offer families a mobile application that provides diverse information in various formats. This will ensure that parents from different backgrounds receive adequate support. Kampouroglou [38] found that some parents prefer visual aids like images and videos to written information, as every individual has unique preferences.

According to this study, a mobile application intervention can effectively assist parents of preschool children in preparing for day surgery. However, further research is

required to gain more up-to-date insights into the experiences of families during pediatric surgery. This will enable the development of interventions that provide families with the appropriate care and support they require. Our study showed that relevant mobile application content, when provided at the right time, enables families to internalize information and adequately prepare their children for the procedure. According to Free [39], it is essential to consider the timeline of a procedure when designing an intervention, i.e., the preparation content should be synchronized with human attention at a time when it is most relevant. mHealth apps improve healthcare delivery and management by transforming information exchange and storage. Furthermore, mobile applications have now been a part of the daily life of adults for some time; this means that they can reach the entire adult population.

Mobile health applications can assist parents in preparing for day surgeries and improve communication between families and healthcare providers [37]. ICT can enhance health promotion services, making healthcare more accessible, effective, equitable, and rational [40, 41]. VR technology can create virtual tours of operating theaters and other relevant areas [10]. The primary goal of preparing parents for their child's pediatric day surgery should be to guide them through the care process. It's necessary to let parents know what to expect during the journey from home to the hospital and back, including the important issues they need to be aware of post-discharge [42]. Families should receive post-procedure recovery information upon returning home to ensure continued care. It's essential to remember that digitalization is not just about converting paper-based information into electronic form but leveraging the range of digital tools that are available for users. While doing so, it is crucial to ensure that digitalization facilitates genuine social interaction.

4.1 Limitation

The study has certain limitations despite its aim to maximize validity and reliability. The sample size was inadequate to detect significant differences between groups due to missing responses from some parents. The statistical difference between the groups regarding the gender of the parents may affect the reliability of the results. Also, blinding was difficult, and some parents decided not to participate. The study results have been reported transparently, including statistically insignificant results. All participants were recruited from a single university hospital, which may limit generalizability of the results. Future studies should assess the effectiveness of mobile interventions over a longer period of time.

4.2 Conclusion

The mobile application intervention did not decrease anxiety, but it did help to reduce stress levels in parents. It seemed that mobile application interventions can be used to prepare preschool children's parents for day surgery as an alternative to the traditional preparation method. Although mobile applications cannot fully replace face-to-face interaction, they could be a cost-effective option in the future. Future developments should consider the individual characteristics and needs of families in pediatric care to offer new effective solutions.

Acknowledgments. The authors express their gratitude to the parents and healthcare professionals who took part in this study. Thanks to the organizations supporting the study: the Research Foundation of the Mannerheim Child Protection Union, the Pediatric Research Foundation, and the Society for Research in Nursing Sciences (HTTS).

Disclosure of Interests. The authors have not declared any conflicts of interest, and there has been no patient or public contribution. There are no declarations of interest, and the authors have no competing interests to declare.

References

1. Erhaze, E.K., Dowling, M., Devane, D.: Parental presence at anaesthesia induction: a systematic review. Int. J. Nurs. Pract. **22**(4), 397–407 (2016). https://doi.org/10.1111/ijn.12449
2. Pomicino, L., Maccacari, E., Buchini, S.: Levels of anxiety in parents in the 24hr before and after their child's surgery: a descriptive study. J. Clin. Nurs. **27**(1–2), 278–287 (2018). https://doi.org/10.1111/jocn.13895
3. Quemby, D.J., Stocker, M.E.: Day surgery development and practice: key factors for a successful pathway. Contin. Educ. Anaesth. Crit. Care Pain **14**(6), 256–261 (2014). https://doi.org/10.1093/bjaceaccp/mkt066
4. Wright, K.D., Stewart, S.H., Finley, G.A.: When are parents helpful? A randomized clinical trial of the efficacy of parental presence for pediatric anesthesia. Can. J. Anaesth. J. Can. Anesth. **57**(8) 751–758 (2010). https://doi.org/10.1007/s12630-010-9333-1
5. Chang, C.M., et al.: A qualitative study exploring Singaporean parents' perceptions on preparing their child for surgery. J. Pediatr. Nurs. **54**, e69–e77 (2020). https://doi.org/10.1016/j.pedn.2020.05.004
6. Healy, K.: A descriptive survey of the information needs of parents of children admitted for same day surgery. J. Pediatr. Nurs. **28**(2) 179–185 (2013). https://doi.org/10.1016/j.pedn.2012.07.010
7. Carlsson, R.N.E., Henningsson, R.N.: Visiting the operating theatre before surgery did not reduce the anxiety in children and their attendant parent. J. Pediatr. Nurs. **38**, e24–e29 (2018). https://doi.org/10.1016/j.pedn.2017.09.005
8. Gabriel, M.G., et al.: The psychosocial experiences and needs of children undergoing surgery and their parents: a systematic review. J. Pediatr. Health Care **32**(2), 133–149 (2018). https://doi.org/10.1016/j.pedhc.2017.08.003
9. Pomicino, L., Maccacari, E., Buchini, S.: Levels of anxiety in parents in the 24 hr before and after their child's surgery: a descriptive study. J. Clin. Nurs. **27**(1–2), 278–287 (2018). https://doi.org/10.1111/jocn.13895
10. Agbayani, C.-J.G., Fortier, M.A., Kain, Z.N.: Non-pharmacological methods of reducing perioperative anxiety in children. BJA Educ. **20**(12), 424–430 (2020). https://doi.org/10.1016/j.bjae.2020.08.003
11. Fortier, M.A., et al.: Web-based tailored intervention for preparation of parents and children for outpatient surgery (WebTIPS): formative evaluation and randomized controlled trial. Anesth. Analg. **120**(4), 915–922 (2015). https://doi.org/10.1213/ANE.0000000000000632

12. Ayenew, N.T., Endalew, N.S., Agegnehu, A.F., Bizuneh, Y.B.: Prevalence and factors associated with preoperative parental anxiety among parents of children undergoing anesthesia and surgery: a cross-sectional study. Int. J. Surg. Open **24**, 18–26 (2020). https://doi.org/10.1016/j.ijso.2020.03.004
13. Hoetzenecker, W., et al.: Parental anxiety and concern for children undergoing dermatological surgery. J. Dermatol. Treat. **25**(5), 367–370 (2014). https://doi.org/10.3109/09546634.2013.814757
14. Tural Buyuk, E., Bolişik, B.: An analysis of the anxiety levels of mothers who participate in education and therapeutic games about their children's surgeries. J. Perianesth. Nurs. **33**(3), 290–295 (2018). https://doi.org/10.1016/j.jopan.2016.09.011
15. Justus, R., Wyles, D., Wilson, J., Rode, D., Walther, V., Lim-Sulit, N.: Preparing children and families for surgery: mount Sinai's multidisciplinary perspective. Pediatr. Nurs. **32**(1), 35–43 (2006). http://pc124152.oulu.fi:8080/login?url=
16. Copanitsanou, P., Valkeapää, K.: Effects of education of paediatric patients undergoing elective surgical procedures on their anxiety - a systematic review. J. Clin. Nurs. **23**(7–8), 940–954 (2014). https://doi.org/10.1111/jocn.12187
17. Lee, J.-H., Jung, H.-K., Lee, G.-G., Kim, H.-Y., Park, S.-G., Woo, S.-C.: Effect of behavioral intervention using smartphone application for preoperative anxiety in pediatric patients. Korean J. Anesthesiol. **65**(6), 508–518 (2013). https://doi.org/10.4097/kjae.2013.65.6.508
18. Boles, J.: Preparing children and families for procedures or surgery. Pediatr. Nurs.**42**, 147–149 (2016)
19. Liu, J., et al.: 'Effects of using WeChat-assisted perioperative care instructions for parents of pediatric patients undergoing day surgery for herniorrhaphy. Patient Educ. Couns. **101**(8), 1433–1438 (2018). https://doi.org/10.1016/j.pec.2018.02.010
20. WHO. Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health. World Health Organ. (2018). https://apps.who.int/iris/handle/10665/260480
21. Santapuram, P., Stone, A.L., Walden, R.L., Alexander, L..: Interventions for parental anxiety in preparation for pediatric surgery: a narrative review. Children **8**(11) (2021). https://doi.org/10.3390/children8111069
22. Majeed-Ariss, R., et al.: Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J. Med. Internet Res. **17**(12), e287 (2015). https://doi.org/10.2196/jmir.5043
23. Rantala, A., Jansson, M.M., Helve, O., Lahdenne, P., Pikkarainen, M., Pölkki, T.: Parental experiences of the pediatric day surgery pathway and the needs for a digital gaming solution: qualitative study. JMIR Med. Inform. **8**(11), e23626 (2020). https://doi.org/10.2196/23626
24. Kain, Z.N., Maclaren, J., Weinberg, M., Huszti, H., Anderson, C., Mayes, L.: How many parents should we let into the operating room? Pediatr. Anesth. **19**(3), 244–249 (2009). https://doi.org/10.1111/j.1460-9592.2008.02889.x
25. Rahul, M., Tewari, N., Goel, S.: How un-random is the randomization. J. South Asian Assoc. Pediatr. Dent. **4**(1), 66–70 (2021). https://doi.org/10.5005/jp-journals-10077-3074
26. Spielberger, C., Gorsuch, R., Lushene, R., Vagg, P., Jacobs, G.: Manual for the State-Trait Anxiety Inventory (Form Y1 – Y2), vol. IV (1983)
27. Gustafson, L.W., et al.: Validity and reliability of State-Trait Anxiety Inventory in Danish women aged 45 years and older with abnormal cervical screening results. BMC Med. Res. Methodol. **20**(1), 89 (2020). https://doi.org/10.1186/s12874-020-00982-4
28. Alfvén, G., Nilsson, S.: Validity and reliability of a new short verbal rating scale for stress for use in clinical practice. Acta Paediatr. **103**(4), e173–e175 (2014). https://doi.org/10.1111/apa.12558

29. Williamson, A., Hoggart, B.: Pain: a review of three commonly used pain rating scales. J. Clin. Nurs. **14**(7), 798–804 (2005). https://doi.org/10.1111/j.1365-2702.2005.01121.x
30. Schulz, K.F., Altman, D.G., Moher, D., and the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med. **8**(1), 18 (2010). https://doi.org/10.1186/1741-7015-8-18
31. Gordon, B.K., et al.: Child and parental surveys about pre-hospitalization information provision: child/parental surveys pre-hospitalization information. Child Care Health Dev. **37**(5), 727–733 (2011). https://doi.org/10.1111/j.1365-2214.2010.01190.x
32. Lööf, G., Lönnqvist, P.: Role of information and preparation for improvement of pediatric perioperative care. Pediatr. Anesth. **32**(5), 600–608 (2022). https://doi.org/10.1111/pan.14419
33. Eldridge, C., Kennedy, R.: Nonpharmacologic techniques for distress reduction during emergency medical care: a review. Clin. Pediatr. Emerg. Med. **11**(4), 244–250 (2010). https://doi.org/10.1016/j.cpem.2010.09.001
34. Kain, Z.N., Wang, S.-M.: Family-centered preparation for surgery improves perioperative outcomes in children **106**(1), 10 (2007)
35. Wray, J., Lee, K., Dearmun, N., Franck, L.: Parental anxiety and stress during children's hospitalisation: the StayClose study. J. Child Health Care Prof. Work. Child. Hosp. Community **15**, 163–74 (2011). https://doi.org/10.1177/1367493511408632
36. Cagiran, E., Sergin, D., Deniz, M.N., Tanattı, B., Emiroglu, N., Alper, I.: Effects of sociodemographic factors and maternal anxiety on preoperative anxiety in children. J. Int. Med. Res. **42**(2), 572–580 (2014). https://doi.org/10.1177/0300060513503758
37. Liang, X., et al.: Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet. Med. **28**(4), 455–463 (2011). https://doi.org/10.1111/j.1464-5491.2010.03180.x
38. Kampouroglou, G., et al.: Parental anxiety in pediatric surgery consultations: the role of health literacy and need for information. J. Pediatr. Surg. **55**(4), 590–596 (2020). https://doi.org/10.1016/j.jpedsurg.2019.07.016
39. Free, C., et al.: The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLOS Med. **10**(1), e1001362 (2013). https://doi.org/10.1371/journal.pmed.1001362
40. Kreps, G. L.: Ehealth Communication. Oxford Research Encyclopedia of Communication (2021). https://doi.org/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-194
41. Sousa, P., et al.: Controlled trial of an mHealth intervention to promote healthy behaviours in adolescence (TeenPower): effectiveness analysis. J. Adv. Nurs. **76**(4), 1057–1068 (2020). https://doi.org/10.1111/jan.14301
42. Bailey, C.R., et al.: Guidelines for day-case surgery 2019. Anaesthesia **74**(6), 778–792 (2019). https://doi.org/10.1111/anae.14639

The Effects of Robotic Training on Walking and Functional Independence of People with Spinal Cord Injury: A Systematic Review, Meta-analysis and Meta-regression

Anna Köyhäjoki[1](✉), Hilkka Korpi[2,3,4], Riku Yli-Ikkelä[3], Harto Hakonen[5], Mirjami Kantola[3], Aki Rintala[6], Sari Honkanen[3], Outi Ilves[3,7], Tuulikki Sjögren[3], Juha Karvanen[8], and Eeva Aartolahti[9]

[1] Assistive Equipment Center, Central Ostrobothnia Wellbeing Service County "Soite", Kokkola, Finland
anna.koyhajoki@gmail.com
[2] Well-Being and Culture Unit, Oulu University of Applied Sciences, Oulu, Finland
[3] Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland
[4] Social and Health Care Unit, Vaasa University of Applied Sciences, Vaasa, Finland
[5] Jamk University of Applied Sciences, LIKES, Jyväskylä, Finland
[6] Physical Activity and Functional Capacity Research Group, Faculty of Health Care and Social Services, LAB University of Applied Sciences, Lahti, Finland
[7] Department of Sport and Rehabilitation, South-Eastern Finland University of Applied Sciences, Savonlinna, Finland
[8] Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland
[9] Institute of Rehabilitation, Jamk University of Applied Sciences, Jyväskylä, Finland

Abstract. Evidence on the effects of robotic technology is required to develop rehabilitation services. This study aimed to evaluate the effects of robot-assisted walking training on walking and functional independence in everyday life in persons with spinal cord injury (SCI) and explore the covariates associated with these effects.

We searched the MEDLINE (Ovid), CINAHL, PsycINFO, and ERIC databases until March 25, 2022. Two reviewers independently assessed the studies for inclusion. We included RCTs on people with SCI receiving robotic training. The Cochrane RoB2, meta-analysis, meta-regression, and Grading of Recommendations Assessment, Development, and Evaluation were performed.

We included 23 RCTs focusing on SCI with outcomes of walking or functional independence, of which 14 were included in the meta-analysis and meta-regression analyses. Small improvements were observed in functional independence in favor of robot-assisted walking training compared to other physical exercises (Hedges' g 0.31, 95% CI 0.02 to 0.59; $I^2 = 19.7\%$, 9 studies, 419 participants, low certainty evidence). There were no significant differences in walking ability, speed, endurance, or independence between the groups.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-59091-7_23.

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 349–365, 2024.
https://doi.org/10.1007/978-3-031-59091-7_23

Robot-assisted walking training may slightly improve functional independence, but its effects on walking ability in SCI patients is uncertain compared to other exercise. Evidence suggests little to no difference in walking independence, and the effects on walking speed and endurance are unclear. No clear evidence exists whether positive effects are linked to personal, clinical, or intervention characteristics. Robot-assisted gait training may be a viable option for improving functional independence in individuals with SCI.

Keywords: Spinal cord injuries · Robotics · Rehabilitation · Exercise · Walking · Functional status · Systematic review · Meta-analysis

1 Introduction

Every year worldwide, 250 000 to 500 000 people sustain a spinal cord injury (SCI) [1]. To reduce health care costs, robotic technology is being used more in care and rehabilitation [2] Depending on the functional ability of the injured person, walking training without robotic technology can be time consuming and requires a lot of human resources, which has promoted the development of technological innovations such as robot-assisted walking devices [3].

One of the most visible consequences of SCI is restrictions in walking function which is a major focus of rehabilitation and affects quality of life[4–6]. Walking ability consists of different aspects: walking speed, walking independence, and walking endurance [7, 8]. The combination of speed and independence is suggested as the most valid measure of improvement in gait and ambulation in individuals with SCI [7, 8]. Walking endurance is also a recommended measure to provide a comprehensive evaluation of the walking performance[7].

Recent reviews and/or meta-analyses have examined different aspects of walking, but the results have been inconclusive. No effect of robot-assisted walking interventions was found for walking speed[9–13], endurance[9, 11], or independence[10, 12] compared with other types of exercise or no intervention, while most recent reviews found significant improvements in walking endurance[10, 13], lower extremity independence[9] and mobility[13].

In addition to walking, the ability to function in everyday activities is an important goal for persons with SCI and changes in this ability are an important indication of the efficacy of rehabilitation efforts [8, 14]. There are very few published meta-analyses covering robot-assisted walking training and functional independence in persons with SCI. The most recent review found improvements in favor of robot-assisted walking training but limited the comparison to overground walking training [15]. Other, previous reviews have not found the superiority of either robot-assisted walking training or other forms of training in improving functional independence [16, 17].

A transparent rating of the certainty of the evidence has been reported only in two previous reviews [11, 16], and none have examined the association of different study factors with the effect of robot-assisted exercise. However, both are important for clinicians interpreting the results of systematic reviews and especially, when moving from evidence to recommendations. Therefore, the effects of robot-assisted walking training

on different aspects of walking function and functional independence should be investigated in more detail. In addition, critical analyses of the certainty of the evidence are required.

The purpose of this systematic review and meta-analysis was to summarize randomized controlled trials (RCT) investigating the effects of robot-assisted walking training on walking and functional independence in persons with SCI because the most recent studies on the topic have been inconclusive, and therefore, high-quality updates on the current evidence are needed [9, 10, 13]. The following questions were addressed: 1) What are the effects of robot-assisted walking training on different aspects of walking ability and functional independence in adults with SCI compared to other exercises and what is the certainty of evidence? 2) Are study factors, such as personal, clinical, or intervention characteristics associated with the effects of robot-assisted walking training on walking and functional independence?

2 Methods

This systematic review and meta-analysis of RCTs was prospectively registered (PROSPERO 2022 CRD42022319235) [18] The reporting corresponds to the PRISMA and Cochrane guidelines [19, 20]. A literature search was conducted in a larger project that studied the effectiveness and meaning of robotics, virtual reality, and augmented reality in medical rehabilitation [21]. The National Library of Medicine (MEDLINE), Cumulative Index to Nursing and Allied Health Literature (CINAHL), Psychological Information Database (PsycINFO), and Education Resources Information Center (ERIC) databases were searched from inception to November 12, 2019. We conducted an updated search for studies published between August 2019 and March 25, 2022. We used MeSH or keyword terms to identify studies describing robotics and exercise combined with the Cochrane filter for RCTs. A full electronic search strategy is provided (Supplementary material). Additionally, we searched the reference lists of previously published systematic reviews.

2.1 Eligibility Criteria

We performed screening for this review in two phases. The first phase served at larger project with a wider scope[21] and included studies using the PICOS (patient, intervention, comparison, outcome, study design) framework as follows: P) adults or children requiring medical rehabilitation; I) any type of robotic device designed for rehabilitation purposes; C) conventional rehabilitation, wait-list-control, or other training modalities different from the experimental group; O) body functions and structures, activities, or participation according to International Classification of Functioning, Disability and Health (ICF), or quality of life; and S) RCT or cross-over RCT. The second phase was carried out after the updated search with more specified PICOS criteria to identify eligible studies of interest in this particular review: P) adults with both SCI and walking impairments; I) robot-assisted lower extremity or walking training intervention; C) a different type of exercise (active control) or no exercise (inactive control) or placebo as comparator; O) validated and standardized measures of walking or functional independence, and S) RCTs and cross-over RCTs. No language or publication date restrictions were imposed.

2.2 Study Selection

The titles, abstracts and full texts of the included studies were independently assessed by two researchers (AK, SH, RY, MK, OI, and EA) according to the eligibility criteria using Covidence software.[22] Disagreements were resolved by discussion or consultation with a third review team member (EA). All eligible RCTs were included in the systematic review. Meta-analyses excluded passive and other type of robot control interventions to control clinical heterogeneity.

2.3 Data Extraction and Quality Assessment

A customized template was designed in Covidence[22] to extract information on participants, interventions, outcomes, and adverse events of the included studies, and to perform quality assessment according to the Cochrane Risk of Bias 2 tool [23]. Two review team members independently extracted data and assessed the quality of the studies (AK, MK, SH, and RY). Disagreements were resolved by discussion or consultation with a third review team member (EA). Researchers of the RCTs were contacted when necessary to acquire missing data or to clarify ambiguities. If adequate data were not received despite three requests, the study or some of the outcomes of the study were excluded from the quantitative analyses. RCTs eligible for this review were included in the meta-analysis, regardless of the risk of bias judgement.

All outcomes measuring walking ability or functional independence in individuals with SCI were extracted from the included studies. A combination of the 10-m walk test (10MWT) measuring walking speed and the Walking Index for Spinal Cord Injury (WISCI), measuring walking independence or the change in the need for a walking aid, is suggested to provide the most valid measure of improvement in gait and ambulation [7, 8]. To provide the most comprehensive battery, a measure of endurance, such as the 6-min walk test (6MWT), is recommended.[7] For the walking ability meta-analysis, all walking outcomes in the included studies were prioritized in accordance[7] the following order: walking speed, walking independence, and walking endurance (Supplementary material).

Both the Spinal Cord Independence Measure (SCIM III) and the Functional Independence Measure (FIM) have been used to measure the broader functioning and independence in everyday life of individuals with SCI. The SCIM was chosen as the primary measure because it was specifically developed for persons with SCI [14, 24].

2.4 Data Analysis

To assess the treatment effect after the intervention, the meta-analysis was conducted using R software with the Metafor package for R.[25] Postintervention mean and standard deviation (SD) values were used in the analyses. Data reported as median or interquartile range (IQR) were converted to mean and SD assuming a normal distribution. A correlated effects model with robust variance estimation (RVE) using the Robumeta package on R and small-sample corrections was used, as it considers the possible dependent effect of the studies used multiple times in the same meta-analysis [26]. This was the case when a study had multiple control groups [27–29] or the study population was

divided according to the level of injury [30]. It is also considered to be a more reliable analysis model for studies with small number of participants [26]. The intervention effect size (Hedges' g), 95% confidence interval (CI), and statistical heterogeneity (I2) were estimated using a forest plot. The scale of Hedges' g was evaluated as small (0.20–0.49), medium (0.50–0.79), or large (0.80 or more) effect [31]. Statistical heterogeneity was assessed as follows:0–40% might not be important heterogeneity, 30–60% may represent moderate heterogeneity, 50–90% may represent substantial heterogeneity, and 75–100% represents considerable het-erogeneity [32]. If a crossover-RCT did not have a washout period, only the first intervention period was included in the meta-analysis.

Meta-regression analysis was performed using the Metafor package for R. We computed the Univariate Mixed effects model with intercept and restricted maxi-mum-likelihood estimation to determine whether covariates related to intervention content (duration of intervention, number of training sessions per week, time of one training session, weekly total volume of training), characteristics of rehabilitees (age, time since injury in months, the baseline WISCI score), and quality of the study (domains of risk of bias) could have an impact on the results. Sensitivity analysis was conducted excluding the studies with a high risk of bias in the domains that were found to be significant in the meta-regression. The certainty of evidence was graded at the outcome level according to the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) guidelines [33–35].

3 Results

3.1 Study Selection

An initial 1 405 abstracts were identified from the electronic databases after duplicates were removed (Fig. 1). After removal of studies considered ineligible according to the PICOS criteria, 23 RCTs were included in this review and 14 in the meta-analyses with all of them studying walking ability and 9 also functional independence. The remaining studies compared two types of robot-assisted walking training [36–40], had the same patient population as another included study [41, 42], had insufficient reporting of results [43], or the comparison group included no exercise [44]. Detailed characteristics of the included studies, justification for full-text exclusions and the information requested from RCTs are provided (Supplementary material).

3.2 Study Characteristics

Participants. The walking ability meta-analysis included 498 individuals with SCI. The average time since injury ranged from 3 months to 11 years (mean 48.7 (SD 65.2) months). The functional independence meta-analysis included 419 individuals with SCI. The average time since injury ranged from 3 months to 4 years (mean 11.5 (SD 15.1) months). In both meta-analyses, the participants' average age ranged from 34 to 59 with mean 45.1 (SD 7.6) years in the walking ability meta-analysis and mean 44.0 (SD 8.6) years in the functional independence meta-analysis.

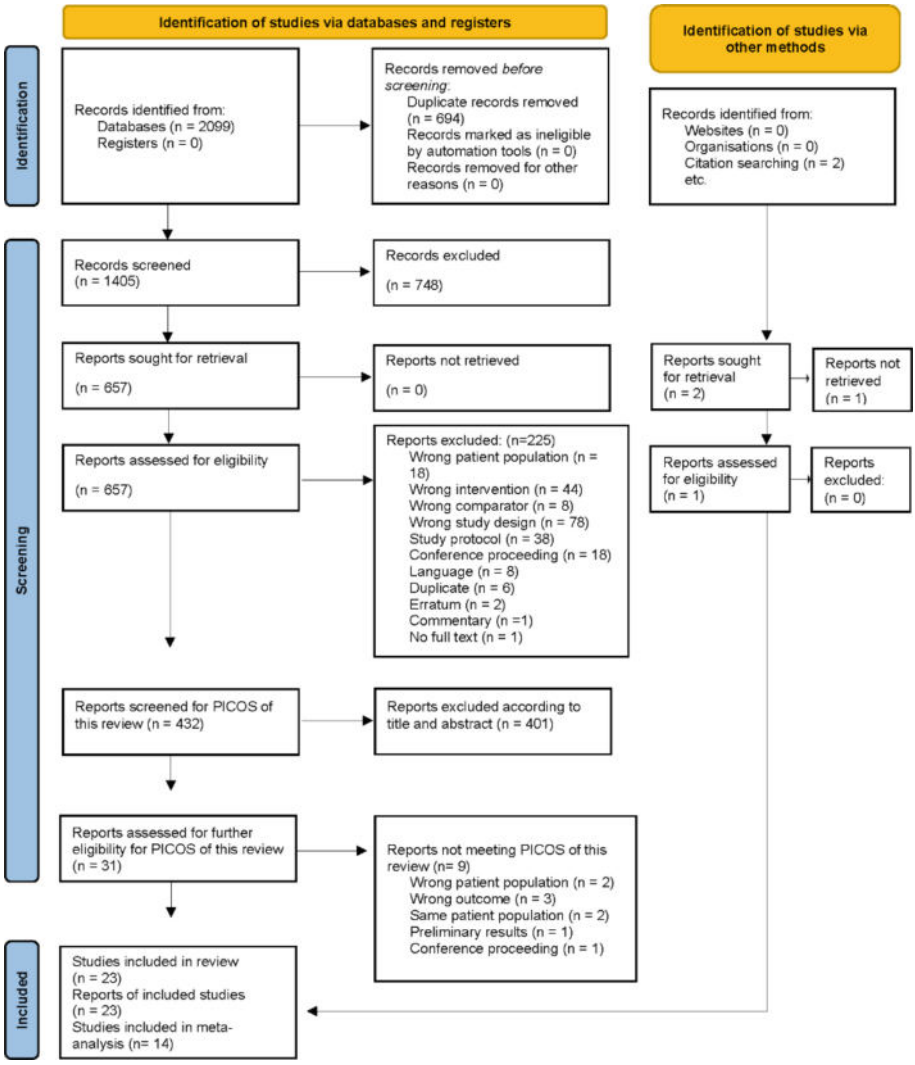


Fig. 1. Prisma flow diagram

Most commonly, the injuries of the participants in both meta-analyses were at the cervical or thoracic level, but there were also participants with lumbar-level injuries. Consequently, the meta-analyses included both paraplegic and tetraplegic participants. Most studies included participants who were grade C or D on the ASIA Impairment Scale (AIS) [45], with the majority being grade D. One study divided the participants into complete and incomplete injuries, without naming the AIS grade [46].

Robotic Interventions. In the walking ability meta-analysis, 11 studies used the exoskeleton device Lokomat (Hocoma; Zurich, Switzerland) [28–30, 46–53], 2 used the exoskeleton device Ekso (Ekso Bionics; CA, USA) [27, 54] and 1 used a 3DCaLT-robot (Shirley Ryan AbilityLab; Chicago, USA) [55]. Intervention durations varied from 3 to 24 weeks (mean 8 (SD 5)), and the duration of one session ranged from 30 to 90 min, 2 to 5 times a week. The average total training time per week was 199 min (SD 97). In the functional independence meta-analysis, all 9 studies used Lokomat. Intervention durations varied from 4 to 8 weeks (mean 7 (SD 2)), and the duration of one session ranged from 30 to 60 min, 2 to 5 times a week. The average total training time per week was 193 min (SD 107).

Body-weight support from robots in the studies in both meta-analyses was mostly utilized according to the person's needs and ranged from 0 to 78%. Less than half of the included studies reported the use of a guidance force (i.e., the assistance provided by the robotic legs to the lower extremities of the person training). The interventions took place in a hospital or university rehabilitation department. Adherence to interventions was rarely reported.

Comparisons. The comparison groups in the meta-analyses received conventional physical rehabilitation[30, 46–52, 54], with passive lower limbs mobilization [52], lower extremity strength training[53], body weight-supported treadmill training [27–29, 55] and/or overground walking training[27–30, 49, 55]. In the walking ability meta-analysis one study compared robot-assisted walking training to treadmill based or overground walking training with nerve stimulation in the control groups [28]. The amount of training in the comparison groups corresponded to that in the intervention groups in most studies.

Outcomes. Ten studies included in the meta-analysis measured walking speed, either self-selected [27, 28, 30, 49, 52–55] or not specified[29, 51], using the 10MWT or other measures, such as GAITRite-analysis. Three studies used the timed up and go (TUG) test [27, 29, 54]. Walking endurance with the 6MWT was measured in seven studies [27, 29, 30, 49, 51, 54, 55], with one study using the 2-min walk test (2MWT) [28]. Walking independence and the change in the need for a walking aid were measured using the WISCI in ten studies [27, 29, 30, 46–50, 52, 53].

The functional independence meta-analysis covered nine studies, of which four utilized the SCIM measure [48, 50, 52, 53] and five the FIM [29, 30, 46, 47, 49]. Only five studies evaluated all the subscales of SCIM [50, 52, 53] or FIM [46, 47].

3.3 Quality Assessment

The overall risk of bias was assessed as unclear [36–39, 46–48, 52, 53, 55] or high [27–30, 40–44, 49–51, 54] in each study (Supplementary material). No studies with a low overall risk of bias were found. High risk originated mainly from deviations from intended interventions but also from missing outcome data. An unclear risk of bias was found in the randomization process, deviations from the intended interventions, and selection of the reported results. Visual inspection of funnel plots suggests that some degree of publication bias is possible, smaller studies seem to favor the comparator (Supplementary material).

3.4 Synthesis of Results

Statistically significant improvements were observed in functional independence in favor of the robot-assisted walking training group compared to the control group (Hedges' g 0.31, 95% CI 0.02 to 0.59; I2 = 19.7%, 9 studies, 419 participants), whereas there were no statistically significant differences between groups in walking ability (Hedges' g 0.02, 95% CI –0.27 to 0.31; I2 = 35.5%, 14 studies, 498 participants), walking speed (Hedges' g –0.09, 95% CI –0.51 to 0.33; I2 = 32.8%, 10 studies, 290 participants), walking endurance (Hedges' g -0.03, 95% CI –0.65 to 0.58; I2 = 63.1%, 8 studies, 259 participants) or walking independence (Hedges' g 0.25, 95% CI –0.14 to 0.64; I2 = 51.3%, 9 studies, 419 participants) (Figs. 2, 3, 4, 5 and 6). Certainty of evidence proved to be low for functional independence and walking independence and very low for walking ability, speed, and endurance (Supplementary material).

In the meta-regression analyses, no relationships were found between the effects of robot-assisted walking training and intervention content or characteristics of rehabilitees. A high risk of bias in selection of the reported results was associated with the effect in functional independence. When excluding the high risk of bias study [29] from the meta-analyses, robot-assisted walking training remained statistically significant in improving functional independence compared to the control group (Hedges' g 0.35, 95% CI 0.05 to 0.64; I2 = 15.5%, 8 studies, 389 participants) (Supplementary material).

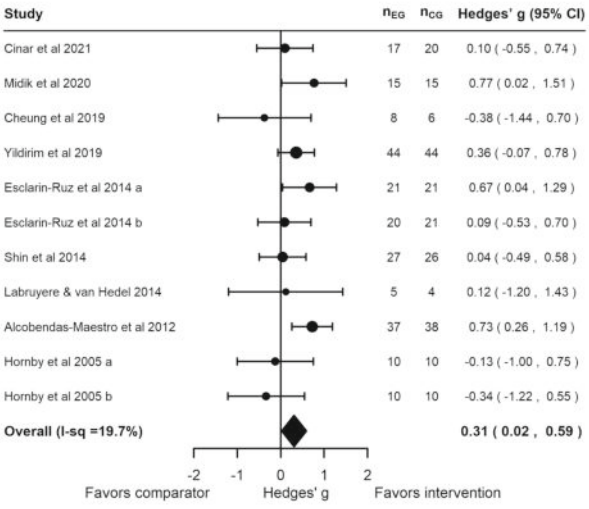


Fig. 2. Results of the meta-analysis comparing robot-assisted walking training and other physical exercise on functional independence of people with SCI.

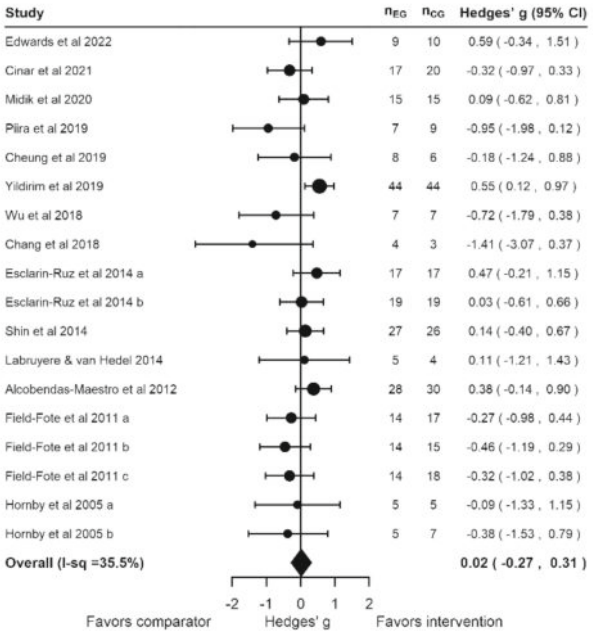


Fig. 3. Results of the meta-analysis comparing robot-assisted walking training and other physical exercise on walking ability of people with SCI.

3.5 Adverse Events

Adverse events were examined in 11 of 23 studies included. Reported adverse events were mostly mild and infrequent and four studies reported no adverse events (Supplementary material).

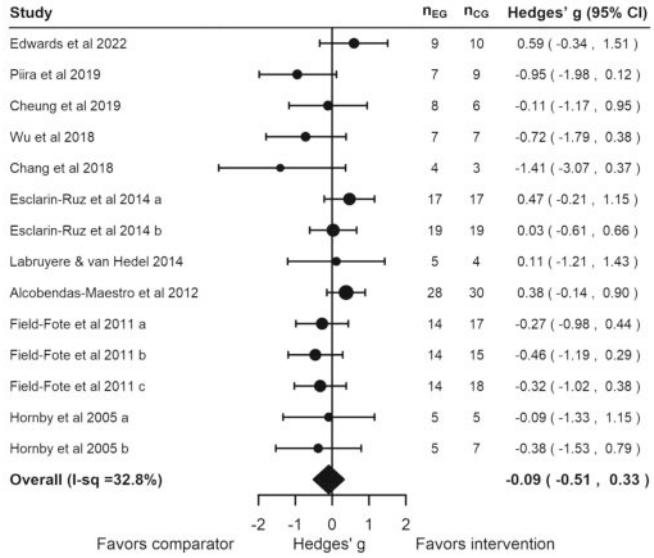


Fig. 4. Results of the meta-analysis comparing robot-assisted walking training and other physical exercise on walking speed of people with SCI.

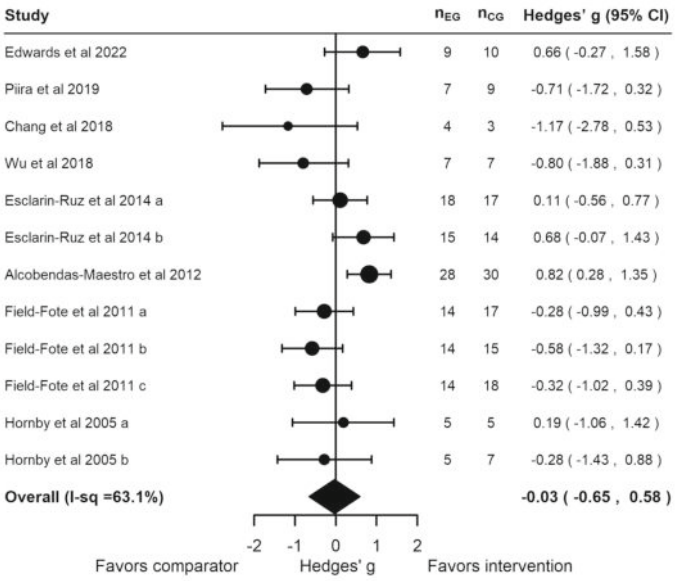


Fig. 5. Results of the meta-analysis comparing robot-assisted walking training and other physical exercise on walking endurance of people with SCI.

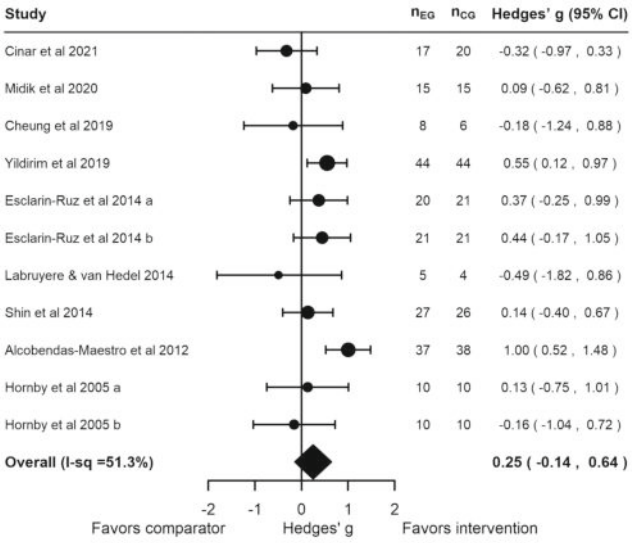


Fig. 6. Results of the meta-analysis comparing robot-assisted walking training and other physical exercise on walking independence of people with SCI.

4 Discussion

This systematic review and meta-analysis summarized evidence from the effects of robot-assisted walking training on walking and functional independence in everyday life in adults with SCI compared to those who had other physical exercises. There was a significant effect of functional independence favoring robot-assisted walking train-ing over other exercises with a small effect size (Hedges' g 0.31). No differences were found in the walking outcomes. Sensitivity analyses based on meta-regression analysis did not affect the results. The certainty of evidence was graded as low for functional independence and walking independence, and very low for walking ability, speed, and endurance. No severe adverse events were found although the reporting of RCTs regarding harms of robot-assisted training was incomplete.

The most recent systematic review of four trials by Harvey et al. [15] found significant improvements with robot-assisted walking training compared to overground walking training using SCIM and FIM. Our meta-analysis included more trials, prob-ably because of the wider scope of possible control interventions and suggests that robot-assisted walking training might be superior to other types of training in contrast to the findings of other reviews [16, 17]. Catz et al. [56] and Itzkovich et al. [57] found SCIM to be more sensitive in detecting functional ability changes than FIM and developed for patients with SCI. Therefore, including FIM in the meta-analysis may underestimate the effects of robot-assisted walking training.

Other reviews have covered various aspects of walking ability. Our review's results are mostly consistent with recent systematic reviews and meta-analyses in which robot-assisted walking interventions did not improve walking speed [9–13], endurance[9, 11] or independence [10, 12] when compared to other exercises or no intervention. Only Fang et al. [10] and Alashram et al. [13] reported significant improvements in walking endurance, and Duan et al. [9] reported significant improvements in lower-extremity independence using WISCI II. However, our meta-analyses included more studies, providing more reliable results.

To the best of our knowledge, our meta-analysis and Yang et al.'s network meta-analysis [58] are the only studies to combine multiple performance measures for walking ability outcome. Yang et al. [58] prioritized 6MWT and the Lower Extremity Motor Score (LEMS), showing significant walking improvements after robot-assisted training. This differs from our review that prioritized the 10MWT and WISCI. Previous studies suggest combining the 10MWT and WISCI to measure improvements in walking and ambulation in persons with SCI [7, 8, 59]. A measure of endurance, such as the 6MWT is also recommended [7], but varying test conditions can cause significant differences [59], hence the preference for the 10MWT and WISCI in our meta-analysis. In addition, Shin et al. [60] found that LEMS, a lower-extremity strength measure, does not significantly correlate with ambulatory function in persons with tetraplegic SCI. This finding demonstrates that different outcome measure priorities can lead to different results. More psychometric research is needed to guarantee SCI-related outcomes' sensitivity. In the future, a meta-analysis may be performed for single outcome measures if high-quality RCTs with similar outcomes are reported.

Publication bias is unlikely, as smaller studies seem to favor the comparator. However, no firm conclusions can be drawn due to the few studies and lack of larger sample sizes. The asymmetry in the funnel plots may have been caused by the high heterogeneity in the studies [61]. Our meta-analyses showed substantial statistical heterogeneity for walking endurance and moderate heterogeneity for walking independence. The meta-regression did not find clinical heterogeneity in the intervention or participant characteristics, such as time since injury, to be associated with the effect of robotic intervention. The RCTs included both paraplegic and tetraplegic individuals with SCI but did not report effects for these separately. So the level of injury could not be used as a covariate. According to Unai et al. [62] regardless of the AIS grade, paraplegic persons gain better results in the SCIM measure than tetraplegic persons; so further studies should differentiate between the two groups.

4.1 Strengths and Limitations

This study provides new information on the effects of robot-assisted walking exercise on functional independence and various aspects of walking ability in individuals with SCI. It is the first to assess the association between personal, clinical, and intervention characteristics and intervention effects. Meta-analyses excluded passive control interventions to control clinical heterogeneity. Meta-regression clarified the results, and GRADE guidelines graded the evidence certainty at the outcome level [33, 34]. To our knowledge, no recent review has provided graded clinical recommendations on this topic.

This review has limitations to consider when interpreting the results and generalizing evidence. The meta-analyses mainly included individuals with AIS grade C or D, and all but one study used an exoskeleton robot (Lokomat or Ekso), so the findings may not be generalizable. More data is needed on the effects of different robots. RCTs' methodological quality limits the reliability of the results. However, sensitivity analyses excluding studies based on the risk of bias did not alter the results. Future studies should pay particular attention to the methodological quality to ensure unbiased results.

4.2 Conclusion

Low level evidence suggests that robot-assisted walking training results in a slight improvement in functional independence, but little to no difference in walking independence in persons with SCI when compared to other exercises. The evidence is very uncertain regarding the effects of robot-assisted walking training on walking ability, walking speed, and walking endurance in persons with SCI when compared to other exercises. Heterogeneity between studies was substantial, and there is no clear evidence if positive effects were associated with age, time since injury, baseline walking independence, intervention programming, or quality of the study. Robot-assisted walking training appears to be a safe rehabilitation method for individuals with SCI. However, additional high-quality RCTs with larger sample sizes, similar outcome measures and differentiation of results between paraplegic and tetraplegic individuals are needed to further evaluate the effects and safety of robot-assisted walking training on functional independence and walking ability in individuals with SCI. When seeking to improve the functional independence of persons with SCI, robot-assisted gait training may be considered as a potential training option.

4.3 Clinical Message

Low-level evidence suggests that in people with SCI robot-assisted walking training may slightly improve functional independence but has little to no effect in walking independence compared to other exercises. Evidence is very uncertain on walking ability, speed, and endurance. Intervention or rehabilitee characteristics, and risk of bias didn't affect results.

Acknowledgements. Thank you to all who were part of the ROVA project at the University of Jyväskylä, Finland and contributed to this study.

Conflict of Interest Statement. The Authors Declare that There is no Conflict of Interest.

Funding. The Authors Disclosed Receipt of the Following Financial Support for the Research, Authorship, and/or Publication of This Article: This Work Was Supported by the Social Insurance Institution of Finland (Kela) [Grant Number 50/26/2019].

Data Availability Statement. The Data that Support the Findings of This Study Are Available from the Corresponding Author, (EA), upon Reasonable Request.

References

1. WHO: Spinal Cord Injury. https://www.who.int/news-room/fact-sheets/detail/spinal-cord-injury
2. EuRobotics: Strategic Research Agenda for Robotics in Europe (2014)
3. Rymer, W.Z., Burt, S., Jayaraman, A.: 13 Advanced Rehabilitation Strategies for Individuals with Traumatic Spinal Cord Injury. In: Vialle, L.R., Fehlings, M., Weidner, N. (eds.) AOSpine Masters Series, pp. 163–178. Thieme (2017)
4. Ditunno, P.L., Patrick, M., Stineman, M., Ditunno, J.F.: Who wants to walk? Preferences for recovery after SCI: a longitudinal and cross-sectional study. Spinal Cord. **46**, 500–506 (2008). https://doi.org/10.1038/sj.sc.3102172
5. Ditunno, P.L., Patrick, M., Stineman, M., Morganti, B., Townson, A.F., Ditunno, J.F.: Cross-cultural differences in preference for recovery of mobility among spinal cord injury rehabilitation professionals. Spinal Cord. **44**, 567–575 (2006). https://doi.org/10.1038/sj.sc.3101876
6. Anderson, K.D.: Targeting recovery: priorities of the spinal cord-injured population. J. Neurotrauma **21**, 1371–1383 (2004). https://doi.org/10.1089/neu.2004.21.1371
7. Anderson, K.D., et al.: Outcome measures for gait and ambulation in the spinal cord injury population. J. Spinal Cord Med. **31**(5), 487–499 (2008). https://doi.org/10.1080/10790268.2008.11753644
8. Lam, T., Noonan, V.K., Eng, J.J.: A systematic review of functional ambulation outcome measures in spinal cord injury. Spinal Cord **46**(4), 246–254 (2008). https://doi.org/10.1038/sj.sc.3102134
9. Duan, R., et al.: Clinical benefit of rehabilitation training in spinal cord injury: a systematic review and meta-analysis. Spine **46**(6), E398–E410 (2021). https://doi.org/10.1097/BRS.0000000000003789
10. Fang, C.Y., Tsai, J.L., Li, G.S., Lien, A.S.Y., Chang, Y.J.: Effects of robot-assisted gait training in individuals with spinal cord injury: a meta-analysis. Biomed. Res. Int. **2020**, 1–13 (2020). https://doi.org/10.1155/2020/2102785
11. Hornby, T.G., et al.: Clinical Practice Guideline to Improve Locomotor Function Following Chronic Stroke, Incomplete Spinal Cord Injury, and Brain Injury (2020)
12. Aguirre-Güemez, A.V., Pérez-Sanpablo, A.I., Quinzaños-Fresnedo, J., Pérez-Zavala, R., Barrera-Ortiz, A.: Walking speed is not the best outcome to evaluate the effect of robotic assisted gait training in people with motor incomplete spinal cord injury: a systematic review with meta-analysis. J. Spinal Cord Med. **42**, 142–154 (2019). https://doi.org/10.1080/10790268.2017.1390644
13. Alashram, A.R., Annino, G., Padua, E.: Robot-assisted gait training in individuals with spinal cord injury: a systematic review for the clinical effectiveness of Lokomat. J. Clin. Neurosci. **91**, 260–269 (2021). https://doi.org/10.1016/j.jocn.2021.07.019
14. Catz, A., Itzkovich, M., Agranov, E., Ring, H., Tamir, A.: SCIM – spinal cord independence measure: a new disability scale for patients with spinal cord lesions. Spinal Cord. **35**, 850–856 (1997). https://doi.org/10.1038/sj.sc.3100504
15. Harvey, L.A., Glinsky, J.V., Chu, J.: Do any physiotherapy interventions increase spinal cord independence measure or functional independence measure scores in people with spinal cord injuries? A systematic review. Spinal Cord **59**(7), 705–715 (2021). https://doi.org/10.1038/s41393-021-00638-0

16. Fisahn, C., et al.: The effectiveness and safety of exoskeletons as assistive and rehabilitation devices in the treatment of neurologic gait disorders in patients with spinal cord injury: a systematic review. Glob. Spine J. **6**(8), 822–841 (2016). https://doi.org/10.1055/s-0036-1593805
17. Wessels, M., Lucas, C., Eriks, I., De Groot, S.: Body weight-supported gait training for restoration of walking in people with an incomplete spinal cord injury: a systematic review. J. Rehabil. Med. **42**(6), 513–519 (2010). https://doi.org/10.2340/16501977-0525
18. Köyhäjoki, A., et al.: The effects of robot-assisted walking training on walking and disability of people with spinal cord injury: a systematic literature review, meta-analysis and meta-regression. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=319235
19. Page, M.J., et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ **132**, 1–9 (2021)
20. Higgins, J., et al.: Cochrane Handbook for Systematic Reviews of Interventions version 6.3, updated February 2022. https://training.cochrane.org/handbook/current
21. Ilves, O., Korpi, H., Honkanen, S., Aartolahti, E.: Effectiveness and meanings of robots, virtual and augmented reality in rehabilitation. Systematic literature reviews. Stud. Soc. Secur. Health **159**, 1–12 (2022)
22. Covidence. https://app.covidence.org
23. Sterne, J.A.C., et al.: RoB 2: A revised tool for assessing risk of bias in randomised trials. The BMJ. **366**, 14898 (2019). https://doi.org/10.1136/bmj.l4898
24. Schuld, C., Weidner, N.: 2 Assessment of functional status and outcomes of individuals with traumatic spinal cord injury. In: Fehlings, M.J., Weidner, N., Vialle, L.R.G. (eds.) AOSpine Masters Series, pp. 11–24. Thieme (2017)
25. Viechtbauer, W.: Conducting Meta-Analyses in R with the metafor Package. J Stat Softw. **36**, 1–48 (2010). https://doi.org/10.18637/jss.v036.i03
26. Pustejovsky, J.E., Tipton, E.: Meta-analysis with Robust Variance Estimation: Expanding the Range of Working Models. Prev. Sci. **23**, 425–438 (2022). https://doi.org/10.1007/s11121-021-01246-3
27. Edwards, D.J., et al.: Walking improvement in chronic incomplete spinal cord injury with exoskeleton robotic training (WISE): a randomized controlled trial. Spinal Cord. **60**, 522–532 (2022). https://doi.org/10.1038/s41393-022-00751-8
28. Field-Fote, E.C., Roach, K.E.: Influence of a locomotor training approach on walking speed and distance in people with chronic spinal cord injury: a randomized clinical trial. Phys. Ther. **91**, 1–48 (2011). https://doi.org/10.2522/ptj.20090359
29. Hornby, T.G., Campbell, D.D., Zemon, D.H., Kahn, J.H.: Clinical and quantitative evaluation of robotic-assisted treadmill walking to retrain ambulation after spinal cord injury. Top Spinal Cord Inj. Rehabil. **11**, 1–17 (2005)
30. Esclarín-Ruz, A., et al.: A comparison of robotic walking therapy and conventional walking therapy in individuals with upper versus lower motor neuron lesions: a randomized controlled trial. Arch. Phys. Med. Rehabil. **95**, 1023–1031 (2014). https://doi.org/10.1016/j.apmr.2013.12.017
31. Cohen, J.: A power primer. Psychol. Bull. **112**, 155–159 (1992). https://doi.org/10.1037/0033-2909.112.1.155
32. Deeks, J.J., Higgins, J.P.T., Altman, D.G.: Analysing data and undertaking meta-analyses: identifying and measuring heterogeneity. In: Higgins, J.P.T., et al. (eds.) Cochrane Handbook for Systematic Reviews of Interventions, version 6.2. (2021)
33. Guyatt, G.H., Oxman, A.D., Schünemann, H.J., Tugwell, P., Knottnerus, A.: GRADE guidelines: a new series of articles. J. Clin. Epidemiol. **64**, 380–382 (2011). https://doi.org/10.1016/j.jclinepi.2010.09.011
34. Schünemann, H., Brozek, J., Gyuatt, G., Oxman, A.: GRADE Handbook. Introduction to GRADE Handbook. https://gradepro.org/resources/#handbook

35. Balshem, H., et al.: GRADE guidelines: 3. Rating the quality of evidence. J. Clin. Epidemiol. **64**, 401–406 (2011). https://doi.org/10.1016/j.jclinepi.2010.07.015
36. Benito-Penalva, J., et al.: Gait training in human spinal cord injury using electromechanical systems: effect of device type and patient characteristics. Arch. Phys. Med. Rehabil. **93**, 404–412 (2012). https://doi.org/10.1016/j.apmr.2011.08.028
37. Lam, T., Pauhl, K., Ferguson, A., Malik, R.N., Krassioukov, A., Eng, J.J.: Training with robot-applied resistance in people with motor-incomplete spinal cord injury: pilot study. J. Rehabil. Res. Dev. **52**, 113–130 (2015). https://doi.org/10.1682/JRRD.2014.03.0090
38. Wirz, M., et al.: The EMSCI network: effectiveness of automated locomotor training in patients with acute incomplete spinal cord injury: a randomized, controlled. Multicenter Trial. J Neurotrauma. **34**, 1891–1896 (2017). https://doi.org/10.1089/neu.2016.4643
39. Wu, M., et al.: Repeat exposure to leg swing perturbations during treadmill training induces long-term retention of increased step length in human SCI. Am. J. Phys. Med. Rehabil. **95**, 911–920 (2016). https://doi.org/10.1097/PHM.0000000000000517
40. Wu, M., Landry, J.M., Schmit, B.D., Hornby, T.G., Yen, S.-C.: Robotic resistance treadmill training improves locomotor function in human spinal cord injury: a pilot study. Arch. Phys. Med. Rehabil. **93**, 782–789 (2012). https://doi.org/10.1016/j.apmr.2011.12.018
41. Nooijen, C., Ter Hoeve, N., Field-Fote, E.: Gait quality is improved by locomotor training in individuals with SCI regardless of training approach. J. Neuroeng. Rehabil. **6**, 36 (2009). https://doi.org/10.1186/1743-0003-6-36
42. Kressler, J., Nash, M.S., Burns, P.A., Field-Fote, E.C.: Metabolic responses to 4 different body weight-supported locomotor training approaches in persons with incomplete spinal cord injury. Arch Phys Med Rehabil. 94, (2013). https://doi.org/10.1016/j.apmr.2013.02.018
43. Tang, Q., Huang, Q., Hu, C.: Research on design theory and compliant control for underactuated lower-extremity rehabilitation robotic systems. J. Phys. Ther. Sci. **26**, 1597–1599 (2014). https://doi.org/10.1589/jpts.26.1597
44. Duffell, L.D., Brown, G.L., Mirbagheri, M.M.: Interventions to reduce spasticity and improve function in people with chronic incomplete spinal cord injury. Neurorehabil. Neural Repair **29**, 566–576 (2015). https://doi.org/10.1177/1545968314558601
45. American Spinal Injury Association ASIA: International Standards for Neurological Classification of SCI (ISNCSCI). https://asia-spinalinjury.org/international-standards-neurological-classification-sci-isncsci-worksheet/
46. Yildirim, M.A., Öneş, K., Gökşenoğlu, G.: Early term effects of robotic assisted gait training on ambulation and functional capacity in patients with spinal cord injury. Turk. J. Med. Sci. **49**, 838–843 (2019). https://doi.org/10.3906/sag-1809-7
47. Çinar, Ç., Yildirim, M.A., Öneş, K., Gökşenoğlu, G.: Effect of robotic-assisted gait training on functional status, walking and quality of life in complete spinal cord injury. Int. J. Rehabil. Res. **44**, 262–268 (2021). https://doi.org/10.1097/MRR.0000000000000486
48. Shin, J.C., Kim, J.Y., Park, H.K., Kim, N.Y.: Effect of robotic-assisted gait training in patients with incomplete spinal cord injury. Ann. Rehabil. Med. **38**, 719 (2014). https://doi.org/10.5535/arm.2014.38.6.719
49. Alcobendas-Maestro, M., et al.: Lokomat robotic-assisted versus overground training within 3 to 6 months of incomplete spinal cord lesion. Neurorehabil. Neural Repair **26**, 1058–1063 (2012). https://doi.org/10.1177/1545968312448232
50. Mıdık, M., Paker, N., Buğdaycı, D., Mıdık, A.C.: Effects of robot-assisted gait training on lower extremity strength, functional independence, and walking function in men with incomplete traumatic spinal cord injury. Turk. J. Phys. Med. Rehabil. **66**, 54–59 (2020). https://doi.org/10.5606/tftrd.2020.3316
51. Piira, A., et al.: Robot-assisted locomotor training did not improve walking function in patients with chronic incomplete spinal cord injury: a randomized clinical trial. J. Rehabil. Med. **51**, 385–389 (2019). https://doi.org/10.2340/16501977-2547

52. Cheung, E.Y.Y., Yu, K.K.K., Kwan, R.L.C., Ng, C.K.M., Chau, R.M.W., Cheing, G.L.Y.: Effect of EMG-biofeedback robotic-assisted body weight supported treadmill training on walking ability and cardiopulmonary function on people with subacute spinal cord injuries – a randomized controlled trial. BMC Neurol. **19**, 140 (2019). https://doi.org/10.1186/s12883-019-1361-z
53. Labruyère, R., van Hedel, H.J.A.: Strength training versus robot-assisted gait training after incomplete spinal cord injury: a randomized pilot study in patients depending on walking assistance. J. Neuroeng. Rehabil. **11**, 4 (2014). https://doi.org/10.1186/1743-0003-11-4
54. Chang, S.-H., Afzal, T., Berliner, J., Francisco, G.E.: Exoskeleton-assisted gait training to improve gait in individuals with spinal cord injury: a pilot randomized study. Pilot Feasibility Stud. **4**, 62 (2018). https://doi.org/10.1186/s40814-018-0247-y
55. Wu, M., Kim, J., Wei, F.: Facilitating weight shifting during treadmill training improves walking function in humans with spinal cord injury. Am. J. Phys. Med. Rehabil. **97**, 585–592 (2018). https://doi.org/10.1097/PHM.0000000000000927
56. Catz, A., Itzkovich, M., Agranov, E., Ring, H., Tamir, A.: The spinal cord independence measure (SCIM): sensitivity to functional changes in subgroups of spinal cord lesion patients. Spinal Cord. **39**, 97–100 (2001). https://doi.org/10.1038/sj.sc.3101118
57. Itzkovich, M., et al.: The spinal cord independence measure (SCIM) version III: reliability and validity in a multi-center international study. Disabil. Rehabil. 29, 1926–1933 (2007). https://doi.org/10.1080/09638280601046302
58. Yang, F.-A., et al.: Body weight-supported gait training for patients with spinal cord injury: a network meta-analysis of randomised controlled trials. Sci. Rep. **12**, 19262 (2022). https://doi.org/10.1038/s41598-022-23873-8
59. Scivoletto, G., Tamburella, F., Laurenza, L., Foti, C., Ditunno, J.F., Molinari, M.: Validity and reliability of the 10-m walk test and the 6-min walk test in spinal cord injury patients. Spinal Cord. **49**, 736–740 (2011). https://doi.org/10.1038/sc.2010.180
60. Shin, J.C., Yoo, J.H., Jung, T.H., Goo, H.R.: Comparison of lower extremity motor score parameters for patients with motor incomplete spinal cord injury using gait parameters. Spinal Cord. **49**, 529–533 (2011). https://doi.org/10.1038/sc.2010.158
61. Sterne, J.A.C., et al.: Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ (Online). **343**, d4002 (2011). https://doi.org/10.1136/bmj.d4002
62. Unai, K., Uemura, O., Takemura, R., Kawakami, M., Liu, M.: Association between SCIM III total scores and individual item scores to predict independence with ADLs in persons with spinal cord injury. Arch. Rehabil. Res. Clin. Transl. **1**, 100029 (2019). https://doi.org/10.1016/j.arrct.2019.100029

Wireless Technologies and Medical Devices

Securing Hybrid Wireless Body Area Networks (HyWBAN): Advancements in Semantic Communications and Jamming Techniques

Simone Soderi[1(✉)], Mariella Särestöniemi[2,3], Syifaul Fuada[3], Matti Hämäläinen[3], Marcos Katz[3], and Jari Iinatti[3]

[1] IMT School for Advanced Studies Lucca, Lucca, Italy
simone.soderi@imtlucca.it
[2] Health Sciences and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland
mariella.sarestoniemi@oulu.fi
[3] Centre for Wireless Communications, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland
{syifaul.fuada,matti.hamalainen,marcos.katz,jari.iinatti}@oulu.fi

Abstract. This paper explores novel strategies to strengthen the security of Hybrid Wireless Body Area Networks (HyWBANs), which are essential in smart healthcare and Internet of Things (IoT) applications. Recognizing the vulnerability of HyWBAN to sophisticated cyberattacks, we propose an innovative combination of semantic communications and jamming receivers. This dual-layered security mechanism protects against unauthorized access and data breaches, particularly in scenarios involving in-body to on-body communication channels. We conduct comprehensive laboratory measurements to understand hybrid (radio and optical) communication propagation through biological tissues. We utilize these insights to refine a dataset for training a Deep Learning (DL) model. These models, in turn, generate semantic concepts linked to cryptographic keys for enhanced data confidentiality and integrity using a jamming receiver. The proposed model significantly reduces energy consumption compared to traditional cryptographic methods, like Elliptic Curve Diffie-Hellman (ECDH), especially when supplemented with jamming. Our approach addresses the primary security concerns and sets the baseline for future secure biomedical communication systems advancements.

Keywords: Heterogeneous · WBAN · energy · security · optical · RF · near-infrared communications

1 Introduction

The advent of wireless and mobile communications technologies has been essential in enhancing healthcare, marking a paradigm shift towards more proactive

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 369–387, 2024.
https://doi.org/10.1007/978-3-031-59091-7_24

and personalized medical interventions. The concept of smart healthcare is at the forefront of this transformation, offering many opportunities to address the growing needs of an ageing population and the increasing prevalence of chronic diseases [1]. Remote health monitoring, a cornerstone of modern healthcare, has emerged as a cost-effective and efficient approach to disease prevention and healthcare provision, especially with the integration of 5G and 6G technologies. These advancements are pivotal in supporting in-body communications with implanted medical devices, enabling real-time health provisioning, virtual consultations, better diagnostics, and telesurgeries, among other benefits [1]. Historically, information transmission through biological tissues has predominantly relied on radio and acoustic waves [2]. However, these conventional methods are fraught with challenges, including security, safety, privacy, and interference, necessitating the exploration of alternative communications media. The vulnerabilities of implantable or in-body devices to hacks and unauthorized access have underscored the urgent need for enhanced security measures [3,4].

Optical Wireless Communications (OWC) has emerged as a promising alternative, utilizing light, especially in the near-infrared range, to transmit information through biological tissues. This method offers many advantages, including high security, privacy, safety and low complexity, as well as low power consumption. It has been used to successfully establish connectivity to electronic devices embedded under the skin [5]. Going further, a hybrid solution which is merging both radio-based and optical-based technologies in Wireless Body Area Network(WBAN) context can open a new, more secure way to implement personalized healthcare services and transfer personal health data. This is also a way to reduce radio signal emission towards the human body.

Hybrid Wireless Body Area Networks (HyWBANs) stand at the forefront of innovation in healthcare and Internet of Things (IoT) applications, merging radio and OWC. These networks offer remarkable advantages such as data throughput enhancement, enhanced security, and improved reliability, making them ideal for critical healthcare applications and various services, from patient monitoring to advanced diagnostics. Reconfigurable HyWBANs takes adaptability to the next level with a dynamic architecture, ensuring consistent performance in diverse environments. Preliminary studies on HyWBANs underscore their potential, showcasing notable performance and energy efficiency improvements [6]. Energy harvesting is crucial to HyWBANs [7], focusing on developing energy-autonomous nodes that enhance sustainability and reduce maintenance. The networks' advanced sensing capabilities support single and dual-mode sensing, enabling comprehensive data collection for diverse applications. Moreover, HyWBANs' design promotes sustainable operations, which is essential in today's environmentally conscious landscape. Optimized data transmission functionality in HyWBANs caters to the high demands of medical and IoT applications. A significant feature is the ability to transfer energy to in-body devices, ensuring continuous operation. Additionally, HyWBANs' advanced sensing capabilities are essential in medical diagnostics, allowing for detailed tissue analysis and health monitoring, thus revolutionizing healthcare and IoT applications [1].

Over the years, academia has shown interest in Physical Layer Security (PLS) solutions that aim to protect communications by exploiting the properties of the communication media [8–11]. These techniques consist of processing the signal sent over a channel in such a way as to obtain certain security properties without resorting to specific primitives, typically cryptography, offered by layers above the physical level. In this paper, we show how to combine PLS techniques with Deep Learning (DL) algorithms to improve the security of HyWBANs.

Fig. 1. Coding strategy for hybrid networks.

Motivation. In this article, we have decided to use a model already used in the literature that involves defining operating modes based on the combinations of the two communications channels [6,12]. Figure 1 depicts how hybrid radio-optical wireless networks utilize Shannon's theory, which defines the maximum channel capacity for communications. It shows the dynamic selection of the device's operating modes based on factors like channel state information (radio/optical) and user context. In our view, HyWBANs can improve *security* by integrating in the radio transmissions the OWC, which is known for the localised and secure transmission of signals. Encoding signals across radio and optical channels maximises secrecy, in line with recent theoretical work on conventional networks. This approach exploits the inherent security features of optical communications, addressing the vulnerabilities of WBANs. This paper uses data measured in the laboratory to implement an *innovative hybrid network security scheme* using semantic communications and intentional interference.

Contribution. In particular, hybrid communications have been of great interest for sensor networks. In a digital healthcare scenario, protecting these communications and doing so effectively while consuming as little energy as possible makes significance. The contributions of this article can be summarised as follows. *(i)* We present a novel concept that exploits the combination of Communications (SC) with a jamming receiver to improve the confidentiality and integrity of these wireless communications. *(ii)* We performed measurements in the laboratory to study the propagation of hybrid (radio and optical) communications in biological tissue. These measurements allowed us to define part of the dataset used for the DL model. Finally, *(iii)* we evaluated and performed a security analysis of the HyWBANs.

The remainder of the paper is organized as follows. Section 2 briefly recalls the concepts useful for understanding the paper. Section 3 discusses the major security threats in this scenario. Then, Sect. 4 presents the proposed scheme to enhance the security of hybrid networks. Section 5 presents the results achieved in terms of performance and energy cost. Finally, Sect. 6 concludes the paper by discussing our findings.

2 Background

This section introduces the radio and optical technologies used for wireless sensor communication. The aim is to provide some notions before discussing how hybrid networks combine these two technologies.

2.1 Radio-Based WBAN Technologies

WBAN is a way to link various wearable sensor nodes wirelessly into one individual and personalized network used to monitor a person's psycho-physiological vital signs. Depending on the need, the vital sensors can be distributed all around the human body. Low-power consumption, small size, and lightweight are the requirements set for the nodes to enable user acceptance. In principle, the amount of connected sensors within one WBAN can be high, but a realistic number is less than five for the sake of usability. The basic idea behind a WBAN is that dedicated sensors are collecting vital information and transmitting it wirelessly to the central node (called a hub), which then pre-process the data or conveys it further. Figure 2 shows the variability of the vital sensor nodes, which can be used in the WBAN context (the list of sensors is not exhaustive) [13]. In addition to sensors which are attached to the skin, so called on-body sensors, WBAN can utilize smart implants, such as pacemakers, or other in-body sensors/devices, such as Wireless Capsule Endoscope (WCE). In WBAN, all the nodes are connected to the on-body hub to enable real-time information transmission towards backbone infrastructure. Typically, WBAN is using a one-hop star network topology.

Currently, *de facto* wireless standard in WBAN is Bluetooth Low Energy (BLE) but there are also other dedicated WBAN standards available, such as ETSI SmartBAN [14], IEEE 802.15.6 [15], or IEEE 802.15.4 [16]. The latter one is better known via its higher layer protocols ZigBee and 6LoWPAN.

From a radio technology point-of-view, WBAN connectivity can be based on narrowband (NB) signals, which are used, e.g., in BLE and SmartBAN, or ultra wideband (UWB), which is adopted by [15]. The most common frequency band at the moment for NB signal is Industrial, Scientific and Medical (ISM) band at about 2.4 GHz. On the other hand, e.g., [15] defines several NB frequency bands for WBAN use also occupying sub-GHz frequencies. As operating in a highly populated frequency range, ISM band is typically subject to high interference originated from other radio equipment nearby. The selected frequency band, as well as signal bandwidth, also have an impact on the observed signal

propagation properties through/along tissues, positioning accuracy, throughput, etc., depending on the application and use-case. If high resolution, real-time data is needed, then UWB can be the best option from radio-based technologies. Lower performance requirements and deeper in-body penetration, however, are favouring NB technology. Reference [15] also defines Human Body Communications(HBC) technology operating around 21 MHz, but this is omitted in this review due to its deviation from the conventional Radio Frequency (RF)-based communications as being a coupling-based solution.

The original network topology in WBAN is based on a star topology, where all the data flows are going through the central node, the hub. In this case, the hub is also a bridge from the body domain to the backbone network. The recent development, e.g., in ETSI SmartBAN has introduced and defined a hub-to-hub communications to transfer information between adjacent WBAN networks [17]. In addition, a two-hop relay functionality is included in the SmartBAN technical specification [18]. From the security viewpoint, all the hops between WBAN nodes, although being short, should be reliable but also secure as the communications chain is as reliable as its weakest link. This highlights the importance of light security protocols to be used also in the WBAN context.

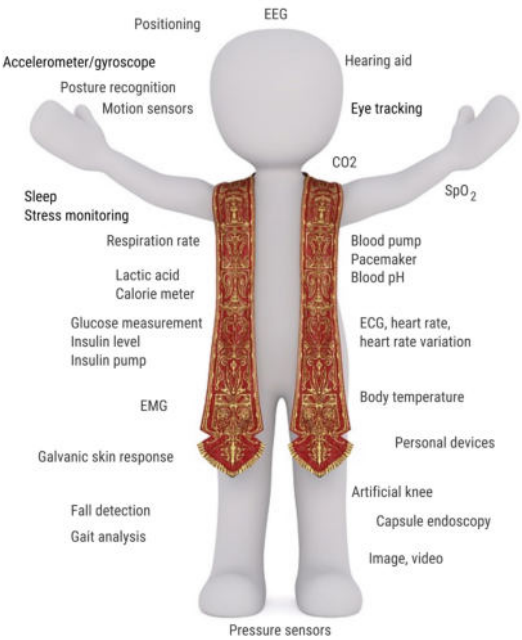


Fig. 2. Variability of the possible sensors that can be used in the WBAN context.

2.2 Optical Communications in Wireless Sensor Networks

The utilization of optical communications, including Visible Light Communication (VLC) and Infrared (IR) technologies, in wireless sensor networks is gaining interest for various IoT and body network applications. Optical communication offers security, bandwidth, and energy efficiency advantages, which are crucial for IoT deployments.

VLC utilizes Light Emitting Diodes (LED) to transmit data using the visible spectrum. This approach is inherently secure due to the limited light propagation and offers high data rates, making it suitable for indoor IoT applications [19]. VLC's potential in hybrid optical-wireless networks for next-generation communications, especially in 5G and beyond, is highlighted in [20]. IR communication leverages the non-visible spectrum for data transmission, offering benefits in terms of device miniaturization and reduced interference with existing RF systems. Its suitability for low-data-rate IoT applications, especially in hybrid networks combining IR and VLC, is explored in [7]. IR in hybrid radio-optical wireless networks offers innovative solutions for versatile IoT applications, as discussed in [21]. Integrating optical and wireless technologies in a hybrid framework opens new avenues for enhancing IoT network performance. The synergy of RF and optical communication technologies in hybrid networks is investigated in [22], which outlines the implementation and advantages of such an approach.

3 Security Analysis of HyWBANs

Developing next-generation networks to support better biomedical applications presents an opportunity. However, cyber-security risks arise mainly from this technology's highly interconnected and ubiquitous nature [1]. Therefore, the cybersecurity analysis of these hybrid communications begins with choosing a system model that best represents the problem.

WBANs are components of cyberspace that assist people in their daily activities and collect data from persons. WBANs and, more broadly, wearable wireless networks (WWNs) have three communication layers, according to the *tier model* [23]. As shown in Fig. 3, wearable sensors capture data in Tier 1 and transmit it to Tier 2 for aggregation and data processing. Finally, data is sent to Tier 3 and made available for remote access. The HyWBANs follow the same system model, where radio link, optical link, or both can be used for each type of communication in Tier 1 and Tier 2. As illustrated in Fig. 3, we have different types of communications in HyWBAN: on-body to in-body devices (labelled as *On-In*) and on-body to on-body (labelled as *On-On*) devices that operate in Tier 1. Instead, at Tier 2, all communications are off-body, including on-body to off-body devices (labelled *On-Off*). We assume that HyWBANs operate up to Tier 2, as depicted in Fig. 3.

One of the main security problems of this communication chain is that Eve, the adversary (attacker) shown in Fig. 3, can carry out several attacks. We can assume that she has complete control to intercept and modify all messages

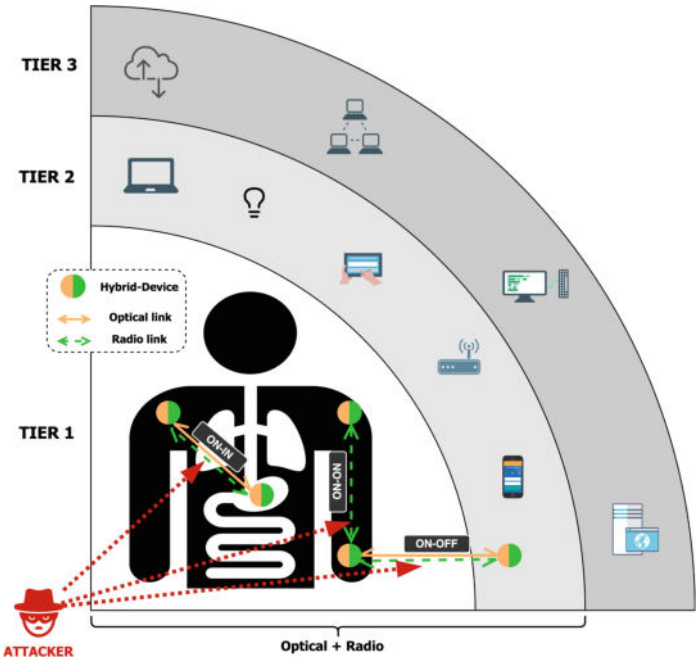


Fig. 3. Communications tiers system model, in which hybrid communications operate in the first two tiers.

exchanged between HyWBAN nodes [24]. In the rest of the paper, we analyse the possible attacks and their mitigation.

3.1 Security Threats Overview

The complexity of HyWBANs, which combines RF and OWC, far exceeds traditional communications systems due to their dual-channel nature. This sophistication poses significant challenges for attackers attempting to compromise the network, as they must navigate radio and optical channels.

In the HyWBANs domain, the security of communication channels is essential, especially when considering transmitting sensitive health data. This section delves into the nuanced vulnerabilities inherent to RF and OWC, laying the groundwork for understanding the superior security posture of optical communications in specific scenarios. RF communications, by their very nature, are susceptible to eavesdropping due to their omnidirectional signal propagation. This characteristic allows malicious entities to intercept signals without necessitating a direct line of sight, thereby posing a significant risk to the confidentiality of transmitted data. Conversely, optical communications demand a line-of-sight for effective transmission, inherently restricting the potential for unauthorized interception. Despite this advantage, optical channels are not impervious to secu-

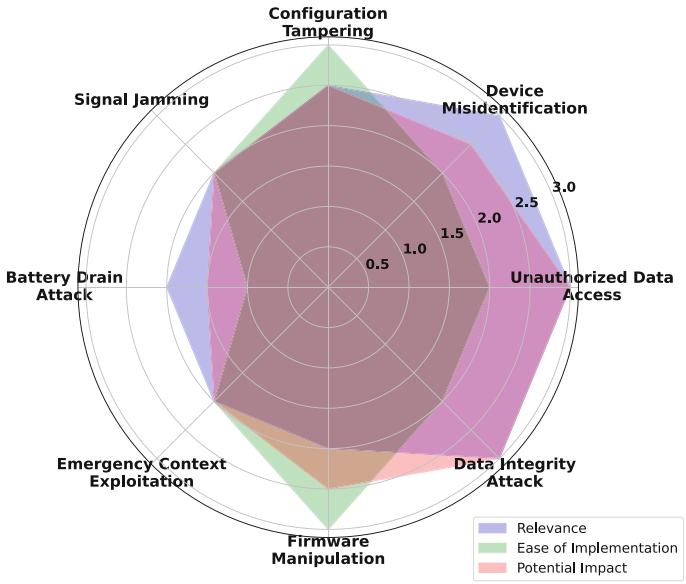


Fig. 4. Security threats to HyWBAN.

rity threats. A breach in the line-of-sight or sophisticated techniques to capture reflected optical signals can compromise data integrity and confidentiality.

In the context of HyWBANs, an attacker would primarily focus on tactics that enable eavesdropping on biomedical device communications. These tactics could include exploiting network security protocol vulnerabilities, conducting Man-in-the-Middle (MitM) attacks to intercept data, or using sophisticated techniques to bypass encryption. Reconnaissance plays a crucial role, as the attacker must gather detailed information about the network's configuration and security mechanisms to successfully deploy malware or other attack vectors.

Figure 4 presents a multifaceted evaluation of security threats in HyWBANs. Each threat is analyzed based on three critical dimensions: *relevance* to the network's security, *ease of implementation* by potential attackers, and the *potential impact* on network integrity and functionality. This brief assessment enables a slight understanding of each threat's effectiveness and helps prioritize security measures.

To fortify the security framework of HyWBANs against these vulnerabilities, we introduce a semantic communication method that significantly enhances the security of transmitted data.

4 Enhancing the Security of HyWBAN Through Semantic Communications

Understanding the implications for security in the dynamic landscape of HyWBANs is essential. Adversaries exploiting vulnerabilities in these networks could potentially gain unauthorized access to the human body, leading to critical threats like hijacking pacemakers, reconfiguring smart pill dispensers, or even creating novel types of diseases. The dual nature of these networks, encompassing radio and optical wireless channels, adds a layer of complexity to potential attacks. This study proposes a novel security mechanism combining the principles of semantic communications with the strategic deployment of a jamming receiver (see Fig. 5), enhancing the confidentiality and integrity of HyWBANs.

Semantic communications [25], an emerging paradigm in network security [26], involves generating semantic concepts related to biomedical applications or patient health status. This approach utilizes a DL model trained on a dataset comprising measured, augmented, and synthetic biological signals [1,27]. During an enrollment phase, assumed free from adversarial presence, each semantic concept is associated with a secret, such as a cryptographic key, stored in the nodes' memory.

The transmission of semantic concepts over the wireless channel, although susceptible to interception by malicious adversaries, is protected through a jamming receiver. As shown in Fig. 5, this receiver introduces intentional interference on either the light or radio channel, or both, effectively interfering with the transmitted data from the in-body device. Consequently, an adversary attempting to decode the data encounters altered signal characteristics, such as decreased Signal-to-Noise Ratio (SNR) for radio and input power for Near-Infrared (NIR) signals; this leads to an *erroneous classification of semantic concepts*.

However, the legitimate receiver, Bob, *knows the jamming pattern* and can reverse the artificially induced bias to correctly decode the transmitted semantic

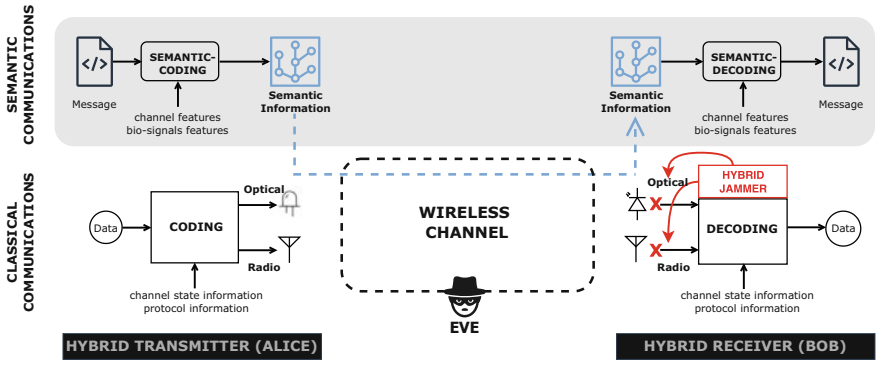


Fig. 5. AI-based data stream encoding in HyWBANs using semantic communications and a jamming receiver to harden the security of wireless communications.

concept. In contrast, an adversary, referred to as Eve, who lacks this knowledge, faces significant challenges in decoding the data accurately [9,11]. This approach, leveraging semantic communication and controlled jamming, offers a dual-layered defence mechanism, enhancing the resilience of HyWBANs against sophisticated cyber threats. The rest of the section describes how the data were prepared and how we propose to use a DL algorithm on devices with constrained resources.

4.1 Radio and Optical Channels Data Measurement

In this study, we investigate the efficacy of hybrid communications that utilize optical and radio channels, explicitly investigating their capacity to penetrate biological tissues. Two distinct experimental setups were designed to assess the performance characteristics necessary for effective and secure communications through such mediums. For optical communications, NIR frequencies were employed, selected for their proven proficiency in penetrating biological tissues. On the other hand, UWB technology, recognized for its superior transmission capabilities, particularly noise-like signals, was chosen for radio communications. This dual-faceted approach allows for a comprehensive evaluation of the potential of hybrid communication systems in medical applications.

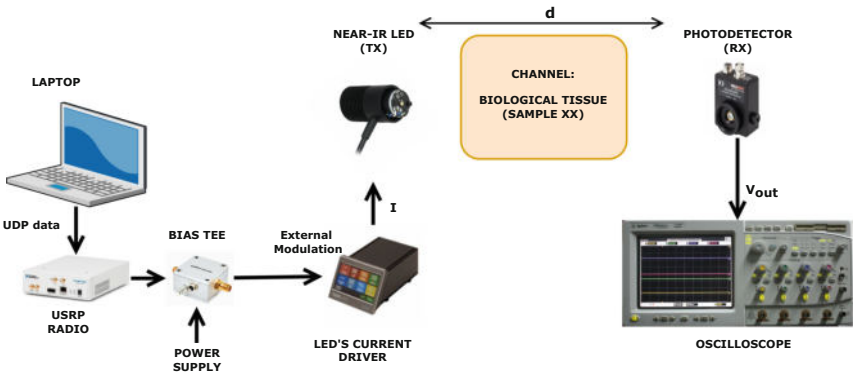


Fig. 6. Measurement set-up with NIR communications (as OWC) through the biological tissue.

The experimental setup, depicted in Fig. 6, is an optical communication part; it comprises various components that can be divided into two subsystems: transmitter and receiver front end. The transmitter unit includes a NIR LED (M810L3, THORLABS) with 810 nm wavelength, a bias-tee, and a LED driver. The receiver unit utilizes a Photo-Detector (PD) (PDA 36A-EC switchable gain detector, THORLABS). A sample of biological tissue was used, acting as the communications channel. The LED is driven by a current driver module (DC2200, THORLABS), which is controlled by an external modulation source.

The modulation of the NIR LED is essential for transmitting data through the biological tissue. The PD, positioned at a specific distance (d is the thickness of the meat sample used in the measurements) from the NIR LED, captures the transmitted light after it has passed through the tissue. The output from the PD (V_{out}) is then analyzed using an oscilloscope (with 50 Ω impedance) to assess the effectiveness of data transmission through the tissue sample. Using a laptop, we sent the same ASCII character with the User Datagram Protocol (UDP) protocol to a software-defined radio USRP that modulated the signal before sending it to the LED driver. Two NI USRPs (2920 model) were employed in this study. We also used a bias-tee (ZFBT-4R2GW-FT+, Mini-Circuits) to combine the modulation signal and the bias current to feed the driver. We measured the peak of the received burst signal for each character sent from the laptop. The V_{out} is then converted into an input power unit by using the equation provided in the datasheet of PDA36A-EC. This setup is crucial for evaluating the feasibility of NIR communications in scenarios where signals need to penetrate biological tissues, such as in implantable medical device applications.

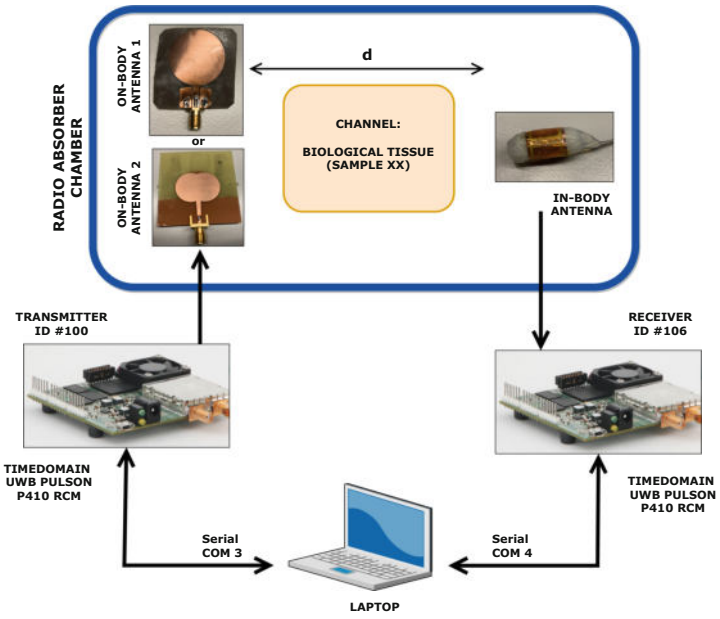


Fig. 7. Measurement set-up with UWB radio through the biological tissue.

As illustrated in Fig. 7, the radio measurement setup was meticulously designed to evaluate the performance of radio communications in HyWBANs. It consists of a UWB transmitter (P410 PulsON, Time Domain) inside the body (i.e., in-body device) that communicates with a UWB receiver (P410 PulsON, Time Domain) that has its antenna positioned on the porcine skin (i.e., on-body).

Using Time Domain's Channel Analysis Tool (CAT) software, we simulated the communications scenario inside the body by sending signals from the transmitter to the receiver. We enclosed the antennas inside a box of RF absorber material to avoid external interference. The received signals were saved on a laptop using CAT software and analyzed later using MATLAB. This system makes it possible to accurately measure the radio signal's capability to penetrate biological tissue and the effectiveness of UWB technology in an in-body communications scenario, essential for developing reliable hybrid WBANs.

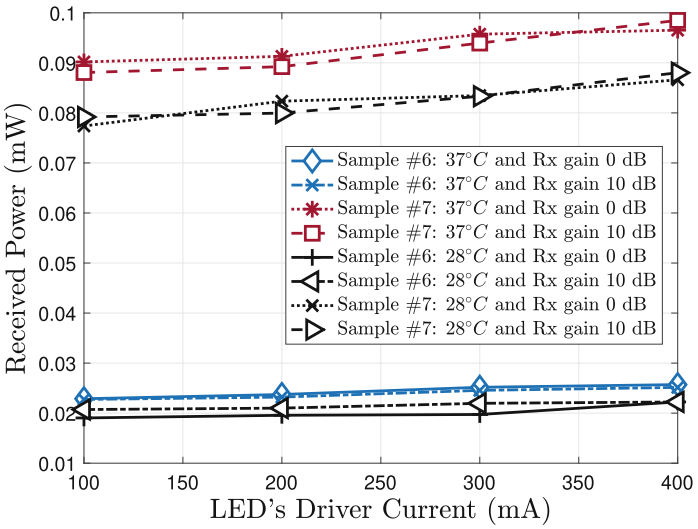
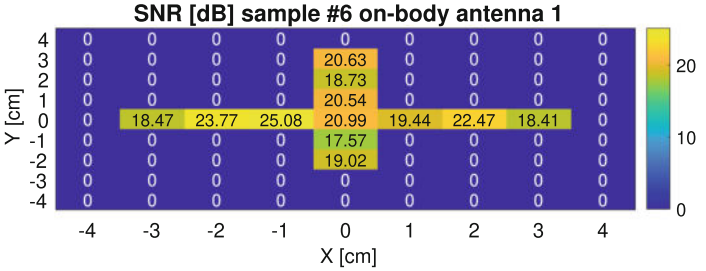


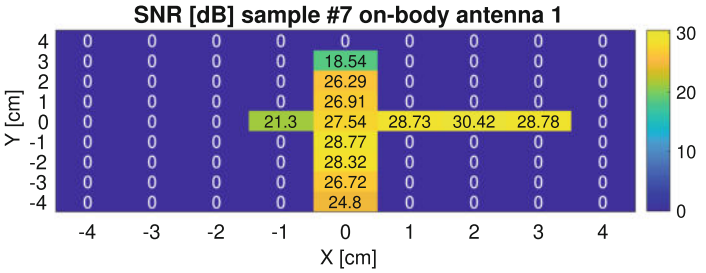
Fig. 8. NIR received power varying the temperature of two biological tissue samples (i.e., sample #6 and #7) and the gain of the PD .

Figure 8 shows the power (expressed in mW) of the NIR signal received after passing through the two biological tissue samples with a maximum thickness of 37 mm and 39 mm for samples #6 and #7, respectively. From Fig. 8, it is evident how the propagation capabilities improve when the temperature reaches $37°C$, which is considered almost typical for a human body. Selecting a higher gain (i.e., from 0 dB to 10 dB) in the receiver (PDA 36A-EC offers this option using a rotary switch) does not lead to a significant advantage regarding receiver sensitivity.

Figure 9 shows the UWB SNR measured by the on-body antenna placed on the skin. Meanwhile, the transmitter and receiver were aligned for light. For the UWB measurements, we left the in-body antenna at the fixed position while we moved the on-body antenna in 1 cm steps to investigate the communication limit.



(a) SNR with on-body antenna 1 and sample #6 at 37°C.



(b) SNR with on-body antenna 1 and sample #7 at 37°C.

Fig. 9. SNR measurements of UWB transmissions through biological tissue.

4.2 Dataset: Measured and Synthetic Data for Medical Applications Using HyWBAN

Developing and optimizing semantic communications and strengthening security within hybrid networks necessitate a comprehensive dataset for training and testing DL models. This dataset encompasses both measured features, such as SNR and the power received by the PD in NIR communications, as well as synthetic features. These synthetic features are conceptualized on the premise that the constituent devices of HyWBAN can acquire and process biological signals from individuals. This dual approach in dataset formulation facilitates a realistic assessment of the HyWBAN's operational capabilities and aids in simulating a wide range of scenarios for advanced medical applications.

To refine the dataset for DL models in the context of HyWBANs for medical applications, we employed a dual-strategy approach involving both the *augmentation of measured data* and the *generation of synthetic data*. The augmentation process for the measured attributes, specifically SNR for UWB and received power for NIR, employs a statistical methodology in which new values are generated based on a Gaussian distribution. This distribution is centred on the measured mean and standard deviation. This statistical rigour ensures the augmented data are closely aligned with realistic measurement variations. We generated values for a few parameters: acceleration, heart rate and body temperature to generate synthetic data to emulate the ability of HyWBAN devices to measure

biological signals. We assumed that a hybrid device could access these quantities (as a knowledge base for semantic communications) in a medical application or at least a part of it. These synthetic values are derived from a Gaussian distribution, adhering to predefined mean and standard deviations to ensure they fall within physiologically plausible ranges. The Table 1 summarises the specification of statistics for the data generation process. Our semantic communication model uses a binary classification approach, simplifying complex data into categories like 'HIGH_SNR' or 'LOW_SNR' using the thresholds defined in Table 1. This method efficiently filters out noise, focusing on key data aspects and significantly reducing computational load. This binary representation accelerates model training and enhances interpretability, facilitating fast and decisive communications analysis. We can then define the labels to be associated with each type of communication in a supervised manner (see Table 2). These labels are the *semantic concepts* that represent data the device measures in a compressed manner. This particular approach to dataset preparation supports the robustness of the developed models. It ensures the simulation of diverse scenarios, which is critical for applying HyWBAN in medical settings.

Table 1. Statistical summary of augmented and synthetic features.

Feature	Mean	Stand. Deviation	[Min, Max]	Threshold
SNR (dB)[a]	23.6	4.23	[17.57, 33.32]	19 dB
Input Power (mW)[b]	0.07	0.03	[0.02, 0.09]	0.05 mW
Acceleration (m/s^2)	0	0.1	[-0.5, 0.5]	0.1 m/s^2
Heart Rate (bpm)	60	25	[50, 120]	<60 bpm, >110 bpm
Body Temperature (°C)	36	2	[34, 42]	37°C

[a] UWB measured data.
[b] NIR measured data.

Table 2. Classification labels for semantic communications

Label	Condition
Full Communications	HIGH_SNR and HIGH_LPW
Wide Communications	HIGH_SNR and LOW_LPW
Communications in Motion	(HIGH_SNR or HIGH_LPW) and HIGH_ACC
Critical Communications	(HIGH_HR or HIGH_TMP) and LOW LPW
Unstable Communications	LOW_SNR or LOW_LPW
Reduced Communications	Other scenarios not covered by above conditions

5 Proposed Model Evaluation

Our study developed a deep learning model for semantic communication in HyWBAN: an autoencoder with a $64 - 32 - 64$ neuron structure and a classification model with dense layers and dropout regularization. The purpose of the autoencoder model within our semantic analysis is to reduce the dimensionality of the input data, including SNR and heart rate, thereby enabling more efficient processing and transmission. The autoencoder helps identify the most significant features crucial for semantic analysis by transforming the data into a lower-dimensional space. This process not only aids in preserving essential information but also contributes to the system's security by minimizing the amount of data exposed to potential threats.

We have conducted a series of experiments to determine the optimal architecture for our model. Our choice was guided by a grid search approach, where we evaluated various configurations and selected the one that minimized the reconstruction error on a validation set: the 64-32-64 structure balanced model complexity and the ability to capture the underlying patterns in the data. We also experimented with different activation functions and learning rates, ultimately choosing a Rectified Linear Unit (ReLU) activation for its efficiency and a learning rate of 0.001 for stable convergence. The training process of our autoencoder was carried out over 50 epochs, with early stopping implemented to prevent overfitting. We used a batch size of 256, which was determined to be optimal through experimentation, balancing the trade-off between training speed and memory constraints. The dataset (after an augmentation of the data by a factor of 50) comprised 2040 samples, split into 80% for training and 20% for validation. This information aims to enhance the transparency and reproducibility of our model evaluation.

To visualize the effectiveness of the autoencoder in capturing semantic relationships, we apply the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique (see Fig. 10). This method is noted for its ability to represent high-dimensional data in lower dimensions while preserving data structures, allowing us to visually inspect the clustering of data points based on their semantic similarities. We have expanded our discussion on the interpretability of clusters formed in the t-SNE visualization. The clusters represent distinct data patterns that the autoencoder has learned to encode. By examining the characteristics of samples within each cluster, we can infer the model's ability to discern different features in the data, which supports its effectiveness.

The model is optimized using *Adam* optimizer and trained to categorize the data into predefined semantic classes, as described in our data preprocessing phase. Performance evaluation using a confusion matrix and accuracy metrics confirmed the model's efficacy. Finally, the model was converted to *TensorFlow Lite* [28] (i.e., a Tiny Machine Learning framework that supports the conversion of ML models into a format that can be run on microcontrollers), aligning with low-power, edge-based IoT device requirements, ensuring privacy, energy efficiency, and real-time processing. This approach signifies a substantial advancement in semantic communication for smart healthcare applications.

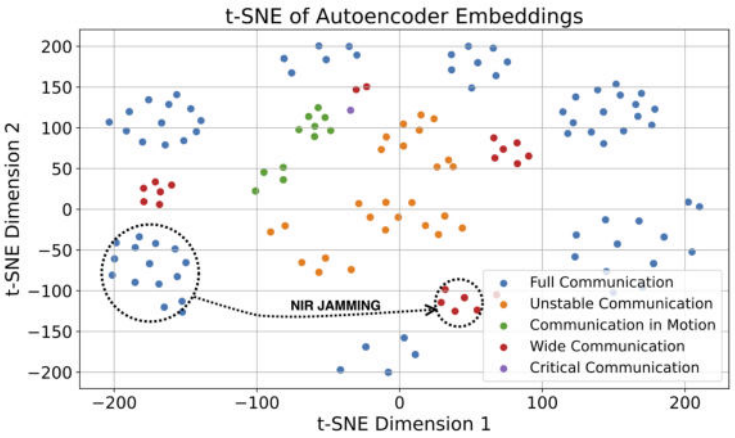


Fig. 10. Low-dimensional representation of data preserving semantic similarities. For example, the figure shows the effect of NIR jamming on the semantic concepts.

In our evaluation, we compared the energy efficiency of our semantic communication with a jamming solution to the Elliptic Curve Diffie-Hellman (ECDH) key exchange [29]. The comparison focused on various configurations, assessing the energy consumption for different key lengths in ECDH (160 and 256 bits) against our semantic communication model that can use 8 or 16 bits to represent the semantic concepts, and it is enhanced with jamming up to 8 and 16 bits (i.e., worst case for our proposal). We assumed 0.1 μJ as the energy per bit and 0.2 μJ as the energy cost to jam a bit. This analysis, crucial for understanding the practicality of deploying these methods in energy-constrained environments like HyWBANs, is visualized Fig. 11, illustrating the total energy consumption of each method. Such comparisons highlight the efficiency of semantic communi-

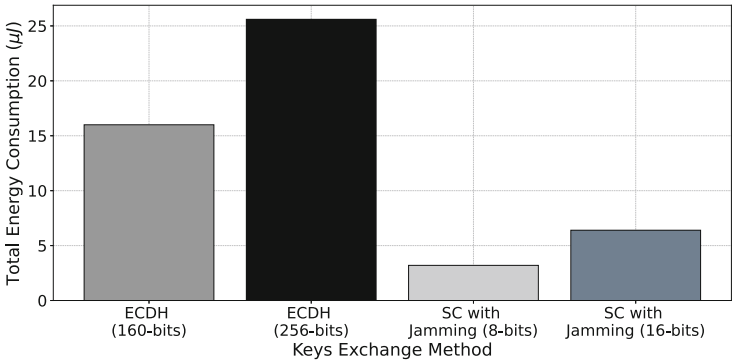


Fig. 11. Energy consumption comparison between SC with jamming and ECDH.

cations, especially when supplemented with jamming, in contrast to traditional cryptographic approaches like ECDH.

The classification performance of our TensorFlow model for semantic communication in HyWBANs demonstrated good precision and recall across most classes, with an overall accuracy of 94%. However, the corresponding TensorFlow Lite model, optimized for low-power devices, showed a variation in performance, particularly in precision and recall for specific communication classes like *Communication in Motion* and *Critical Communication*, resulting in an overall accuracy of 86%. This variation underscores the challenges in collecting more data with measurements and balancing model complexity with the constraints of edge computing devices.

6 Conclusions

The research presented in this paper marks a significant stride in enhancing the security of HyWBANs. By integrating semantic communications with jamming receivers, we demonstrate a robust method to protect sensitive health data and biomedical devices within the HyWBANs framework. Our experimental analysis provides valuable insights into the propagation characteristics of hybrid communications in biological tissues, forming the basis for an advanced DL model. This model's ability to generate and interpret semantic concepts, coupled with a strategic jamming mechanism, ensures the reliable transmission of encrypted data, thereby mitigating potential cybersecurity threats. Notably, our approach outperforms traditional cryptographic methods in energy efficiency, making it a viable solution for the energy-sensitive environment of HyWBANs.

The semantic strategy enhances security by transmitting only the necessary and relevant data, reducing the attack surface. The deep learning model contributes to this by learning to identify and filter out non-essential information, thus streamlining the communication process and making it more secure. The inherent security advantages of optical communications, such as the line-of-sight requirement for the interception, are exploited in our hybrid system to strengthen overall security further.

The findings and methodologies outlined in this study improve the security of current HyWBAN systems and pave the way for their broader adoption in smart healthcare services, aligning with the evolving landscape of 6G technology.

Acknowledgments. This research is partially funded by the Research Council of Finland Profi6 funded 6G Enabling Sustainable Society (University of Oulu) and Research Council of Finland 6G Flagship (grant 318927). and the European Union's Horizon 2020 programme under the Marie Sklodowska-Curie grant H2020-MSCA-RISE-2019 agreement No. 872752.

References

1. Batista, E., Lopez-Aguilar, P., Solanas, A.: Smart health in the 6G Era: bringing security to future smart health services. IEEE Commun. Mag. 1–7 (2023)
2. Ahmed, I., Halder, S., Bykov, A., Popov, A., Meglinski, I.V., Katz, M.: In-body communications exploiting light: A proof-of-concept study using ex vivo tissue samples. IEEE Access **8**, 190378–190389 (2020)
3. Zafar, S., et al.: A systematic review of bio-cyber interface technologies and security issues for internet of bio-nano things. IEEE Access **9**, 93529–93566 (2021)
4. Ransford, B., et al.: Cybersecurity and medical devices: a practical guide for cardiac electrophysiologists. Pacing Clin. Electrophysiol. **40**(8), 913–917 (2017)
5. Halder, S., Särestöniemi, M., Ahmed, I., Katz, M.: Providing connectivity to implanted electronics devices: experimental results on optical communications over biological tissues with comparisons against UWB. In: Alam, M.M., Hämäläinen, M., Mucchi, L., Niazi, I.K., Le Moullec, Y. (eds.) BODYNETS 2020. LNICST, vol. 330, pp. 3–17. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-64991-3_1
6. Saud, M.S., Ahmed, I., Kumpuniemi, T., Katz, M.: Reconfigurable optical-radio wireless networks: meeting the most stringent requirements of future communication systems. Trans. Emerg. Telecommun. Technol. **30**(2), e3562 (2019). https://doi.org/10.1002/ett.3562
7. Kamalakis, T., Ghassemlooy, Z., Zvanovec, S., Nero Alves, L.: Analysis and simulation of a hybrid visible-light/infrared optical wireless network for IoT applications. J. Opt. Commun. Netw. **14**(3), 69–78 (2022)
8. Bloch, M., Barros, J.: Physical-Layer Security: from Information Theory to Security Engineering. Cambridge University Press, Cambridge (2011)
9. Soderi, S., Mucchi, L., Hämäläinen, M., Piva, A., Iinatti, J.: Physical layer security based on spread-spectrum watermarking and jamming receiver. Trans. Emerg. Telecommun. Technol. **28**(7), e3142 (2017)
10. Soderi, S., Dainelli, G., Iinatti, J., Hämäläinen, M.: Signal fingerprinting in cognitive wireless networks. In: 2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), pp. 266–270 (2014)
11. Soderi, S., De Nicola, R.: 6G networks physical layer security using RGB visible light communications. IEEE Access **10**, 5482–5496 (2022)
12. Katz, M., Ahmed, I.: Opportunities and challenges for visible light communications in 6G. In: 2020 2nd 6G Wireless Summit (6G SUMMIT), pp. 1–5 (2020)
13. Hämäläinen, M., et al.: Recent Progress in ETSI TC SmartBAN Standardization. In: 2023 IEEE 17th International Symposium on Medical Information and Communication Technology (ISMICT), pp. 1–6 (2023)
14. Hämäläinen, M., et al.: ETSI SmartBAN architecture: the global vision for smart body area networks. IEEE Access **8**, 150611–150625 (2020)
15. IEEE Computer Society: IEEE standard for local and metropolitan area networks - part 15.6: Wireless body area networks. IEEE standard (2012)
16. IEEE Computer Society: IEEE standard for low-rate wireless networks. IEEE standard (2015)
17. ETSI Tc SmartBAN: Smart body area network (SmartBAN) Hub to Hub communication for SmartBAN medium access control (MAC). ETSI standard TS **103**, 806 (2023)

18. ETSI Tc SmartBAN: Smart body area network (SmartBAN) Relay functionality for SmartBAN medium access control (MAC). ETSI standard TS 103, 805 (2024)
19. Perera, A., Katz, M., Godaliyadda, R., Häkkinen, J., Strömmer, E.: Light-based internet of things: implementation of an optically connected energy-autonomous node. In: 2021 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–7 (2021)
20. Chowdhury, M.Z., Hasan, M.K., Shahjalal, M., Hossan, M.T., Min Jang, Y.: Optical wireless hybrid networks for 5g and beyond communications. In: 2018 International Conference on Information and Communication Technology Convergence (ICTC), pp. 709–712 (2018)
21. Teli, S.R., et al.: Hybrid optical wireless communication for versatile IoT applications: data rate improvement and analysis. IEEE Access **11**, 55107–55116 (2023)
22. Saud, M.S., Chowdhury, H., Katz, M.: Heterogeneous software-defined networks: implementation of a hybrid radio-optical wireless network. In: 2017 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6 (2017)
23. Otto, C., Milenković, A., Sanders, C., Jovanov, E.: System architecture of a wireless body area sensor network for ubiquitous health monitoring. J. Mob. Multimed. **1**(4), 307–326 (2005)
24. Dolev, D., Yao, A.: On the security of public key protocols. IEEE Trans. Inf. Theory **29**(2), 198–208 (1983)
25. Calvanese Strinati, E., Barbarossa, S.: 6G networks: beyond Shannon towards semantic and goal-oriented communications. Comput. Netw. **190**, 107930 (2021). https://www.sciencedirect.com/science/article/pii/S1389128621000773
26. Du, H., Wang, J., Niyato, D., Kang, J., Xiong, Z., Guizani, M., Kim, D.I.: Rethinking wireless communication security in semantic internet of things. IEEE Wirel. Commun. **30**(3), 36–43 (2023)
27. Zafar, S., et al.: A systematic review of bio-cyber interface technologies and security issues for internet of bio-nano things. IEEE Access **9**, 93529–93566 (2021)
28. Dutta, D.L., Bharali, S.: TinyML meets IoT: a comprehensive survey. Internet Things. **16**, 100461 (2021). https://www.sciencedirect.com/science/article/pii/S2542660521001025
29. Stallings, W.: Cryptography and Network Security: Principles and Practice, 7th edn. Pearson Education Ltd., London (2017)

Optical Wireless Power Transmission Through Biological Tissue Using Commercial Photovoltaic Cells Under 810 nm LEDs: Feasibility Study

Syifaul Fuada[1(✉)], Malalgodage Amila Nilantha Perera[1], Mariella Särestöniemi[1,2], and Marcos Katz[1]

[1] Centre for Wireless Communications, University of Oulu, 90570 Oulu, Finland
{syifaul.fuada,malalgodage.perera,mariella.sarestoniemi, marcos.katz}@oulu.fi

[2] Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, 90570 Oulu, Finland

Abstract. Ensuring the provision of sustainable and secure electrical power for ingestible/implantable medical devices (IMDs) is crucial for facilitating the multifaceted capabilities of these IMDs and preventing the need for recurrent battery replacements. Using photovoltaic (PV) energy harvesting in conjunction with an external light source can be advantageous for an optical wireless power transfer (OWPT) system to enable energy self-sufficiency in IMDs. This study investigates the performance of OWPT using commercial monocrystalline silicon PV cells exposed to an 810 nm Near-infrared (NIR) LED light. The ethical concerns are addressed by utilizing porcine samples (*ex vivo* approach), eliminating the need for live animal experimentation. The experimental setup employs porcine meat samples with several compositions, e.g., pure fat, pure muscle, and different layers of fat-muscle. The primary goal of this initial study is to analyze the open-circuit voltage output (V_{OC}) of the PV against received optical power in the presence of biological tissue. Our study demonstrates that PV cells can generate voltage even when exposed to light passing through porcine samples with a thickness of up to 30 mm. Furthermore, the V_{OC} values of PV cells attained in this study meet the required voltage input level for supplying current IMDs, typically ranging from 2V to 3V. The findings of this study provide valuable insights into OWPT systems in the future, where monocrystalline silicon PV cells can be employed as energy harvester devices to supply various IMDs utilizing NIR light.

Keywords: Optical Wireless Power Transmission · Biological Tissue · Photovoltaic Cells · Near-infrared Light · Implantable Medical Devices

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 388–400, 2024.
https://doi.org/10.1007/978-3-031-59091-7_25

1 Introduction

IMDs offer significant advantages in real-time health monitoring and targeted treatments within the human body [1–3]. So far, a massive amount of research has been conducted to enhance wireless implantable medical devices (IMDs) dedicated to improving patients' well-being, for instance, in [4–10]. Various types of IMDs have been developed, each with unique functions and designs [11–13]. A crucial aspect in making these resource-limited IMDs more practical is providing sustainable electrical power [14], eliminating the need for frequent surgical procedures to replace batteries, typically occurring every 3 – 7 years [15, 16].

Emerging technologies for delivering sustainable and reliable electrical power to IMDs encompass piezoelectric or triboelectric generators, biofuel cells, inductive radio frequency (RF), photovoltaics (PV), and other approaches. Inductive RF as an OWPT system stands out due to its ability to provide relatively higher power levels [17, 18]. However, efficiency may be compromised when the transceiver is reduced in size or needs to be correctly aligned [14]. The PV energy harvesting also offers a viable solution by utilizing ambient or external light sources to generate sufficient electrical power for IMDs. Nevertheless, to the best of our understanding, most of the studies in the literature were conducted using visible light spectrum (e.g., sunlight) as a light source to power PV cell, for instance, as done by [19, 20]; as a result, the light beam can not penetrate deeply to the human body (limited to human skin layer) as visible light does not propagate efficiently across biological tissue [21–23]. This is due to the fact that the skin absorbs the majority of the visible light spectrum; particularly light with a wavelength of 1,300 nm, as it is almost absorbed by the water content in the skin layer [24]. Near-infrared (NIR) light which is a part of the light spectrums has the ability to penetrate deeper into biological tissue [4]. Unlike other wavelengths, NIR radiation is less affected by the absorption and scattering properties of biological tissues. Studies have confirmed that NIR wavelengths >700 nm, referred to as long waves, do not pose any harm to the human body and can effectively penetrate tissues [24]. To this end, there needs to be a more comprehensive investigation regarding the electrical performance of commercial PV cells when subjected to NIR through biological tissue.

Up to the authors' knowledge, there is still few research on OWPT using NIR. Authors in [25] studied efficient enhancement strategies for OWPT systems using NIR and PV cells; showing that up to 48% of OWPT efficiency can be reached. However, this research is conducted on a free space [25]. In [26], evaluation of OWPT employing a 750 nm 5 mW laser is investigated; it has been shown that the proposed system could recharge a 150 mAh battery even when situated beneath a skin tissue and regulated the power provided to a low power IMDs, which is less than 10 mW. Nevertheless, the efficacy of LED power level usage diminishes when applied to thicker biological tissue due to the substantial loss (caused by absorption, scattering, and reflectance factors) of optical power during the propagation of light through tissue [26].

This study analyzes the electrical performance attributes of a commercial PV that can be implanted in the human body for the OWPT system employing NIR (810 nm 375 mW), focusing on electrical voltage. We considered NIR light as it can propagate relatively well through biological tissue. Our experimental trials involved testing a PV placed under a porcine sample with the surface emitted by an NIR light source in an

aligned setting. We varied the transmitted optical power, and afterward, we measured the received power density and the output PV cells in an open circuit (V_{OC}). The findings of our study proved that harvesting energy using NIR and PV cells across the biological tissue is promising. On the other hand, our study offers unique insights into the design and implementation of OWPT using NIR for IMDs by studying the received power density and V_{OC} of commercial PV cells.

Contribution. This paper is the first time to exploit the feasibility of employing commercial monocrystalline silicon PV cells for OWPT through biological tissue under NIR LED under 810 nm, focusing on analyzing the V_{OC} characteristics. We considered porcine samples with different compositions, including pure fat tissue, pure muscle tissue, and porcine samples with thicknesses up to 30 mm. Moreover, we also considered much deeper provision of optical energy that in previous literature that focused under-skin cases.

2 Methodology

Figure 1 depicts the experimental setup used in this study; it consists of a LED driver (DC2200, *Thorlabs*), NIR LED (M810L4, *Thorlabs*), optical power meter (PM100D, *Thorlabs*) to measure received optical power (in mW) and power density (in mW/cm^2 unit), and multimeter (MM400, *Klein tools*) to measure V_{OC} of commercial PV cells. Figure 2 shows the commercial monocrystalline silicon PV cells used in this study. Three different PV cells were used: SM710K12L R3.0, SM111K04L R3.0, SM141K08LV R3.0 (Ixolar, *ANYSOLAR*), then denoted as #PV1, #PV2, #PV3, respectively. These PV cells are suitable for both outdoor and indoor applications. Table 1 summarizes the specifications of PV cells used. The number of cells of PV#1, #PV2, and #PV3 are 12, 4, and 8, respectively.

In this study, we considered only V_{OC}, that is open circuit performance. Therefore, we did not connect the output of PV cell to any load. Figure 3 shows the porcine samples used in this study, we used four samples: #1 (15 mm of thickness, fat layer), #2 (15 mm of thickness, muscle layer), #3 (25 mm of thickness), and #4 (30 mm of thickness). In the experimental procedure, all fresh meat samples were heated to a temperature of 37 °C using an air heater within a chamber, aiming to align with the typical body temperature of humans.

The dimensions of each sample are approximately 5 cm × 5 cm, and the LED is positioned in alignment with the receiver through the sample without any free space between them. Despite the small area of the used samples, the optical signal does not pass through the sides of the sample. The LED as a transmitter source typically has a limited field-of-view (FOV), leading to a very narrow optical beam. The propagation of optical light is more directional compared to radio waves. In the case of our study, the NIR LED has a confined viewing angle of 80° maximum, resulting in a narrow beam and very short optical distance. A similar study conducted by [14] also employed a small tissue area.

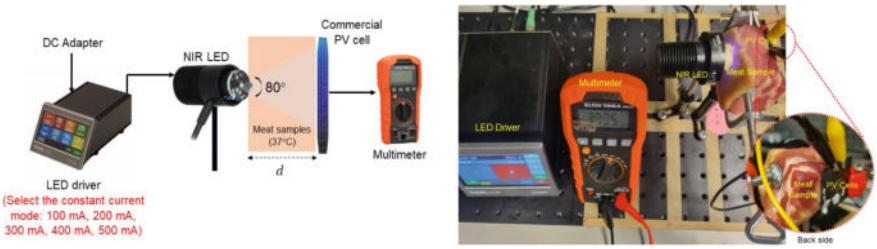


Fig. 1. Experimental setup.

We first measure the received power in the air propagation medium (free space). In the experimental setup, the NIR LED is positioned in direct line-of-sight (aligned) to the optical sensor (S121C, *Thorlabs*) and exposed towards 4 cm of free space optical distance. The optical sensor is connected to an optical power meter with settings as follows: attenuation = 0 dB, input aperture = Ø9500 nm, and wavelength = 810 nm. By adjusting the LED current through the LED, the subsequent endeavor entails measuring the received optical power and power density.

The results of measurements conducted in a 4 cm free space using LED driver settings of 500 mA, 400 mA, 300 mA, 200 mA, and 100 mA yielded power outputs of 47.3 mW, 38.5 mW, 29.1 mW, 19.4 mW, and 9.37 mW, respectively (Fig. 4a). Correspondingly, the power densities were calculated to be 66.7 mW/cm^2, 54.2 mW/cm^2, 41.1 mW/cm^2, 27.4 mW/cm^2, and 13.21 mW/cm^2, when subjected to 500 mA, 400 mA, 300 mA, 200 mA, and 100 mA of LED current, respectively (Fig. 4b). It was observed that an increase in LED current resulted in higher light intensity emitted by LED and consequently, higher power received by the receiver. The relationship between light intensity and received power was found to be linear, confirming previous findings by [27, 28].

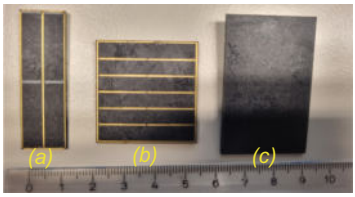


Fig. 2. Commercial Monocrystalline Si PV cells: (a) #PV2, (b) #PV1, (c) #PV3.

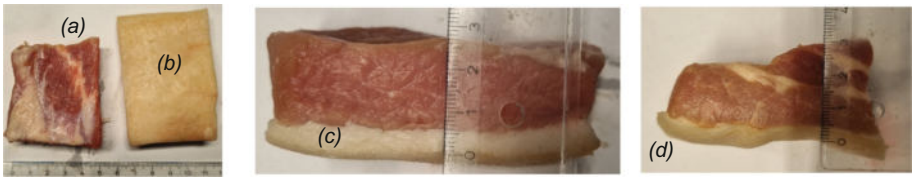
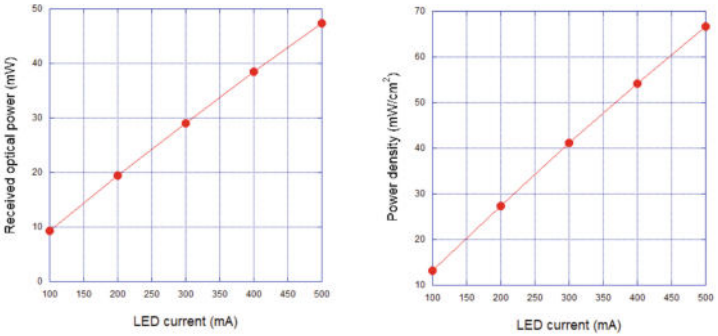


Fig. 3. Porcine samples: (a) sample #2, (b) sample #1, (c) sample #3, (d) sample #4.



(a) (b)

Fig. 4. Characteristics of NIR LED at 4 cm of optical distance (free space): (a) received power in mW; (b) power density in mW/cm^2.

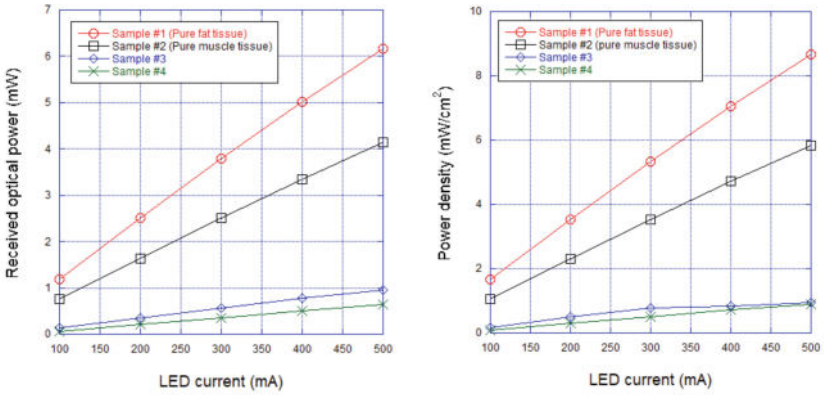
Table 1. Electrical characteristics of commercial PV cells used in the study

Cell Parameter Typical Ratings	#PV1	#PV2	#PV3
Open circuit voltage (V_{OC})	**8.29V**	**2.76V**	**5.53V**
Short circuit current (I_{SC})	29.2 mA	46.7 mA	58.6 mA
Voltage at max. Power point (V_{MPP})	6.70V	2.23V	4.46V
Current at max. Power point (I_{MPP})	27.4 mA	43.9 mA	55.1 mA

3 Results and Analysis

3.1 Received Optical Power

In this section, the optical power received after the NIR optical signal passes through each sample is measured. The experimental scenario refers to Fig. 1, where the NIR beam is directed towards the sample, while a sensor positioned behind the sample captures the optical power. The measurement of LED power reception is crucial as it determines the level of intensity received by the PV cells. The PV cell is capable of converting light into electricity as long as the given light can penetrate the skin's thickness, with the conversion process being directly proportional to the intensity of the light. The experimental findings, as presented in Fig. 5(a) and (b) reveal that the power density and optical power received by each sample are much lower compared to free space conditions, indicating a reduction in power due to the optical properties of the tissues. The percentages of power density and optical power received in relation to free space are 13%, 9%, 2%, and 1% for sample #1, sample #2, sample #3, and sample #4 respectively. These results suggest that fat tissue (sample #1) is a suitable signal propagation medium compared to muscle tissue (sample #2), in line with previous research on RF wave cases [10, 29]. We found that the results in optical waves coincide with those in RF waves on biological tissue. It is observed that thicker tissues lead to a decrease in optical power and power density, which is consistent with the results found in [30].



(a) (b)

Fig. 5. Measurement of received power on various samples: (a) in mW; (b) power density in mW/cm^2.

3.2 V_{OC} Measurement of Each PV Cells

As mentioned in the introduction, the main objective of this research was more highlighted on measuring the electrical voltage generated by PV cells under the OWPT cases. The V_{OC} of a single PV cell is a reliable indicator for assessing the depth of light penetration into the human body and the amount of external light collected by the PV cell implanted within the body; this is because the PV cell exhibits high sensitivity to variations in low-light environments [31]. Figures 6(a), (b), (c) show the results of PV#1, PV#2, and PV#3, respectively. The measurement results of #PV1 on a sample #1 show the V_{OC} for LED currents 500 mA, 400 mA, 300 mA, 200 mA, and 100 mA are 5.24 V, 5.04 V, 4.77 V, 4.40 V, and 3.75 V, respectively. Upon comparison with #PV1, it is observed that the V_{OC} values for #PV2 and #PV3 are lower, with average reductions of 32% and 71% respectively. This average reduction is determined based on the comparison between sample #1 and sample #4. The typical V_{OC} of #PV1 based on the datasheet (Table 1) is higher than #PV2 and #PV3, the voltage rating remains higher even when NIR is utilized through biological tissue. It can be summarized that the commercial PV cells are capable of operating when exposed to NIR light, particularly in applications involving biological tissue. The V_{OC} of #PV1 and #PV2 meets the IMD's voltage input requirements. The typical voltage demand for IMDs (e.g., pacemakers) is generally 2 V–3 V to operate [32]. These mentioned voltages are the ideal threshold of the voltage received in the PV cells that indicate a satisfactory for OWPT purposes. However, the average V_{OC} of #PV2 does not meet this standard, which is below 2 V. Possible solutions that can be addressed further include connecting multiple PV cells either in series or parallel to attain adequate V_{OC} level, or keep using single PV cells but incorporating storage (e.g., supercapacitors or rechargeable batteries) to store the generated energy – a similar approach has been suggested by [33].

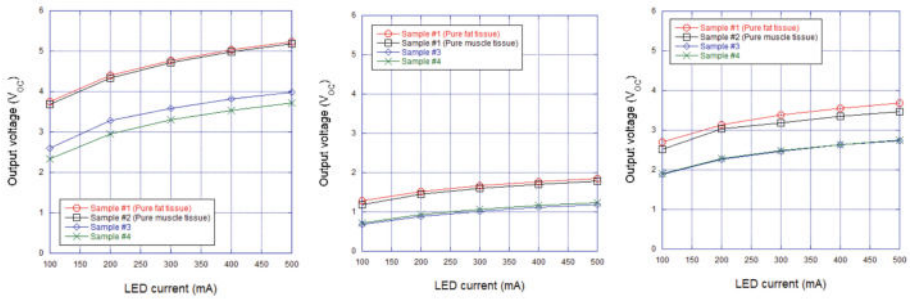


Fig. 6. Measurement result of V_{OC} of each PV cells on different samples: (a) #PV1; (b) #PV2; (c) #PV3.

Table 2 provides a summary of the voltage (V_{OC}) generated by PV cells. Several research studies have been conducted to design PV cells specifically for IMDs. However, it is important to note that these custom-designed PV cells as references [14, 34–36] are not currently available in the market. Instead, our study investigates the commercially available PV cells for IMDs. The V_{OC} level can be enhanced by configuring PV cells in parallel as proved in Table 2. The results show that the maximum attainable V_{OC} is 5.25 V (at an LED current of 500 mA on #PV1 and sample #1), while the lowest recorded V_{OC} is 0.67 V (at an LED current of 100 mA on #PV2 and sample #3). Using single PV cells, the V_{OC} attained in this study is relatively higher compared to the provided literature as in Table 2.

Table 2. Profile of V_{OC} of PV cell for IMDs on similar works

References	Year	Wavelength	PV Cell type	Open circuit voltage (V_{OC})
[19]	2014	Visible light	Commercial	1.89 V (single cell) and 5.67 V (parallel 3 ×)
[20]	2015	Visible light	Commercial	4.08 V
[34]	2016	NIR 532 nm	Own design	2.3 V (single cell) & 4.6 V (parallel 2 ×)
[35]	2017	Halogen lamp	Own design	0.53 V
[36]	2018	NIR 780 nm	Own design	4.25 V
[14]	2022	NIR 670 nm	Own design	0.957 V (single cell) & 3.81 V (parallel 5 ×)
[37]	2023	NIR 808 nm	Own design	0.45 V
This study		NIR 810 nm	Commercial	0.67 V (single cell, lowest) 5.24 V (single cell, highest)

3.3 Impact of Tissue Thickness

This section focuses on observing the influence of tissue thickness on the V_{OC} of each PV cell. Performing measurements on *ex-vivo* meat samples offer a more convenient option

compared to measurements on anaesthetized animals, as it eliminates the requirement for a clinical setting as typically conducted in hospital environments. Furthermore, the tissue dielectric characteristics of adult pigs closely mimic those of humans, rendering a commonly employed substitute for simulating the human body in medical studies [10]. In addition, adjusting the meat to temperatures closer to the average human body temperature, specifically 37 °C, is important as it will yield more realistic results, as concerned by [9, 38].

Figures 7(a) – (d) show the measurement results of samples #1 to #4, respectively. When measuring the V_{OC} of #PV2 in sample #1 (fat tissue), the recorded values were 1.85 V, 1.77 V, 1.67 V, 1.51 V, and 1.29 V, with corresponding LED currents of 500 mA, 400 mA, 300 mA, 200 mA, and 100 mA, respectively. Subsequently, when measuring V_{OC} of #PV2 in samples #2, #3, and #4, the average reductions were 95%, 62%, and 60% respectively. Meanwhile, for #PV1 in samples #2, #3, and #4, the V_{OC} were 98%,

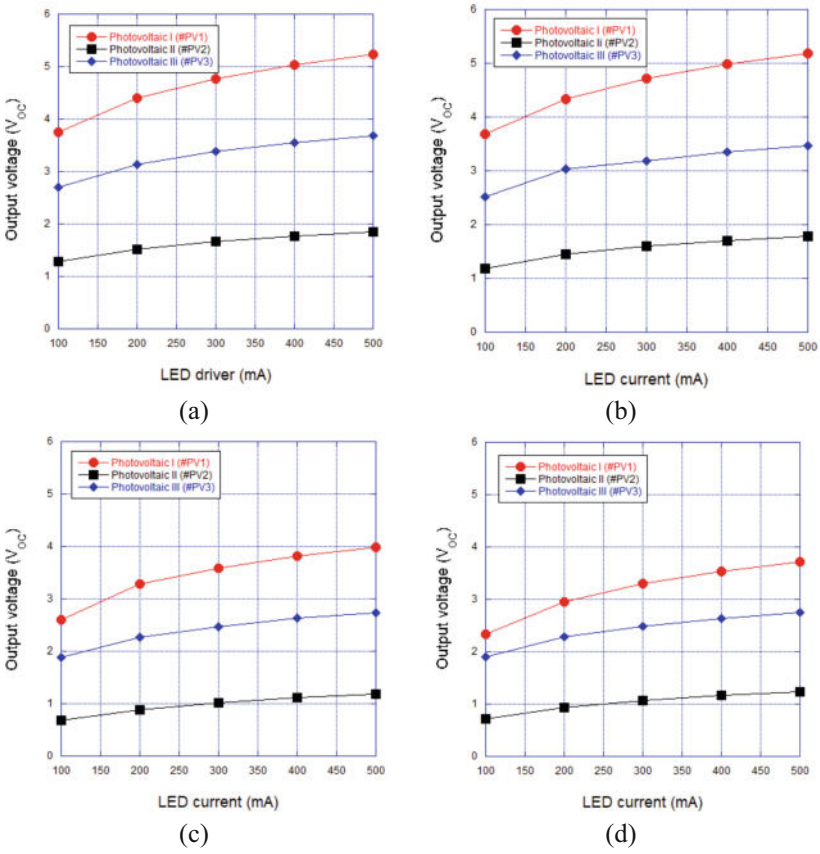


Fig. 7. Measurement result of V_{OC} of each PV cells on different samples: (a) sample #1; (b) sample #2; (c) sample #3; and (d) sample #4.

74%, and 68%, respectively. Similarly, the V_{OC} for #PV3 in samples #2, #3, and #4 were 95%, 74%, and 72%, respectively.

It is evident that as tissue thickness increases, the intensity received by PV cells decreases, resulting in smaller V_{OC} values for each PV cell. Notably, sample #4, the thickest, exhibited the most significant reduction of 60%–72%. Conversely, muscle tissue (sample #2), which has the same thickness as fat tissue (15 mm), experienced the most minor reduction, which is 95%–98%. This study did not include repeat measurements; thus, future studies should consider conducting repeated measurements and analyzing them statistically. This is crucial due to the potential deviations between initial and final measurements, as highlighted by [31].

4 Conclusion

The findings presented in this paper provide an understanding of the performance attributes of the electrical power delivery system within biological tissues; this includes scenarios where devices such as NIR LED 810 nm and PV cells. This study considered 375 mW of transmitter power maximum (supplied by 500 mA of LED current) and conducted on *ex-vivo* experiments using porcine samples. OWPT through biological tissue employing commercial PV cells and single beam NIR LED 810 nm is promising. PV cells can generate voltage despite the light being highly attenuated thick porcine samples up to 30 mm. Fat tissue is a better medium for light propagation than muscle tissue, as it results in higher optical power received by the receiver and, consequently, greater V_{OC} of the PV cell. The experimental measurements in this study furnish essential and foundational data for developing robust OWPT tailored to various types of IMDs. The V_{OC} value obtained in this study is greater than that reported in comparable literature utilizing commercial PV cells. The V_{OC} value obtained in this study is greater than that reported in comparable literature utilizing commercial PV cells. This study only relies on V_{OC} parameter since it is initial form of extensive study, hence, more analyses are required (for instance, electrical current, power, and energy).

In the future, we will consider integrating the presented approach with an energy harvesting circuit to analyze its electrical power and energy charging level against time. Thus, we will able to determine the duration of charging periods using PV cells and storage (e.g., coin rechargeable battery or supercapacitor) and its operating time when connected to IMDs (e.g., pacemakers). When connecting PV cells to a commercial power management integrated circuit (PMIC) development kit, it is crucial to be aware of the V_{OC} value of the PV cells. This is because the kit has a specific input voltage rating, for instance, the E-peas PMIC AEM10330 or AEM10300 has a voltage input rated from 100 mV – 4.5 V; exceeding these maximum limits for voltage input could result in damage. The V_{OC} measured in this study satisfy the minimum requirements for current IMDs devices, which is typically 2V – 3V. The #PV2 and #PV3 can be seamlessly integrated with these mentioned PMIC for further analysis as it falls within safe limits, which is lower than 4.5V. The PMIC with a higher maximum input voltage, like the AEM10941 which has an input voltage range of 50mV to 5V, provide an alternative option for #PV2 and #PV3. However, caution should be exercised before integrating to PMIC when using #PV1 as the measured voltage level for some samples exceeds 4.5 V. The compatible

PMIC designated for #PV1 could be BQ25504RGTT, which supports a maximum input voltage of 5.5V [39]. One of the challenges is that PV cells are naturally designed for wide spectrum operation (e.g., outdoor light coming from the sun or solar light, indoor light, and so on), while in our case, we used a narrowband light source, which is NIR light. Employing commercial PV cells for NIR light can be a fascinating issue in future.

Acknowledgments. This study was funded by the University of Oulu's Infotech (CWC-NS 2406124). The authors would like to acknowledge to the Research Council of Finland (former Academy of Finland) Profi6 funding, 6G-Enabling Sustainable Society (6GESS) the University of Oulu and the Research Council of Finland 6G Flagship Programme (Grant Number: 346208).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Fuada, S., Särestöniemi, M., Katz, M.: Analyzing the trends and global growth of energy harvesting for implantable medical devices (IMDs) research – a bibliometric approach. Int. J. Online Biomed. Eng. (iJOE) **20**, 115–135 (2024). https://doi.org/10.3991/ijoe.v20i03.45681
2. Fuada, S., Ma, G., Katz, M.: Global growth and trends of in-body communication research—insight from bibliometric analysis. Int. J. Online Biomed. Eng. (iJOE) **20**, 128–149 (2024). https://doi.org/10.3991/ijoe.v20i01.44967
3. Fuada, S., Särestöniemi, M., Katz, M.: Analyzing emerging trends in wireless implantable medical devices (IMDs): a bibliometric study. Int. J. Online Biomed. Eng. (iJOE) **20**, 115–143 (2024). https://doi.org/10.3991/ijoe.v20i04.46559
4. Ahmed, I., Halder, S., Bykov, A., Popov, A., Meglinski, I.V., Katz, M.: In-body communications exploiting light: a proof-of-concept study using ex vivo tissue samples. IEEE Access **8**, 190378–190389 (2020). https://doi.org/10.1109/ACCESS.2020.3031574
5. Boulogeorgos, A.-A.A., Trevlakis, S.E., Chatzidiamantis, N.D.: Optical wireless communications for in-body and transdermal biomedical applications. IEEE Commun. Mag. **59**, 119–125 (2021). https://doi.org/10.1109/MCOM.001.2000280
6. De Marcellis, A., Palange, E., Faccio, M., Di Patrizio Stanchieri, G., Constandinou, T.G.: A 250Mbps 24pJ/bit UWB-inspired optical communication system for bioimplants. In: 2017 IEEE Biomedical Circuits and Systems Conference (BioCAS), pp. 1–4 (2017). https://doi.org/10.1109/BIOCAS.2017.8325081
7. De Marcellis, A., Stanchieri, G.D.P., Faccio, M., Palange, E., Constandinou, T.G.: A 300 Mbps 37 PJ/bit pulsed optical biotelemetry. IEEE Trans. Biomed. Circuits Syst. **14**, 441–451 (2020). https://doi.org/10.1109/TBCAS.2020.2972733
8. Di Patrizio Stanchieri, G., De Marcellis, A., Battisti, G., Faccio, M., Palange, E., Constandinou, T.G.: A multilevel synchronized optical pulsed modulation for high efficiency biotelemetry. IEEE Trans. Biomed. Circuits Syst. **16**, 1313–1324 (2022). https://doi.org/10.1109/TBCAS.2022.3209542
9. Halder, S., Särestöniemi, M., Ahmed, I., Katz, M.: Providing connectivity to implanted electronics devices: experimental results on optical communications over biological tissues with comparisons against UWB. In: Alam, M.M., Hämäläinen, M., Mucchi, L., Niazi, I.K., Le

Moullec, Y. (eds.) Body Area Networks. Smart IoT and Big Data for Intelligent Health. LNICSSITE, vol. 330, pp. 3–17. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-64991-3_1
10. Särestöniemi, M., Pomalaza-Raez, C., Kissi, C., Iinatti, J.: On the UWB in-body propagation measurements using pork meat. In: Alam, M.M., Hämäläinen, M., Mucchi, L., Niazi, I.K., Le Moullec, Y. (eds.) Body Area Networks. Smart IoT and Big Data for Intelligent Health. LNICSSITE, vol. 330, pp. 18–33. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-64991-3_2
11. Liu, C., et al.: A wireless, implantable optoelectrochemical probe for optogenetic stimulation and dopamine detection. Microsyst. Nanoeng. **6**, 1–12 (2020). https://doi.org/10.1038/s41378-020-0176-9
12. Shi, Z., et al.: Silk-enabled conformal multifunctional bioelectronics for investigation of spatiotemporal epileptiform activities and multimodal neural encoding/decoding. Adv. Sci. **6**, 1801617 (2019). https://doi.org/10.1002/advs.201801617
13. Wang, L., Jiang, K., Shen, G.: Wearable, implantable, and interventional medical devices based on smart electronic skins. Adv. Mater. Technol. **6**, 2100107 (2021). https://doi.org/10.1002/admt.202100107
14. Seo, J., et al.: Wireless electrical power delivery using light through soft skin tissues under misalignment and deformation. Adv. Mater. Interfaces **9**, 2102586 (2022). https://doi.org/10.1002/admi.202102586
15. Kurtz, S.M., et al.: Implantation trends and patient profiles for pacemakers and implantable cardioverter defibrillators in the United States: 1993–2006. Pacing Clin. Electrophysiol. **33**, 705–711 (2010). https://doi.org/10.1111/j.1540-8159.2009.02670.x
16. Timmermann, L., et al.: A new rechargeable device for deep brain stimulation: a prospective patient satisfaction survey. Eur. Neurol. **69**, 193–199 (2013). https://doi.org/10.1159/000342236
17. Singer, A., Robinson, J.T.: Wireless power delivery techniques for miniature implantable bioelectronics. Adv. Healthcare Mater. **10**, 2100664 (2021). https://doi.org/10.1002/adhm.202100664
18. Xue, R.-F., Cheng, K.-W., Je, M.: High-efficiency wireless power transfer for biomedical implants by optimal resonant load transformation. IEEE Trans. Circuits Syst. I Regul. Pap. **60**, 867–874 (2013). https://doi.org/10.1109/TCSI.2012.2209297
19. Haeberlin, A., et al.: Successful pacing using a batteryless sunlight-powered pacemaker. EP Europace. **16**, 1534–1539 (2014). https://doi.org/10.1093/europace/euu127
20. Haeberlin, A., et al.: The first batteryless, solar-powered cardiac pacemaker. Heart Rhythm **12**, 1317–1323 (2015). https://doi.org/10.1016/j.hrthm.2015.02.032
21. Dremin, V., Zherebtsov, E., Bykov, A., Popov, A., Doronin, A., Meglinski, I.: Influence of blood pulsation on diagnostic volume in pulse oximetry and photoplethysmography measurements. Appl. Opt. **58**(34), 9398 (2019). https://doi.org/10.1364/AO.58.009398
22. Sohn, I., Rahman, M.F., Jang, Y.H., Lee, S.H.: An optical implant for biotelemetry: design, in vivo verification, and challenges. IEEE Commun. Mag. **60**, 50–56 (2022). https://doi.org/10.1109/MCOM.001.2100784
23. Sohn, I., Jang, Y.H., Lee, S.H.: Ultra-low-power implantable medical devices: optical wireless communication approach. IEEE Commun. Mag. **58**, 77–83 (2020). https://doi.org/10.1109/MCOM.001.1900609
24. Lee, J.-M., Kim, K.-H.: An experimental study on the penetration of 850nm and 940nm infrared radiation into porcine tissues. Indian J. Public Health Res. Dev. **10**, 1086–1089 (2019). https://doi.org/10.5958/0976-5506.2019.01217.8
25. Kim, S.M., Park, H.: Optimization of optical wireless power transfer using near-infrared laser diodes. Chin. Opt. Lett. COL. **18**, 042603 (2020)

26. Saha, A., Iqbal, S., Karmaker, M., Zinnat, S.F., Ali, M.T.: A wireless optical power system for medical implants using low power near-IR laser. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1978–1981. IEEE, JeJu, Korea (South) (2017). https://doi.org/10.1109/EMBC.2017.8037238
27. Nguyen, D.H.: Optical wireless power transfer for implanted and wearable devices. Sustainability **15**, 8146 (2023). https://doi.org/10.3390/su15108146
28. Putra, A.W.S., Kato, H., Maruyama, T.: Hybrid optical wireless power and data transmission system. In: 2020 IEEE PELS Workshop on Emerging Technologies: Wireless Power Transfer (WoW), pp. 374–376. IEEE, Seoul, Korea (South) (2020). https://doi.org/10.1109/WoW47795.2020.9291276
29. Särestöniemi, M., et al.: Fat in the abdomen area as a propagation medium in WBAN applications. In: Mucchi, L., Hämäläinen, M., Jayousi, S., Morosi, S. (eds.) Body Area Networks: Smart IoT and Big Data for Intelligent Health Management: 14th EAI International Conference, BODYNETS 2019, Florence, Italy, October 2-3, 2019, Proceedings, pp. 175–187. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-34833-5_15
30. Ahmed, I., Bykov, A., Popov, A., Meglinski, I., Katz, M.: Optical wireless data transfer through biotissues: practical evidence and initial results. In: Mucchi, L., Hämäläinen, M., Jayousi, S., Morosi, S. (eds.) Body Area Networks: Smart IoT and Big Data for Intelligent Health Management: 14th EAI International Conference, BODYNETS 2019, Florence, Italy, October 2-3, 2019, Proceedings, pp. 191–205. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-34833-5_16
31. Tholl, M., et al.: Implications of wound healing on subcutaneous photovoltaic energy harvesting. IEEE Trans. Biomed. Eng. **69**, 23–31 (2022). https://doi.org/10.1109/TBME.2021.3086671
32. Shi, B., Li, Z., Fan, Y.: Implantable energy-harvesting devices. Adv. Mater. **30**, 1801511 (2018). https://doi.org/10.1002/adma.201801511
33. MacVittie, K., et al.: From "cyborg" lobsters to a pacemaker powered by implantable biofuel cells. Energy Environ. Sci. **6**, 81–86 (2012). https://doi.org/10.1039/C2EE23209J
34. Song, K., et al.: Subdermal flexible solar cell arrays for powering medical electronic implants. Adv. Healthcare Mater. **5**, 1572–1580 (2016). https://doi.org/10.1002/adhm.201600222
35. Chen, Z., Law, M.-K., Mak, P.-I., Martins, R.P.: A single-chip solar energy harvesting IC using integrated photodiodes for biomedical implant applications. IEEE Trans. Biomed. Circuits Syst. **11**, 44–53 (2017). https://doi.org/10.1109/TBCAS.2016.2553152
36. Lu, L., et al.: Biodegradable Monocrystalline Silicon Photovoltaic Microcells as Power Supplies for Transient Biomedical Implants. Adv. Energy Mater. **8**, 1703035 (2018). https://doi.org/10.1002/aenm.201703035
37. Zhao, J., et al.: Self-powered implantable CMOS photovoltaic cell with 18.6% efficiency. IEEE Trans. Elect. Dev. **70**(6), 3149–3154 (2023). https://doi.org/10.1109/TED.2023.3268630
38. Halder, S.: Measurements and Characterization of Optical Wireless Communications Through Biological Tissues (2020). http://jultika.oulu.fi/Record/nbnfioulu-202007042732
39. Chen, S., Rodriguez-Villegas, E.: Solar energy harvesting node for battery-free physiological monitoring wearable wristband*. In: IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society. pp. 1–6. IEEE, Singapore, Singapore (2023). https://doi.org/10.1109/IECON51785.2023.10312620

Method to Monitor Cough by Employing Piezoelectric Energy Harvesting Configurations

Jaakko Palosaari, Eetu Virta, Miika Miinala, and Yang Bai(✉)

Microelectronics Research Unit, Faculty of Information Technology and Electrical Engineering, University of Oulu, 90570 Oulu, Finland
yang.bai@oulu.fi

Abstract. Cough is the most common symptom prompting individuals to seek medical advice. However, the widespread adoption of autonomous cough monitoring using wearable devices remains limited. This paper introduces a wireless cough monitoring device utilizing piezoelectric energy harvesting technology. The design emphasizes cost-effectiveness and energy efficiency, allowing simple attachment onto human skin using medical-grade tapes. The device's standout feature lies in its departure from continuously recording real-time acoustic data at a high sampling rate, as commonly employed in prior works. Instead, it capitalizes on the energy harvesting capability, utilizing harvested energy from muscle movements induced by coughing as crucial information. The energy harvested within specific intervals translates into a historical record of cough occurrences during that timeframe. This Energy-as-Data protocol substantially reduces the device's duty cycle, resulting in a remarkable extension of battery life by up to 2100%. Notably, this extension is achieved while maintaining reasonable accuracy in cough monitoring. With this capability, the device can autonomously monitor and analyze cough data from both in- and outpatients, serving daily, research, and clinical purposes. Its potential extends to enhancing prediction and management of severe respiratory diseases.

Keywords: energy-as-data · battery life extension · energy harvesting

1 Introduction

Respiratory conditions impact over 700 million patients globally. Despite technological advancements in disease management and treatment, more than half of the affected population remains undiagnosed or experiences inadequate disease management [1–3].

Coughing signifies an underlying issue in an individual's respiratory system, yet it remains an underutilized diagnostic tool. Currently, there's a lack of reliable and user-friendly cough monitors. Most existing solutions for smart cough sensing primarily rely on sound detection and recording methods [4, 5]. Prior studies have described the utilization of smartphone microphones to capture cough sounds, followed by the application of specific algorithms within mobile apps for processing and analyzing the recorded sounds. These algorithms were typically designed to differentiate coughs from speech

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 401–412, 2024.
https://doi.org/10.1007/978-3-031-59091-7_26

and background noises, utilizing AI-assisted protocols and machine learning. Furthermore, these applications were capable of distinguishing a user's coughs from those of others, providing analyzed statistics and information such as the timing, frequency, and severity of coughs, while also maintaining a relatively extensive and continuous history of the user's coughing behavior [4, 5].

However, research has highlighted that the previously mentioned solution heavily depends on the user keeping the smartphone in close proximity due to the nature of sound propagation. A notable drawback is that as the distance between the user and the smartphone increases, both the accuracy and reliability of cough measurement significantly diminish [6]. Moreover, the application requires the microphone to continuously operate, resulting in rapid battery drain and raising privacy concerns.

In this study, a wearable cough monitoring device that utilizes the piezoelectric energy harvesting concept to accurately measure coughs regardless of the distance between the user and the data receiver is introduced. This device can be directly affixed to the user's throat, enabling on-site storage of all cough-related data. Consequently, it operates independently of any smartphone or digital data receiver. This independence significantly expands its utility to individuals who either do not use smartphones or do not consistently remain in proximity to their phones. This includes demographics such as the elderly, athletes, children, individuals with disabilities, in-patients, and more. Additionally, a substantial extension of battery life is a significant benefit derived from this technology.

2 Methodology

2.1 Architecture of Piezoelectric Cough Sensing System

Generally, piezoelectric materials possess the capability to convert mechanical stress or strain into electric charge, or conversely, convert electric potential into material deformation. Piezoelectric energy harvesters harness the former attribute of these materials. When exposed to ambient kinetic energy like vibrations, impacts, or acceleration/deceleration, the piezoelectric component experiences mechanical stress or strain. In a closed external circuit, this results in the flow of electric charges, generating current, and subsequently producing usable electricity [7].

In fact, the mechanism utilized by piezoelectric sensors for decades involves continuously monitoring the output of the piezoelectric component, reflecting real-time mechanical stimuli on the sensor. However, in this study, employing this sensing mode was unfeasible. Continuous monitoring of the piezoelectric output necessitates incessant readings by the microcontroller in the circuitry. This continuous procedure would rapidly deplete the battery, which proves impractical for long-term cough monitoring, requiring an extended power source lifespan.

Conversely, the conventional piezoelectric energy harvesting approach was not utilized in this study either. Typically, the generated electricity from this method is employed to charge the battery, thereby extending its lifespan. However, in this particular work, the muscle movement induced by coughing is highly intermittent. The power generated per cough, typically ranging from nanowatts to microwatts, is insufficient to effectively charge the battery.

Hence, this study integrates the principles of piezoelectric sensing and energy harvesting. Figure 1 illustrates the operational mechanism of the cough sensing system. The system comprises a piezoelectric energy harvester connected in series with an AC-DC rectifier and an energy storage component. The piezoelectric energy harvester can adopt various structures optimized for off-resonance harvesting of kinetic vibrational signals, such as unimorph and bimorph cantilevers or diaphragms.

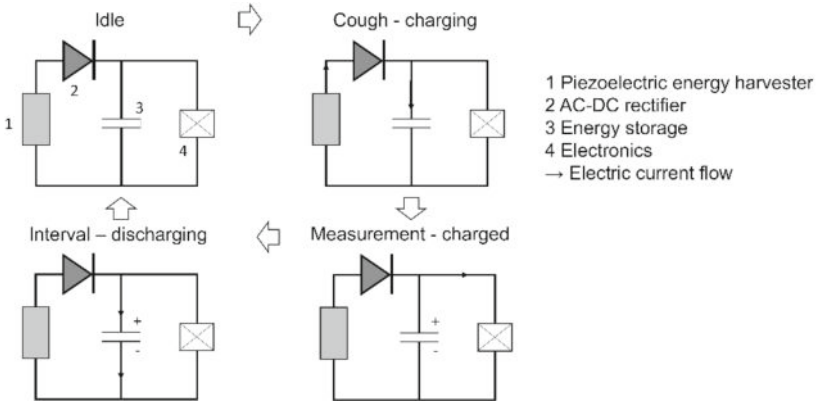


Fig. 1. Schematics of the working mechanism for the cough monitoring system.

The AC-DC rectifier may take the form of a diode or a full diode bridge rectifier. Its role is to convert the AC electric field generated by the piezoelectric energy harvester into a DC field for charging the energy storage. The energy storage element could be any component facilitating rapid charging and discharging, like a capacitor, thin-film battery, supercapacitor, and similar devices. Importantly, it is essential to note that the energy storage utilized here is not designed to serve as the primary power supply for the entire cough sensing device. Instead, it functions as an approach to significantly reduce the system's power consumption, which is elaborated in subsequent sections.

The choice of forward voltage for the rectifier should hinge on the sensitivity of the piezoelectric energy harvester concerning the ratio between the input cough strength and background noise, including body movements, speech, and other acoustic signals. Essentially, the rectifier should effectively distinguish genuine cough signals from non-cough signals.

Regarding the selection of capacitance for the energy storage, it should be based on the intended duration of the monitoring period. For prolonged cough monitoring, a larger capacitance with low leakage is preferable, while shorter-term monitoring favors a smaller capacitance with higher leakage.

2.2 Working Mechanism of Cough Sensing by Piezoelectric Energy Harvester

During the idle stage (Fig. 1), the system operates in standby mode without any energy generation from the piezoelectric energy harvester. However, when the system user

coughs, this action triggers the piezoelectric energy harvester to convert the kinetic energy from the muscle movement into electrical energy. This electrical energy then charges the capacitor, constituting what is termed as the 'Cough-charging' stage.

Within a defined measurement interval, the voltage level of the energy storage after charging is proportionate to the accumulation in number or intensity of coughs which occurred during that timeframe. Subsequently, the electronics measure the voltage of the charged capacitor at the end of this interval, known as the 'Measurement-charged' stage.

In the interval between two consecutive measurements, the energy storage undergoes a discharge process either through natural leakage or intentional discharge via programming. This discharged energy storage prepares for the next operational cycle, termed the 'Interval-discharging' stage. Following this stage, the system reverts to the idle stage, and this cyclical process continues over time. Figure 2 illustrates the block diagram of the electronics depicted in Fig. 1, where the energy storage is additionally connected in parallel with the analog input GPIO (general-purpose input/output).

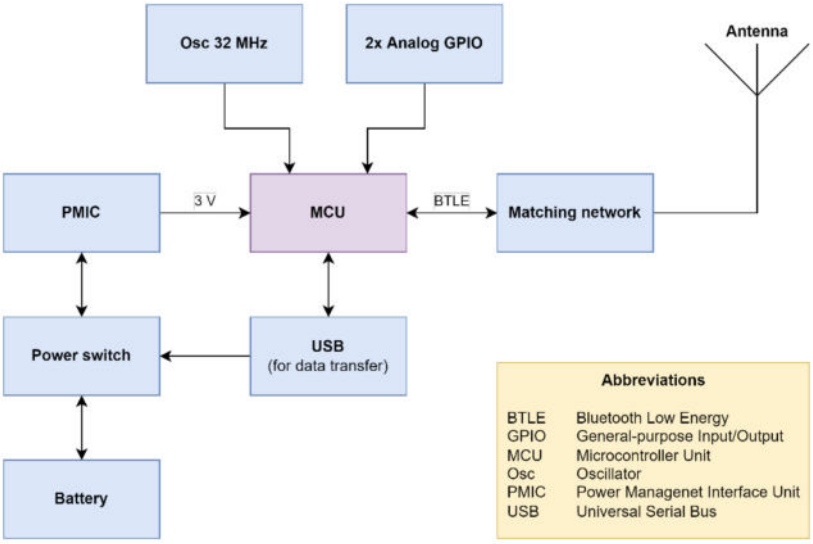


Fig. 2. Block diagram of the electronics.

2.3 System Integration and Device Packaging

The complete system was devised and compactly encapsulated within a small-sized device. Figure 3 illustrates two examples of the system's packaging. Figure 3(a) demonstrates the application of a piston-type mechanism, while Fig. 3(b) depicts a ball-triggering mechanism implemented onto the piezoelectric component. Biocompatible

materials were carefully selected for the packaging of the device. Figure 3(c) demonstrates the outlook of a fabricated device with diameter of 25 mm and thickness of 5 mm.

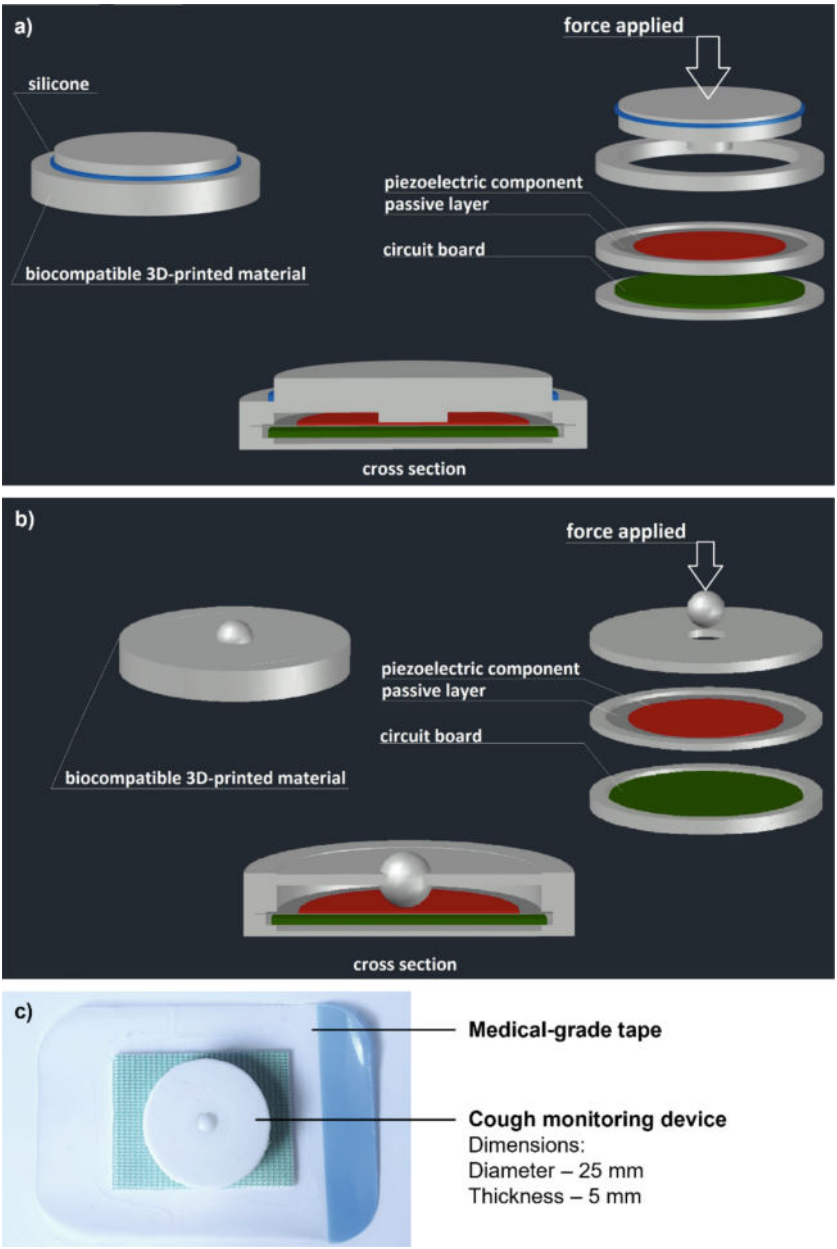


Fig. 3. Schematics of two examples for device structures and packaging: a) piston-type and b) ball-induced sensor element; c) Photo of a fabricated device placed on a medical-grade tape.

2.4 Program for Data Acquisition, Storage, and Transmission

Figure 4 illustrates the logical flowchart of the program utilized in this study. At the commencement of the monitoring procedure, the programmed microcontroller manages voltage measurements from the energy storage and subsequently stores the obtained data in a mass data storage unit. Furthermore, the program has the capability to enable wireless data transmission, although this specific function was not within the scope of this study.

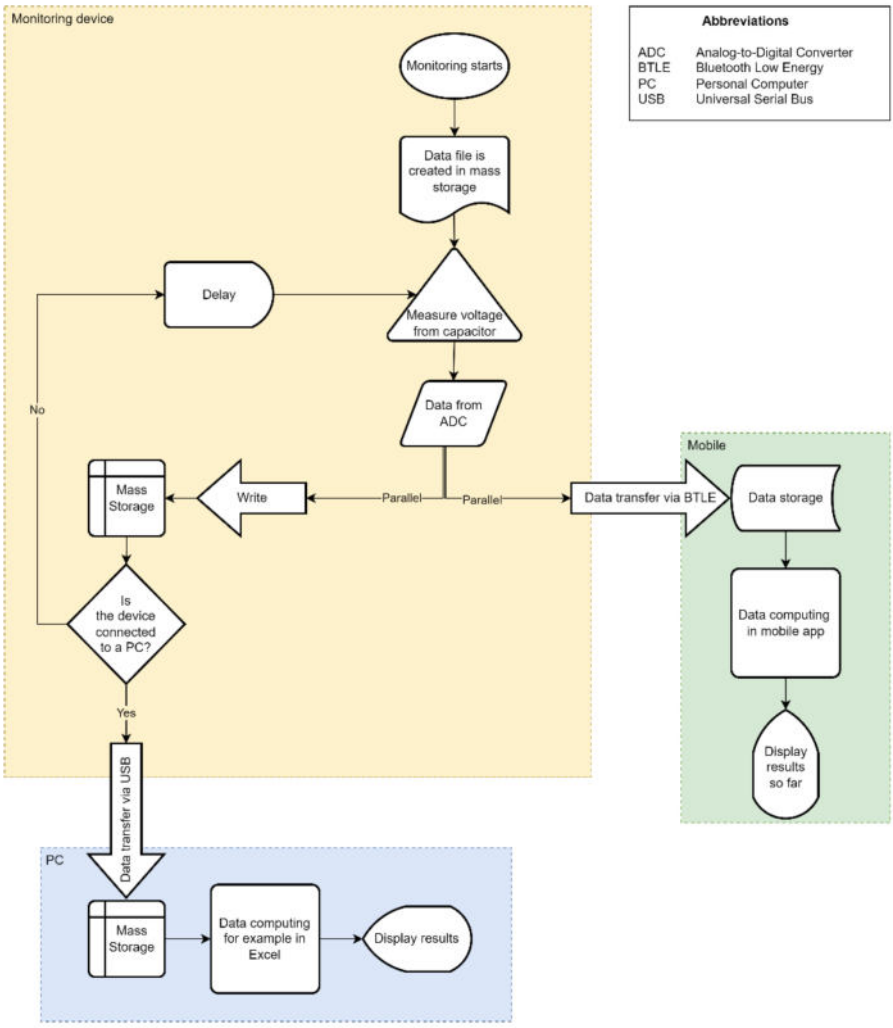


Fig. 4. Logic flow chart of the program used in this work.

3 Results and Discussions

3.1 Optimization of Sensitivity for Cough Detection

The placement of the device on the user's throat skin is a critical factor that determines the sensitivity and selectivity of cough signals. Figure 5 demonstrates the distinction between data collected at an improper position and data collected at an ideal position[1]. It should be noted that the data were unrectified and were collected before the energy storage.

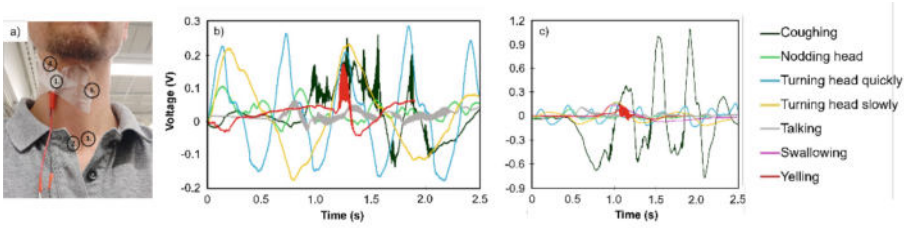


Fig. 5. a) Demonstration of measurement positions and raw output voltage before rectification detected from the piezoelectric energy harvester when the cough sensing device was attached at b) an improper location (position 4. in a)) and c) an ideal location on throat (position 1. in a)).

In Fig. 5 (b), it is evident that when the device was improperly attached to the throat, the piezoelectric energy harvester exhibited sensitivity to various types of signals beyond coughing. This sensitivity was due to muscle movements triggered by diverse actions, leading to the undesired outcome where the intended cough signals couldn't be distinguished from other signals categorized as noise.

Contrastingly, in Fig. 5 (c), when the device was positioned at an ideal location, the majority of unwanted signals could be effectively filtered out, emphasizing the amplified representation of the anticipated cough signal. Determining the correct positioning on the throat may vary among individuals and should be evaluated through trials before implementing the device in clinical use.

Impedance in the voltage measurements of Fig. 5 was fixed at 1 MΩ, giving the maximum output current of approximately 1.1 μA. However, it should be noted that for the methodology used in this study, the output current from the energy harvester was no longer essential. The charged capacitor voltage replaces the instantaneous output voltage and current and then becomes a crucial factor in the measurement procedure.

[1] Ethical issues were addressed before testing the device on human subjects in University of Oulu Hospital Testbed via contractual clinical trial.

3.2 Correction of Influence of Self-discharging on Data Acquisition from Energy Storage Unit

Understanding the self-discharge behavior of the chosen energy storage unit is crucial, as depicted in Fig. 1 and Fig. 4. This entails comparing the actual voltage increase after each 'Cough-charging' stage against that subsequent to the 'Interval-discharging' stage. Figure 6 exhibits an illustration of the self-discharge curve for a 100 μF commercial capacitor (C1210C107M8PAC7800, KEMET), showcasing its behavior without any external contribution of charging from the piezoelectric energy harvester. Additionally, the real-time voltage drop derived from the self-discharge curve is graphically presented.

When programming the system, it is essential to utilize the self-discharge curve as a dynamic baseline and as a reference for voltage calculation and analysis. This analysis involves estimating the expected decrease in voltage over specific periods to ascertain whether the capacitor has indeed been charged during those intervals.

Equation 1 expresses the gained voltage (V_{gain}) of the capacitor in each operation cycle (Fig. 1) with data acquisition interval (t_i), where t is time, V_t and V_{t+ti} are capacitor voltages at the beginning and end of a data acquisition cycle, respectively, and V_d is instantaneous voltage drop shown in Fig. 6. $V_{gain} > 0$ within t_i indicates that the capacitor has been charged by the piezoelectric energy harvester and the value of V_{gain} can then be translated to number or intensity of the coughs for the period of t_i.

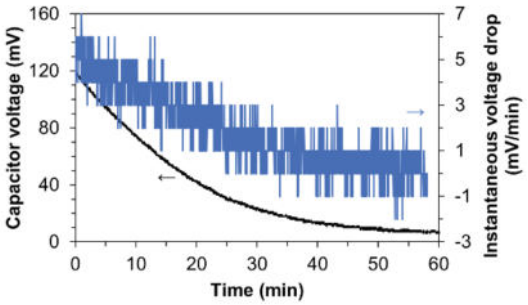


Fig. 6. Dependence of measured capacitor voltage and calculated instantaneous voltage drop on time for a 100 μF capacitor connected in the cough sensing system.

$$V_{\text{gain}} = V_{t+t_i} - V_t + \int_t^{t+t_i} dV_d \cdot t_i \tag{1}$$

3.3 Functionality Demonstration of Cough Sensing and Data Interpretation

Figure 5 previously detailed the device's optimization and positioning at an optimal throat location, effectively filtering out most non-cough signals during data collection. Considering that non-cough signals were relatively smaller compared to cough signals,

a threshold value for V_{gain} was established during the programming phase (Fig. 4) to exclude the recording of undesired harvested energy. Figure 7 displays the data gathered from a volunteer wearing the device during routine activities.

Setting V_{gain} to 100 mV before data collection ensured that only values above 100 mV were exhibited in Fig. 7. Values below this threshold, unlikely to represent a genuine cough, were omitted. In Fig. 7(a), the individual stayed indoors for approximately 6 h during the test. At the test's start and end, notably large V_{gain} values were recorded due to attaching and detaching the device, causing substantial impacts on the sensor. These anomalies are easily identifiable in practical scenarios and symmetrically appear at both ends, posing no significant concern for the accuracy of cough signal detection.

The dashed line in Fig. 7(a) represents the minimum level (150 mV) for successful cough signal detection. Among the 13 points surpassing this level, 10 were identified as true cough signals. If this minimum threshold was increased to 200 mV, all nine points above this level were accurate cough signals, but one genuine cough signal fell below the threshold, resulting in a missed detection. Therefore, basic post-data acquisition statistical analysis yielded a detection accuracy range of approximately 77% (10 out of 13) to 90% (9 out of 10) for indoor cough monitoring. It should be noted that the detection accuracy may also be affected by the frequency of cough during the wearing period of the user.

However, detection accuracy notably declined during outdoor activities, as seen in Fig. 7(b). Besides the substantial signals at the test's start and end due to device attachment/detachment, numerous non-cough signals were recorded. Regardless of the threshold level set for statistical analysis, either more non-cough data points were recorded than true cough data points, or a majority of authentic cough data points were disregarded.

In Fig. 7(b), non-cough signals were primarily from clothing adjustments and outdoor activities such as walking and driving. The major reason for these false signals was a scarf worn by the volunteer rubbing against the sensor. As a scarf is likely the closest possible object to the wearer's throat during outdoor activities, the case in Fig. 7(b) hence represents the worst possible scenario that may appear in practice.

The volunteer deliberately tapped the sensor during the test, with the finger-tapping signals reaching comparable levels to non-cough signals, indicating that outdoor clothes impacting the sensor generated a similar energy amount and transferred it to the capacitor.

While advanced data analysis, like deep learning, might differentiate between cough and non-cough data by analyzing data shapes or correlations [8, 9], this aspect exceeds the study's scope. Despite potential advanced analytical tools, optimizing the piezoelectric energy harvester and device structure to be less susceptible to external stimuli is pivotal. Future works should focus on designing a more sensitive piezoelectric energy harvester, possibly functioning in resonance mode, for improved cough detection.

Another possible challenge could be defining a standard set of thresholds that can be applied to a particular demographic based on gender, age, size, etc. Nevertheless, through clinical trials, it was proven that the position of the device did not need to be adjusted throughout the day for the same device wearer.

Despite the need for enhancements in outdoor cough monitoring, the primary advantage exhibited by the Energy-as-Data protocol implemented in the device—where muscle movement energy harvested through piezoelectric means served as data regarding cough

history within specific intervals—was the remarkable extension of battery life. The Li-ion battery (CR1225, Reneta Batteries, Switzerland) powering the entire monitoring system would typically last only overnight when operating in the traditional piezoelectric sensing mode with a high sampling rate. However, in the Energy-as-Data mode employed in this study, the battery could sustain operation for over a week, marking a 2100% increase in battery life thanks to the significantly lowered duty cycle. Considering the fact that possible addition of data analysis protocols can increase the energy consumption, the above-mentioned device structure optimization will be preferred to using on-site data analysis since there is a good chance that an optimized device structure is already able to screen cough data from the input end.

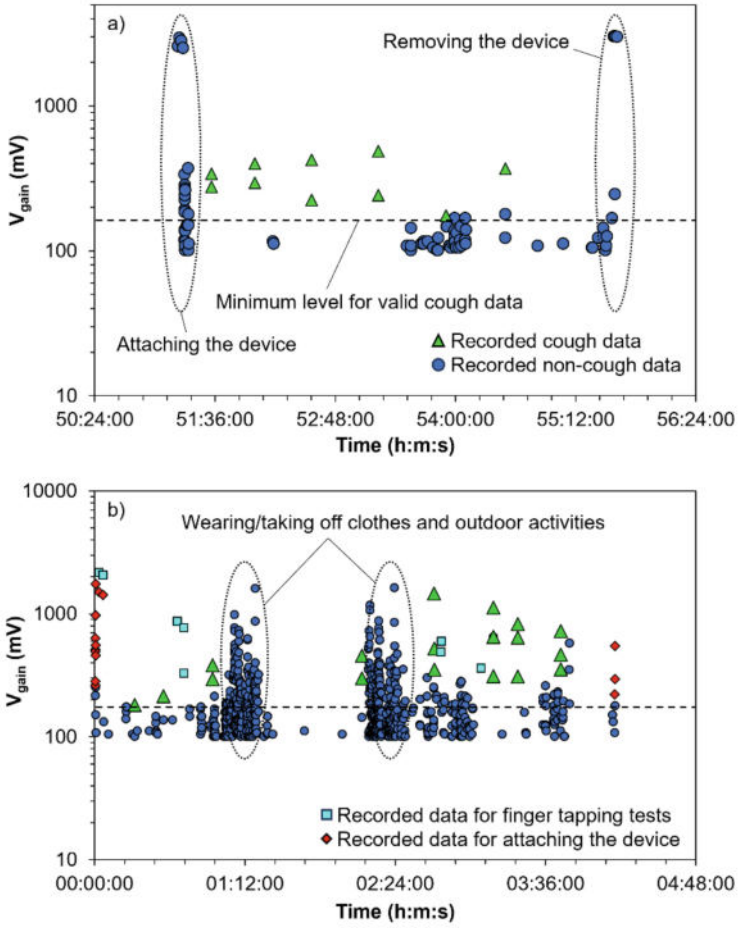


Fig. 7. Data captured by the cough monitoring device worn by a volunteer during regular activities, covering a) indoor sessions exclusively and b) combined indoor and outdoor scenarios. The V_{gain} threshold for data acquisition was established at 100 mV.

4 Conclusions

This research has introduced the operational mechanism, architectural design, data acquisition methodology, and performance evaluation of an independent cough monitoring system using a wearable piezoelectric energy harvesting device. Leveraging the Energy-as-Data concept, the battery life has been magnified by a factor of 21. Given the reasonably accurate cough detection rates of 77–90% for indoor monitoring, this novel device has showcased suitability for clinical trials aimed at forecasting and handling respiratory illnesses in patients predominantly indoors. Future endeavors will focus on refining cough detection accuracy in both indoor and outdoor settings. Additionally, study should be conducted for comfort level of wearing the device and hence possible strategies for improving the practicality of wearing the device for cough monitoring in real scenarios.

Acknowledgments. This work was funded by Business Finland (grant number 1428/31/2020). The authors also received funding from Riitta ja Jorma J. Takanen Foundation for carrying out part of this work. The authors acknowledge Nordic Semiconductor Oy for technical supports. The authors also acknowledge Prof. Heli Jantunen, Dr. Jari Juuti, Prof. Janne Haverinen, Pekka Räsänen, Toni Soini, Jyrki Still, and Christian Haller for insightful discussions related to this work.

Disclosure of Interests. Authors J.P., E.V. and M.M. own stocks in company Emedpatch Oy, Finland.

References

1. Ebrahimi, H., et al.: Global, regional, and national burden of respiratory tract cancers and associated risk factors from 1990 to 2019: a systematic analysis for the global burden of disease study 2019. Lancet Respir. Med. **9**(9), 1030–1049 (2021). https://doi.org/10.1016/S2213-2600(21)00164-8
2. Levine, S.M., Marciniuk, D.D.: Global impact of respiratory disease. Chest **161**(5), 1153–1154 (2022). https://doi.org/10.1016/j.chest.2022.01.014
3. Adams, E.J., et al.: Estimating the health and economic impact of improved management in prevalent chronic obstructive pulmonary disease populations in England, Germany, Canada, and Japan: a modelling study. Int. J. Chron. Obstruct. Pulmon. Dis. **18**, 2127–2146 (2023). https://doi.org/10.2147/COPD.S416988
4. Peltonen, V.T.K., Keating, A.J., Partridge, N.K.: Methods and apparatus for cough detection in background noise environments. Patent WO2018141013A3 (2018)
5. Peltonen, V.T.K., Keating, A.J., Partridge, N.K.: Methods and apparatus for cough detection in background noise environments. Patent US20200015709A1 (2018)
6. The Hyfe Team: Effects of distance on Hyfe performance. https://global-uploads.webflow.com/601331581ba868154325e525/611535225a2e852abfdd829c_wp5-hyfe-distance-tests.pdf. Accessed 27 Dec 2023
7. Bai, Y., Jantunen, H., Juuti, J.: Energy harvesting research: the road from single source to multisource. Adv. Mater. **30**(34), 1707271 (2018). https://doi.org/10.1002/adma.201707271

8. Babu, A., Ranpariya, S., Sinha, D.K., Chatterjee, A., Mandal, D.: Deep learning enabled early predicting cardiovascular status using highly sensitive piezoelectric sensor of solution-processable nylon-11. Adv. Mater. Technol. **8**(10), 2202021 (2023). https://doi.org/10.1002/admt.202202021
9. Babu, A., Ranpariya, S., Sinha, D.K., Mandal, D.: Deep learning enabled perceptive wearable sensor: an interactive gadget for tracking movement disorder. Adv. Mater. Technol. **8**(14), 2300046 (2023). https://doi.org/10.1002/admt.202300046

Microwave Technique Based Noninvasive Monitoring of Intracranial Pressure Using Realistic Phantom Models

Daljeet Singh[1(✉)], Erkki Vihriälä[2], Mariella Särestöniemi[1,3], and Teemu Myllylä[1,2,4]

[1] Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland
{daljeet.singh,mariella.sarestoniemi,teemu.myllyla}@oulu.fi
[2] Optoelectronics and Measurement Technique Unit, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland
erkki.vihriala@oulu.fi
[3] Centre for Wireless Communications, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland
[4] Medical Research Center, Oulu, Finland

Abstract. Microwave technology is emerging as a promising candidate in the field of medical diagnosis and imaging and has paved the way for a transition from invasive to non-invasive methods of monitoring various biological phenomena inside the human body. Intracranial Pressure (ICP) is considered to be a very important parameter by medical practitioners for assessing the health of a subject. Accurate, prolonged, and noninvasive measurement of ICP is still an open area of research with no clinical success so far. Therefore, in this paper, a microwave-based method for non-invasive monitoring of ICP is proposed. The setup utilizes flexible, thin, small, and lightweight planner antennas that are very suitable for non-invasive monitoring of ICP from the skin without compromising the comfort of subject. The proposed microwave method is tested on a realistic head phantom model which imitates the functioning of hydrodynamics in a real human head. The measurement results from the proposed method are verified using invasive pressure sensors. It is deduced from numerous trials that the proposed microwave system can detect small changes in ICP pressure and its response is analogous to actual pressure values measured by invasive pressure sensors.

Keywords: Intracranial Pressure · Microwave · non-invasive · brain monitoring · Cerebrospinal Fluid · hemorrhage · stroke

1 Introduction

The brain is one of the sovereign organs in the human body responsible for a variety of intricate biological processes that affect a person's overall functioning and well-being. Thus monitoring the different phenomena associated with brain

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 413–425, 2024.
https://doi.org/10.1007/978-3-031-59091-7_27

activity and functioning is of paramount importance to doctors and researchers working in the field of medicine and bioengineering [32]. The measurement of Intracranial Pressure (ICP) is one of these phenomena that can give deep insights into brain health and performance. Numerous neurological disorders such as swelling in the brain, intracranial hemorrhage, stroke, brain tumor, traumatic brain injury (TBI), and/or hydrocephalus have an impact on ICP [3,18,24,30]. The human brain is surrounded by a rigid bone structure that maintains a constant pressure inside the skull by optimizing the volume of its contents. ICP is the pressure within the craniospinal compartment constituted of brain, blood, and Cerebrospinal Fluid (CSF) and is governed by the Monro-Kellie doctrine [25]. The mean ICP for human adults is in the range of 5–15 mmHg when the subject is lying down with face and body looking upwards [4].

The current methods for ICP measurement in clinical conditions are mainly invasive and involve inserting either an intraventricular catheter, micro transducer, external ventricular drain (EVD), or lumbar puncture inside the skull [6,38]. These methods can produce fairly accurate ICP results but are bound by numerous limitations in terms of infections, malposition, time to set, and the requirement of precise neurosurgical expertise. To add to this, the ICP measuring device can cause hemorrhage of its own. Further, the invasive methods cannot be used for a prolonged time and are only suitable for surgical procedures in hospital settings [37]. The field of wearable brain monitoring technologies has seen a massive upsurge in clinical trials and research over the past two decades due to its ease of access and safety. In recent years, a lot of non-invasive methods have been proposed as a solution to the problems of invasive methods. A non-invasive device can eliminate the problems associated with invasive devices and is a suitable candidate for both clinical applications as well as prolonged monitoring of ICP outside hospital settings. The use of current 5G and 6G technologies in healthcare is also supporting the usage of non-invasive devices outside standard hospital settings. Understanding the trend of ICP values can benefit the diagnosis of less critical illnesses such as headaches, migraines, and sight issues, for which ICP readings are typically not considered required [38].

The non-invasive methods for ICP measurements include ultrasound time of flight technique [27,28], Transcranial Doppler (TCD) ultrasonography [2,34], otoacoustic emission [8], Magnetic Resonance Imaging (MRI) [11], Electroencephalography (EEG) [7], tympanic membrane displacement [9], acoustic methods [8,21], optic nerve sheath ultrasonography [20], ophthalmodynamometry [22], optical coherence tomography of retina [35] and jugular vein measurement [36]. Table 1 gives an overview of different techniques for ICP measurements based on the type of sensors and phantom models. Apart from these methods, the microwave technique is also proposed in some of the studies for ICP measurement [1,12,13,16,26,29]. When compared with other techniques, the use of microwaves in ICP measurement offers several advantageous features such as safety due to the use of a non-ionizing electromagnetic (EM) field, higher penetration depth compared to optical modalities, mobility of equipment due to low power small size transducers and transceivers, ease of application due to its

Table 1. Different techniques for ICP measurements based on the type of sensors and phantom models.

Ref.	Senor used	Phantom/dataset
[19]	PZT sensor and MEMS capacitive sensor	Dry tests in a sealed chamber, canine model (beagle), and specimen from dura mata
[16]	Sub-dural ICP implant and planar inverted F antenna	In-vitro and in-vivo canine test set-up
[23]	CC2500 2.4 GHz transceiver and MSP430 microcontroller	dry test and wet test in sealed chamber
[15]	Model-based approach on dataset	Dataset from comatose patients with severe closed head injury
[13]	Annular slot antennas (ASA) at 0.9 GHz	Phantom of cubic shape plastic container box containing a balloon
[17]	MEMS capacitive sensor	Animal model of blast-induced brain injury
[3]	MEMS pressure sensor	5 mm layer of pig skin
[29]	Microwave SRR sensor	Phantom of the upper part of head
[12]	EM resonant sensor patch	Human Subjects and phantom
[5]	B4C sensor	Human Subjects
[10]	NellcorTM SPO2 Forehead Sensor	Human Subjects

non-invasive nature and the possibility to be used from a distance in wireless measurements from the bedside without requirements of moving the patient, fast signal acquisition, lower cost of equipment and usage, etc. These benefits make microwaves an ideal solution for the measurement of ICP.

Motivated by the advantages of non-invasive methods for ICP measurement, this paper presents a method for accurately monitoring ICP changes in the head using microwave technology. One unique feature of this work is the testing of the proposed microwave method on realistic head phantom models. Most of the related studies in the literature are based on very simplified phantom models that fail to resemble the actual hydrodynamics of the human head. In this regard, a realistic phantom model is developed in this paper which is more suitable to study the changes in ICP. Another important feature of this work is the comparison between different antenna configurations which aids in selecting the best suitable position for placement of antennas around the skull. Further, unlike most of the previous studies, the present work is not confined to the study of only S11 but also considers other S parameters. This provides a better insight into the relationship between the S parameters and changing ICP values. Measurements

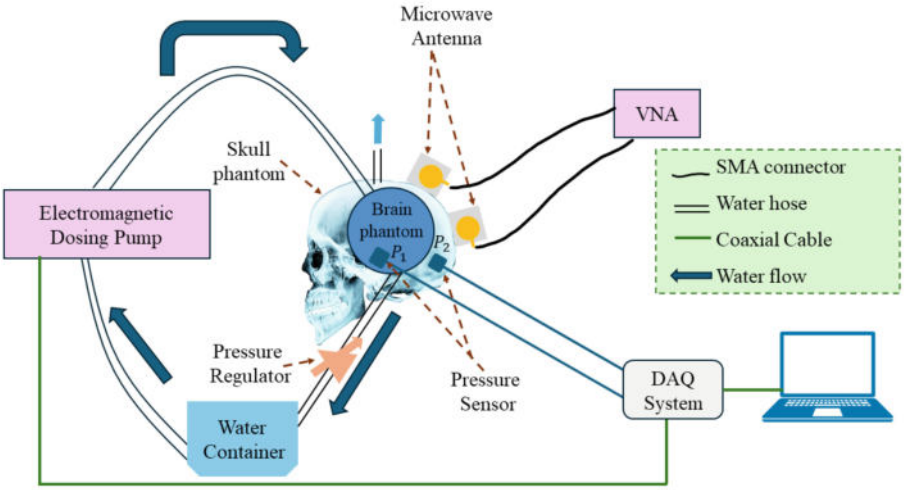


Fig. 1. Block diagram of ICP Measurement Setup.

made on realistic phantoms and simulation results showcase that the proposed system can be utilized for accurate and efficient non-invasive measurement of ICP.

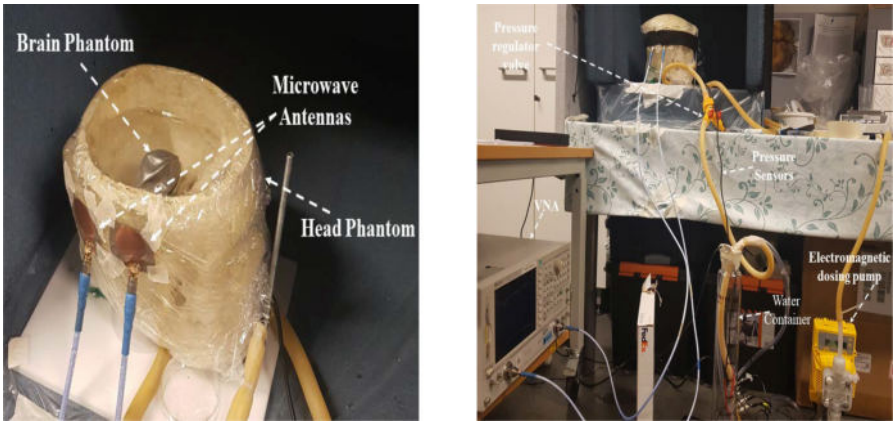
The rest of the paper is organized as follows: Sect. 2 presents the Material and Methods utilized in the study which constitutes the description of the brain phantom model and microwave method for ICP measurement. The results and discussion is presented in Sect. 3 and Sect. 4 holds the concluding remarks of the paper.

2 Material and Methods

The block diagram depicting the setup utilized in this study for ICP measurements is shown in Fig. 1. The setup comprises two blocks i.e. the phantom system and the system for ICP measurement. The phantom system consists of a head phantom, an electromagnetic dosing pump, and a water container. The ICP measurement system consists of a Vector Network Analyzer (VNA) connected to two microstrip patch antennas using SMA connectors, two pressure sensors, a Data Acquisition (DAQ) System, and a personal computer. A detailed description of both sub-systems is presented in the following subsections.

2.1 Realistic Phantom Model

The phantom model developed for ICP measurement consists of a brain phantom inside a skull phantom. The skull phantom is acquired from True Phantom Solutions [33]. This human skull phantom is made from epoxy-based bone material and the dimensions are taken from a Computed Tomography (CT) scan of



(a) Setup with microwave antennas attached to head phantom and brain phantom.

(b) Complete Setup with antennas, VNA, pump, hoses and pressure sensors.

Fig. 2. Microwave-based setup for ICP measurement using a realistic phantom model.

an average human male head. The dielectric properties of the skull phantom correspond to an actual human skull. The skull is horizontally cut into two parts which can be joined for easy accessibility of space inside the skull. A customized stand is built to maintain the skull phantom in an upright position. The brain phantom is made from a nonporous flexible balloon of dimensions similar to the average human brain when inflated. The skull phantom is partially filled with water. The brain phantom is carefully placed inside the skull phantom.

The skull phantom has a hole at the bottom through which a Y-shaped two-pronged connector is installed for pumping liquid inside the skull. One end of the pronged connector is connected to the brain phantom using a water hose of suitable length. The other ends of the connector are used to pump water in and out of the brain phantom. The inlet of the brain phantom is connected to a water container using one end of the Y-shaped two-pronged connector and hose which acts as an outlet for the phantom system. Similarly, the other end of the connector is connected to the water pump which works as the inlet of liquid inside the brain phantom. A very precise electromagnetic dosing pump 'Athena 4' from Injecta [14] is utilized for pumping the liquid in and out of the brain phantom. A pressure regulator is installed between the hose connecting the outlet of the brain phantom with the water container. The outflow of liquid from the brain phantom can be controlled using this regulator and thus the required value of ICP can be maintained inside the phantom system. The snapshot of the microwave-based setup for ICP measurement using a realistic phantom model is shown in Fig. 2. The setup with microwave antennas attached to the head phantom is shown in Fig. 2(a) and the complete setup with antennas, Vector Network Analyser (VNA), pump, hoses, and pressure sensors can be visualized in Fig. 2(b).

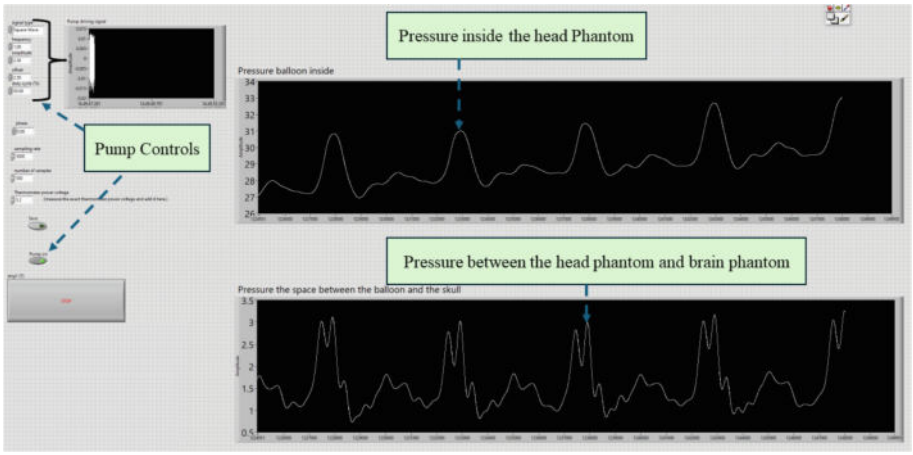


Fig. 3. LabVIEW program for monitoring pressure values and controlling the electromagnetic dosing pump.

The pressure and pulsation generated by the pump are controlled by using a LabVIEW program on a computer. A pressure sensor is installed inside the brain phantom (P_1). Another pressure sensor is mounted inside the skull phantom which measures the pressure between the skull and brain phantom (P_2). The measured pressure readings from both sensors are captured by a DAQ system which is then transferred to a personal computer. A LabVIEW program is created for monitoring and recording the pressure values from both sensors. The snapshot of the LabVIEW program is shown in Fig. 3.

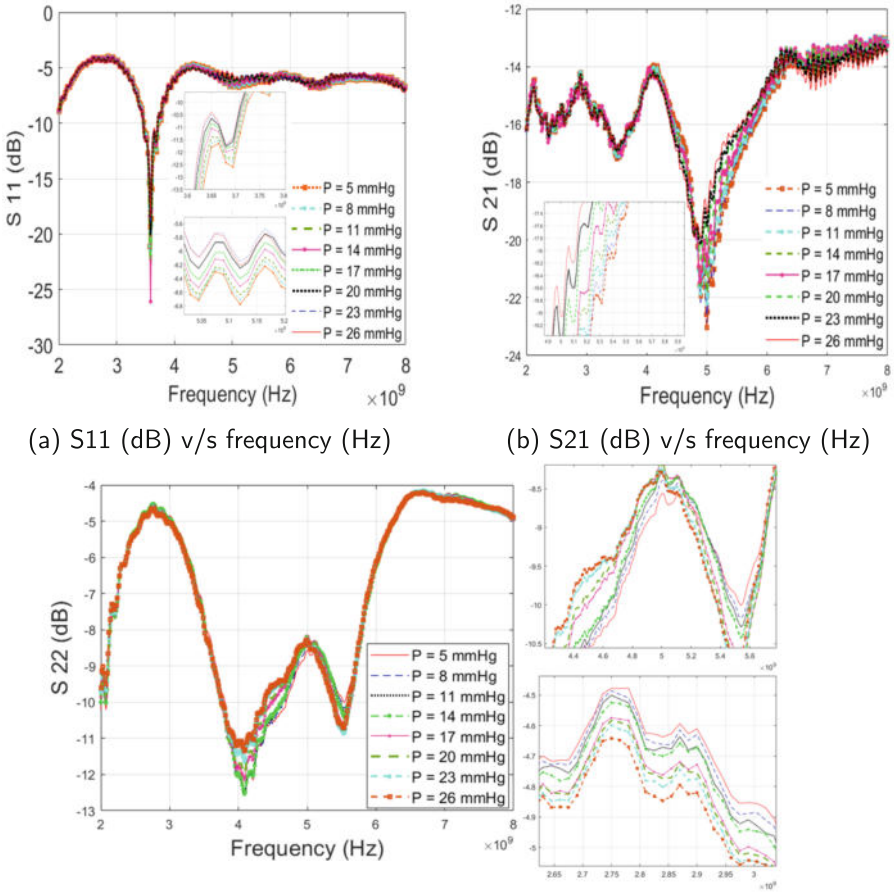
2.2 Microwave-Based System for ICP Measurement

The proposed microwave-based system for ICP measurement consists of two small flexible microstrip patch antennas designed on Rogers5880. The flexibility offered by these antennas makes them suitable for on-body measurements. The antennas operate in two bands i.e. 2.5 GHz (ISM band) and 3.1–10.6 GHz (UWB band). The overall dimensions of the antenna are 40 × 40 mm. Further details on this antenna design are available in [31]. Two different configurations are tested for antenna placement and compared in terms of their accuracy for ICP measurement. In the first configuration, the antennas are placed in a vertical orientation with a horizontal spacing of 1 cm from each other on the same side of the skull as shown in Fig. 2(a). The other configuration involves placing the antenna in a vertical orientation on opposite sides of the skull 13 cm apart. The antennas are directly connected to VNA and S parameters are computed for numerous scenarios. Two pressure sensors P_1 and P_2 are utilized to measure real-time pressure readings of the setup. These pressure values are used to verify the results obtained from the microwave-based ICP measurement system. The pressure sensor P_1 is installed inside the brain phantom and the pressure sensor

P_2 is installed between the head phantom and brain phantom. The pressure sensors are connected to a Data Acquisition (DAQ) system using coaxial cables which are further connected to a computer for processing the data.

3 Results and Discussion

The microwave technique-based setup explained in Sect. 2 is utilized for measuring ICP in different physiological conditions. One of the primary aims of this study is to test the feasibility of microwave techniques for ICP monitoring in realistic scenarios. The pressure sensors installed in the skull phantom and brain phantom are utilized to extract very precise real-time pressure values. These



(a) S11 (dB) v/s frequency (Hz)

(b) S21 (dB) v/s frequency (Hz)

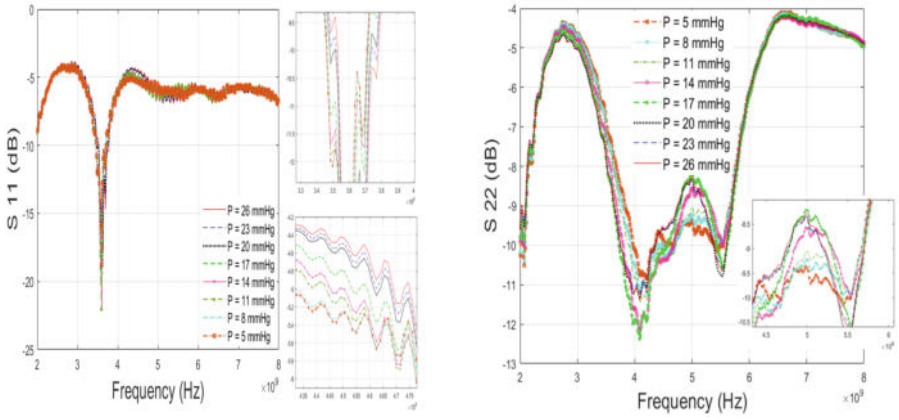
(c) S22 (dB) v/s frequency (Hz)

Fig. 4. S parameters when both antennas are placed on the same side with increasing values of pressure from 5 mmHg to 26 mmHg.

pressure values are utilized as a reference to verify the results obtained from the proposed microwave method. The relationship between variation in S parameters at different frequencies and ICP values is developed using numerous trials.

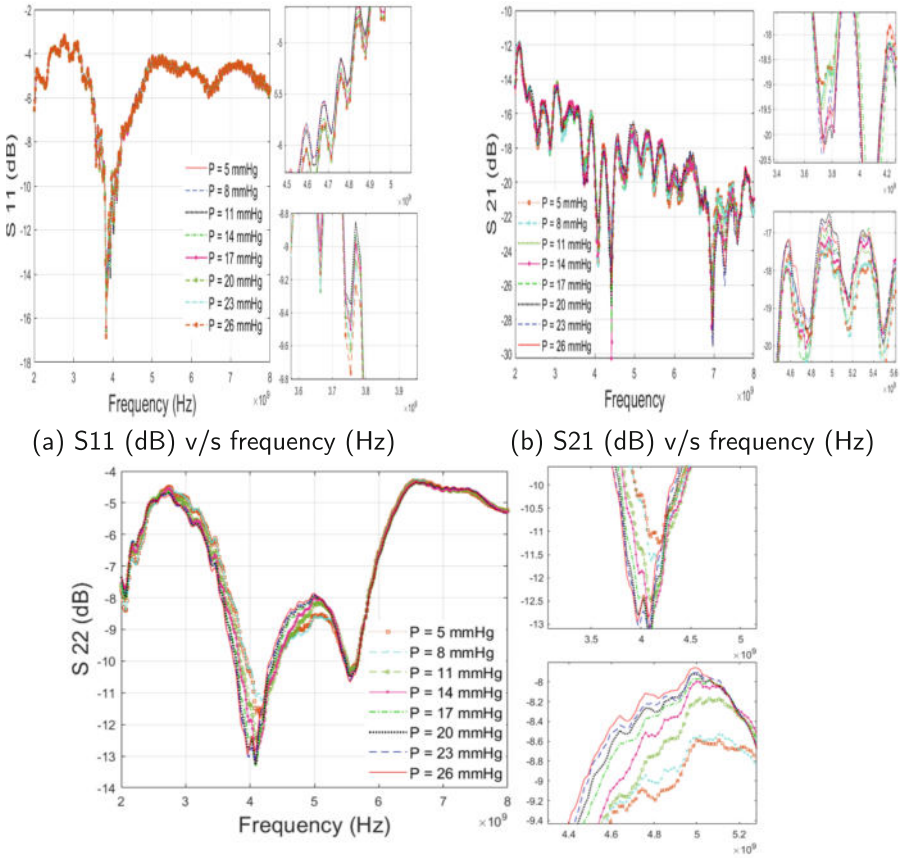
Figure 4(a) shows the S11 plot for the first configuration of the antenna (when both antennas are placed on the same side). The measurements are taken with increasing ICP values from 5 mmHg to 26 mmHg. The subfigures are provided in each graph to present a zoomed view of a particular frequency range of interest. It can be visible from the subfigures of Fig. 4(a) that change in ICP has a distinctive effect on the S11 parameter at some specific frequency bands. The results are shown for 3.6–3.8 GHz and 5.05–5.2 GHz. It can be observed that the S11 curves show a distinctive pattern, especially in the 5.05–5.2 GHz band with changing ICP. The S21 results for this antenna configuration are presented in Fig. 4(b). Similar to the case of S11, S21 also shows a cognitive trend at some frequency bands (4.9–5.4 GHz) with increasing values of ICP. The S22 results for microwave setup with both antennas on the same side and increasing values of pressure are shown in Fig. 4(c). An interesting observation from Fig. 4(c) is that the trend of S22 curves is similar to S11 and S21. This gives a better insight into the system response regarding changing ICP values especially when the difference in S parameters is small for minute changes in ICP.

In order to verify the trend and establish a thorough understanding of the relationship between S parameters and ICP values, the investigation carried out for the case described in Fig. 4 is repeated but with decreasing values of ICP. For this case, the balloon was initially filled to achieve the maximum value of ICP, and the pressure was then gradually reduced in small steps to study the trends of S parameters. The results of S parameters for the case of decreasing values of pressure when both antennas are placed on the same side of the skull are shown



(a) S11 (dB) v/s frequency (Hz) (b) S22 (dB) v/s frequency (Hz)

Fig. 5. S parameters for the case of decreasing values of pressure from 26 mmHg to 5 mmHg when both antennas are placed on the same side of the skull.



(a) S11 (dB) v/s frequency (Hz)

(b) S21 (dB) v/s frequency (Hz)

(c) S22 (dB) v/s frequency (Hz)

Fig. 6. Curves of S parameters v/s frequency for the case when both antennas are placed on the opposite sides of the skull with increasing values of pressure from 5 mmHg to 26 mmHg.

in Fig. 5. It can be visualized from Fig. 5(a) that the S11 parameters showcase a similar trend as observed during increasing values of ICP but not in the same frequency bands. In this case, a strong trend is visible for the frequency bands of 3.3–3.8 GHz and 4.35–4.75 GHz. Similarly, the results of S22 (dB) v/s frequency for this configuration are shown in Fig. 5(b).

Further, Fig. 6 shows the results of S parameters for another antenna configuration wherein the antennas are placed on opposite sides of the skull 13 cm apart. The S11 v/s frequency plot for this configuration is shown in Fig. 6(a). It can be visualized from Fig. 6(a) that the S11 parameter shows very subtle changes wrt. frequency. The reason for such a response is the penetration losses which become higher in this case due to signal transmission through the complete head phantom. Similar results for S21 and S22 are shown in Fig. 6(b) and

(c). It is also observed that the S parameter curves show some outliers in the trend with change in ICP values which is again due to the higher signal losses as compared to earlier cases when antennas were placed on the same side of the skull.

4 Conclusion

A microwave-based method is proposed in this work for non-invasive monitoring of ICP. The proposed method is based on flexible small-sized antennas and a realistic phantom model. The method is tested for both the cases of increasing as well as decreasing ICP values. Two different antenna configurations have been evaluated and it is observed that placing the antennas on the same side of the skull produces more favorable results in terms of accurate tracking of ICP as compared to when the antennas are placed on opposite sides. The interrelation between the S-parameters and ICP values is visible from the results. However, further research is required to accurately translate the antenna coefficients to the corresponding ICP values. This work can be further extended by testing with different antenna setups to determine the frequency band and optimal distance between the antennas.

Acknowledgments. This work was supported by the 6GBRIDGE - Next generation healthcare and wearable diagnostics utilizing 6G project (11146/31/2022), Academy of Finland Profi6 funding, 6G-Enabling Sustainable Society (University of Oulu, Finland) under the Emerging project, and 6G-enabled sustainable society (6GESS) program: 6GESS6. A part of this research is also funded by the Global Pilots financed by the Finnish Ministry of Education and Culture as part of the project INDFICORE. The authors thank Erkki Vihriälä from the University of Oulu for helping with the LabVIEW program and sensor setup.

Disclosure of Interest. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alruwaili, F., Cluff, K., Griffith, J., Farhoud, H.: Passive self resonant skin patch sensor to monitor cardiac intraventricular stroke volume using electromagnetic properties of blood. IEEE J. Transl. Eng. Health Med. **6**, 1–9 (2018)
2. Badillo, S.P.J., Navarro, J.C., Qureshi, A.I., Sharma, V.K.: Transcranial doppler ultrasonography in tuberculous meningitis: A systematic review. J. Neuroimaging **33**, 501–510 (2023)
3. Behfar, M.H., Björninen, T., Moradi, E., Sydänheimo, L., Ukkonen, L., et al.: Biotelemetric wireless intracranial pressure monitoring: an in vitro study. Int. J. Antennas Propag. **2015**, 1–11 (2015)
4. Benson, J., Madhavan, A., Cutsforth-Gregory, J., Johnson, D., Carr, C.: The Monro-Kellie doctrine: a review and call for revision. Am. J. Neuroradiol. **44**(1), 2–6 (2023)

5. Brasil, S., Solla, D.J.F., Nogueira, R.C., Teixeira, M.J., Malbouisson, L.M.S., Paiva, W.S.: A novel noninvasive technique for intracranial pressure waveform monitoring in critical care. J. Personal. Med. **11**(12), 1302 (2021)
6. Czosnyka, M., Pickard, J.D.: Monitoring and interpretation of intracranial pressure. J. Neurol. Neurosurg. Psychiatry **75**(6), 813–821 (2004)
7. Dan, W., Chai, W.: Application of multimodal EEG in ICP monitoring. In: Wang, X., Li, F., Pan, S. (eds.) Multi-Modal EEG Monitoring of Severely Neurologically Ill Patients, pp. 287–293. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-4493-1_12
8. Dhar, R., Sandler, R., Manwaring, K., Cosby, J., Mansy, H.: Non-invasive ICP monitoring by auditory system measurements. In: Obeid, I., Picone, J., Selesnick, I. (eds.) Signal Processing in Medicine and Biology, pp. 121–147. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-21236-9_5
9. Dhar, R., Sandler, R.H., Manwaring, K., Kostick, N., Mansy, H.A.: Noninvasive detection of elevated ICP using spontaneous tympanic membrane pulsation. Sci. Rep. **11**(1), 21957 (2021)
10. Dixon, B., et al.: Assessment of a non-invasive brain pulse monitor to measure intra-cranial pressure following acute brain injury. In: Medical Devices: Evidence and Research, pp. 15–26 (2023)
11. Geeraerts, T., et al.: Use of T2-weighted magnetic resonance imaging of the optic nerve sheath to detect raised intracranial pressure. Crit. Care **12**(5), 1–7 (2008)
12. Griffith, J., et al.: Non-invasive electromagnetic skin patch sensor to measure intracranial fluid-volume shifts. Sensors **18**(4), 1022 (2018)
13. Huang, S.M., Tofighi, M.R., Rosen, A.: Novel microwave techniques for non-invasive intracranial pressure monitoring following traumatic brain injury. In: 2014 IEEE Benjamin Franklin Symposium on Microwave and Antenna Sub-systems for Radar, Telecommunications, and Biomedical Applications (BenMAS), pp. 1–3. IEEE (2014)
14. Injecta: https://www.injecta.eu/en/
15. Kashif, F.M., Verghese, G.C., Novak, V., Czosnyka, M., Heldt, T.: Model-based noninvasive estimation of intracranial pressure from cerebral blood flow velocity and arterial pressure. Sci. Transl. Med. **4**(129), 129ra44–129ra44 (2012)
16. Kawoos, U., et al.: Embedded microwave system for monitoring of intracranial pressure. In: 2009 IEEE Radio and Wireless Symposium, pp. 119–122. IEEE (2009)
17. Kawoos, U., Meng, X., Huang, S.M., Rosen, A., McCarron, R.M., Chavko, M.: Telemetric intracranial pressure monitoring in blast-induced traumatic brain injury. IEEE Trans. Biomed. Eng. **61**(3), 841–847 (2013)
18. Kawoos, U., Meng, X., Tofighi, M.R., Rosen, A.: Too much pressure: wireless intracranial pressure monitoring and its application in traumatic brain injuries. IEEE Microwave Mag. **16**(2), 39–53 (2015)
19. Kawoos, U., Tofighi, M.R., Warty, R., Kralick, F.A., Rosen, A.: In-vitro and in-vivo trans-scalp evaluation of an intracranial pressure implant at 2.4 GHz. IEEE Trans. Microwave Theory Techn. **56**(10), 2356–2365 (2008)
20. Kim, D.Y., et al.: Comparison of ultrasonography and computed tomography for measuring optic nerve sheath diameter for the detection of elevated intracranial pressure. Clin. Neurol. Neurosurg. **204**, 106609 (2021)
21. Levinsky, A., Papyan, S., Weinberg, G., Stadheim, T., Eide, P.K.: Non-invasive estimation of static and pulsatile intracranial pressure from transcranial acoustic signals. Med. Eng. Phys. **38**(5), 477–484 (2016)

22. Lo, L., et al.: Non-invasive measurement of intracranial pressure through application of venous ophthalmodynamometry. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 6771–6774. IEEE (2021)
23. Meng, X., Tofighi, M., Rosen, A.: Digital microwave system for monitoring intracranial pressure in hydrocephalic and traumatic brain injury patients. In: 2011 IEEE MTT-S International Microwave Symposium, pp. 1–4. IEEE (2011)
24. Meng, X., et al.: Dynamic evaluation of a digital wireless intracranial pressure sensor for the assessment of traumatic brain injury in a swine model. IEEE Trans. Microw. Theory Tech. **61**(1), 316–325 (2012)
25. Mokri, B.: The Monro-Kellie hypothesis: applications in CSF volume depletion. Neurology **56**(12), 1746–1748 (2001)
26. Perez, M.D., Avetisyan, E., Mandal, B., Monorchio, A., Léwen, A., Augustine, R.: Phantom-based evaluation of a planar microwave sensor for non-invasive intracranial pressure monitoring. In: 2023 IEEE MTT-S International Microwave Biomedical Conference (IMBioC), pp. 1–3. IEEE (2023)
27. Petkus, V., Ragauskas, A., Jurkonis, R.: Investigation of intracranial media ultrasonic monitoring model. Ultrasonics **40**(1–8), 829–833 (2002)
28. Popovic, D., Khoo, M., Lee, S.: Noninvasive monitoring of intracranial pressure. Recent Patents Biomed. Eng. (Discontin.) **2**(3), 165–179 (2009)
29. Redzwan, S., et al.: Initial in-vitro trial for intra-cranial pressure monitoring using subdermal proximity-coupled split-ring resonator. In: 2018 IEEE International Microwave Biomedical Conference (IMBioC), pp. 73–75. IEEE (2018)
30. Särestöniemi, M., Singh, D., Dessai, R., Heredia, C., Myllymäki, S., Myllylä, T.: Realistic 3d phantoms for validation of microwave sensing in health monitoring applications (2024)
31. Särestöniemi, M., Sonkki, M., Myllymäki, S., Pomalaza-Raez, C.: Wearable flexible antenna for UWB on-body and implant communications. In: Telecom, vol. 2, pp. 285–301. MDPI (2021)
32. Scapaticci, R., Di Donato, L., Catapano, I., Crocco, L.: A feasibility study on microwave imaging for brain stroke monitoring. Progr. Electromagn. Res. B **40**, 305–324 (2012)
33. Solutions, T.P.: https://truephantom.com/product/adult-skull/
34. Tsivgoulis, G., Alexandrov, A.V., Sloan, M.A.: Advances in transcranial doppler ultrasonography. Curr. Neurol. Neurosci. Rep. **9**(1), 46–54 (2009)
35. Vijay, V., et al.: Using optical coherence tomography as a surrogate of measurements of intracranial pressure in idiopathic intracranial hypertension. JAMA Ophthalmol. **138**(12), 1264–1271 (2020)
36. Yang, B., Li, M., Liang, J., Tang, X., Chen, Q.: Effect of internal jugular vein catheterization on intracranial pressure and postoperative cognitive function in patients undergoing robot-assisted laparoscopic surgery. Front. Med. **10**, 1199931 (2023)
37. Zhang, X., et al.: Invasive and noninvasive means of measuring intracranial pressure: a review. Physiol. Meas. **38**(8), R143 (2017)
38. Zhong, J., Dujovny, M., Park, H.K., Perez, E., Perlin, A.R., Diaz, F.G.: Advances in ICP monitoring techniques. Neurol. Res. **25**(4), 339–350 (2003)

Detection of Intestinal Tumors Outside the Visibility of Capsule Endoscopy Camera Utilizing Radio Signal Recognition

Mariella Särestöniemi[1,2(✉)], Attaphongse Taparugssanagorn[3], Jari Iinatti[2], and Teemu Myllylä[1,4,5]

[1] Health Science and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland
mariella.sarestoniemi@oulu.fi
[2] Centre for Wireless Communications, University of Oulu, Oulu, Finland
[3] Department of ICT, School of Engineering and Technology, Asian Institute of Technology, Klong Luang, Phatum Thani, Thailand
[4] Optoelectronics and Measurements, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland
[5] Medical Research Center, Oulu, Finland

Abstract. Early cancer detection is crucial, especially for intestinal cancer with subtle early symptoms. While camera-based Wireless Capsule Endoscopy (WCE) systems are efficient, patient-friendly, and safe investigating gastrointestinal (GI) track thoroughly, some limitations persist in visualizing only the inner part of the GI regions. Our study introduces a radio channel analysis -based approach to detect intestinal/abdominal tumors which are not visible for the WCE camera, i.e., the tumors which have started to grow on the outer parts of the intestinal track. Focused on S-parameter patterns in realistic human voxel models, our simulation-based method discerns dielectric property variations in normal and tumorous tissues, replicating intricate tissue characteristics. Preliminary simulation results in different intestine locations demonstrate our technique's efficacy in differentiating normal and tumor cases based on S-parameter patterns. With a 98% accuracy rate, simple logistic regression classification model excels in distinguishing normal from tumor tissues, significantly enhancing diagnostic precision in GI health monitoring showcasing its potential to revolutionize early cancer detection and advance diagnostic accuracy within simulated human anatomy. This represents a substantial stride toward improving healthcare outcomes through cutting-edge technology.

Keywords: Early detection of tumors · Gastrointestinal monitoring · implant communications · ultra-wideband

1 Introduction

Gastrointestinal (GI) tumors encompass a diverse spectrum of lesions, ranging from benign polyps to aggressive malignancies which may represent a significant health burden globally especially in developed countries [1, 2]. Colorectal cancer ranks as the

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 426–440, 2024.
https://doi.org/10.1007/978-3-031-59091-7_28

most prevalent gastrointestinal (GI) malignancy [2] whereas small intestine cancer is rare even though small intestine forms a major part of the digestive tract [3]. Small bowl cancers, often originating in inner lining, can extend through various layers [3].

The early detection of intestinal tumors stands as a critical challenge in modern oncology, as timely intervention significantly improves patient outcomes and survival rates [2]. Traditional diagnostic modalities such as computed tomography (CT) [4], magnetic resonance imaging (MRI) [5], and conventional endoscopy with the flexible tube [6] are effective, although often pose limitations in terms of invasiveness, patient discomfort, efficiency, and potential side-effects [6].

Wireless capsule endoscopy (WCE) is attracting attention due to its simplicity and ability to comprehensively examine the gastrointestinal (GI) tract, especially the small intestine, which poses challenges for conventional endoscopy [7]. Persistent challenges include issues such as frame rate, battery life, and automated anomaly detection. Artificial intelligence (AI) assists in image recognition during GI endoscopy, enhancing screening quality and reducing unnecessary costs [8]. Convolutional neural network (CNN) based systems are employed for cancer detection, adding value to colonoscopy based colorectal screening [8].

One main challenge both with conventional endoscopy and capsule endoscopy is that tumors deeply infiltrating the intestinal wall or invading surrounding structures may not be detected since camera's field of view is limited to the mucosal surface of the gastrointestinal tract. Hence, deeper lesions may be beyond its reach. This paper proposes a novel idea of analyzing microwave radio channel between the capsule and on-body antennas which could also reveal tumors which are not visible for capsule camera.

In this paper, we propose WCE radio channel analysis-based detection of tumors which are not visible for capsule endoscopy cameras. The radio channel analysis could be used as an additional feature to the traditional WCE. To the best of our knowledge, none of the previous studies have explored the detectability of tumors, which are outside the visibility of the capsule endoscopy camera, using radio channel-based analysis. Our methodology revolves around radio signal recognition, specifically analyzing channel transfer parameters S_{N1} between the capsule and the on-body antennas, with N being the number of on-body antennas. In contrast to conventional visual-based methods, our simulation-centric approach utilizes S_{N1} patterns to discern variations in dielectric properties, such as relative permittivity and conductivity, between normal and tumorous tissues. By leveraging realistic human voxel models, our approach aims to faithfully replicate intricate tissue characteristics, providing a nuanced understanding of the subtle differences that indicate the presence of tumors.

Our initial simulation findings, conducted with a realistic voxel model, underscore the effectiveness of our method in distinguishing between normal and tumor cases based on SN1 patterns. By utilizing SN1 pattern analysis, we not only enhance precision in the WCE system but also highlight its potential to redefine strategies for early cancer detection of the tumors non-visible for capsule cameras. This advancement contributes to improved diagnostic accuracy within a simulated realistic human anatomy, paving the way for transformative developments in the field.

Our study focuses on automatically distinguishing between normal and tumorous tissues in colon and small intestine areas using SN1 patterns. Besides of analyzing differences in SN1 patterns, our further objective is to implement a classification approach by employing Logistic Regression as a straightforward yet effective method. The classification model is trained on a labeled dataset, where SN1 patterns serve as features, and corresponding labels indicate tissue status (0 for normal, 1 for tumorous). Through this approach, we aim to achieve automated discrimination, thereby advancing early cancer detection strategies.

The paper is structured as follows: Sect. 2 provides description of simulation models, dielectric properties of tumor and normal tissues, as well as details about antennas and their locations. Additionally, evaluated capsule and tumor locations are illustrated in the small intestine and colon regions. Section 3 presents the radio channel evaluations in the presence and absence of tumors in small intestine and colon areas. Section 4 presents Logistic Regression Classification results. The paper concludes with a summary and outlines the potential future works in Sect. 5.

2 Methodology

2.1 Simulation Models

The investigation employs computer systems technology (CST), an electromagnetic simulation software utilizing the finite integration technique [9]. An anatomical voxel model is utilized, and WCE-model, equipped with a dipole antenna, is strategically positioned within various segments of the small intestine in the voxel model. To establish an in-to-out wireless body area network (WBAN), a highly directive on-body antenna is integrated. Notably, the radio channel characteristics undergo significant variations depending on the placement of the on-body antenna within the intestines and its proximity to the WCE-model. Therefore, careful selection of on-body antenna locations becomes essential to ensure comprehensive coverage across the entire intestinal area.

It is important to highlight that path loss emerges as a primary constraint in the investigation. This constraint is influenced by both the distance between the capsule and the on-body antenna, and the tissues situated between them. Hence, a nuanced consideration of path loss, incorporating both distance and tissue characteristics, is crucial for the successful execution and interpretation of the investigation.

Among the voxel models provided by CST, the anatomical voxel model named Laura, depicting a middle-aged female body with a resolution of $1.87 \times 1.87 \times 1.25$ mm, is chosen due to its accurate modeling of the intestinal region including subcutaneous and visceral fat, muscles, small intestine (wall and content), and colon (large intestine). Table 1 furnishes the dielectric properties of various human body tissues at 4 GHz [10] including dielectric tumors of intestinal tumor retrieved from [11]. Table 1 also includes tissue thicknesses in capsule location A.

2.2 Antennas and Antenna Locations

In conventional WCE, the integrated camera captures images while navigating the GI tract, transmitting them to a monitoring device on the user's belt. The patient returns the

Table 1. Voxel model thickness at the selected crosscut and tissues' dielectric properties at 4 GHz.

Tissues	Thickness [mm]	Permittivity [F/m]	Conductivity [S/m]
Skin	1.4	36.6	2.34
Fat (subcutaneous)	15	5.31	0.18
Muscle	9	50.8	3.01
Fat (visceral)	4	5.31	0.18
Small intestine (wall)	8	50.8	3.11
Small intestine (content)	20	51.7	4.62
Intestinal tumor [11]	1 cm/2 cm/3 cm	57	5.2

monitoring device to the doctor the following day for image review. Nevertheless, the potential for doctors to remotely monitor real-time images presents a valuable opportunity. Our proposed idea facilitates comprehensive GI track and surrounding area monitoring through radio signal transmission. The integration of an in-to-out WBAN, 5G functionalities and a straightforward classification approach makes real-time tumor detection achievable [12, 13]. The IEEE802.15.6 standard for WBAN defines the frequency range in Ultra-Wide Band (UWB), specifically 3.1–10.6 GHz [14]. Considering propagation losses, our study utilizes the lower segment of the UWB band for capsule endoscopy (3.75–4.25 GHz).

In our research, on-body antennas are designed for 3.75–4.25 GHz, meeting IEEE802.15.6 standard requirements and falling within the 5G frequency range 3.3–4.2 GHz) used in the USA and partly in Europe. A cavity-backed low-band UWB directive antenna type (Fig. 1a) is chosen for on-body antennas due to its good directivity, aligned with IEEE 802.15.6 standard requirements [14]. Details of the antenna characteristics and its radiation patterns are presented in [15]. Five on-body antennas are used to cover small and large intestinal areas thoroughly as shown in Fig. 1b. The antennas are numbered according to their port number in the simulation model (port number 1 for the capsule antenna, port numbers 2–6 for on-body antennas), as described in our WCE channel modeling paper in [16, 17].

As a capsule model, we used a small dipole embedded inside a realistic shaped and sized capsule shell, as described in [16]. The evaluated capsule and tumor locations are presented in Fig. 2a-b in small intestine and in Fig. 2c-h in large intestine areas. Location in small intestine is named as MI00, similarly to our previous study presenting radio channel modeling with realistic models [17]. Locations in large intestines are named as "Loc. D" (Fig. 2c-d), "Loc. C" (Fig. 2e-f), and "Loc. A" (Fig. 2g-h), also similarly to [13]. These capsule locations are chosen to present different propagation conditions in abdominal area between the capsule and the on-body antenna - in terms of different thicknesses of fat and muscle tissues as well as capsule location respect to the closest on-body antennas.

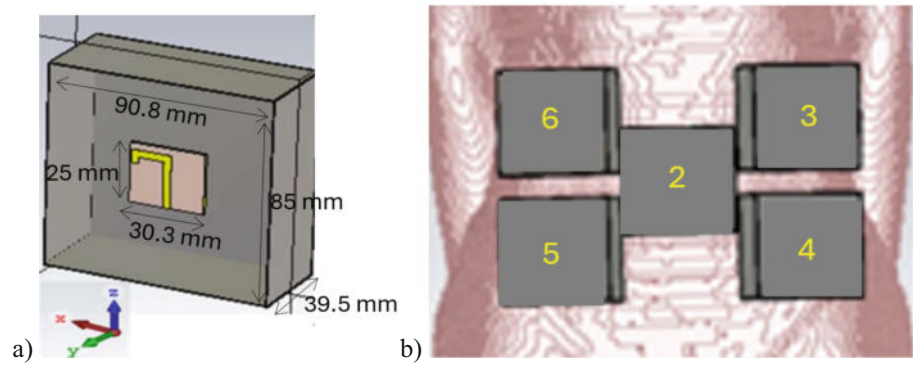


Fig. 1. a) The cavity-backed low-band UWB directive on-body antenna designed for in-body communications, b) locations of five on-body antennas [16].

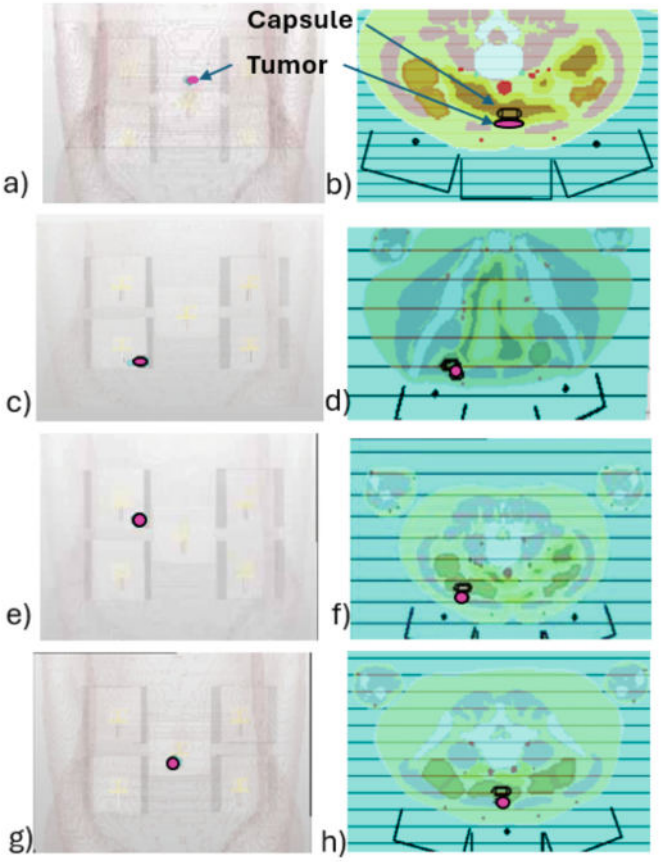


Fig. 2. The locations of tumors and capsule respect the on-body antennas and corresponding cross-section illustration a-b) Loc. MI00 (small intestine), c-d) Loc. D (colon), e-f) Loc. C (colon) and g-h) Loc. A (colon).

3 Results: Impact of the Tumors on Channel Characteristics

In this section, the channel characteristics between the capsule model and the on-body antennas are evaluated in several locations of Laura-voxel's intestine. Due to brevity, the focus is showing the S-parameters in the presence and absence of tumors especially in the colon area, where most of the GI tumors usually grow. Additionally, an example case for small intestine area is also presented to demonstrate the efficiency of the method.

3.1 Impact of the Tumor in Small Intestinal Area

Firstly, radio channel characteristics are evaluated in small intestinal tumor case, in location MI00. In this case, three tumor sizes are evaluated: with the widths 1 cm (small), 2 cm (medium), and 3 cm (large). All of them are located inside the small intestine wall without visibility to the interior of the small intestine where WCE moves.

The channel parameters between the capsule and on-body antenna S_{21}, S_{31}, S_{41}, S_{51} and S_{61} are presented in Fig. 3a-e. As can be seen, the impact of the tumors is visible in all the simulated channel parameters even with the smallest tumor. The impact of the tumor varies significantly with the frequency. The changes due to small tumor vary from 0.01–4 dB whereas due to the large tumor 0.1–25 dB.

3.2 Impact of Tumors on Channel Characteristics in Colon Area

Location D

Next, the evaluations are carried out in different locations of the colon area, first in Loc. D illustrated in Fig. 4a-b. This centralized location is demonstrated first to show how tumors which are non-visible for the capsule camera, may change channel characteristics between the capsule and all the surrounding antennas. The results for S_{21}, S_{31}, and S_{41} parameters are shown in Fig. 6a and for S_{51} and S_{61} in Fig. 6b. It is found that tumor located on the outer surface of the colon, causes clear differences in S-parameters. The smallest differences are found in the channel response, which is closest to the capsule, in this case S_{21} parameter: maximum difference is 5 dB at 5 GHz, but within the antenna's specific operational frequency range 3.75–4.25 GHz, the maximum difference is only 1 dB. Instead with S_{61}, the maximum difference within the antenna's operational range is even 7 dB. This phenomenon is due to the capsule's location exactly in the same horizontal line as the radiator of the on-body antenna 2 as well as due to the large size on the on-body antenna which captures multipath signal components from large area around the capsule. Hence the tumor which is located exactly on the front of the capsule in this scenario, does not affect significantly. However, it is assumed that with a smaller cavity, or with the same antenna without the cavity, the impact would be more significant.

Location C

Next, the channel characteristics are evaluated in Location C which is on the down right corner of the on-body antenna 6's cavity. In this case, we evaluated the impact of the tumor of two sizes 1 cm (small) and 3 cm (large). The results are presented in Fig. 5a-e. Also in this case, the tumor has an impact on the channel characteristics of all the on-body antennas. Now even the impact is larger than in the on-body antenna location D

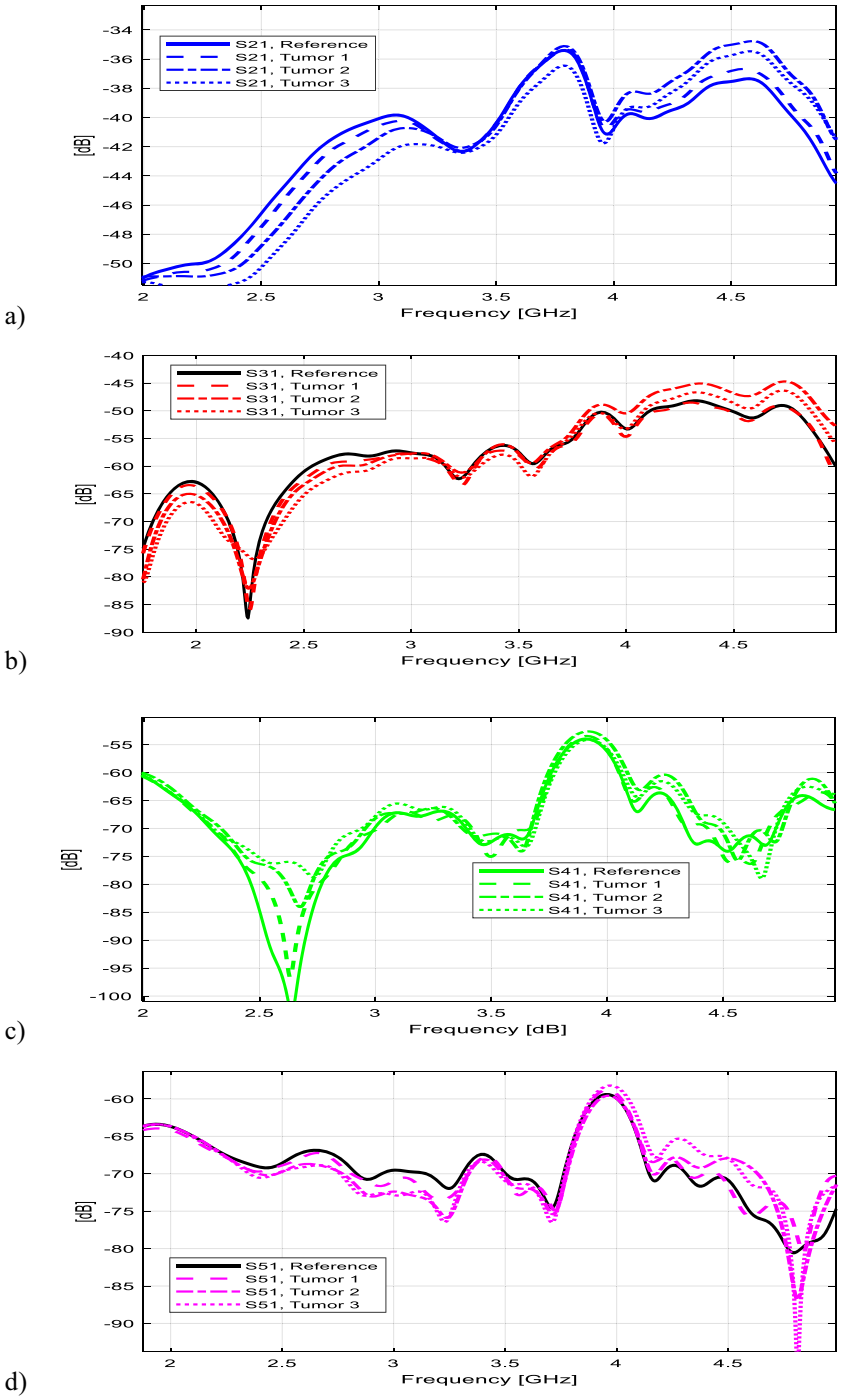
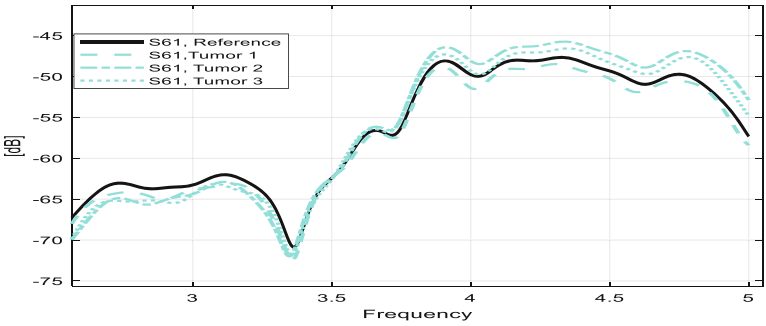


Fig. 3. Channel characteristics between the capsule and the on-body antennas in the small intestine region in the presence of tumors having sizes 1cm, 2 cm and 3cm: a) S_{21}, b) S_{31}, c) S_{41}, d) S_{51}, e) S_{61} results.



e)

Fig. 3. (*continued*)

since the capsule is not located in the middle of the antenna's cavity. The larger tumor has naturally larger impact on the channel characteristics. The maximum difference between the larger tumor and reference case is over 20 dB whereas with smaller tumor the maximum difference is few decibels.

Location A

Finally, an assessment of the channel characteristics is conducted at Location A, and the results are illustrated in Fig. 6a-c. This analysis includes the examination of capsule antenna reflection coefficients S_{11}, a commonly utilized parameter in tumor detection. The findings reveal that, in this specific case of tumor and capsule placement, tumors induce only negligible changes in the S_{11} parameters. Consequently, it is determined that S_{11} is not conducive to effective tumor detection under these circumstances. Instead, variation with channel parameters is obvious and clearly detectable: the difference is even over 20 dB in several frequencies.

It is crucial to note that manually or visually inspecting a substantial amount of data for such differences is impractical. Therefore, in the pursuit of a more systematic approach, the next subsection introduces automatic classification modeling. This approach allows for a comprehensive analysis and categorization of the data, ensuring efficient identification of patterns indicative of normal and tumor tissue conditions.

4 Logistic Regression Classification Results

Logistic regression emerges as a robust statistical model tailored for binary classification challenges, precisely aligning with our objective of distinguishing between normal tissue and tumor tissue. Comprising key components, the logistic regression model incorporates extracted features from the S_{21} as input features ($\boldsymbol{X}$), encapsulating vital channel response characteristics. The output variable ($\boldsymbol{Y}$) denotes the binary class assignment (normal or tumor). Employing the sigmoid activation function $\boldsymbol{\sigma}(z) = \frac{1}{1+\mathbf{exp}(-z)}$, the model transforms the linear combination of input features into probabilities, ensuring outputs fall within the 0 to 1 range, signifying the likelihood of belonging to the positive class (tumor) [18–20]. The linear combination, expressed as the logit function [20]

$$\hat{p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n, \tag{1}$$

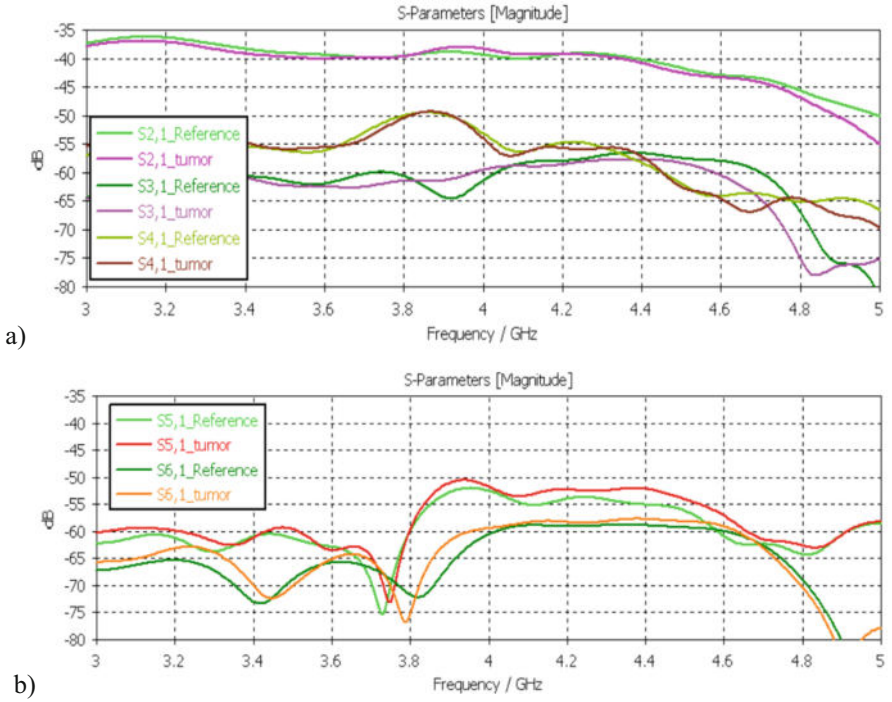


Fig. 4. S_{21}-S_{61} parameters in the presence and absence of small and large tumors and reference case in capsule and tumor locations D in colon.

computes the log-odds, with coefficients ($\boldsymbol{\beta}$) associated with input features determining the probability of the positive class.

In the training phase of logistic regression, the model learns from a labeled dataset to optimize its coefficients (β) for accurate predictions. The objective is to find the values of β that minimize the difference between the predicted probabilities and the actual class labels. This process is often achieved through optimization techniques, with two common methods being Maximum Likelihood Estimation (MLE) and gradient descent.

Maximum Likelihood Estimation (MLE): This statistical method aims to maximize the likelihood function, which measures the probability of observing the given dataset under the assumed statistical model. In logistic regression, MLE finds the set of coefficients (β) that maximizes the likelihood of observing the actual outcomes given the input features [20].

Gradient Descent: An iterative optimization algorithm, gradient descent adjusts the coefficients (β) by moving towards the minimum of the cost function. The cost function quantifies the difference between predicted probabilities and actual labels. By calculating the gradient of the cost function with respect to the coefficients, the algorithm updates β in the direction that minimizes the cost, gradually converging towards the optimal values [20].

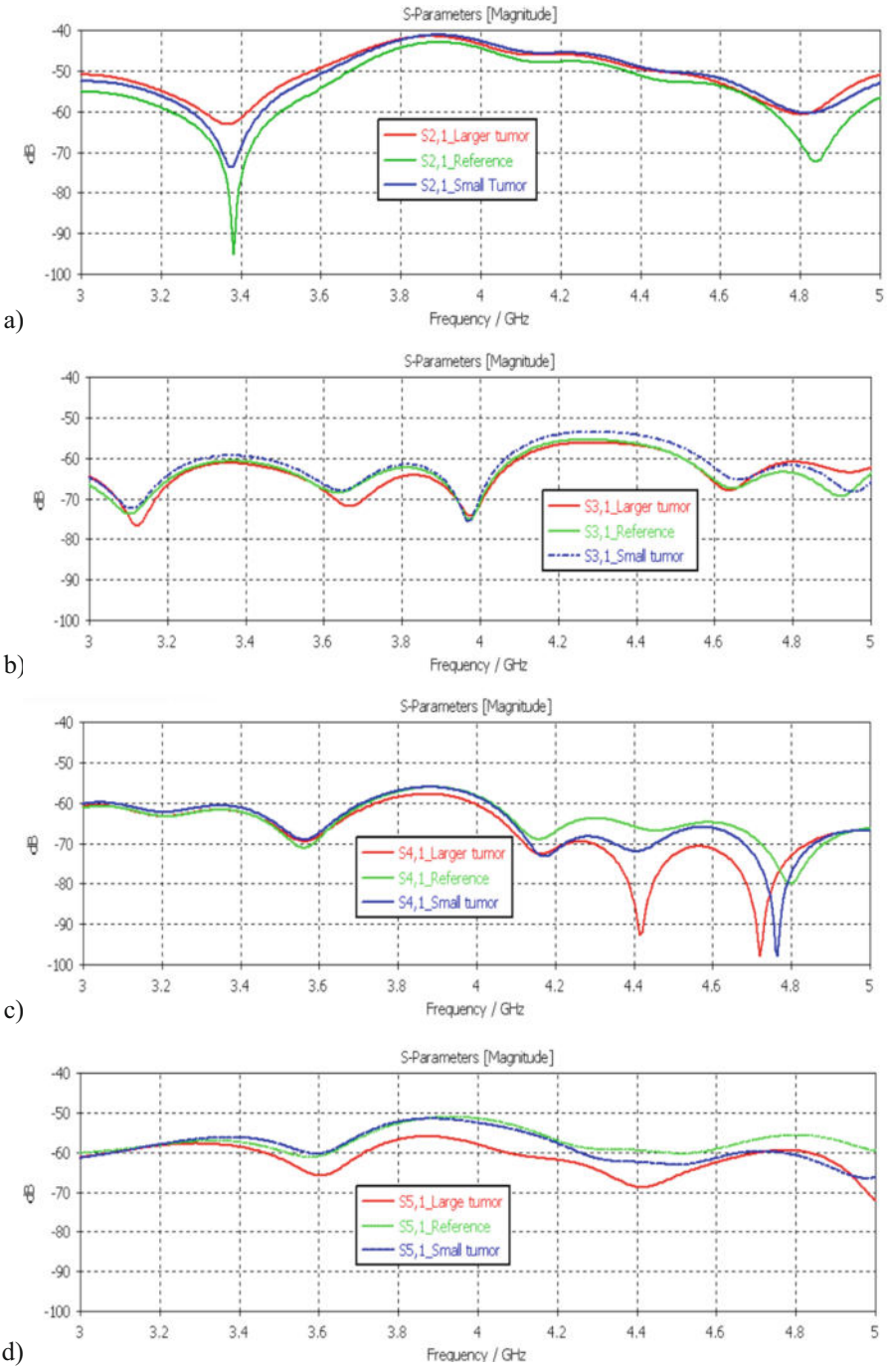
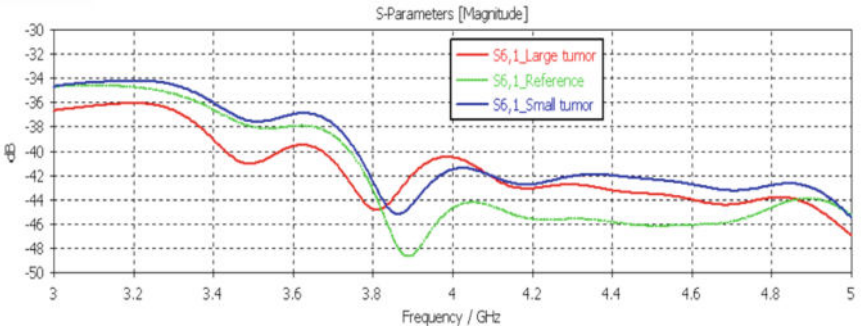


Fig. 5. S_{21}-S_{61} parameters in the presence and absence of small and large tumors and reference case in capsule and tumor locations C.



e)

Fig. 5. (*continued*)

Both MLE and gradient descent play crucial roles in refining the logistic regression model during training, ensuring it captures the underlying patterns in the data and can make accurate predictions on new, unseen examples. This iterative process of adjusting coefficients enhances the model's ability to discern between normal and tumor tissue cases based on the extracted features from S_{21}.Evaluation metrics, including accuracy, precision, recall, and F1-score, play a crucial role in assessing the proficiency of the logistic regression model in classifying normal and tumor cases. Accuracy, represented by [10].

$$Accuracy = (TP + TN)/(TP + TN + FP + FN), \tag{2}$$

gauges the overall correctness of the model's predictions, considering true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Precision expressed as [20].

$$Precision = TP/(TP + FP), \tag{3}$$

evaluates the model's ability to correctly identify positive cases among all predicted positives. Recall, denoted by [20].

$$Recall = TP/(TP + FN), \tag{4}$$

$$F1score = (2Precision\,Recall)/(Precision + Recall), \tag{5}$$

provides a balanced assessment, considering both false positives and false negatives. These metrics collectively offer a comprehensive evaluation of the logistic regression model's performance, highlighting its strengths in binary classification tasks. In summary, logistic regression emerges as a powerful and reliable solution for intestinal tumor detection, even those outside the visibility of WCE camera, excelling in simplicity, interpretability, and efficacy.

Building upon the foundations laid in the preceding sections, our investigation into Channel Frequency Response (CFR) using the S_{N1} parameter revealed compelling insights. In the CFR analysis, distinctive patterns emerged between normal and tumor

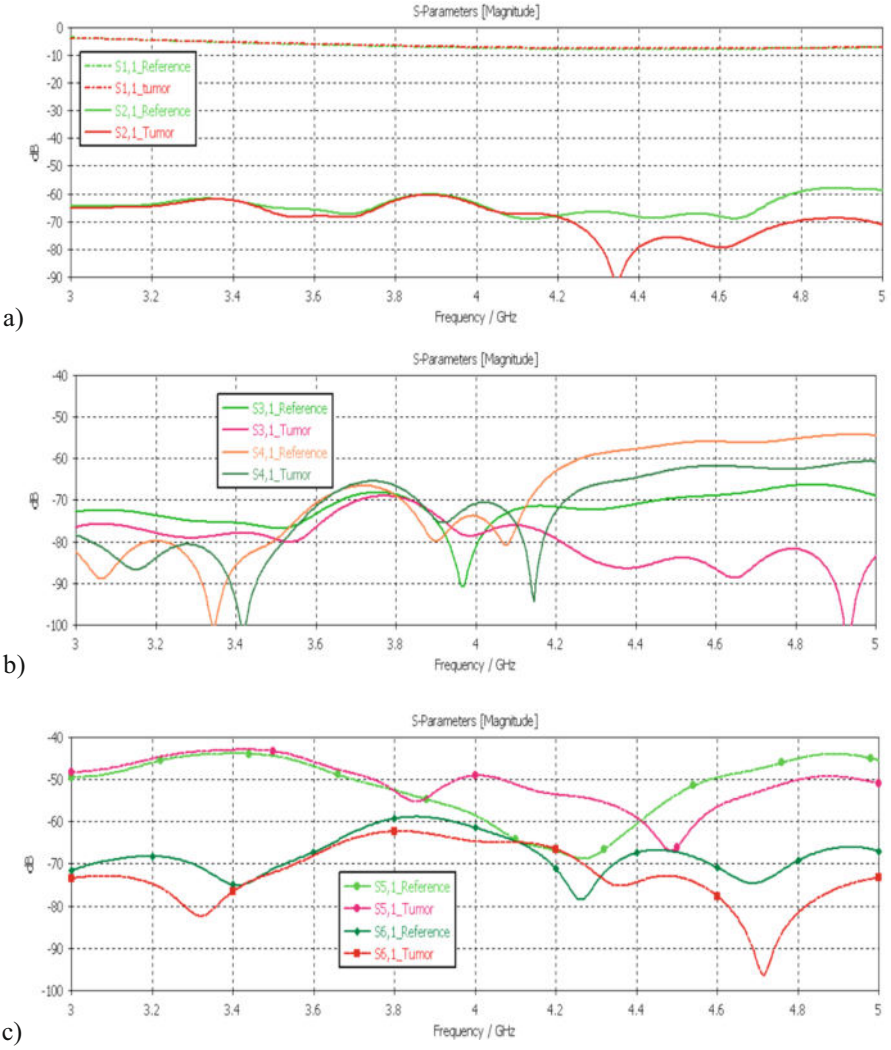


Fig. 6. a) S_{11} and S_{21} parameters, b) S_{31}, S_{41} parameters and c) S_{51} and S_{61} parameters in the presence and absence of tumors in capsule and tumor location A.

cases, affirming the capability of wireless capsule endoscopy data to capture nuanced differences in electromagnetic interactions within the small intestine. These findings set the stage for a detailed examination of the results derived from logistic regression modeling.

The application of logistic regression for intestinal tumor detection showcased excellent outcomes, achieving a remarkable accuracy of 98% with this initial data set consisting of 20 samples. The precision of the model, indicating its ability to correctly identify tumor cases among the positive predictions, stood at an outstanding 96%. Recall, reflecting the model's capacity to capture all actual tumor cases, demonstrated an impressive

rate of 97%. Additionally, the F1-score, a balanced metric considering both precision and recall, reached an excellent level of 96%. Trained on labeled datasets, the model not only displayed commendable performance across multiple metrics but also provided interpretability, enhancing our understanding of the impact of extracted features on classification decisions. Logistic regression emerged as a practical tool for our proposed methodology.

The results obtained with this initial data set, highlighted by an outstanding 98% accuracy, coupled with impressive precision (96%), recall (97%), and F1-score (96%), validate the robustness of our methodology. In addition to effectively distinguishing between normal and tumor tissues, these outcomes hold promise for advancing diagnostic accuracy in GI health monitoring. The interpretability and computational efficiency of logistic regression significantly contribute to the practicality of our approach. Future research endeavors may focus on refining the model, incorporating additional features, and expanding the dataset to enhance its clinical applicability. In essence, our findings represent a substantial advancement in the evolution of wireless capsule endoscope localization and small intestinal polyp detection.

5 Conclusions and Future Works

This study introduces a groundbreaking approach to detecting non-visible intestinal using WCE with radio channel analysis feature. Through the integration of CFR analysis and logistic regression modeling, the study achieves remarkable results. The presence of tumors effects clearly on the channel characteristics between the capsule and the closest on-body antennas although the tumors are not visible for WCE. The logistic regression model carried out to initial data set exhibits an impressive accuracy of 98%, demonstrating outstanding precision, recall, and F1-score in distinguishing between normal and tumor tissues. This achievement holds great promise for advancing diagnostic accuracy in gastrointestinal health monitoring even for the cases where tumors are not visible for capsule camera.

As a future work, we will evaluate a more comprehensive study with different tumor types, different tumor locations, as well as using voxel models having different body constitutions. Additionally, exploration of deep learning models, capitalizing on their ability to discern intricate patterns within more extensive datasets. Models like CNN or Recurrent Neural Networks (RNNs) could elevate the sophistication of our current model, potentially enhancing its performance across diverse and expansive datasets. Moreover, ongoing refinement and expansion of the dataset could significantly bolster the model's robustness. The inclusion of additional features, such as patient-specific information or real-time physiological data, has the potential to provide a more comprehensive context for improved classifications. Collaborative efforts with medical professionals to validate the model's outcomes in clinical settings would further establish its practical applicability. In essence, the success achieved in this study lays a solid foundation for future advancements in WCE localization and intestinal tumor detection. The integration of cutting-edge technologies, particularly deep learning, holds the promise of pushing the boundaries of accuracy and reliability in GI monitoring.

Acknowledgments. This research is funded by Academy of Finland Profi6 funding, 6G-Enabling Sustainable Society (University of Oulu, Finland).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Stewart, W.B.W., Christopher, P.: World cancer report from world health organization (WHO) (2014)
2. Jardim, S.R., de Souza, L.M.P., de Souza, H.S.P.: The rise of gastrointestinal cancers as a global phenomenon: unhealthy behavior or progress? Int. J. Environ. Res. Public Health **20**(4), 3640 (2023). https://doi.org/10.3390/ijerph20043640
3. Chamberlain, R.S., Krishnaraj, M., Shah, S.A.: Chapter 54: cancer of the small bowel. In: DeVita, V.T., Lawrence, T.S., Rosenberg, S.A. (eds.) DeVita, Hellman, and Rosenberg's Cancer: Principles and Practice of Oncology. 10th ed. Lippincott Williams & Wilkins, Philadelphia (2015)
4. Aaron, F.: The history, development and impact of computed imaging in Neurological Diagnosis and Neurosurgery: CT, MRI, and DTI. Nat. Proc. **1**(1), 76 (2009)
5. Haacke, E.M., Brown, R.F., Thompson, M., Venkatesan, R.: Magnetic resonance imaging: physical principles and sequence design. Wiley, New York (1999) ISBN 978–0–471–35128–3
6. Yuce, M.R., Dissanayake, T.: Easy-to-swallow wireless telemetry. IEEE Microwave Mag. **13**(6), 90–101 (2012)
7. Ciuti, G., Menciassi, A., Dario, P.: Capsule endoscopy: from current achievements to open challenges. IEEE Rev. Biomed. Eng. **4**, 59–72 (2011)
8. Gao, H., Lin, S., Yang, Y., Li, C., Yang, M.: Convolution neural network based on two-dimensional spectrum for hyperspectral image classification. J. Sens. **2018**, 13 (2018)
9. CST Microwave Studio. http://www.cst.com
10. https://www.itis.ethz.ch/virtual-population/tissue-properties/database
11. Guardiola, M., et al.: Dielectric properties of colon polyps, cancer, and normal mucosa: ex vivo measurements from 0.5 to 20 GHz. Med. Phys. **45**(8), 3768–3782 (2018)
12. Teshome, A.K., Kibret, B., Lai, D.T.H.: A review of implant communication technology in WBAN: progress and challenges. IEEE Rev. Biomed. Eng. **12**, 88–99 (2019). https://doi.org/10.1109/RBME.2018.2848228
13. Leelatien, P., Ito, K., Saito, K., Sharma, M., Alomainy, A.: Channel characteristics and wireless telemetry performance of transplanted organ monitoring system using ultrawideband communication. IEEE J. Electromagnet. RF Microwaves Med. Biol. **2**(2), 94–101 (2018). https://doi.org/10.1109/JERM.2018.2827779
14. IEEE Standard for Local and metropolitan area networks Part 15.6: Wireless Body Are Networks, pp. 1–271. IEEE (2012)
15. Kissi, C., Särestöniemi, M., Pomalaza-Raez, C., Sonkki, M., Srifi, M.N.: Low-UWB directive antenna for wireless capsule endoscopy localization. In: Proceedings of the 13th EAI International Conference on Body Area Networks (2018)
16. Särestöniemi, M., Pomalaza-Raez, C., Kissi, C., Berg, M., Hämäläinen, M., Iinatti, J.: WBAN channel characteristics between capsule endoscope and receiving directive UWB on-body antennas. IEEE Access Spec. Sect. Body Area Netw. **8**, 55953–55968 (2020)

17. Särestöniemi, M., Taparugssanagorn, A., Wisanmongkol, J., Hämäläinen, M., Iinatti, J.: Comprehensive analysis of wireless capsule endoscopy radio channel characteristics using anatomically realistic gastrointestinal simulation model. IEEE Access **11**, 35649–35669 (2023). https://doi.org/10.1109/ACCESS.2023.3263555
18. Eren, L.: Bearing fault detection by one-dimensional convolutional neural networks. Math. Probl. Eng. **2017**, 1–9 (2017). https://doi.org/10.1155/2017/8617315
19. Venna, S.R., Tavanaei, A., Gottumukkala, R.N., Raghavan, V.V., Maida, A.S., Nichols, S.: A novel data-driven model for real-time influenza forecasting. IEEE Access **7**, 7691–7701 (2018)
20. Hosmer, D.W., Lemeshow, S., Sturdivant, R.X.: Applied Logistic Regression. Wiley (2013)

Inter- and Intra-Day Precision of a Low-Cost and Wearable Bioelectrical Impedance Analysis Device

Leon Robertz[1(✉)], Lassi Rieppo[2], Seppo Korkala[2], Tommi Jaako[2], and Simo Saarakkala[1]

[1] Research Unit of Health Sciences and Technology, University of Oulu, 90014 Oulu, Finland
leon.robertz@oulu.fi
[2] Polar Electro Oy, Professorintie 5, 90440 Kempele, Finland

Abstract. Bioimpedance analysis (BIA) is a non-invasive and safe method to measure body composition. Nowadays, due to technological progress, smaller and cheaper devices allow the implementation of BIA into wearable devices. In this pilot study, we analyzed the measurement precision of a cheap BIA solution for wearable devices. Intra-session, intra-day, and inter-day reproducibility of raw impedance values from three subjects at three different body locations (hand-to-hand, hand-to-torso, torso-to-torso), and for three different frequencies (6, 54, and 500 kHz) were analyzed using the coefficient of variation (CV%). Hand-to-hand and hand-to-torso measurements resulted, on average, in high intra-session (CV% = 0.14% and CV% = 0.11%, respectively), intra-day (CV% = 1.67% and CV% = 1.26%, respectively), and inter-day (CV% = 1.53% and CV% = 1.31%) precision. Absolute impedance values for the torso-to-torso measurements showed a larger mean variation (intra-session CV% = 0.68%; intra-day CV% = 5.53%, inter-day CV% = 3.13%). Overall, this cheap BIA solution shows high precision and promising usability for further integration into a wearable measurement environment.

Keywords: Bioelectrical Impedance Analysis (BIA) · Body Composition · Repeatability

1 Introduction

Body composition, including fat mass, fat-free mass (e.g., muscle mass), and hydration status, is of interest to both the clinical and the general populations. The information from body composition measurements can be used, e.g., to predict the outcome and appropriateness of clinical interventions [1], while self-monitoring via digital health solutions, in general, has a positive influence on weight loss [2].

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 441–450, 2024.
https://doi.org/10.1007/978-3-031-59091-7_29

Several methods for measuring body composition exist, including computed tomography (CT), ultrasound, dual-energy X-ray absorptiometry, and dilution-measured total-body water. However, most of these methods are laboratory-bound, invasive, or need a considerable amount of resources [1, 3]. Bioelectrical impedance analysis (BIA) overcomes several of those limitations [4, 5]. In BIA, the impedance generated by sending weak alternating currents through the body at defined frequencies gives indirect information about the composition of the body [6]. Nowadays, with increasing technological progress, cheaper and smaller BIA solutions have entered the market making the method available to a broad audience while simultaneously improving usability. This will improve the implementation of BIA measurements into wearable devices. The prerequisite for any BIA solution within the wearable device is high precision. While the here used device's measurement accuracy between 1 and 340 kHz has previously been validated against a reference device using several circuits modeling human tissue [7], its reproducibility in human subjects has not been investigated, yet. Thus, in this pilot study, we analyzed the measurement precision regarding the reproducibility of this cheap and commercial BIA solution suitable for integration into a wearable device. Intra-session, intra-day, and inter-day reproducibility of raw impedance values from three subjects at three different body locations (hand-to-hand, hand-to-torso, torso-to-torso), and for three different frequencies (6, 54, and 500 kHz) were reported.

2 Methods

Three healthy subjects (1 female: 178cm, 68.4kg, 25 years; 2 male: 182/185 cm, 86.0/88.7 kg, 31/28 years) participated in the pilot study after they were informed about the scope of the study and gave informed written consent. We collected data over three frequencies (6, 54, and 500 kHz at 63.99 μA) using MAX30009EVKIT (Analog Devices, Inc., Wilmington, USA) for three different electrode locations (wrist-to-wrist, wrist-to-torso, and torso-to-torso), replicating typical location for wearable devices (e.g., watch, belt). The frequencies were selected to cover a wide frequency band to investigate the device's potential for bioimpedance spectroscopy, multi-frequency, and single-frequency BIA considering its capabilities, and were set up as suggested by the manufacturer. Because setting up each frequency also affected the sampling rate independently, the sampling rate between frequencies differed. The device was calibrated daily using an internal resistor of 600 Ω according to the manufacturer's guidelines.

We placed the electrode pairs (BlueSensor L, Ambu A/S, Ballerup, Denmark) on the skin with reference to defined bony landmarks (Fig. 1A). For the left wrist, the drive and sense electrodes were placed on the dorsal side of the wrist at the level of the ulnar styloid and five centimeters proximally to the drive electrode, respectively. For the right hand, the sensing electrode was placed at the wrist and the drive electrode was five centimeters distal to the sensing electrode. The sensing electrodes on the torso were placed five centimeters cranial from the anterior superior iliac spine and the driving electrodes were placed five centimeters laterally with respect to the sensing electrodes. For the wrist-to-torso measurements, we used the electrodes on the left hand and the contralateral side of the torso.

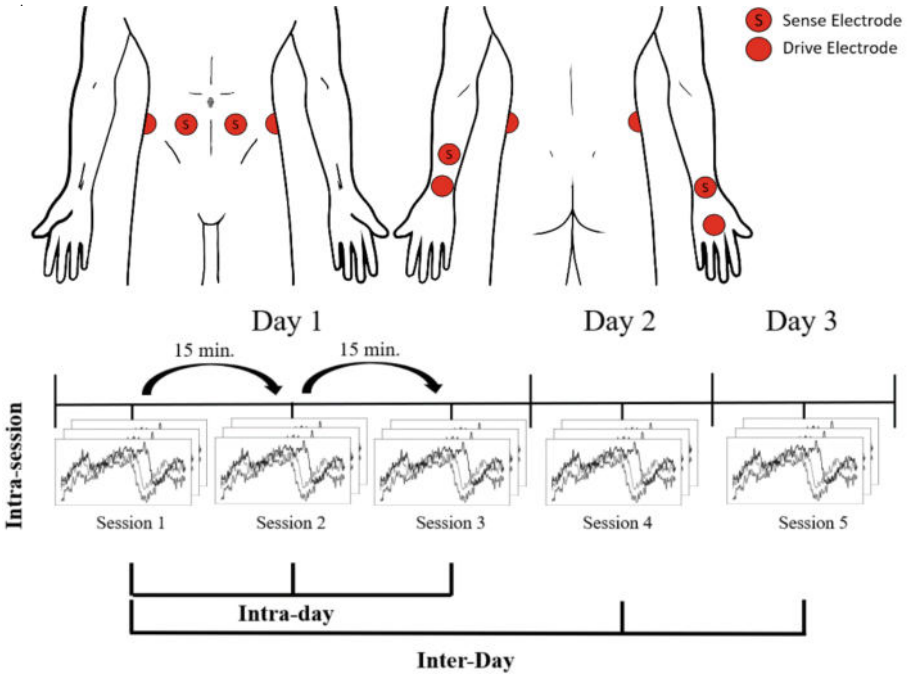


Fig. 1. Schematic drawing of the electrode locations (A) and overview of the data collection procedure (B). During each session, three consecutive measurements per measurement location (hand-to-hand, hand-to-torso, torso-to-torso) were conducted.

We collected data for the three previously defined body locations and in the named order five times in total (Fig. 1B). Within each of the five sessions, we collected three consecutive time series per frequency and location, resulting in a total of nine measurements per session. The first three sessions were conducted on the first day of data collection to measure the intra-day reproducibility. Between each session, we took a 15-min break, in which we reattached new electrodes. Further, we measured inter-day reproducibility by conducting two sessions on the two following days and comparing the data with the data from the first session. During all measurements, we followed the measurement standards by Kyle et al. [5], including e.g., data collection in the morning after overnight fasting in a supine position with abducted limbs, preparation of the skin, and exercise recommendations. Furthermore, we marked the outlines of the electrodes with a waterproof marker to ensure the electrode reattachment at the same location. To minimize the influence of breathing, subjects were told to completely exhale and keep their breath for eight seconds when collecting the data after the signal had been settled. Additionally, the subjects were told not to move during the sessions to minimize the influence of motion artifacts.

We post-processed the data using Python (V 3.11.5). In the first step, we cut and synchronized the raw signals semi-automatically (Fig. 2). To accomplish this, we plotted the raw signals and visually selected the area where the signals were settled after starting the measurement. From the beginning of the selection window, the local minimum was detected automatically, and the next 350 frames were subsequently exported. Signals were not filtered for the following analysis. To evaluate the intra-session precision, we calculated the mean coefficient of variation (CV% = standard deviation/mean*100) of all five sessions combined. Here, we first calculated the mean impedance of each of the time series. Using these values, we then calculated the CV% for each frequency and electrode location per session. The final CV% was then calculated by taking the mean values of all sessions combined. For the intra- and inter-day precision, the mean impedance for the time series of the corresponding frequencies and locations was used to calculate the overall mean and SD and consequently the CV%.

3 Results

Impedance values were the largest for measuring from one hand to another, and they decreased with increasing frequency (Table 1, Fig. 2). Overall, the female subject demonstrated higher impedance values compared to the two male subjects independent of electrode location and frequency, while the electrode position on the torso throughout all subjects and frequencies demonstrated the smallest impedance values.

The intra-session variation, regardless of electrode location and frequency, was small with CV% $\leq$ 1.44% (Table 2). While measurements from one hand to another (meanCV% = 0.14%; minCV% = 0.04% maxCV% = 0.28%) and from the hand to the torso (meanCV% = 0.11%; minCV% = 0.04% maxCV% = 0.22%) overall showed CV% $\leq$ 0.28%, larger means and standard deviations (SD) for CV% were found for the electrodes placed on both sides of the torso (meanCV% = 0.68%; minCV% = 0.24% maxCV% = 1.44%).

Regarding the intra-day and inter-day measurements (Table 3), the CV% for both hand-to-hand (intra-day: meanCV% = 1.67%; minCV% = 0.30% maxCV% = 2.69%; inter-day: meanCV% = 1.53%; minCV% = 0.75% maxCV% = 2.8%) and hand- to-torso (intra-day: meanCV% = 1.26%; minCV% = 0.77% maxCV% = 2.05%; inter-day: meanCV% = 1.31%; minCV% = 0.53% maxCV% = 1.89%) indicate good performance (CV% $\leq$ 2.8%). Overall, the torso-to-torso measurements demonstrated larger variations (intra-day: meanCV% = 5.53%; minCV% = 1.71% maxCV% = 24.64%; inter-day: meanCV% = 3.13%; minCV% = 2.20% maxCV% = 4.82%).

Table 1. Mean ± standard deviation (SD) intra-day, inter-day, and combined impedance (Ω) for each frequency and electrode location. Sub-columns represent the individual subjects. Grey shading indicates the values of the female subject.

Impedance (Ω)									
intraday	hand-to-hand			hand-to-torso			torso-to-torso		
6 kHz	670 ±11.4	423 ±1.3	350 ±9.4	311 ±2.4	191 ±1.9	167 ±3.3	35 ±1.4	25 ±0.5	20 ±4.8
54 kHz	592 ±12.0	348 ±3.9	287 ±6.2	271 ±3.1	153 ±1.8	136 ±1.9	29 ±1.5	18 ±0.3	16 ±0.6
500 kHz	530 ±10.5	299 ±3.7	245 ±4.6	242 ±4.9	131 ±1.0	116 ±1.3	24 ±0.8	15 ±0.4	12 ±0.3
inter-day									
6 kHz	643 ±10.9	415 ±6.6	336 ±3.1	304 ±2.8	189 ±3.6	165 ±2.8	35 ±1.3	24 ±0.6	22 ±0.5
54 kHz	568 ±7.0	341 ±8.7	278 ±2.1	267 ±1.4	152 ±2.7	135 ±1.8	29 ±0.8	18 ±0.5	17 ±0.7
500 kHz	509 ±6.9	292 ±8.2	239 ±2.0	237 ±1.7	129 ±1.6	115 ±1.9	24 ±0.6	15 ±0.4	13 ±0.6
combined									
6 kHz	656 ±19.8	418 ±6.3	334 ±10.7	307 ±4.6	189 ±2.7	167 ±3.5	35 ±1.2	24 ±0.5	21 ±3.8
54 kHz	581 ±16.9	343 ±7.0	284 ±6.9	269 ±3.5	152 ±2.1	135 ±2.1	29 ±1.2	19 ±0.6	16 ±0.7
500 kHz	520 ±15.0	294 ±6.9	242 ±5.1	240 ±4.4	130 ±1.4	116 ±1.8	24 ±0.7	15 ±0.5	12 ±0.5

4 Discussion

In this pilot study, we analyzed the measurement precision of a cheap and commercial BIA solution suitable for integration into a wearable device. Overall, the solution showed high precision in different body locations over a large range of frequencies and, therefore, suggests good usability for this cheap and wearable device.

Mean impedance values for the hand-to-hand and the torso-to-torso measurements ranged on average from 242–670 Ω and 12–35 Ω, respectively. Despite differences in electrode locations, these values are within the range of previous research for hand-to-hand measurements [8, 9] and for the latter setting [9]. While the extremities only account for a small fraction of the body volume, they contribute to the biggest part of the whole-body impedance contrary to the torso [10, 11]. We, therefore, expected the values for the measurement from one hand to the contralateral side of the torso to

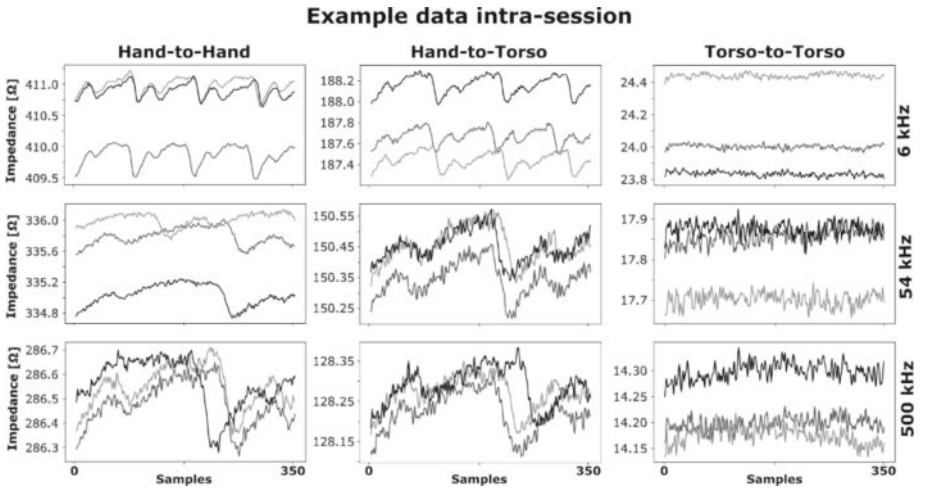


Fig. 2. Synchronized raw signals from one subject during one session. The columns represent the different electrode locations, and the rows represent the used frequencies.

lay in between the impedance values for the other two settings. In accordance with the previous literature, we found higher impedance values for the female subject independent of electrode position and frequency compared to the two male subjects [12, 13].

The high intra-session precision with CV% values smaller than 0.3% for the hand-to-hand and hand-to-torso measurements is in accordance with results from Hamilton-James et al. [14] who showed similar results comparing percentage fat mass within three measurements using the same clinical hand-to-foot device. Additionally, the intra-day (hand-to-hand-CV% = 1.67%; hand-to-torso-CV% = 1.26%) and inter-day (hand-to-hand-CV% = 1.53%; hand-to-torso-CV% = 1.31%) measurements showed, on average, similar precision as previously reported with intra-day CV% of around 1–2% and inter-day CV% of around 2–3.5% [6]. This underlines that the here-used device can achieve similar precision for those two settings compared to the earlier established and more expensive devices. We were not able to confirm previously observed higher inter-day variation for frequencies lower than 50 kHz [15] on an individual basis, which might be due to the small sample size.

Compared to hand-to-hand and hand-to-torso, the torso-to-torso measurements had comparably larger mean CV% and SD both for intra-session measurements (meanCV% = 0.68%; minCV% = 0.24% maxCV% = 1.44%), within one day (meanCV% = 5.53%; minCV% = 1.71% maxCV% = 24.64%), and between days (meanCV% = 3.13%; minCV% = 2.20% maxCV% = 4.82%). Nevertheless, intra-session and inter-day CV% on average meet similar precision as previously reported for whole-body measurements [14] and [6], respectively. Still, the large SDs for the intra-session measurements and the overall larger values for the intra-day and inter-day CV% compared to the other settings indicate that the measurements on the torso only are more prone to measurement errors than the other locations. Interestingly, we were not able to observe noticeable differences between hand-to-hand and hand-to-torso measurement CV%, even though the latter also

Table 2. Mean intra-session CV% and SD for each frequency and electrode position. The first sub-columns of each location indicate the values of the female subject.

CV% intra-session	hand-to-hand			hand-to-torso			torso-to-torso		
6 kHz	0.24 ± 0.12	0.07 ± 0.05	0.11 ± 0.05	0.15 ± 0.06	0.12 ± 0.09	0.10 ± 0.14	0.61 ± 0.56	1.44 ± 0.48	0.84 ± 1.23
54 kHz	0.28 ± 0.16	0.04 ± 0.03	0.05 ± 0.03	0.11 ± 0.12	0.06 ± 0.04	0.04 ± 0.03	0.90 ± 1.47	0.50 ± 0.2	0.24 ± 0.10
500 kHz	0.20 ± 0.19	0.17 ± 0.06	0.10 ± 0.02	0.22 ± 0.18	0.15 ± 0.04	0.08 ± 0.05	0.42 ± 0.31	0.95 ± 0.86	0.26 ± 0.16
Mean ± SD	0.24 ± 0.04	0.09 ± 0.07	0.09 ± 0.03	0.16 ± 0.06	0.11 ± 0.05	0.07 ± 0.03	0.64 ± 0.24	0.96 ± 0.47	0.45 ± 0.34

Table 3. Mean intra- and inter-day CV% for each frequency and electrode position. Sub-columns represent the individual subjects. The first sub-columns of each location indicate the values of the female subject.

CV% intra-day	hand-to-hand			hand-to-torso			torso-to-torso		
6 kHz	1.70	0.30	2.69	0.77	1.00	1.94	4.11	1.86	24.64
54 kHz	2.02	1.11	2.15	1.14	1.17	1.40	5.41	1.71	3.60
500 kHz	1.98	1.23	1.87	2.05	0.78	1.09	3.49	2.74	2.17
Mean ± SD	1.90 ± 0.17	0.88 ± 0.51	2.24 ± 0.42	1.32 ± 0.66	0.98 ± 0.20	1.48 ± 0.43	4.43 ± 0.98	2.10 ± 0.56	10.14 ± 12.6
inter-day									
6 kHz	1.70	1.59	0.92	0.91	1.89	1.72	3.59	2.29	2.20
54 kHz	1.23	2.55	0.75	0.53	1.75	1.33	2.79	2.73	4.44
500 kHz	1.36	2.80	0.85	0.71	1.27	1.66	2.57	2.71	4.82
Mean ± SD	1.43 ± 0.24	2.31 ± 0.64	0.84 ± 0.09	0.72 ± 0.19	1.64 ± 0.33	1.57 ± 0.21	2.98 ± 0.54	2.58 ± 0.25	3.82 ± 1.42

included the torso. Differences in absolute impedance due to, e.g., electrode position or small movements, therefore, seem to affect the torso more and these seem to be averaged out for longer conductors.

For the intra-day measurement, the CV% for the lowest frequency and the last subject (CV% = 24.64) was more than four times higher than the second-largest value (CV% = 5.41). We tried to explore the reason for this unexpected outlier, but measurement errors (e.g., wrong electrode placements) seem unlikely to have caused this because the other frequencies should have then been equally affected. When excluding this outlier, both the intra-day (meanCV% = 3.14 ± 1.26%) and inter-day (meanCV% = 3.24 ± 0.94%) variations were close to each other.

Several limitations must be considered when interpreting the results of this pilot study. One limitation is the small sample size which only allows us for a descriptive analysis of the results. Measurement errors and other deviations can, therefore, lead to large variations in data, which can affect the interpretation of the results. In the future, we will collect data from a larger sample size to draw stronger conclusions and to analyze if the trends observed in this study hold true. Another limitation is that we used only three frequencies in the measurements. However, the used frequencies still covered a wide range from low to high frequencies. Consequently, we assume that the device works likewise for frequencies in between. Due to limitations of the device regarding current regulations, we were not able to increase the frequency band without shifting it toward lower or higher frequencies. While the device itself gives stable outcomes for the collected frequencies with only small intra-session variation and has previously been validated against circuits [7], it still needs to be validated using subjects to determine the overall measurement accuracy of the device. Since our results are in accordance with the previous literature, we are confident that the current cheap BIA device provides

reasonable results. Finally, all analyses in this study were based on unfiltered data that were cut semi-automatically. Therefore, we expect the results to improve when filtering out the random noise and physical signals, e.g., the heart rate.

5 Conclusion

On average, in this small subject group, the intra-session, intra-day, and inter-day precision of this commercial BIA solution is high and in accordance with the established devices. This suggests good usability for this cheap and wearable device which could allow for a comprehensive integration of BIA-based body composition measurements in clinical and non-clinical settings. For the torso-to-torso measurements, those results were only met for the intra-session and inter-day measurements, while overall showing a larger variation due to one unexpectedly different measurement result. Future studies need to validate the device and to show if the here-found trends hold for a larger sample.

Acknowledgments. Co-funded by the European Union under the Marie Skłodowska-Curie Action I4WORLD 101081280. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency [REA]. Neither the European Union nor REA can be held responsible for them.

Disclosure of Interests. The research project was conducted in cooperation with Polar Electro Oy.

References

1. Price, K.L., Earthman, C.P.: Update on body composition tools in clinical settings: computed tomography, ultrasound, and bioimpedance applications for assessment and monitoring. Eur. J. Clin. Nutr. **73**, 187–193 (2019). https://doi.org/10.1038/s41430-018-0360-2
2. Patel, M.L., Wakayama, L.N., Bennett, G.G.: Self-monitoring via digital health in weight loss interventions: a systematic review among adults with overweight or obesity. Obesity **29**, 478–499 (2021). https://doi.org/10.1002/oby.23088
3. Sergi, G., De Rui, M., Stubbs, B., et al.: Measurement of lean body mass using bioelectrical impedance analysis: a consideration of the pros and cons. Aging Clin. Exp. Res. **29**, 591–597 (2017). https://doi.org/10.1007/s40520-016-0622-6
4. Jaffrin, M.Y., Morel, H.: Body fluid volumes measurements by impedance: a review of bioimpedance spectroscopy (BIS) and bioimpedance analysis (BIA) methods. Med. Eng. Phys. **30**, 1257–1269 (2008). https://doi.org/10.1016/j.medengphy.2008.06.009
5. Kyle, U.G., Bosaeus, I., De Lorenzo, A.D., et al.: Bioelectrical impedance analysis - Part II: Utilization in clinical practice. Clin. Nutr. **23**, 1430–1453 (2004). https://doi.org/10.1016/j.clnu.2004.09.012
6. Kyle, U.G., Bosaeus, I., De Lorenzo, A.D., et al.: Bioelectrical impedance analysis - Part I: review of principles and methods. Clin. Nutr. **23**, 1226–1243 (2004). https://doi.org/10.1016/j.clnu.2004.06.004

7. Crandall, H., Burt, A., Sanchez, B.: Characterization of the analog device inc (ADI) MAX30009 bioimpedance analog front end chip. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2022-July, pp. 2502–2505 (2022). https://doi.org/10.1109/EMBC48229.2022.9872000
8. Polokhin, A., Pronina, A., Boev, A., Gorbunov, S.: Validation of non-empirical fat-free mass estimation model for a wrist-worn device. J. Electr. Bioimpedance **13**, 31–38 (2022). https://doi.org/10.2478/joeb-2022-0006
9. Tanaka, N.I., Miyatani, M., Masuo, Y., et al.: Applicability of a segmental bioelectrical impedance analysis for predicting the whole body skeletal muscle volume. J. Appl. Physiol. **103**, 1688–1695 (2007). https://doi.org/10.1152/japplphysiol.00255.2007
10. Foster, K.R., Lukaski, H.C.: Whole-body impedance - What does it measure? Am. J. Clin. Nutr. **64**, 388S-396S (1996). https://doi.org/10.1093/ajcn/64.3.388s
11. Baumgartner, R.N., Chumlea, W.C., Roche, A.F.: Original estimation of body of body segments3 communications-general impedance composition. Am. J. Clin. Nutr. **50**, 221–226 (1989)
12. Dittmar, M.: Reliability and variability of bioimpedance measures in normal adults: effects of age, gender, and body mass. Am. J. Phys. Anthropol. **122**, 361–370 (2003). https://doi.org/10.1002/ajpa.10301
13. Kyle, U.G., Genton, L., Karsegard, L., et al.: Single prediction equation for bioelectrical impedance analysis in adults aged 20–94 years. Nutrition **17**, 248–253 (2001). https://doi.org/10.1016/S0899-9007(00)00553-0
14. Hamilton-James, K., Collet, T.H., Pichard, C., et al.: Precision and accuracy of bioelectrical impedance analysis devices in supine versus standing position with or without retractable handle in caucasian subjects. Clin. Nutr. ESPEN **45**, 267–274 (2021). https://doi.org/10.1016/j.clnesp.2021.08.010
15. Jansen, D.F., Korbijn, C.M., Deurenberg, P.: Variability of body density and body impedance at different frequencies. Eur. J. Clin. Nutr.Nutr. **46**, 865–871 (1992)

Experimental Study of In-Body Devices Misalignment Impact on Light-Based In-Body Communications

Syifaul Fuada[1(✉)], Mariella Särestöniemi[1,2], Marcos Katz[1], Simone Soderi[3], and Matti Hämäläinen[1]

[1] Centre for Wireless Communications, University of Oulu, 90570 Oulu, Finland
{syifaul.fuada,mariella.sarestoniemi,marcos.katz,matti.hamalainen}@oulu.fi
[2] Faculty of Medicine, Research Unit of Health Sciences and Technology, University of Oulu, 90570 Oulu, Finland
[3] IMT School for Advanced Studies, 55100 Lucca, Italy
simone.soderi@imtlucca.it

Abstract. Optical wireless communication (OWC) has emerged as a promising technology for implantable medical devices because it provides private and secure wireless links for patients, low-power consumption, and high-speed data transmission. The OWC system's receiving end typically relies on a photodetector with a limited field-of-view, necessitating direct line-of-sight connections for effective transmission. The directional nature of light-tissue interaction on the in-body communication can be problematic as the quality of the optical signal is rapidly deteriorated due to the properties of biological tissues, including scattering, absorption, and reflection, leading to a substantial loss of optical beam power reaching the photodetector's sensitive area. In this sense, any misalignment that occurs in the in-body device can directly impact the power level and further degrade the received signal quality. Numerous studies have been conducted on this topic in free-space environments; nevertheless, only a few results have been found for in-body cases. In this work, we experimentally demonstrate the impact of the in-body device misalignment on the OWC-based in-body communication system. Three cases were investigated: aligned systems, as well as lateral and angular misalignments. We considered an 810 nm Near-infrared (NIR) LED as a transmitter because the optical signal of the mentioned wavelength propagates better than other wavelengths through biological tissues. For the experiments, we used pure muscle and fat tissues with 15 mm thickness at different temperatures (23 °C and 37 °C). We also tested with thicker meat samples (30 mm, 38 mm, and 40 mm, consisting of muscle + fat layers) at 37 °C. This study adhered to ANSI.Z136.1–2007 safety standards. First, the results reveal that optical power still reaches the receiver in an aligned reference case at a meat thickness of 40 mm. Second, the in-body device misalignment significantly degrades the optical power density received, which is more pronounced under lateral than angular conditions. These misalignment effects must be carefully considered for further system enhancement when using OWC for the in-body communication system.

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 451–466, 2024.
https://doi.org/10.1007/978-3-031-59091-7_30

Keywords: Biological tissue · Optical wireless communication · in-body device misalignment · In-body communication · Near-infrared

1 Introduction

An implantable medical device (IMD) is one of the in-body medical devices that provides various benefits to patients, including real-time health monitoring and precise body treatment purposes [1–4]. For this reason, it is crucial to conduct extensive research in advancing wireless IMD systems to enhance the quality of life for patients. Light-based communication, also known as optical wireless communication (OWC), is well-suited technology for various IMDs because it facilitates energy-efficient and high-speed data communication for nerve recording and prostheses [5]. Numerous researchers have investigated OWC for in-body data transmission, and experimental evidence has confirmed its feasibility [6–13]. OWC presents an attractive option for in-body communication due to its low power consumption, typically ranging from a few microwatts to less than 10 milliwatts, even at high data rates; in contrast, conventional radio frequency (RF) requires power consumption in tens of milliwatts which higher than OWC [5]. OWC works typically under aligned connections, where the transmitter and receiver are in the same line (direct line-of-light), enhancing the protection of medical implants against unauthorized access and ensuring the patient's comfort and well-being [14]; this attribute of OWC contributes to its efficacy in addressing privacy concerns [15]. Besides, OWC offers advantages over traditional RF communication, including avoiding radio interference [16]. OWC covers light spectrums, including ultraviolet, visible light, and infrared.

OWC commonly employs a photodetector with a limited field-of-view (FOV) at the receiving end [17], which presents a significant challenge in providing seamless wireless network connectivity. In essence, OWC transmissions heavily rely on line-of-sight links [18]. Consequently, propagation through OWC channels is usually configured to be highly directional [19]. In the context of in-body communication applications, the performance of OWC links for transmitting data within the human body can be significantly degraded due to the high level of signal loss (attenuation) caused by natural phenomena in the biological tissue such as absorption, scattering, and reflection, along with the occurrence of random misalignment between the transmitting and receiving ends. In the OWC system, generating rather narrow beams for the optical links is a common practice, though their strength rapidly diminishes as it propagates across the tissue, affecting the quality of the received signal [20]. Accordingly, it is imperative to account for variations in receiver/transmitter orientation when employing OWC for the in-body communication system. The position of in-body devices might change and it creates two events of misalignment, i.e., lateral (the device is shifted) or angular (the device is tilted) from its original position due to undesirable events, for instance, an inadequately set transmitter/receiver of IMDs. The in-body device contains transmitter and receiver parts to communicate with out-body counterpart device. This changing position can significantly impact the signal quality in the OWC link. Improving OWC for in-body communication remains a prominent research subject due to the need to address certain limitations, such as how to overcome signal losses due to factors on light-biological tissue interactions and misalignment event [5]. Misalignment is one of

the crucial reasons in degrading the OWC systems' performance which will be elaborated in this paper.

Investigating the influence of receiver position and orientation is essential to understanding how the OWC system will operate in a realistic environment [17–19]. Nonetheless, there is still little research on this topic, specifically in the in-body communication context. A seminal work in [21] observed the efficacy of light-based in-body communication and found that the dependability and effectiveness of the transdermal connection are significantly affected by transmitter (out-body device) misalignment. However, the research in [21], it was carried out through simulated scenarios rather than realistic environments and it was considered as a shorter range application (transdermal application). For this study, we consider a deeper link compared to [21], where the effect of misalignment will be more pronounced than across very short links.

Figure 1 illustrates the type of receiver's misalignment in the context of OWC-based in-body communications, i.e., lateral and angular, that will be investigated in this paper. The scenario is derived from literature [22]; they focused on the transmitter part (external or out-body device), while this study will more emphasize the in-body device, thereby reasoning a novelty aspect and worth being carried out. The signal quality received by in-body device is greatly influenced by the alignment between positions of on-body and in-body devices. Nevertheless, ensuring the operation in ideal conditions is challenging due to the possibility of misalignment on the in-body device, for example inadequately set the antenna of IMDs, as addressed by various researchers done in RF technology [23–26]. To the best of our knowledge, this is the first paper that elaborates on the in-body device misalignment effect on OWC performance conducted in a realistic setting.

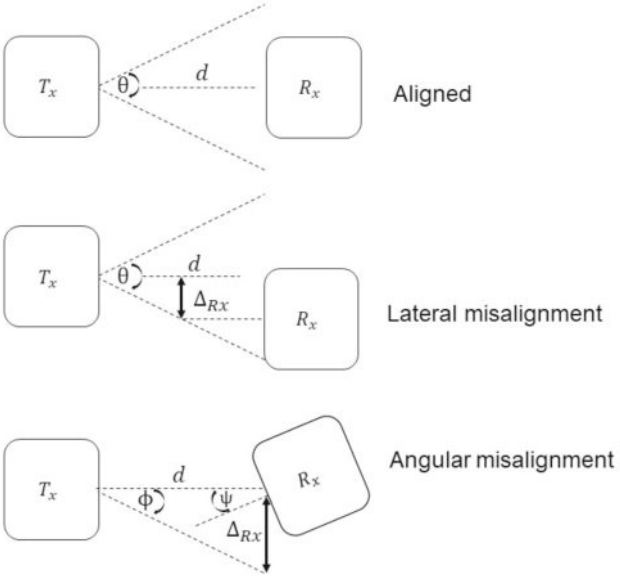


Fig. 1. In-body device's misalignment modelling, modified from [22]. The communication system consists of transmitter (T_x) and receiver (R_x) where d denotes as distance between T_x and R_x, Δ_{Rx} represents the distance of misalignment, ψ denotes a half of receiver's FOV, ϕ denotes a half of transmitter's FOV, and θ denotes transmitter's FOV.

For this reason, this paper will fill the gap by investigating the impact of misalignment on the performance of OWC links for in-body communication. Specifically, this study focuses on the angular or lateral misalignment that may occur at the in-body device in practical applications. Our experimental investigation was based on trials using *ex vivo* fresh pork meat samples (fat and muscle tissues with 15 mm thickness). Measurements were taken at two different temperatures, i.e., 23 °C and 37 °C. A Near-infrared (NIR) LED with a wavelength of 810 nm was chosen due to the light's favorable penetration capabilities in biological tissues [16]. The rationale behind presenting the results at 23 °C and 37 °C was to emphasize the significance of preheating meat samples in the measurement, as previous research mostly neglect this when they examine samples at room temperature without prior heating [27]. In addition, thicker meat samples composed of fat and muscle layers were also used (30 mm, 38 mm, and 40 mm), which were measured when the sample temperature was set at 37 °C.

Contribution: (i) This study contributes to explore the potential risk factors associated with postoperative misalignment: lateral and angular cases. (ii) We showed the importance of temperature matching to the human body (around 37 °C) for *ex vivo* experiment on light-based in-body communications. (iii) Compared to [21] that used thin pork meat samples, our study considered using thicker meat sample. (iv) fat has been found to be a good propagation channel compared to muscle for light-based in-body communications.

2 Methodology

Figure 2 shows the experimental setup in OWC-based in-body communication employed in this study to represent the in-body device's misalignment. This setup refers to Fig. 1, which encompasses three cases: aligned configuration (transmitter–receiver in the same line), angular (receiver is slightly tilted), and lateral (receiver is slightly shifted). In the aligned reference case, the transmitter was directed or exposed towards the surface of the meat sample, while the receiver was positioned precisely on the opposite side. In the angular misalignment case, the receiver was inclined at an angle of $\psi = 30°$ from its original position. While in the lateral misalignment case, the receiver was displaced by $\Delta_{Rx} = 2$ cm from its original position. In this experiment case, the transmitter is considered as out-body device, whereas receiver is considered as in-body device.

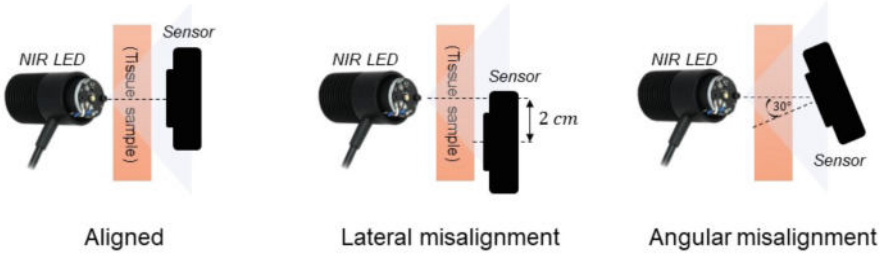


Fig. 2. Experimental setup of three different cases.

An experimental test-bed was constructed primarily using commercially available components produced by Thorlabs (Fig. 3). The test-bed comprises transmitter units (Thorlabs LED driver DC2200 and M810L3 LED modules) and a receiver unit (PM100D optical power meter and S121C optical sensor). The LED has $\theta = 80°$ (narrow beam angle) and it operates at 810 nm and is driven by 500 mA maximum current. The LED driver is fed to the LED through the provided port. The LED driver can be controlled using the provided digital display on its front panel. A constant current mode of the LED driver was used for this study. By controlling the LED driver, the input current to the LED was varied, i.e., 20% (100 mA), 40% (200 mA), 60% (300 mA), 80% (400 mA), and 100% (500 mA). The transmitted power of the LED, depending on the applied electrical current, was set to be 74.2 mW, 153 mW, 230 mW, 303 mW, and 372 mW for 100 mA, 200 mA, 300 mA, 400 mA, and 500 mA, respectively. Data modulation over a power carrier can also increase signal bandwidth, enabling higher data rates transmission [28]. However, the transmitted power must be set below a certain limit as it can create heat and damage biological tissues. In this study, the fully transmitted optical and incident power of LEDs were measured at 372 mW and 525 mW/cm^2, respectively, driven by 500 mA. It should be emphasized that the LED remains within the safe range, as it falls below the maximum permissible limit of the LED's incident power specified in the ANSI.Z136.1–2007 safety standard, which is 2W/cm^2 for a one second exposure.

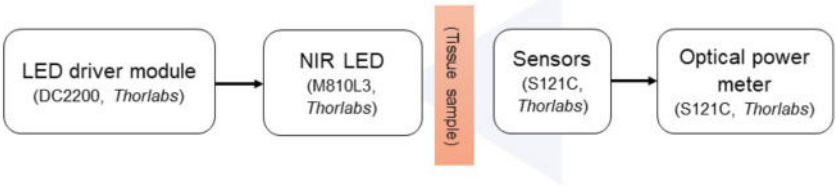


Fig. 3. Test-bed for experiment.

This study utilized two fresh pork meat samples as the optical medium, i.e., pure muscle tissue and pure fat tissue. Each sample had dimensions and thickness of approximately 50 mm × 50 mm and 15 mm, respectively (Fig. 4). When initially purchased from the market, the fresh samples were at 11 °C (measured using a thermometer). Subsequently, they were heated to 23 °C and 37 °C for measurements. To do this, we designed a small chamber heated by an off-the-shelf heater; it is plexiglass box equipped with temperature control (STC-1000) and a blower. To preserve meat sample quality from potential detrimental effects such as excessive evaporation and damage caused by high temperatures, it was imperative to carefully control and maintain their temperature below 40 °C [3]. The received optical power density of the meat sample was measured using the mentioned optical power meter. The attenuation was manually adjusted to 0 dB, and the mentioned optical power meter provides various measurement modes. However, only a single parameter, specifically power density (in W/cm^2 unit), was utilized for this study.

Figure 5 shows the pork meat samples with different thicknesses. Figure 5(a), (b), (c), are then denoted as sample #1, #2, and #3, respectively. The #sample 1 was composed by 25 mm fat + 5 mm muscle tissues. The #sample 2 is composed by 15 mm muscle +

23 mm fat; therefore, it can be denoted by more fatty tissue. The #sample 3 is composed by 20 mm muscle + 20 mm fat, that can be denoted as musculus tissue. The objective of this scenario is to clearly present the results in OWC for in-body communication, explicitly highlighting how fat is a suitable medium for propagation, similar to RF communication schemes [27, 29, 30]. Additionally, our investigation reveals that a thicker fat sample can yield a satisfactory reception power level, whereas a sample with more muscle (musculus) may not even transmit a signal, viewing from three different cases: aligned, lateral, and angular.

It is affirmed that ethical aspects are not applicable to this study, as it did not involve any human or live animal subjects. The fresh pork meats utilized in the study were procured from a local market selling various meat cuts, including those derived from pork, thereby exempting the study from being classified as an animal experiment.

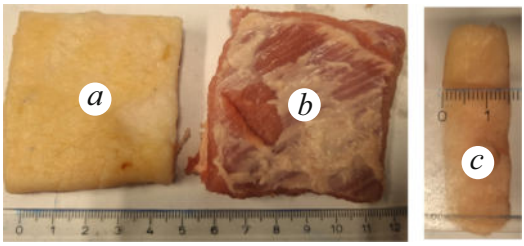


Fig. 4. Photographs of pork meat samples: (a) fat tissue, (b) muscle tissue, and its (c) thickness.

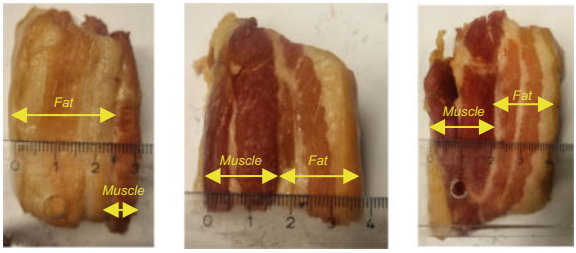


Fig. 5. Photographs of the used pork meat samples with different thicknesses: (a) 30 mm, (b) 38 mm, and (c) 40 mm.

3 Results and Analysis

3.1 Fat and Muscle Tissues Comparison

Figure 6(a) and (b) show the results of optical power comparison measurements observed on a power density scale (mW/cm^2) for fat and muscle tissue samples, respectively. The power density was measured after the NIR light passed these pork meat samples. The graph encompasses both samples under cold (23 °C) and warm (37 °C) conditions. In

top-level analysis, the amount of received power is changes linearly with the transmitted optical power: the higher the transmitted power, the higher the received power, and vice versa. On the other hand, meat samples were heated to a particular temperature close to the human body (around 37 °C), resulting in an increased transparency of the biological tissue, allowing for better light propagation through the tissue. These findings suggest that when conducting experiments involving meat samples in the context of the in-body communication system employing OWC, the meat samples should be heated or warmed to match human body temperature conditions rather than kept at room temperature because the power results are better at 37 °C where this temperature is more realistic than 23 °C. Most of the papers in the literature did not consider tissue temperature matching in which they used the temperature room. Note that in a realistic scenario, a temperature of 23 °C is basically not possible as the temperature is well below human life-sustaining temperatures. For this reason, matching to ~ 37 °C for any experiments using *ex vivo* samples should be considered.

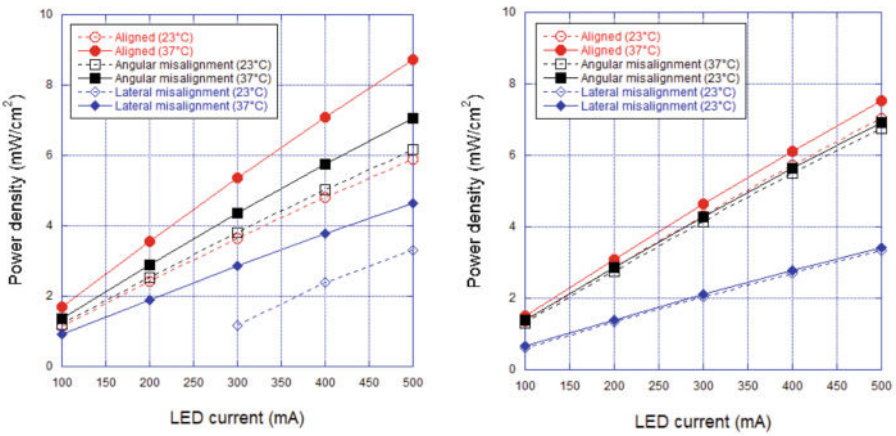


Fig. 6. Results of measurement on different tissue samples: (a) fat tissue, (b) muscle tissue.

3.2 In-Body Device's Misalignment Comparison

The results of the measurements for the aligned, angular, and lateral cases, can be seen in Fig. 7(a), (b), and (c), respectively. In the aligned scenario, the receiver is positioned perpendicular to the transmitter, resulting in the attenuated received power density being solely influenced by the natural characteristics of the biological tissue, such as absorption, scattering, and reflection. Meanwhile, in the misalignment cases, the received power density will be lower than the aligned configuration, not only due to the natural properties of the biological tissue but also affected by the imperfection of the transmitter–receiver positioning. The misalignment in the OWC's receiving end leads to a suboptimal arrangement, and it can contribute to weakening the received power density compared to the aligned link.

According to the results, the receiver experiences higher power loss in the lateral misalignment case compared to the angular misalignment one. In the lateral misalignment case, when it delineates a realistic situation, the transmitter unit (an out-body device operated by the doctors or nurses) can be adjusted manually by shifting the transmitter over the patient's skin to keep aligned with the receiver end (in-body device), establishing an aligned connection. In contrast, the angular misalignment situation is more complex than the lateral misalignment as it may necessitate surgical intervention because the receiver's position is not ideal, resulting in financial burdens and potential psychological risks for the patient. However, this measurement demonstrates that further surgery is unnecessary, as the power loss remains within acceptable limits when the device is angular misalignment approximately 30° from its original position (aligned); this situation is accepted for fat and muscle tissues with 15 mm of thickness. The FOV of the photodiode is still tolerable in this case. Further investigation should be conducted, such as varying the photodiode's angle, for instance, 45°, 60°, 75°, etc.

This study has confirmed existing literature that communication through biological tissue using OWC is feasible, allowing for secure data transmission due to its limited aligned configuration [16]. This technology can be applied to various in-body devices such as pacemakers, defibrillators, insulin pumps, brain implants, cochlear implants, etc. Nevertheless, this advantage comes with a trade-off. Supposing there are physical disruptions that can impact the positioning of the in-body device (receiver end), resulting in misalignment occurrences (lateral and angular misalignments). These misalignments subsequently lead to a decline in the signal quality received. Further, a study addressing this factor is crucial to developing a reliable receiver-end device, for instance, by incorporating the automatic gain controller feature, which has been successfully implemented in many OWC free-space scenarios [31–34]. Besides, the results of this study could also be beneficial for the in-body device's positioning idea as in the capsule endoscopy use case [23], we could suggest setting optical sensors where muscle tissue is thinnest where abdominal muscles have this "six-pack form," so that they do not have a constant thickness.

The findings depicted in Fig. 7 also confirm that fat tissue serves as a suitable propagation medium compared to muscle at approximating the human body's average temperature of 37 °C [30]. Fat tissue is more vulnerable to temperature changes than muscle tissue. The received power density measured in these three scenarios remains within a safe range as defined by the ANSI.Z136.1–2007 safety standard.

Further investigation should address the characterization of out-body device misalignment (transmitter side). It is important to differentiate whether the results are identical to those observed on the receiver side when in angular or lateral misalignment cases. Accordingly, it is advisable to conduct experiments involving shifting and tilting on both the transmitter and receiver to distinguish how far the impact of their misalignments, assuming their equivalence.

3.3 Experiments on Different Thicknesses of Meat Sample

After obtaining individual comparison data between fat and muscle layers with a thickness of 15 mm each, the subsequent experiment involved three samples with different thicknesses (i.e., 30 mm, 38 mm, and 40 mm) consisting of fat and muscle layers. To

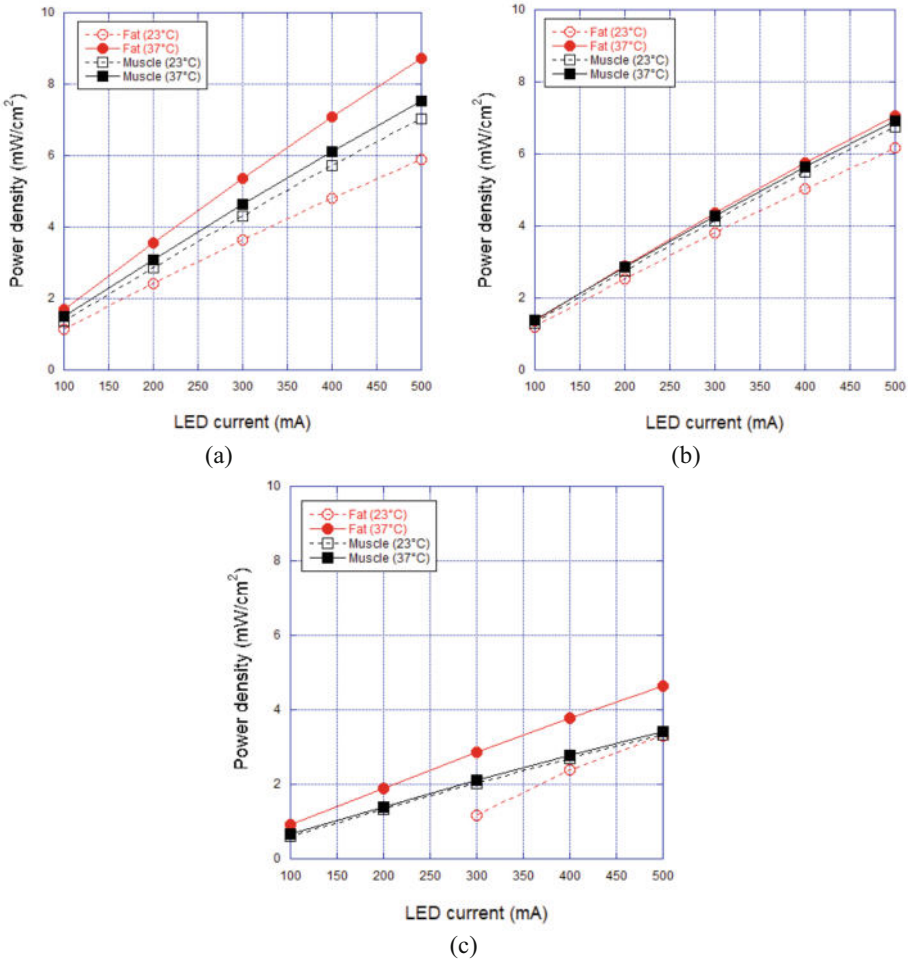


Fig. 7. Measurement results in different scenarios: (a) aligned, (b) angular, and (c) lateral.

better understand how much misalignments affects the OWC performance, measurements were also conducted in a free space channel with a separation distance of 40 mm. In alignment cases, the NIR light beam was directed towards the sensor (receiver). The receiver position is changed from the origin for misalignment cases.

As shown in Fig. 8, the maximum power density in the free space scenario (LED's driving current = 500 mA) is 63.7 mW/cm^2 in which the power density observed in the free space test remains within the safe limits outlined in the standard. The corresponding power densities for lateral and angular misalignments were 59.4 mW/cm^2 and 7.25 mW/cm^2, respectively. On average, the received power in lateral and angular misalignments amounted to 93% and 11% of the aligned situation, respectively. These findings suggest that the received power density loss in lateral misalignment is more significant than that caused by angular misalignment.

Figure 9 depicts a photograph of the experimental setup in this study, wherein an 810 nm NIR LED emits optical power to the sensor as a receiver through a pork meat sample. The sensor was used to measure the received power density. The photograph visualizes experiment on aligned reference case under 30 mm of tissue thickness.

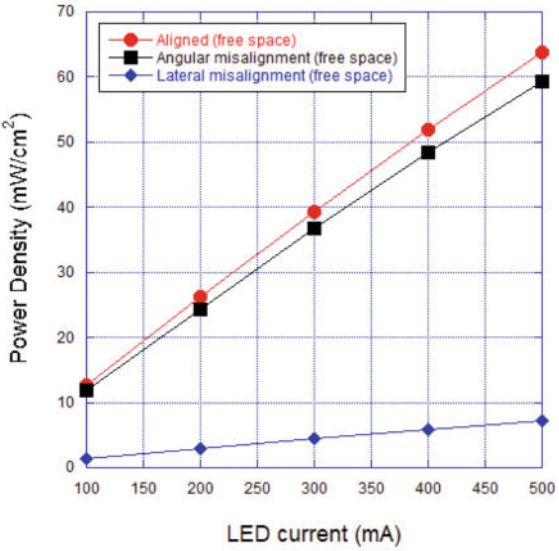


Fig. 8. Measurement results in a free space experiment (40 mm of distance).

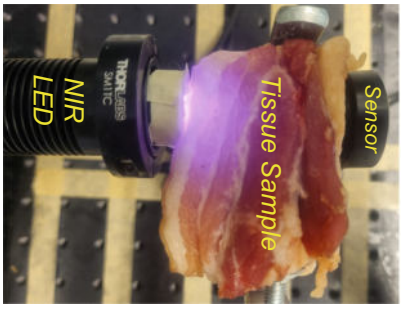


Fig. 9. A photograph of experimental setup (aligned reference case).

As shown in Fig. 10, optical power can still be received at a depth of 40 mm in the aligned cases for #samples 1, 2, and 3. However, the power received in #sample 1 is higher than #samples 2 and 3 due to a more significant proportion of fat layer composition. Conversely, #sample 3 exhibits the lowest optical power reception due to a higher percentage of muscle layer composition. Significant optical power attenuation is evident in angular misalignment for #sample 1. Moreover, #samples 2 and 3 do not receive any optical power in cases of misalignment. Accordingly, misalignment factors

should be considered when designing OWC-based in-body device system (or later can be called as optical implants) for tissue thicknesses up to 40 mm.

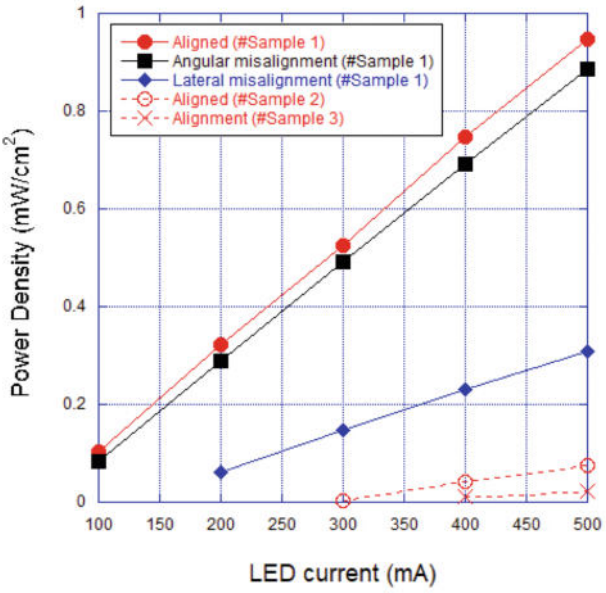


Fig. 10. Measurement results in different sample thickness.

The result shows that fat tissue has better propagation than muscle tissue for optical channels (Fig. 6), a finding consistent with observations conducted in RF case as done by [29]; where this study explored the potential use of the fat layer as a propagation medium for ultra-wideband (UWB) based medical applications through experiment and simulation approaches. The fat layer demonstrated less decreased RF signal loss than other tissues under investigation [29]; RF waves propagated through the fat tissue from the abdomen to the back of an individual, with a power loss of 60 dB.

By seeing overall measurements, we have clearly investigated the aligned, lateral, and angular misalignments in pure muscle and fat tissues at different temperatures and varying meat thicknesses.

3.4 Limitations of the Study

In this study, we have explored the impact of misalignment in in-body OWC systems using an 810 nm NIR LED transmitter on *ex-vivo* testing. We only focus on specific scenarios and conditions, such as postoperative misalignments, temperature matching, and tissue thickness. The experiment only used two samples (fat and muscle) with a thickness of 15 mm each, which may oversimplify the complexity of in-body tissue environments. On the other hand, concerns may arise regarding the generalizability of the findings to diverse clinical or real-world settings, potentially limiting the broader

applicability of the results. Future studies should consider more thickness of meat sample in order to capture the full spectrum of conditions encountered in actual clinical scenarios, and using meat temperatures ranging 36 – 40 °C to match the average of human body. Therefore, the results will be more realistic. Our investigation only justified the choice of the 810 nm peak wavelength for its penetration capabilities. Hence, concerns are raised about the exclusive focus on a single wavelength. In future, we can employ NIR LED with different wavelength ranging NIR I (λ = 700–900 nm) and NIR II (λ = 1000–1700 nm) windows [35]. One potential broadband NIR LED that can be used further experiments including MBB1L3 (λ = 470–850 nm), MBB2L1 (λ = 770 nm, 860 nm, and 940 nm), MBB2LP1 (λ = 770 nm, 860 nm, and 940 nm), which are provided by Thorlabs. The last limitation of this study is the experiment relies on static conditions and may not account for real-time biological dynamics, such as movement or deformation of tissues during normal bodily activities. These aspects should be considered in the future works, which is involve dynamics situation.

The initial findings of this study also highlight the significance of considering the losses incurred from misalignment when designing robust OWC systems. Moreover, it is recommended to incorporate a digital system for subsequent analysis. It transmits data in bitstreams to determine the threshold at which the optical communication link can still be reliable while considering the optical signal losses resulting from misalignment. This approach takes into account not only the allowable limit of received power as adhered by ANSI.Z136.1–2007 safety standard, but also acknowledges the trade-off between received power and sensitivity. Furthermore, future investigation should consider the impact of misalignment on the OWC link by analyzing parameters such as throughput, signal-to-noise ratio (SNR), bit-error rate (BER), and other quality of service (QoS) indicators.

Previous studies have demonstrated that optical communication links can still be maintained with a tissue thickness up to 40 mm in aligned position under a received power of tens μW [16]. Additionally, other studies have revealed that OWC system can be demonstrated at extremely low-intensity levels with the communication speed trade-off [36]. Based on prior research, it is hypothesized that communication can persist even though misalignment at the in-body device occurs, with a trade-off resulting in a decrease in the wireless data communication speed. However, there is a threshold of communication link loss due to excessive misalignments where the optical signal received is very weak in which we will address it in future studies.

4 Conclusion

This study has investigated the in-body device misalignment impact on the performance of OWC-based in-body communication in different realistic scenarios. The experiment utilized a NIR LED with a wavelength of 810 nm, as it is known to have better penetration capabilities through biological tissue than other wavelengths. Two samples (fat and muscle) with a thickness of 15 mm each were used in the experiment at a temperature of 23 °C. The meat sample was also heated to a temperature closer to the average human body, 37 °C, for comparison. The study evaluated three cases of misalignment on the receiver side: aligned (considered the ideal or baseline condition), angular, and lateral

misalignments. Experiments on different thicknesses of meat samples were conducted carefully as well. The results showed that a meat sample with a fatty layer has the potential to achieve a desirable level of reception power density, as evidenced in #sample 1 and #sample 2, while a sample with a higher proportion of muscle does not possess the capability to transmit a signal properly, even though it is an aligned case, as proved in #sample 3.

The findings indicate that a misalignment situation on the in-body device point-of-view can negatively impact the performance of OWC for an in-body communication system, as the light that propagates through biological tissue may not reach the photodetector's sensitive area on the in-body device due to limited FOV. Furthermore, the signal quality received in the lateral misalignment case was poorer than in the angular misalignment case, primarily due to decreased received power density. Future studies will consider the tissue thickness, misalignments in the transmitter side, combination of lateral – angular misalignment, and practical methods to find alignment position.

Acknowledgments. This study was funded by the University of Oulu's Infotech (CWC-NS 2406124). The authors would like to acknowledge to the Research Council of Finland (former Academy of Finland) Profi6 funding, 6G-Enabling Sustainable Society (6GESS) the University of Oulu and the Research Council of Finland 6G Flagship Programme (Grant Number: 346208).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Fuada, S., Särestöniemi, M., Katz, M.: Analyzing the trends and global growth of energy harvesting for implantable medical devices (IMDs) research – a bibliometric approach. Int. J. Online Biomed. Eng. **20**, 115–135 (2024). https://doi.org/10.3991/ijoe.v20i03.45681
2. Fuada, S., Ma, G., Katz, M.: Global growth and trends of in-body communication research – insight from bibliometric analysis. Int. J. Online Biomed. Eng. **20**, 128–149 (2024). https://doi.org/10.3991/ijoe.v20i01.44967
3. Seo, J., et al.: Wireless electrical power delivery using light through soft skin tissues under misalignment and deformation. Adv. Mater. Interfaces **9**, 2102586 (2022). https://doi.org/10.1002/admi.202102586
4. Fuada, S., Särestöniemi, M., Katz, M.: Analyzing emerging trends in wireless implantable medical devices (IMDs): a bibliometric study. Int. J. Online Biomed. Eng. **20**, 115–143 (2024). https://doi.org/10.3991/ijoe.v20i04.46559
5. Varotsos, G.K., Nistazakis, H.E., Aidinis, K., Jaber, F., Rahman, K.K.M.: Transdermal optical wireless links with multiple receivers in the presence of skin-induced attenuation and pointing errors. Computation. **7**, 33 (2019). https://doi.org/10.3390/computation7030033
6. Ahmed, I., Bykov, A., Popov, A., Meglinski, I., Katz, M.: Optical wireless data transfer through biotissues: practical evidence and initial results. In: Mucchi, L., Hämäläinen, M., Jayousi, S., Morosi, S. (eds.) Body Area Networks: Smart IoT and Big Data for Intelligent Health Management: 14th EAI International Conference, BODYNETS 2019, Florence, Italy, October 2-3, 2019, Proceedings, pp. 191–205. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-34833-5_16

7. Ahmed, I., Bykov, A., Popov, A., Meglinski, I., Katz, M.: Wireless data transfer through biological tissues using near-infrared light: testing skull and skin phantoms. In: Neural Imaging and Sensing 2020, pp. 50–54. SPIE, San Francisco, California, United States (2020). https://doi.org/10.1117/12.2545221
8. Alizadeh, H., Koolivand, Y., Sodagar, A.M.: Pulse-based, multi-beam optical link for data telemetry to implantable biomedical microsystems. In: 2022 20th IEEE Interregional NEWCAS Conference (NEWCAS), pp. 529–532. IEEE, Quebec City, QC, Canada (2022). https://doi.org/10.1109/NEWCAS52662.2022.9842007
9. De Marcellis, A., Palange, E., Faccio, M., Di Patrizio Stanchieri, G., Constandinou, T.G.: A 250Mbps 24pJ/bit UWB-inspired optical communication system for bioimplants. In: 2017 IEEE Biomedical Circuits and Systems Conference (BioCAS), pp. 1–4 (2017). https://doi.org/10.1109/BIOCAS.2017.8325081
10. De Marcellis, A., Stanchieri, G.D.P., Faccio, M., Palange, E., Constandinou, T.G.: A 300 Mbps 37 pJ/bit pulsed optical biotelemetry. IEEE Trans. Biomed. Circuits Syst. **14**, 441–451 (2020). https://doi.org/10.1109/TBCAS.2020.2972733
11. Kayani, M., Ahmed, I., Amila Perera, Bykov, A., Katz, M.: A proof of concept for in-body implants for longevity and selfcare. In: Mejías, E., Kouri, P., Ahonen, O., and Reponen, J. (eds.) The 26th Finnish National Conference on Telemedicine and eHealth, p. 57. Finnish Society of Telemedicine and eHealth, Oulu Finland (2021)
12. Sohn, I., Rahman, M.F., Jang, Y.H., Lee, S.H.: An optical implant for biotelemetry: design, in vivo verification, and challenges. IEEE Commun. Mag.Commun. Mag. **60**, 50–56 (2022). https://doi.org/10.1109/MCOM.001.2100784
13. Sohn, I., Jang, Y.H., Lee, S.H.: Ultra-low-power implantable medical devices: optical wireless communication approach. IEEE Commun. Mag.Commun. Mag. **58**, 77–83 (2020). https://doi.org/10.1109/MCOM.001.1900609
14. Halder, S.: Measurements and characterization of optical wireless communications through biological tissues (2020). http://jultika.oulu.fi/Record/nbnfioulu-202007042732
15. Halder, S., Särestöniemi, M., Ahmed, I., Katz, M.: Providing connectivity to implanted electronics devices: experimental results on optical communications over biological tissues with comparisons against UWB. In: Alam, M.M., Hämäläinen, M., Mucchi, L., Niazi, I.K., Le Moullec, Y. (eds.) Body Area Networks. Smart IoT and Big Data for Intelligent Health: 15th EAI International Conference, BODYNETS 2020, Tallinn, Estonia, October 21, 2020, Proceedings, pp. 3–17. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-64991-3_1
16. Ahmed, I., Halder, S., Bykov, A., Popov, A., Meglinski, I.V., Katz, M.: In-body communications exploiting light: a proof-of-concept study using ex vivo tissue samples. IEEE Access. **8**, 190378–190389 (2020). https://doi.org/10.1109/ACCESS.2020.3031574
17. Abdalla, I., Rahaim, M., Little, T.: Impact of receiver FOV and orientation on dense optical networks. In: 2018 IEEE Global Communications Conference (GLOBECOM), pp. 1–6. IEEE, Abu Dhabi, United Arab Emirates (2018). https://doi.org/10.1109/GLOCOM.2018.8647842
18. Le Bas, C., Sahuguede, S., Julien-Vergonjanne, A., Behlouli, A., Combeau, P., Aveneau, L.: Impact of receiver orientation and position on visible light communication link performance. In: 2015 4th International Workshop on Optical Wireless Communications (IWOW), pp. 1–5. IEEE, Istanbul, Turkey (2015). https://doi.org/10.1109/IWOW.2015.7342254
19. Eroğlu, Y.S., Yapıcı, Y., Güvenç, İ: Impact of random receiver orientation on visible light communications channel. IEEE Trans. Commun.Commun. **67**, 1313–1325 (2019). https://doi.org/10.1109/TCOMM.2018.2879093
20. Fuada, S., Pratama, A., Adiono, T.: Analysis of received power characteristics of commercial photodiodes in indoor LOS channel visible light communication. Int. J. Adv. Comput. Sci. Appl.Comput. Sci. Appl. **8**(7), 2017 (2017). https://doi.org/10.14569/IJACSA.2017.080722

21. Trevlakis, S.E., Boulogeorgos, A.-A.A., Karagiannidis, G.K.: On the impact of misalignment fading in transdermal optical wireless communications. In: 2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST), pp. 1–4. IEEE, Thessaloniki, Greece (2018). https://doi.org/10.1109/MOCAST.2018.8376613
22. Ghassemlooy, Z., Alves, L.N., Zvánovec, S., Khalighi, M.A. (eds.): Visible Light Communications: Theory and Applications. CRC Press (2017). https://doi.org/10.1201/9781315367330
23. Blauert, J., Kiourti, A.: Quarter-wave plates to improve rotational misalignment robustness in medical telemetry. Bioelectromagnetics **42**, 583–592 (2021). https://doi.org/10.1002/bem.22365
24. Waldeck, S., et al.: New ultra-fast algorithm for cochlear implant misalignment detection. Eur. J. Radiol.Radiol. **151**, 110283 (2022). https://doi.org/10.1016/j.ejrad.2022.110283
25. Wang, Q., Che, W., Mongiardo, M., Monti, G.: Wireless power transfer system with high misalignment tolerance for bio-medical implants. IEEE Trans. Circuits Syst. II Express Briefs **67**, 3023–3027 (2020). https://doi.org/10.1109/TCSII.2020.2985056
26. Wei, P., Li, J., Jiao, X., Yu, Z., Song, H.: Short-term clinic observation of misalignment and rotational stability after implantable collamer lens implantation. Graefes Arch. Clin. Exp. Ophthalmol. Arch. Clin. Exp. Ophthalmol. **261**, 1473–1481 (2023). https://doi.org/10.1007/s00417-022-05929-7
27. Särestöniemi, M., Pomalaza-Raez, C., Kissi, C., Iinatti, J.: On the UWB in-body propagation measurements using pork meat. In: Alam, M.M., Hämäläinen, M., Mucchi, L., Niazi, I.K., Le Moullec, Y. (eds.) Body Area Networks. Smart IoT and Big Data for Intelligent Health: 15th EAI International Conference, BODYNETS 2020, Tallinn, Estonia, October 21, 2020, Proceedings, pp. 18–33. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-64991-3_2
28. Poon, A.S.Y., O'Driscoll, S., Meng, T.H.: Optimal frequency for wireless power transmission into dispersive tissue. IEEE Trans. Antennas Propag.Propag. **58**, 1739–1750 (2010). https://doi.org/10.1109/TAP.2010.2044310
29. Särestöniemi, M., et al.: Fat in the abdomen area as a propagation medium in WBAN applications. In: Mucchi, L., Hämäläinen, M., Jayousi, S., Morosi, S. (eds.) Body Area Networks: Smart IoT and Big Data for Intelligent Health Management: 14th EAI International Conference, BODYNETS 2019, Florence, Italy, October 2-3, 2019, Proceedings, pp. 175–187. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-34833-5_15
30. Särestöniemi, M., Pomalaza-Ráez, C., Kissi, C., Iinatti, J.: Simulation and measurement data-based study on fat as propagation medium in WBAN abdominal implant communication systems. IEEE Access. **9**, 46240–46259 (2021). https://doi.org/10.1109/ACCESS.2021.3068116
31. Adiono, T., Fuada, S., Pradana, A.: A circuit for robust visible light communication systems in indoor environment. In: 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), pp. 1–5. IEEE, Bali, Indonesia (2018). https://doi.org/10.1109/ICITEED.2018.8534829
32. Fuada, S., Adiono, T., Pradana, A.: Employing LM13700 as AGC for mobile visible light communication system. IJEETC **9**, 88–93 (2020). https://doi.org/10.18178/ijeetc.9.2.88-93
33. Ibrahim, M.H., Irfani, M., Adriyanto, F., Maghfiroh, H., Ramelan, A.: Design and performance evaluation of visible light communication analog front-end (AFE) receiver using IC LM741. In: 2020 3rd International Conference on Information and Communications Technology (ICOIACT), pp. 509–512. IEEE, Yogyakarta, Indonesia (2020). https://doi.org/10.1109/ICOIACT50329.2020.9332064

34. Wu, H.-K., Kang, J.-W., Yang, J.-C., Wu, B.-C., Tang, Z.-M.: Implementation of visible light communication system with auto gain control. In: 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), pp. 100–101. IEEE, Osaka, Japan (2019). https://doi.org/10.1109/GCCE46687.2019.9015589
35. Shen, Q., Wang, S., Yang, N.-D., Zhang, C., Wu, Q., Yu, C.: Recent development of small-molecule organic fluorophores for multifunctional bioimaging in the second near-infrared window. J. Lumin.Lumin. **225**, 117338 (2020). https://doi.org/10.1016/j.jlumin.2020.117338
36. Tian, Z., Wright, K., Zhou, X.: The darkLight rises: visible light communication in the dark. In: Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, pp. 2–15. Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2973750.2973772

Study on Fat as the Propagation Medium in Optical-Based In-Body Communications

Syifaul Fuada[1](✉), Mariella Särestöniemi[1,2], Marcos Katz[1], Simone Soderi[3], and Matti Hämäläinen[1]

[1] Centre for Wireless Communications, University of Oulu, 90570 Oulu, Finland
{syifaul.fuada,mariella.sarestoniemi,marcos.katz,matti.hamalainen}@oulu.fi
[2] Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, 90570 Oulu, Finland
[3] IMT School for Advanced Studies, 55100 Lucca, Italy
simone.soderi@imtlucca.it

Abstract. This paper investigates fat tissue as a medium for communication in implantable/ingestible medical device (IMD) systems based on optical wireless communication (OWC). The findings emphasize the importance of tissue characteristics (temperature in particular) for optimizing OWC performance. This study considered Near-infrared (NIR) light with 810 nm wavelength and fresh porcine samples to mimic the human tissue. The study employs a realistic measurement approach in an *ex vivo* setting using various porcine samples: pure fat and flesh tissues and samples with different thicknesses. This study also investigates the influence of porcine temperature on the optical communication channels, which are measured by comparing the received optical power at 23 °C and 37 °C. In general, tissue samples at warmer temperatures (37 °C) receive higher optical power than colder samples. The results also demonstrate the superior optical power transmission capabilities of pure fat compared to pure flesh in porcine tissue samples in warm conditions. We also found that porcine with multiple layers of fat (fatty sample) yields higher received optical power than porcine with multiple layers of flesh (muscular). The results of this study provide valuable insights and relevant considerations for OWC-based in-body communication conducted using porcine samples.

Keywords: Fat · Flesh · Porcine Sample · Optical Wireless Communication · Tissue Temperature · In-body Communications

1 Introduction

Medical in-body devices, or in the most literature called as implantable/ingestible medical devices (IMD), such as implants, smart pills, and biosensors, play an essential role in diagnosing, treating, and monitoring various clinical conditions of patients. The incorporation of wireless communication within these devices serves crucial functions, such as transmitting data, regulating device operations, and facilitating immediate communication with healthcare professionals [1]. Optical wireless communication (OWC) is

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 467–479, 2024.
https://doi.org/10.1007/978-3-031-59091-7_31

emerging as a viable technology for in-body communication cases alongside established radio frequency (RF) and ultrasound technologies [2]. Numerous studies indicate that OWC offers advantages in the context of the in-body application, including its capacity to provide a substantial security level as the communication range is limited to a few millimeters [3], low power consumption [4, 5], electromagnetic interference-free operation, avoiding RF emission which might cause harm to tissues (if high transmission power or long exposure times are used) [6]. On the other hand, OWC has the ability to enable high-speed data transmission [7–9], making it an attractive option for the sixth generation (6G) of communication technology [10]. OWC also facilitates simultaneous energy and data transmission for IMD [11].

Having a thorough understanding of the optical channel's characteristics is essential to achieve a seamless design of in-body communication systems. It is widely acknowledged that signal propagation differs among various tissues, primarily due to variations in their optical properties [12]. In the context of the RF use case, research has shown that fat tissue offers the most advantageous conditions for signal propagation in terms of velocity and loss [12]. Significant studies have specifically explored the potential of fat tissue as a medium for medical applications; these cases were based on RF waves [13–16]; these seminal works have verified fat tissue's feasibility through simulation and measurement studies.

When brought to OWC, measurements conducted on anaesthetized animals yield the most realistic results as the optical properties of tissues begin to alter immediately after the animal's death. Nevertheless, conducting these measurements is challenging due to the need for a hospital environment with a strict clinical procedure. As an alternative, measurement using porcine samples are more practical to do. Porcine sample is frequently utilized in in-body communication studies, as it has similar optical properties to those of human beings [17]. There are specific considerations when using porcine samples: tissue temperature and tissue composition. The tissue temperature can affect its properties, thus influencing the characteristics of the optical channel within the body. The composition of fat and flesh layers in the porcine sample is believed to have a significant impact on the results, as fat is known to be a better conductive medium for signal transmission [17].

Despite the fat layer has been explored by various researchers on the RF domain, to the best of the authors' knowledge, no studies have been conducted that exploit optical channel characteristics, especially on porcine samples containing fat and flesh composition with realistic measurement. In addition, there is a lack of research examining the influence of porcine temperature on optical channel characteristics.

The objective of this study is to investigate the factors above: (*i*) The optical channel characteristics are assessed using porcine: pure fat and flesh tissue samples and samples with flesh and fat layers. (*ii*) The effect of the porcine sample's temperature on the received power. The specifications of the porcine samples were determined, and the measurements were conducted using two parameters: received power in milliwatts and power density in W/cm^2. These measurements were carried out using an 810 nm NIR LED, as NIR waves experience less attenuation compared to visible light wavelengths [18] and other wavelengths [19]. The optical power received in the OWC system is a crucial factor in determining its performance, as it is closely related to the signal-to-noise

ratio (SNR) of the received signal [20–22]. Accordingly, before implementing an OWC system for in-body communication, understanding the tissue characteristics is one of the top priorities.

This paper is organized as follows: Section 2 presents brief description of OWC technology for in-body communication to bridge concepts for a wide audience, from the biomedical engineering, wireless communication, and healthcare technology standpoints. Section 3 described methodology followed in this study, including measurement tools and experiment procedures. Section 4 presents the Results and Analysis. Section 5 discusses the finding. Conclusions are given in Sect. 6.

2 OWC Technology for In-Body Communication

OWC is feasible approach to enable a wireless link through biological tissue as the receiver is able to receive optical power, offering highly secure communication to in-body devices that may be used in the future IMD such as modern pacemakers, defibrillators, insulin pumps, cochlear implants, brain implants, etc. [3].

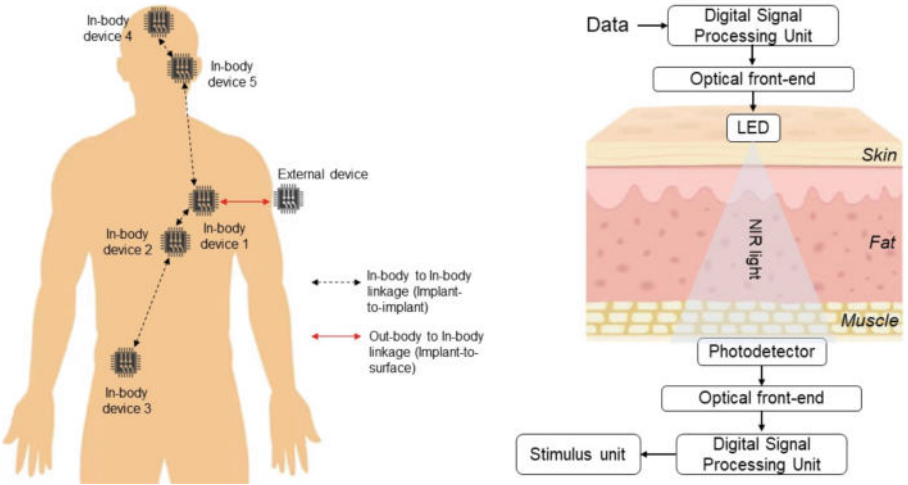


Fig. 1. OWC technology for in-body communication: (a) generic architecture; (b) used architecture

Communication modalities on in-body communication can be categorized into two distinct types: in-body to out-body linkage and in-body to in-body linkage communication (Fig. 1a). Surface-to-implant communication (out-body to in-body linkage) involves communication between in-body and external devices (out-body device) [23]. This type of communication is appropriate for situations where there is a need to transmit data collected from in-body devices to out-body devices for processing or to centralized control and diagnostics centers implant-to-implant communication (in-body to in-body linkage) refers to the communication between two in-body devices or more. This type

of communication is commonly used in applications involving implants that operate in a closed-loop control system [24].

We will use the results of this study for the out-body to in-body communication, as a reference using OWC as a mechanism for providing information from outside to the human body. The system architecture for this purpose is illustrated in Fig. 2(b), which was adopted from [25, 26]. The system contains a digital processing unit transceiver, an optical front-end transceiver, a NIR LED as a transmitter, and a photodetector as a receiver device. The data comes from out-body device or host computer and is processed by a digital signal processing unit; the modulated data are emitted optically using the NIR LED. The data passed the biological tissue are captured by the photodetector and then is demodulated by the digital signal processing subsystem fed to the stimulus device. The implementation of this system is highly possible be applied to modern IMD, such as pacemakers, allowing for wireless control and monitoring through an OWC-based telemetry link.

3 Methodology

The measurements were conducted using commercially available equipment provided by *Thorlabs*. An experimental test-bed comprised an optical transmitter and an optical receiver (Fig. 2). The biological tissue was used as the optical medium. NIR light was chosen to illuminate the biological tissue as it has better propagation properties across tissues than other wavelengths [27], specifically between 800 nm and 900 nm [28]. The transmitter side implemented a driver (*Thorlabs* DC2200) and a mounted NIR LED (*Thorlabs* M810L3).

The LED driver module can be controlled easily using the front panel and a digital display, we used constant current mode in this study. The maximum current for the 810 nm LED was 500 mA, resulting in a maximum transmitted optical power of 372 mW and a maximum incident power density of 525 mW/cm^2, based on actual measurements using an optical sensor (*Thorlabs* S121C) connected to a power meter (*Thorlabs* PM100D). This power level was considered safe as it is below the maximum allowable limit (2 W/cm^2 in 1 sec using $\lambda = 830$ nm) according to the ANSI.Z136.1-2007 standard [29, 30]. The optical power meter was used to measure the received optical power. In this study, only two parameters were used: received power in milliwatts and power density in W/cm^2.

Pure fat and flesh tissues and thicker tissues were used for measurement. Each sample had dimensions of approximately 5 cm $\times$ 5 cm. The thickness of pure fat and flesh tissues are 1.5 cm. Three different thicknesses were used; it was composed of fat and flesh. Afterward, we labeled the thicker tissues as follow: sample #1 (30 mm), sample #2 (38 mm), and sample #3 (40 mm). All samples were freshly purchased from the local market at an initial temperature of 11 °C but were subsequently in a heat chamber to temperatures of 23 °C and 37 °C for measurement purposes.

The LED input current was varied (100 mA, 200 mA, 300 mA, 400 mA, and 500 mA) by controlling the driver module, and the corresponding optical power employed on the porcine sample was measured. The transmitted power of the LED varied depending on the electrical current, with values of 74.2 mW, 153 mW, 230 mW, 303 mW, and 372 mW

observed for the given currents of 100 mA, 200 mA, 300 mA, 400 mA, and 500 mA, respectively.

The porcine sample's surface was illuminated by an NIR LED, while the optical sensor was positioned on the opposing side. The porcine sample heating was performed at different places to protect the test-bed from the heating process. In our measurement, once the sample has been heated and reached a temperature of 37 °C, it is promptly placed in a provided holder where the alignment of the NIR LED and sensor remains unchanged. This positioning remains consistent with previous measurements conducted at a temperature of 23 °C. The temperature of the porcine samples was kept below 44 °C to prevent excessive evaporation and potential harm caused by excessively high temperatures [31]. During the experiment, we strictly adhered to protective eyeglasses (certified laser safety glasses provided by Thorlabs) to minimize the potential risk of eye injuries from exposure to high radiation levels [32–34].

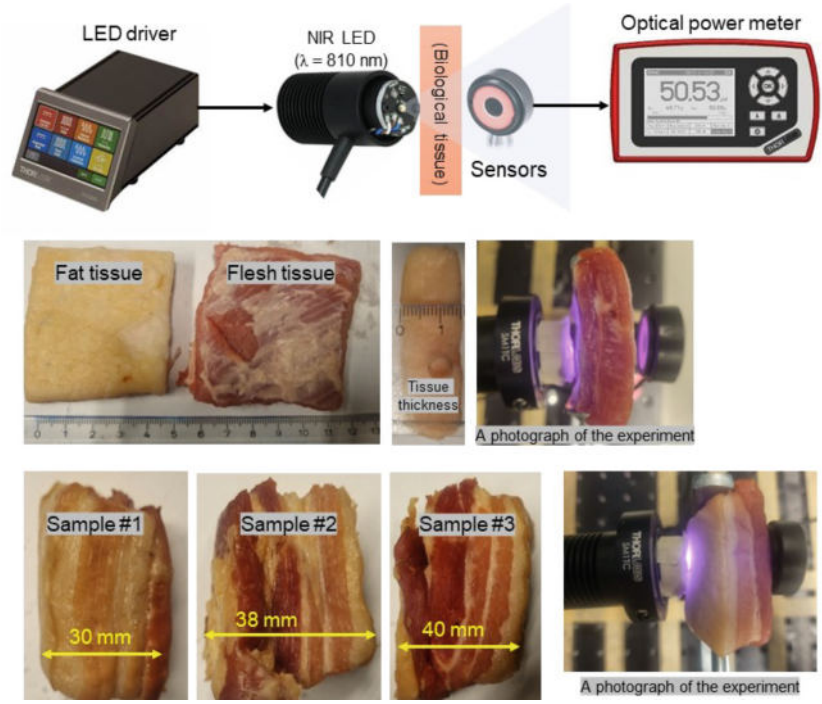
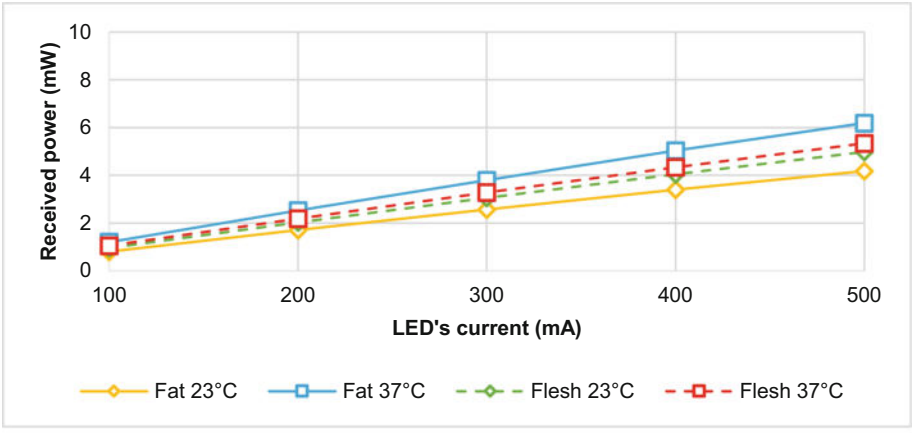


Fig. 2. Experimental setup and details of the employed porcine samples.

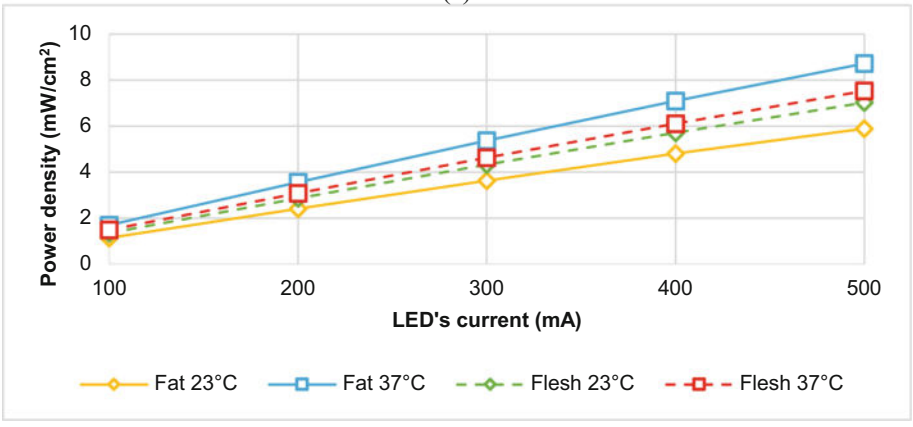
4 Results and Analysis

4.1 Measurement on Fat and Flesh Tissues under Different Temperatures

The results of the measurement of optical power received in mW and power density units for fat and flesh samples under cold (23 °C) and warm (37 °C) conditions are presented in Figs. 3(a) and (b), respectively. It should be noted that the power received increases proportionally with the optical power emitted [35]. For instance, in Fig. 3(a), when the temperature is 37 °C (LED current = 500 mA), the power received in pure fat and flesh samples is 6.19 mW and 5.34 mW, respectively. Similarly, at the same temperature (37 °C), the power received by pure fat and flesh samples is 1.20 mW and 1.05 mW, respectively, when the LED current is changed to 100 mA. The optical power received after passing through the biological tissue is very important parameter as it determines the SNR value.



(a)



(b)

Fig. 3. Measurement results of fat and flesh tissues under different temperatures: (a) received power in mW; (b) power density in mW/cm^2.

The experiment on porcine samples revealed that the optical characteristics of biological tissues are influenced by temperature, leading to changes in transparency and the amount of power received. Our finding suggests that higher levels of transparency on the tissue promote better light propagation through tissue. Figure 3(b) shows that the optical power received by the sample during optical transmission remains lower than the designated safety threshold (below 2 W/cm^2) [29, 30]. For instance, when considering a sample at a temperature of 37 °C with an LED current of 500 mA, the power density received in fat and flesh tissues is 8.72 mW/cm^2 and 7.53 mW/cm^2, respectively. Similarly, at the same temperature (37 °C), fat and flesh receive power densities of 1.70 mW/cm^2 and 1.49 mW/cm^2 with LED current of 100 mA, respectively.

The results indicate that at temperatures close to the human body's temperature (37 °C), the optical penetration of fat tissue is better than that of flesh tissue. Flesh tissue contains the most significant amount of water compared to other constituents, e.g., bone, fat, and skin [36]. Fat tissue demonstrates a better propagation medium when compared to other human tissues, such as flesh tissue, particularly in the case of radio communications, such as ultrawideband (UWB), in terms of signal loss and propagation speed [13, 17].

In the context of meat on cold temperatures (23 °C), it has been observed that optical penetration through flesh tissue is better than that through fat tissue. In contrast, the opposite occurs under warmer conditions (37 °C). One possible reason for the reduced optical penetration in fat tissue under cold conditions is the relatively higher level of light reflection in fat tissue, which can be attributed to its denser nature [3]. Additionally, fat exhibits a higher susceptibility to temperature effects than flesh. Compared to radio waves, particularly UWB technology, optical waves (especially in 810 nm) are less affected by changes in tissue temperature [37], even though the level of optical power received is affected by the temperature of the biological tissue.

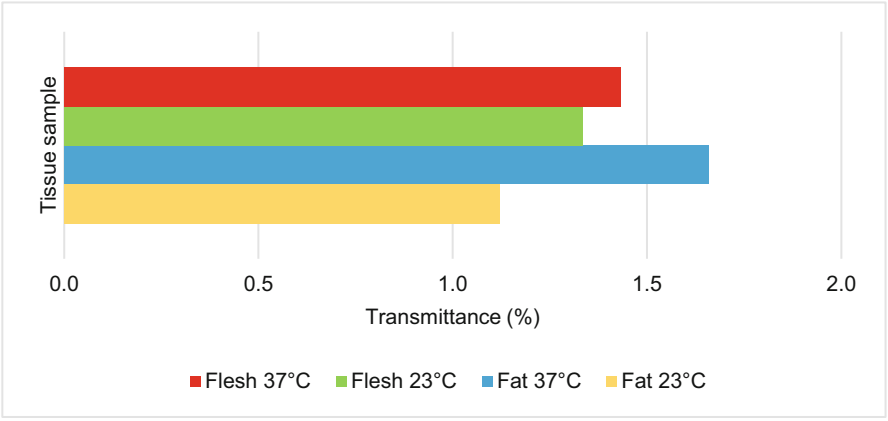


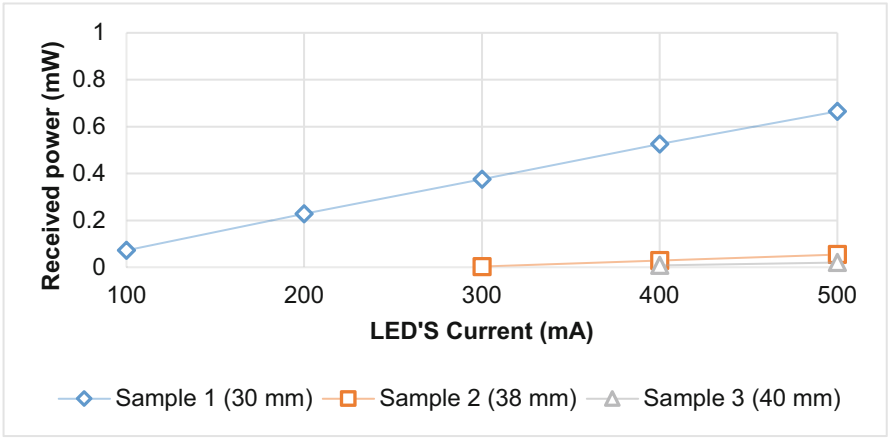
Fig. 4. Transmittance (%) of flesh and fat tissues at 23 °C and 37 °C.

The transmittance rate of porcine samples was obtained to understand better how light penetrates through biological tissues. The transmittance rates were calculated and

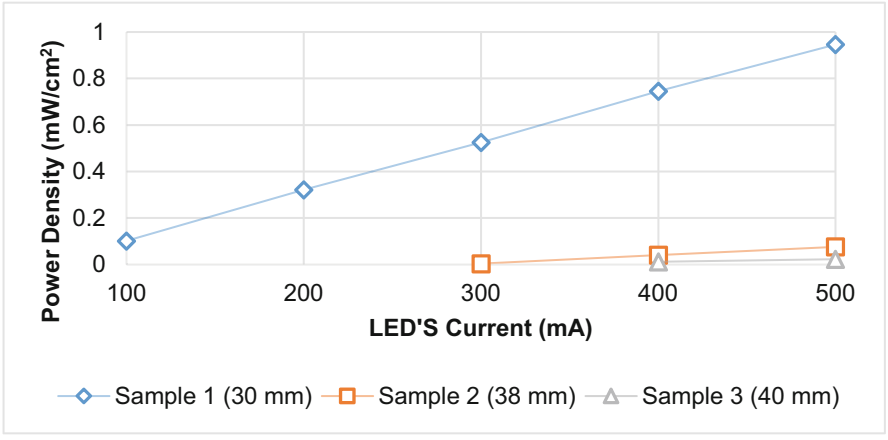
compared for different LED currents across both tissue samples. Transmittance is a measure of the total amount of light that typically passes through a specific medium and is calculated by dividing the transmitted light's power density by the received light's power density in %, as shown in Fig. 4. The transmittance of fat at temperatures of 23 °C and 37 °C was 1.1% and 1.7% respectively. Similarly, the transmittance of flesh tissue at temperatures of 23 °C and 37 °C was 1.3% and 1.4% respectively.

4.2 Experiments Using Various Thicknesses of Samples

Figure 5(a) and (b) show the measurements results on thicker samples of received power and power density, respectively. Sample #1 denotes a fatty tissue, whereas sample #2 and



(a)



(b)

Fig. 5. Measurement results of experiments using various thicknesses of samples: (a) received power in mW; (b) power density in mW/cm^2.

#sample 3 is considered as musculus tissue. Measurements were conducted at a temperature of 37 °C. The received power measured in sample #1 for LED currents 100 mA, 200 mA, 300 mA, and 500 mA were 0.072 mW, 0.228 mW, 0.376 mW, 0.526 mW, and 0.665 mW, respectively. Correspondingly, the power densities were 0.102 mW/cm^2, 0.321 mW/cm^2, 0.526 mW/cm^2, 0.746 mW/cm^2, and 0.946 mW/cm^2. Samples #2 and #3 exhibited only 5% and 1% power reception relative to sample #1.

The findings suggest that tissue thickness influences the received power and power density level, with samples #2 and #3 not receiving any power when LED power levels were set at 75 and 150 mW. The 810 nm NIR light can penetrate fatty tissue (sample #1) up to 30 mm. However, in the case of thicker tissue, as in samples #2 and #3, 810 nm NIR LED penetration requires a power level of 375 mW. This significant decrease in penetration is attributed to the flesh composition in the tissue, which attenuates the NIR light. Fatty tissue is observed to be a better medium for the propagation of NIR light than musculus tissue.

5 Discussion

OWC is an emerging technology that holds promise as a viable and attractive technology for in-body communication, connecting with modern in-body devices, e.g., pacemakers, cardiac defibrillators, insulin pumps, smart pills, and bio-sensors, instead of relying on RF and acoustic technologies. OWC is a viable communication technology to provide wireless connectivity to in-body and on-body devices, as the optical signal can penetrate biological tissues based on observations of received optical power. According to the literature, OWC is deemed to be a more secure method compared to RF as it uses light waves for data transmission which has limited coverage area and offers faster data transmission speeds than acoustic [3, 26]. This study can support future brain-machine communications as light could be used to securely connect certain parts of the brain to the external world.

We have conducted *ex-vivo* experiments on porcine samples (e.g., pure fat tissue, pure flesh tissue, musculus tissue, and fatty tissue). The porcine serves as a general model for human tissue. The experiments involved using an 810 nm LED as a transmitter, an LED driver to control the LED's current, and an optical power meter to measure the received power after the NIR optical light passes the porcine samples. The optical power is one of the critical factors that can impact the performance of OWC systems within in-body devices; it is closely associated with the SNR. For this reason, measuring received optical factor is very crucial. Pure flesh and fat tissues were compared at different temperatures (23 °C and 37 °C). On the other hand, we also conducted on different thicknesses of porcine sample (fatty and musculus tissues) at fixed temperature close to the human body (37 °C). Using thick sample, it was clear to conclude that the muscular tissue received lower optical power than fatty tissue.

The temperature of porcine sample significantly impacted the optical power received by fat tissue but have minimal effect on flesh tissue. The optical power that went through fat tissue at 23 °C and 37 °C was higher than flesh tissue. The optical power received after the fat tissue experiences a substantial decrease of 60% compared to its power at a temperature of 37 °C, while the reduction in optical power in the flesh tissue is

approximately 90%. At a temperature of 37 °C, the optical power after the flesh tissue is 80% of the power after the fat tissue.

This paper provides novel findings over earlier efforts, showing that fat tissue benefits more from heating than in the case of flesh. The study contributes to potential advancements in wireless medical device design and remote healthcare. It is essential to acknowledge that this study was restricted to examining only two varieties of porcine samples (fat and flesh only), different thicknesses were considered, and two temperature levels (23 °C and 37 °C). Future investigations should encompass a wider variety of porcine samples, including different layers with varying compositions of fat, skin, flesh, and bone and different thicknesses. Additionally, exploring a variety of realistic body temperatures (e.g., from 36 °C to 41 °C) holds significant value in the pursuit of further research. However, it is imperative to exercise meticulousness and caution in controlling the temperature of porcine samples using a heater, as excessive heat can cause harm (e.g., exceeding the limits, also sample surface may get dry and then changes in the optical properties). This study focused solely on constant light conditions and did not address achievable rates. Feasibility assessments were based on the received optical power. A subsequent study will integrate the optical front-end to digital signal processing to assess the quality of service (e.g., throughput, bit-error-rate, etc.) on fat tissue propagation under NIR light.

6 Conclusion

The propagation of light through pure fat tissue for optical-based in-body communication has been conducted and we have compared its received power with pure flesh tissue. The experiments also used porcine samples with different thicknesses composed of flesh and fat layers. The impact of sample temperature (cold and warm) was also investigated. The study suggests that heating the meat to 37 °C would be beneficial for a more realistic evaluation of scenarios. The findings of this study provide evidence that the presence of fat layers in porcine sample results in higher received optical power than flesh layers. Furthermore, the study highlights the importance of carefully selecting porcine samples for OWC-based in-body propagation studies, considering the potential impact of meat composition on optical channel characteristics.

Acknowledgments. This study was funded by the University of Oulu's Infotech (CWC-NS 2406124). The authors would like to acknowledge to the Research Council of Finland (former Academy of Finland) Profi6 funding, 6G-Enabling Sustainable Society (6GESS) the University of Oulu and the Research Council of Finland 6G Flagship Programme (Grant Number: 346208).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article

References

1. Fuada, S., Ma, G., Katz, M.: Global growth and trends of in-body communication research – insight from bibliometric analysis. Int. J. Online Biomed. Eng. (iJOE) **20**, 1–22 (2024)
2. Fuada, S., Särestöniemi, M., Katz, M.: Analyzing emerging trends in wireless implantable medical devices (IMDs): a bibliometric study. Int. J. Online Biomed. Eng. (iJOE) **20**, 115–143 (2024). https://doi.org/10.3991/ijoe.v20i04.46559
3. Ahmed, I., Halder, S., Bykov, A., Popov, A., Meglinski, I.V., Katz, M.: In-body communications exploiting light: a proof-of-concept study using ex vivo tissue samples. IEEE Access **8**, 190378–190389 (2020). https://doi.org/10.1109/ACCESS.2020.3031574
4. Sohn, I., Jang, Y.H., Lee, S.H.: Ultra-low-power implantable medical devices: optical wireless communication approach. IEEE Commun. Mag. **58**, 77–83 (2020). https://doi.org/10.1109/MCOM.001.1900609
5. Sohn, I., Rahman, M.F., Jang, Y.H., Lee, S.H.: An optical implant for biotelemetry: design, in vivo verification, and challenges. IEEE Commun. Mag. **60**, 50–56 (2022). https://doi.org/10.1109/MCOM.001.2100784
6. Alizadeh, H., Koolivand, Y., Sodagar, A.M.: Pulse-based, multi-beam optical link for data telemetry to implantable biomedical microsystems. In: 2022 20th IEEE Interregional NEWCAS Conference (NEWCAS), pp. 529–532. IEEE, Quebec City, QC, Canada (2022). https://doi.org/10.1109/NEWCAS52662.2022.9842007
7. De Marcellis, A., Stanchieri, G.D.P., Faccio, M., Palange, E., Constandinou, T.G.: A 300 Mbps 37 pJ/bit pulsed optical biotelemetry. IEEE Trans. Biomed. Circuits Syst. **14**, 441–451 (2020). https://doi.org/10.1109/TBCAS.2020.2972733
8. De Marcellis, A., Palange, E., Faccio, M., Di Patrizio Stanchieri, G., Constandinou, T.G.: A 250 Mbps 24 pJ/bit UWB-inspired optical communication system for bioimplants. In: 2017 IEEE Biomedical Circuits and Systems Conference (BioCAS), pp. 1–4 (2017). https://doi.org/10.1109/BIOCAS.2017.8325081
9. Di Patrizio Stanchieri, G., De Marcellis, A., Battisti, G., Faccio, M., Palange, E., Constandinou, T.G.: A multilevel synchronized optical pulsed modulation for high efficiency biotelemetry. IEEE Trans. Biomed. Circuits Syst. **16**, 1313–1324 (2022). https://doi.org/10.1109/TBCAS.2022.3209542
10. Katz, M., Ahmed, I.: Opportunities and challenges for visible light communications in 6G. In: 2020 2nd 6G Wireless Summit (6G SUMMIT), pp. 1–5. IEEE, Levi, Finland (2020). https://doi.org/10.1109/6GSUMMIT49458.2020.9083805
11. Kayani, M., Ahmed, I., Perera, A., Bykov, A., Katz, M.: A proof of concept for in-body implants for longevity and selfcare. In: Mejías, E., Kouri, P., Ahonen, O., Reponen, J. (eds.) The 26th Finnish National Conference on Telemedicine and eHealth, p. 57. Finnish Society of Telemedicine and eHealth, Oulu, Finland (2021)
12. Särestöniemi, M., et al.: Fat in the abdomen area as a propagation medium in WBAN applications. In: Mucchi, L., Hämäläinen, M., Jayousi, S., Morosi, S. (eds.) Body Area Networks. Smart IoT and Big Data for Intelligent Health Management. LNICSSITE, vol. 297, pp. 175–187. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-34833-5_15
13. Särestöniemi, M., Pomalaza-Ráez, C., Kissi, C., Iinatti, J.: Simulation and measurement data-based study on fat as propagation medium in WBAN abdominal implant communication systems. IEEE Access **9**, 46240–46259 (2021). https://doi.org/10.1109/ACCESS.2021.3068116
14. Asan, N.B., et al.: Reliability of the fat tissue channel for intra-body microwave communication. In: 2017 IEEE Conference on Antenna Measurements & Applications (CAMA), pp. 310–313. IEEE, Tsukuba, Japan (2017). https://doi.org/10.1109/CAMA.2017.8273435

15. Asan, N.B., et al.: Characterization of the fat channel for intra-body communication at R-band frequencies. Sensors **18**, 2752 (2018). https://doi.org/10.3390/s18092752
16. Rangaiah, P.K.B., Engstrand, J., Johansson, T., Perez, M.D., Augustine, R.: 92 Mb/s fat-intrabody communication (Fat-IBC) with low-cost WLAN hardware. IEEE Trans. Biomed. Eng. **71**, 89–96 (2024). https://doi.org/10.1109/TBME.2023.3292405
17. Särestöniemi, M., Pomalaza-Raez, C., Kissi, C., Iinatti, J.: On the UWB in-body propagation measurements using pork meat. In: Alam, M.M., Hämäläinen, M., Mucchi, L., Niazi, I.K., Le Moullec, Y. (eds.) Body Area Networks. Smart IoT and Big Data for Intelligent Health. LNICSSITE, vol. 330, pp. 18–33. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-64991-3_2
18. Chen, Y., Wang, S., Zhang, F.: Near-infrared luminescence high-contrast in vivo biomedical imaging. Nat. Rev. Bioeng. **1**, 60–78 (2023). https://doi.org/10.1038/s44222-022-00002-8
19. Wu, S., Butt, H.-J.: Near-infrared photochemistry at interfaces based on upconverting nanoparticles. Phys. Chem. Chem. Phys. **19**, 23585–23596 (2017). https://doi.org/10.1039/C7CP01838J
20. Fuada, S., Putra, A.P., Adiono, T.: Analysis of received power characteristics of commercial photodiodes in indoor Los channel visible light communication. Int. J. Adv. Comput. Sci. Appl. (IJACSA) **8** (2017). https://doi.org/10.14569/IJACSA.2017.080722
21. Lee, S.-J., Jung, S.-Y.: A SNR analysis of the visible light channel environment for visible light communication. In: 2012 18th Asia-Pacific Conference on Communications (APCC), pp. 709–712. IEEE, Jeju Island (2012). https://doi.org/10.1109/APCC.2012.6388286
22. Wang, L., Wang, C., Chi, X., Zhao, L., Dong, X.: Optimizing SNR for indoor visible light communication via selecting communicating LEDs. Opt. Commun. **387**, 174–181 (2017). https://doi.org/10.1016/j.optcom.2016.11.024
23. Caputo, S., Biotti, L., Mucchi, L.: Bio-optical communication: a case-study of out-to-in body interface. In: Proceedings of the Eight Annual ACM International Conference on Nanoscale Computing and Communication, pp. 1–2. Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3477206.3477470
24. Mohamed, M., Maiseli, B.J., Ai, Y., Mkocha, K., Al-Saman, A.: In-body sensor communication: trends and challenges. IEEE Electromagn. Compat. Mag. **10**, 47–52 (2021). https://doi.org/10.1109/MEMC.2021.9477235
25. Boulogeorgos, A.-A.A., Trevlakis, S.E., Chatzidiamantis, N.D.: Optical wireless communications for in-body and transdermal biomedical applications. IEEE Commun. Mag. **59**, 119–125 (2021). https://doi.org/10.1109/MCOM.001.2000280
26. Xu, Z., Truong, N.D., Nikpour, A., Kavehei, O.: A miniaturized and low-energy subcutaneous optical telemetry module for neurotechnology. J. Neural Eng. **20**, 036017 (2023). https://doi.org/10.1088/1741-2552/acd147
27. Ash, C., Dubec, M., Donne, K., Bashford, T.: Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods. Lasers Med. Sci. **32**, 1909–1918 (2017). https://doi.org/10.1007/s10103-017-2317-4
28. Dremin, V., Zherebtsov, E., Bykov, A., Popov, A., Doronin, A., Meglinski, I.: Influence of blood pulsation on diagnostic volume in pulse oximetry and photoplethysmography measurements. Appl. Opt. AO **58**, 9398–9405 (2019). https://doi.org/10.1364/AO.58.009398
29. Ahmed, I., Bykov, A., Popov, A., Meglinski, I., Katz, M.: Wireless data transfer through biological tissues using near-infrared light: testing skull and skin phantoms. In: Neural Imaging and Sensing 2020, pp. 50–54. SPIE, San Francisco, California, United States (2020). https://doi.org/10.1117/12.2545221
30. Ahmed, I., Bykov, A., Popov, A., Meglinski, I., Katz, M.: Optical wireless data transfer through biotissues: practical evidence and initial results. In: Mucchi, L., Hämäläinen, M., Jayousi, S., Morosi, S. (eds.) Body Area Networks. Smart IoT and Big Data for Intelligent

Health Management. LNICSSITE, vol. 297, pp. 191–205. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-34833-5_16
31. Seo, J., et al.: Wireless electrical power delivery using light through soft skin tissues under misalignment and deformation. Adv. Mater. Interfaces **9**, 2102586 (2022). https://doi.org/10.1002/admi.202102586
32. Kourkoumelis, N., Tzaphlidou, M.: Eye safety related to near infrared radiation exposure to biometric devices. Sci. World J. **11**, 520–528 (2010). https://doi.org/10.1100/tsw.2011.52
33. Wamsley, C.E., Hoopman, J., Kenkel, J.M.: Safety guidelines concerning the use of protective eyewear and gauze during laser procedures. Aesthetic Surg. J. **41**, 1179–1185 (2021). https://doi.org/10.1093/asj/sjaa233
34. Glover, C., Richer, V.: Preventing eye injuries from light and laser-based dermatologic procedures: a practical review. J. Cutan. Med. Surg. **27**, 509–515 (2023). https://doi.org/10.1177/12034754231191064
35. Algora, C., et al.: Beaming power: photovoltaic laser power converters for power-by-light. Joule. **6**, 340–368 (2022). https://doi.org/10.1016/j.joule.2021.11.014
36. Peng, L., et al.: A fused learning and enhancing method for accurate and noninvasive hydration status monitoring with UWB microwave based on phantom. IEEE Trans. Microw. Theory Tech. **71**, 4027–4036 (2023). https://doi.org/10.1109/TMTT.2023.3248788
37. Halder, S., Särestöniemi, M., Ahmed, I., Katz, M.: Providing connectivity to implanted electronics devices: experimental results on optical communications over biological tissues with comparisons against UWB. In: Alam, M.M., Hämäläinen, M., Mucchi, L., Niazi, I.K., Le Moullec, Y. (eds.) Body Area Networks. Smart IoT and Big Data for Intelligent Health. LNICSSITE, vol. 330, pp. 3–17. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-64991-3_1

Preliminary Studies on mm-Wave Radar for Vital Sign Monitoring of Driver in Vehicular Environment

Daljeet Singh[1(✉)], Theresa Eleonye[2], Lukasz Surazynski[1], Hany Ferdinando[1], Atul Kumar[3], Hem Dutt Joshi[4], Mariella Särestöniemi[1,2], and Teemu Myllylä[1,5,6]

[1] Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland
{daljeet.singh,lukasz.surazynski,hany.ferdinando,mariella.sarestoniemi, teemu.myllyla}@oulu.fi

[2] Centre for Wireless Communications, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland
theresa.eleonye@oulu.fi

[3] Department of Electronics Engineering, IIT (BHU) Varanasi, Varanasi, India
atul.ece@iitbhu.ac.in

[4] Department of Electronics and Communication Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab, India
hemdutt.joshi@thapar.edu

[5] Medical Research Center, Oulu, Finland

[6] Optoelectronics and Measurement Technique Unit, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland

Abstract. The last decade has witnessed significant improvements in vehicular technology, especially in providing a safer and more enjoyable environment for drivers and passengers. Fully autonomous vehicles are no longer a dream but are now a successful technology across the globe. Features such as autopilot, assisted parking, speed warning, and lane change assistance have improved the quality of user experience while using an automobile. Apart from this, e-health services have also become a prime aspect of the modern vehicular industry. Therefore, this research presents preliminary studies on mm-wave radar setup based on Frequency Modulated Continuous Wave (FMCW) technology in the 76 to 81 GHz band for vital sign monitoring of drivers and passengers in a vehicular environment. The effect of system parameters and the driver's location with respect to radar is studied using human subjects to determine the optimum setup for vital sign monitoring. Measurement results showcase that mm-wave radars can be utilized for accurate and efficient measurement of the vital signs of drivers in vehicular environments.

Keywords: Breathing rate · Heart rate · In-cabin sensing · mm-wave radar · Vital Sign Monitoring

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 480–493, 2024.
https://doi.org/10.1007/978-3-031-59091-7_32

1 Introduction

The system architecture of 6G for e-healthcare includes communication and sensing as one of its primary fronts. A collaboration of wireless communication and radar technology resulting in Integrated Sensing and Communication (ISAC) is envisioned to govern the beyond 5G and 6G systems. The ISAC systems will play a crucial role in advancing vehicular technology making it fully smart in terms of in-vehicle and outside sensing as well as data processing and communication [1]. Automobiles are the primary transportation method for billions of people worldwide. Cars went through an extraordinarily long road of systematic improvements. Currently, we are observing rapid development of electric and hybrid cars, where modified engines are replacing standard petrol engines. Another branch of development is focused on automated driving, where computers or steering units navigate and drive the car from point to point or even parks [11].

Nowadays, cars are packed with different kinds of sensors, which assist the driver and serve multiple purposes. Such sensors could be divided into two groups: contact and contactless sensors. Contact sensors are placed in locations that usually make constant contact during travel such as seats, bolster, steering wheel, seatbelt in passenger cars, and driving gear (helmet, suit) in specialized cases. This attribute results in restricting their usage and accuracy. Contactless or wireless sensors can sense from a distance and therefore have the advantage of no real restrictions regarding the location of the sensors, other than not to interfere with the driver's sight. Such sensors could be also placed on or in close proximity to previously mentioned locations such as seats or steering wheel [8].

Initial systems for vital sign monitoring in vehicles were primarily based on imaging (cameras) which have the inherent drawback of security and privacy. Other wireless sensors, which in the majority are radar-based, can provide sets of valuable information regarding the position of the driver/passengers inside the vehicle without compromising privacy. This allows inter alia to evaluate driver conditions such as tiredness or consciousness [3]. Moreover, such sensors can be used to monitor the vital signs of the driver as well as passengers such as respiration, and heart rate. In some instances, this information is encoded simultaneously allowing constant measurement with high sampling rates. Some key research articles based on vital sign minoring using mm-wave radar technology are [2,3,7,8,10–12].

2 Key Challenges in Radar Based In-cabin Sensing

Radar-based in-cabin sensing has many advantages but it has its challenges too. This section presents the key challenges associated with radar-based in-cabin sensing from the perspective of future joint communication and sensing system scenarios.

2.1 Interference from Other Devices

With the advancements in technology, the vehicles in the market now come fully equipped with smart sensors capable of communicating with the driver as well as with each other for seamless operation of the vehicle. Apart from this, other major causes of interference include cellular devices carried by the driver, in-built radars in a vehicle to support auto-pilot/driving assistance, and other communicating devices in the vicinity of the vehicle. Spectrum overlapping between these applications create even severe problems which tend to grow with each day. One straightforward solution to this problem is to use highly directional antennas for in-cabin sensing. But it comes at the cost of losing spatial coverage of the radar [5].

2.2 Optimum Placement and Location of Radar Chip

Even though designed for comfort, vehicles have limited space, and finding a practical yet optimum location for radar placement inside the vehicle is challenging. There are some locations that have been extensively studied in the literature such as behind the steering wheel on the speedometer, on the rooftop, pasted on the rear mirror, inside the car seat targeting from the back, on the side door, etc. The effectiveness of these locations depends largely on the build (shape) of the car, the location of other electronic components, and user preference. Therefore, choosing the optimal location, position, and radar's angle of attack are very important parameters.

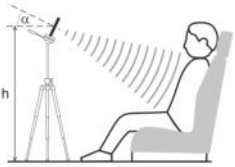
2.3 Movement of Subjects with Respect to Radar

Most of the studies in the literature present the problem of in-cabin sensing as a very simplified problem in which the subject/driver is sitting in an ideal position without any movement with respect to radar. These assumptions are rarely met in practical scenarios wherein the driver usually has restricted movements in his seat. Especially during the first few minutes of starting the journey, these movements are found to be more dominant, and they gradually decrease once the driver settles down [10].

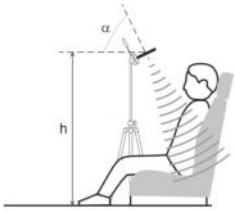
2.4 Selection of Frequency Band of Operation for Radar

The problem of vital sign monitoring of humans with radar technology has been studied extensively in the last ten years. Researchers have proposed a variety of frequency bands starting from a few GHz to the Tera Hz range. The selection of the optimal frequency band of operation for radar has a dominant effect on the detection accuracy of the system. Especially due to the opening of the mm-wave spectrum for both communication and sensing, most of the current standards are adopting mm-wave and it is expected to rule the future sensing and communication applications.

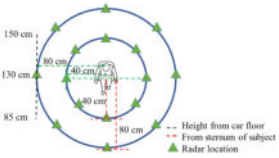
(a) Setup with subject positioned in front of AWR 1642 mm-wave radar chip placed on a tripod



(b) Measurement of heart rate and breathing rate with angle α



(c) Measurement of heart rate and breathing rate with different angle and radar locations



(d) Scenarios for placement of radar with respect to the subject

(e) Location of Zephyr BioHarness attached with chest strap [9]

(f) Actual setup with human subject

Fig. 1. mm-wave Radar setup for vital sign monitoring of driver in a vehicular environment.

2.5 Cabin Shape and Reflections Due to Vehicle Body

A car cabin, being a metallic hollow body, creates challenges in the effective working of radar. The reflections caused by the car body create interference for received signals thus deteriorating the performance of radar. Apart from that, the vibrations in the car body during its motion also create strong artifacts in the captured signals from radar [5].

2.6 Multi-persons In-cabin Scenarios

Usually, the driver in the car is accompanied by co-passengers whose vital sign monitoring is also equally important but challenging. Some articles [4,8] showcase the implementation of multi-person monitoring using a single radar system. Nevertheless, the task of separating dopplers from multiple targets and then extracting vital sign information from each one of them is undoubtedly a tedious task.

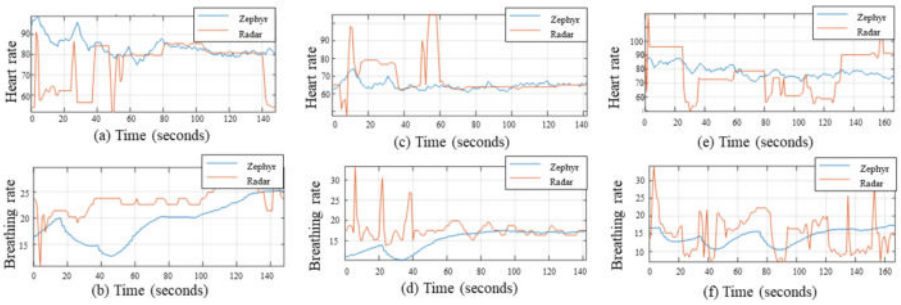


Fig. 2. Breathing rate and heart rate measurements of Subject 1 for three scenarios: (a), (b) Radar in front (80 cm from sternum at height= 85 cm); (c), (d) Radar in back (20 cm from sternum at height= 85 cm); (e), (f) Radar at side (90 degrees to left 80 cm from sternum at height= 85 cm).

3 Material and Methods

The hardware setup utilized in this study consists of a car seat with adjustable height along with a tripod stand to hold the radar chip which allows movement of the sensor along roll, pitch, and yaw. The measurements are taken with different radar placement options based on which the optimized position of the radar in a vehicular environment is selected. The radar setup utilized in this study is FMCW-based TI 1642 Boost which operates in 76–81 GHz W-band and offers excellent performance in terms of detection accuracy [6]. Figure 1(a),(b) and (c) showcase the measurement of heart rate and breathing rate with AWR 1642 mm-wave radar chip with different angles and radar locations. The measurement settings with possible scenarios chosen for placement of radar with respect to the subject are shown in Fig. 1(d) which has a total of 24 possible locations for radar at two distinct heights of 85 cm and 130 cm from the car floor.

The sensors are located at a horizontal distance of 20 cm and 150 cm from the sternum of the subject. These locations are chosen in accordance with the structure of a typical car environment. The subject under study is to sit on the car seat with hands resting on laps or in a position of holding a driving wheel. The radar setup is moved in three-dimensional space in between the measurements of 120 s each. There are 2–4 transmit/receiver antennae on the radar chip with peak gain >9 dBi across the operating frequency. The setup measures the rate of chest displacement to calculate the heart rate and breathing rate of a person. The results measured from the proposed setup are compared with the BioHarness 3.0 module developed by Zephyr Technology which is worn by the subject during the measurements [9]. Figure 1(e) shows the location of the Zephyr BioHarness Module attached to the chest strap of the subject. The actual setup with a human subject is shown in Fig. 1(f).

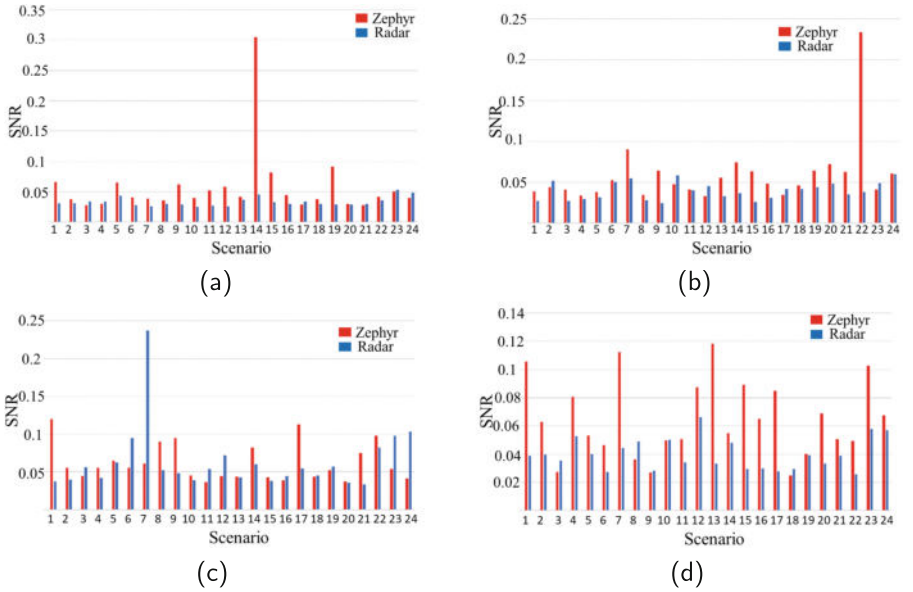


Fig. 3. SNR plots of (a) heart rate, (b) breathing rate for Subject 2 and (c) heart rate, (d) breathing rate for Subject 5 in different scenarios.

4 Data Collection and Processing

The data collection for this work was conducted on 5 adult subjects consisting of 3 males and 2 females for variability in the dataset. The first 5 scenarios lasted for 3 min each while the remaining scenarios lasted for 2 min each. The data collection procedure lasted for approximately 2 h for a total of 25 measurements per participant. Each participant was notified about the measurement scenario. A detailed description of measurement scenarios taken in this study is depicted in Table 1. Instructions were provided concerning the procedure and informed consent was duly signed by each of the participants. The data collected from each subject was stored simultaneously in both the Zephyr device and the Radar attached to a computer.

Each subject is instructed to tap the chest 3 times which is used as a marker for the manual alignment of the signals from the Zephyr and Radar devices. The raw ECG data with the sampling frequency of 250 Hz from the Zephyr was used which clearly shows the sharp bursts of chest tap for each scenario for the time of occurrence. The respective time of occurrence of the chest tap on the raw ECG waveform is used for segmenting the physiological parameters of the heart rate and the breath rate signals from the Zephyr device. In other to merge the two datasets from the two devices, both devices must have the same sampling frequency.

Therefore, the data from the radar was resampled to match the sampling frequency of data from Zephyr. The averaged Radar data were manually merged

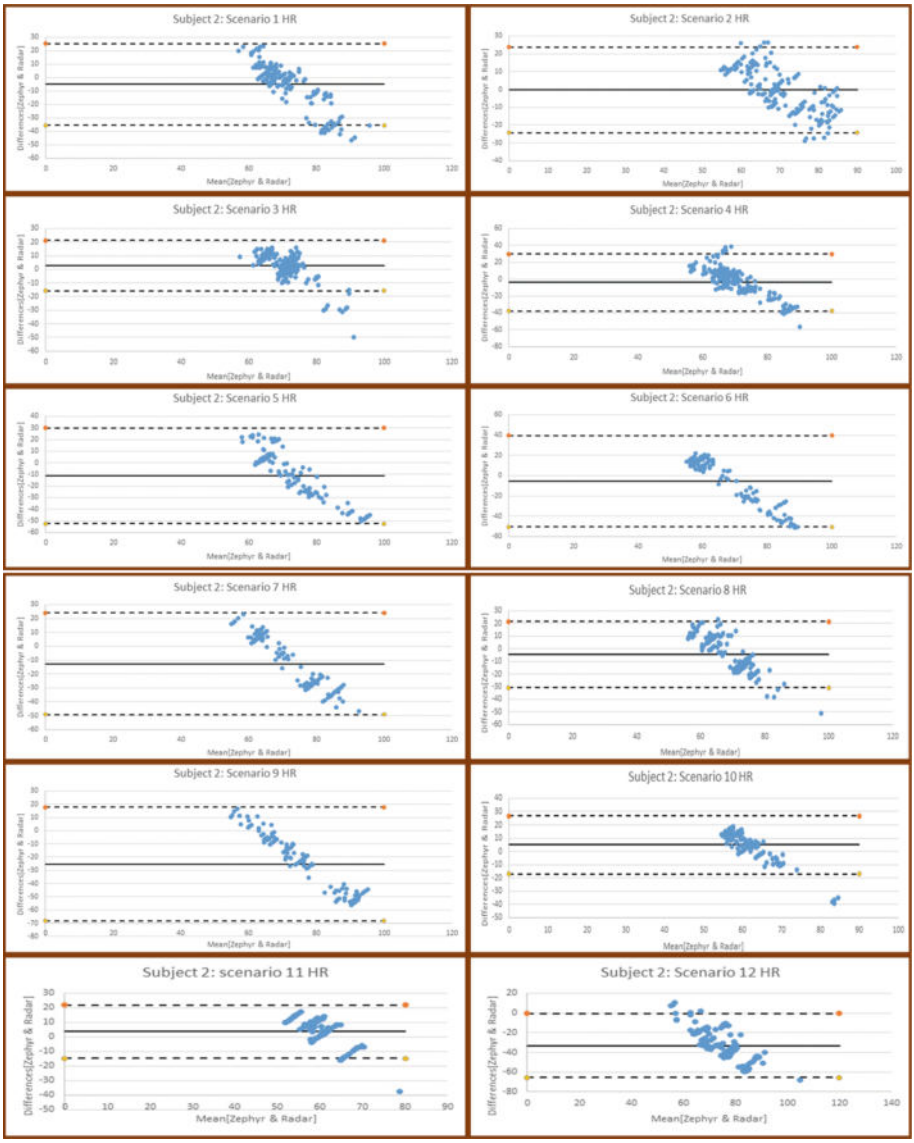


Fig. 4. Bland Altman's plots of subject 2 for Scenario 1–12.

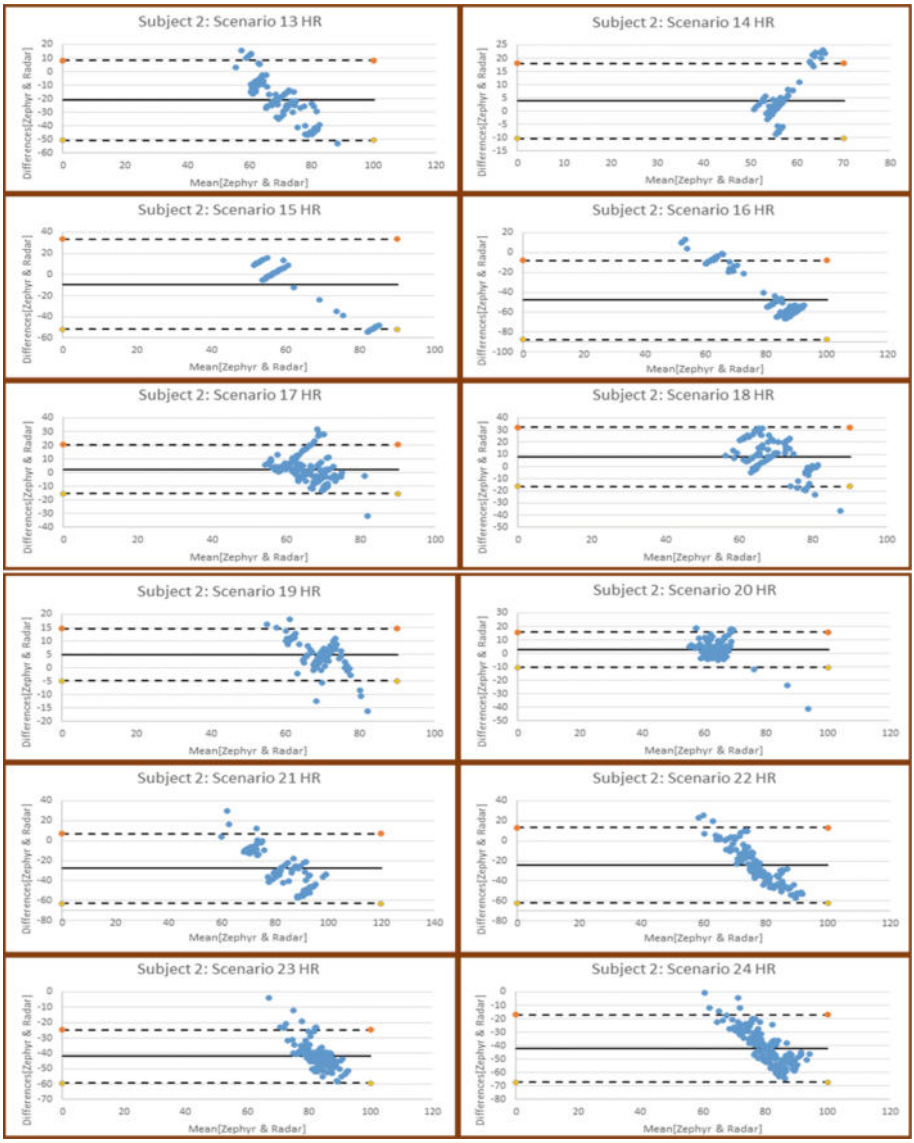


Fig. 5. Bland Altman's plots of subject 2 for Scenario 13–24.

Table 1. Description of measurement Scenarios.

Scenario	Azimuth angle ($\cdot^0$)	Range (cm)	height of radar (cm)	tilt angle ($\cdot^0$)
1	0°	80	85	0°
2	45°	80	85	0°
3	90°	40	85	0°
4	270°	40	85	0°
5	315°	80	85	0°
6	135°	80	85	0°
7	180°	0	85	0°
8	225°	80	85	0°
9	0°	80	130	45°
10	45°	80	130	45°
11	90°	40	130	45°
12	270°	40	130	45°
13	315°	80	130	45°
14	135°	80	130	45°
15	180°	0	130	45°
16	225°	80	130	45°
17	0°	80	150	30°
18	45°	80	150	30°
19	90°	40	150	30°
20	270°	40	150	30°
21	315°	80	150	30°
22	135°	80	150	30°
23	180°	0	150	30°
24	225°	80	150	30°
25	0°	80	85	0°

or aligned with the Zephyr data for the time of occurrence of the chest tap for each scenario. The parameters of measurement which include the range, height of the radar device from the floor, the tilt angle of the radar, and the azimuthal angles are captured in the measurement setup. Scenarios 1, 2, 3, 4, 5, 9, 10, 11, 12, 13, 17, 18, 19, 20, 21, and 25 (with no subject) have the Radar positioned in the front while scenarios 6, 7, 8, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24 have the Radar positioned at the back. Scenario 25 with no human subject was measured once during the measurement procedure of each subject as a reference case.

5 Results and Discussion

Statistical analysis is applied for comparison of results from radar setup with reference Zephyr device. Various performance matrices are calculated including Bland Altman plots, Pearson's correlation coefficient (r) with statistical significance of p-values, and the measurement accuracy is determined using the signal-to-noise ratio (SNR) plots. Moreover, the boxplots of individual human subjects for each scenario versus the physiological parameters of heart rate and breath rate are compared amongst all 5 adult subjects to ascertain if the results are comparable. Figure 2 shows the comparison of measured results with benchmark results from Zephyr. The heart rate and breathing rate results for the three cases are presented in Fig. 2. All results are verified by data collected from Zypher and are found to be fairly accurate. To study the accuracy of the proposed radar-based setup, the coefficient of variation is calculated for each scenario. The coefficient of variation measures the dispersion of points from the mean which is given as a ratio of the standard deviation to the mean in percentage. Further, the signal-to-noise ratio (SNR) is calculated as the inverse of the coefficient of variation. The evident occurrence of a lower coefficient of variation in heart rate and breath rate in the Radar device than in the Zephyr device is reflected in the high signal-to-noise ratio (SNR) plots in the different scenarios. The SNR plots of heart rate and breathing rate for Subject 2 and 5 for different scenarios are shown in Fig. 3.

Table 2. Correlation analysis of heart rate and breathing rate of subjects from radar and Zephyr in different scenarios.

Scenarios	p≤0.05	Subject	Correlation
Scenario1	yes	4	Both negative
Scenario2	yes	3	Both positive
Scenario3	yes	4	Positive and negative
Scenario4	yes	1	Positive and negative
Scenario5	yes	1	Both positive
		4	Both negative
		5	Both positive
Scenario6	yes	4	Both negative
		5	Both negative
Scenario7	yes	2	Both positive
Scenario8	yes	4	Both negative
Scenario9	yes	5	Positive and negative
Scenario10	None		
Scenario11	yes	4	Both negative
Scenario12	yes	5	Both positive

continued

Table 2. continued

Scenarios	p≤0.05	Subject	Correlation
Scenario13	yes	1	Positive and negative
Scenario14	yes	4	Positive and negative
Scenario15	None		
Scenario16	yes	1	Both negative
Scenario17	yes	1	Both positive
Scenario18	yes	1	Both positive
		3	Positive and negative
Scenario19	yes	5	Negative and positive
Scenario20	yes	5	Negative and positive
Scenario21	yes	1	Negative and positive
		3	Negative and positive
Scenario22	yes	1	Both positive
		2	Both negative
		3	Negative and positive
Scenario23	None		
Scenario24	None		

The Bland Altman's plots of subject 2 for scenarios 1–12 are shown in Fig. 4. Further, Bland Altman's plots of subject 2 for scenarios 13–24 are shown in Fig. 5. It can be visualized from Fig. 4, 5 and similar plots obtained for other subjects that the scatter plots lie within the 95% confidence level alongside some outliers on either or both sides of the limits of agreement with small systematic bias. The acceptable bias limit is taken to be ±10 bpm and exceptions are visible in some scenarios where the heart rate and breathing rate have high systematic bias. In subject 1, the scatter plots of the heart rate event exhibit a trend from high values to low values with small bias in different scenarios while the breath rate exhibits random scatter plots reflecting the consistent differences in the measurement between the two devices with low bias occurring in several scenarios. Similarly, this general trend in the scatter plots is replicated in the remaining human subjects for most scenarios except for scenario 7 (heart rate) in subject 3 and subject 4 which show the trend from lowest value to highest values, scenario 14 (heart rate) in subject 2 and subject 3 display the same trend. Hence, the Bland Altman plots showed that the Radar device is in strong agreement with the Zephyr device with scatter plots closer to the bias and are different or divergent at extremely low or high values.

It is worthy of note that for all 5 subjects, scenario 7 with the Radar device having the parameters of measurement, displays a large bias for the heart rate and a small bias for the breath rate while scenario 11 displays a low bias for both physiological parameters. However, scenario 12 displays a small bias in both heart rate and breath rate for subject 1 and subject 4 while the remaining

three subjects had a high bias in the heart rate and a low bias in the breath rate respectively.

Further, Pearson's correlation coefficient is used to determine the linear relation as per the strength of association and direction between the results obtained from the Zephyr and Radar device. In addition, the interpretation of the correlation coefficient is based on the decision rule of statistical significance of the p-values where $p \leq 0.05$ to satisfy the null hypothesis of a significant relationship between the results from both devices. The correlation plots with statistical significance on both heart rate and breath rate for all the subjects are observed in several scenarios ranging from the front side to the back side of each subject except for scenarios 10, 15, 23, and 24 where no simultaneous statistical significant values were obtained on both physiological parameters. Table 2 provides a summary of the occurrences of the p-values (heart rate and breath rate) with the corresponding correlation for the subject.

There are some disparities in the correlation coefficient for all 5 subjects for all the scenarios such that there are scenarios where the heart rate has a positive correlation and the breath rate has a negative correlation and vice versa. The probable reason can be attributed to the differences in height and body type of each subject which can be linked to the radar cross-section (RCS). In addition, the field of view (FOV) of the radar sensor is affected by the vertical height which determines the angular tilt orientation of the radar. Hence as the vertical height is increased the radar sensor with the antenna patch should be tilted to an angle that can accommodate the FOV for good signal quality.

6 Conclusion

Our results suggest that radars offer an effective method for vital sign monitoring of drivers and passengers in vehicular environments which should be explored further. Radars have the advantage of penetration through clothes and other materials to provide better accuracy than imaging-based methods. Further, the privacy of the subject is not compromised in this method. It is observed that the radar placed in front at a height equivalent to the sternum of the subject (Scenario 1) provides the best sensing results as compared to when it is placed at any other location. Apart from this, good results are obtained for Scenario 11 for all subjects and Scenario 12 for two subjects. A maximum deviation between standard Zephyr data and radar data is observed in Scenario 7. It can be concluded that the radar placed at a distance of 40 cm from the chest of the driver at a height of 130 cm from the car floor can be an optimum location of radar. This location corresponds to the area behind the driving wheel in the car dashboard.

Acknowledgments. This work was supported by the Global Pilots financed by the Finnish Ministry of Education and Culture as part of the project INDFICORE, University of Oulu Emerging Program Project, 6G-enabled sustainable society (6GESS) program: 6GESS6 and 6GBRIDGE - Next generation healthcare and wearable diagnostics utilizing 6G project (11146/31/2022). A part of this research is also funded by

the Academy of Finland Profi6 funding, 6G-Enabling Sustainable Society (University of Oulu, Finland) under the Emerging project for which we acknowledge the University of Oulu, Focus Institute Infotech Oulu for supporting our project. The authors also thank INDFICORE (India-Finland cooperation on 6G) under the 6G Flagship group at the University of Oulu, Finland, and its Indian partners for providing valuable discussions regarding this project.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Daljeet, S., et al.: Generalized adaptive spreading modulation: a novel waveform for integrated sensing and communication oriented vehicular applications. Authorea Preprints (2023)
2. Daljeet, S., Lukasz, S., Hany, F., Mariella, S., Teemu, M.: Radar setup for vital sign monitoring of driver in vehicular environment. In: The 28th Finnish National Conference on Telemedicine and eHealth. Finnish Society of Telemedicine and eHealth (2023)
3. Du, F., Wang, H., Zhu, H., Cao, Q.: Vital sign signal extraction based on mmwave radar. J. Comput. Commun. **10**(3), 141–150 (2022)
4. Fernandes, J.M., Silva, J.S., Rodrigues, A., Boavida, F.: A survey of approaches to unobtrusive sensing of humans. ACM Comput. Surv. (CSUR) **55**(2), 1–28 (2022)
5. Gharamohammadi, A., Khajepour, A., Shaker, G.: In-vehicle monitoring by radar: A review. IEEE Sensors Journal (2023)
6. Instruments, T.: Awr1642 single-chip 77-and 79-ghz fmcw radar sensor. Datasheet AWR1642, Rev p. 60 (2017)
7. Iyer, S., et al.: mm-wave radar-based vital signs monitoring and arrhythmia detection using machine learning. Sensors **22**(9), 3106 (2022)
8. Marty, S., Pantanella, F., Ronco, A., Dheman, K., Magno, M.: Investigation of mmwave radar technology for non-contact vital sign monitoring. In: 2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1–6. IEEE (2023)
9. Nazari, G.: Psychometric Parameters of Zephyr Bioharness & Fitbit Charge. Ph.D. thesis (2016)
10. Wang, F., Zeng, X., Wu, C., Wang, B., Liu, K.R.: Driver vital signs monitoring using millimeter wave radio. IEEE Internet Things J. **9**(13), 11283–11298 (2021)
11. Wang, Y., Wang, Z., Zhang, J.A., Zhang, H., Xu, M.: Vital sign monitoring in dynamic environment via mmwave radar and camera fusion. IEEE Transactions on Mobile Computing (2023)
12. Zhang, B., Jiang, B., Zheng, R., Zhang, X., Li, J., Xu, Q.: Pi-vimo: physiology-inspired robust vital sign monitoring using mmwave radars. ACM Trans. Internet Things **4**(2), 1–27 (2023)

Abstracts

Pipeline for Automatic Brain Stimulation with Real-Time Machine-Learning-Based Decoding of EEG Responses

Matilda Makkonen(✉), Olli-Pekka Kahilakoski, Kyösti Alkio, Ivan Zubarev, Risto J. Ilmoniemi, Tuomas P. Mutanen, and Pantelis Lioumis

Department of Neuroscience and Biomedical Engineering, Aalto University, 02150 Espoo, Finland
matilda.makkonen@aalto.fi

Abstract. We introduce a flexible pipeline for brain-state-dependent transcranial magnetic stimulation, which operates with customizable real-time electroencephalogram analysis and preprocessing algorithms that support classification with machine-learning tools.

Keywords: Electroencephalography · Transcranial magnetic stimulation · Brain-state-dependent brain stimulation · Brain-activity decoding

1 Introduction

Transcranial magnetic stimulation (TMS) is a non-invasive brain stimulation method for activating cortical neuronal populations. TMS is increasingly used as a therapeutic tool in neurology and psychiatry. The application of TMS for diagnostic and therapeutic purposes continues to be driven by group-level results, with little regard to patient-specific characteristics. However, the neuronal signaling and its breakdown during disease are highly individual [1], promoting the need for personalized TMS-treatment procedures.

Here, we discuss the Brain-State-Dependent-Stimulation approach (BSDS), in which the timing and/or location of the stimulation is automatically controlled based on the instantaneous brain state, measured with electroencephalography (EEG) with the help of machine-learning algorithms to detect complex brain states. BSDS is a promising method for personalized TMS treatment [1], but its implementation is hindered by the lack of software pipelines enabling real-time EEG processing and simultaneous TMS control.

2 Materials and Methods

We introduce a flexible pipeline for BSDS, which operates with customizable real-time EEG-analysis and preprocessing algorithms, written in Python, that support classification with machine-learning algorithms. The pipeline controls and monitors the TMS

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 497–499, 2024.
https://doi.org/10.1007/978-3-031-59091-7

device with the outcome of the real-time EEG analysis. The pipeline currently supports EEG devices from two manufacturers (Bittium NeurOne and Brain Products acti-CHamp with TurboLink) and provides TTL trigger signal output to an arbitrary TMS device through a BNC connection, thus controlling the timing of TMS pulses with millisecond precision. In addition, the pipeline can control a state-of-the-art multi-locus TMS device [2], allowing the system to change the stimulation location without physically moving the TMS coil or the need for multiple TMS devices.

The pipeline, implemented with the Robot Operating System (ROS2), can be controlled and monitored through a user-friendly graphical user interface. The pipeline can be customized with algorithms for EEG preprocessing and brain-activity decoding for TMS-control decisions. Adding custom algorithms is straightforward, as a Python template is provided into which scientists can add their own processing algorithms in a similar way as in offline processing. The pipeline runs on a laptop or desktop computer, and no expensive hardware is needed.

We have integrated a novel real-time noise-removal algorithm as a preprocessing step before the EEG signals are fed into the brain-activity-decoding algorithms. The real-time noise-removal is performed with the SOUND algorithm [3] and is currently being validated (publication in preparation). The real-time version of the SOUND algorithm processes each streamed data sample with a spatial filter that suppresses signals unlikely to originate from cortical current sources. The spatial filter is based on the estimated spatial distribution of noise, and is constantly updated in an asynchronous process, which estimates the noise level from the latest data buffer. The real-time SOUND has so far been validated with TMS–EEG datasets measured from 20 healthy subjects.

Many types of experiments can be performed with the presented pipeline. It is currently employed, for example, in paradigms for timing TMS to epileptic spikes, or to motor imagery of the hands. These applications are intended for the treatment of epilepsy and spinal cord injury, respectively. In both applications, a pre-trained machine-learning algorithm detects features of the brain states of interest, namely epileptic spikes and motor imagery, in real-time. TMS can thus be timed to epileptic spikes, with the goal of suppressing epileptogenic activity, or to motor imagery, with the goal of strengthening neuronal connections between the brain and the limb [4]. The real-time detection of spikes has been tested with an EEG dataset from an 11-year-old epilepsy patient.

For the machine-learning approach in these applications, we used a compact convolutional-neural-network-based classifier, LF-CNN, designed specifically to fit the spatiotemporal properties of EEG and MEG signals [5]. LF-CNN architecture is informed by the generative model of these signals, allowing to achieve high representational capacity of deep learning while keeping the number of training parameters low. Thus, the model can be trained using a limited amount of training data, run efficiently in real-time, and be easily adapted for various experiments. Measurements and analysis for the epileptic-spike-timed and motor-imagery-timed TMS are currently ongoing.

3 Results

The pipeline is currently under development, and therefore, many aspects of the system still require further testing. However, the tentative validation results of the real-time SOUND approach are promising. Independent-component-analysis-based analysis showed that the signal-to-noise ratio was improved in all the subjects studied, on average by 200%. In the first benchmark tests, the SOUND algorithm processed streamed samples in, on average, 38 μs per sample, easily handling EEG sampled with 5 kHz.

The preliminary results of the real-time epileptic-spike-detection algorithm showed specificity and sensitivity up to 99% and 83%, respectively.

4 Discussion

The advantages of the presented BSDS pipeline are its easy customizability with algorithms written in Python, applicability to different EEG and TMS devices, as well as the need for only a desktop or laptop computer without expensive custom hardware. The pipeline can be applied to a wide spectrum of research and clinical use cases extending beyond the example cases of epilepsy and spinal cord injury. BSDS protocols are predicted to greatly impact the treatment of psychiatric and neurological disorders as well as rehabilitation.

Disclosure of Interests. TPM currently works for Bittium Ltd. PL has a common project with Bittium Ltd. ("Brain–Computer Interface for Automated EEG-guided Brain Stimulation", Tandem Industry Academia). The other authors have no competing interests to declare that are relevant to the content of this article.

References

1. Rösch, J., et al.: Individualized treatment of motor stroke: a perspective on open-loop, closed-loop and adaptive closed-loop brain state-dependent TMS. Clin. Neurophysiol. **158**, 204–211 (2024)
2. Koponen, L.M., Nieminen, J.O., Ilmoniemi, R.J.: Multi-locus transcranial magnetic stimulation—theory and implementation. Brain Stimul. **11**(4), 849–855 (2018)
3. Mutanen, T.P., Metsomaa, J., Makkonen, M., Varone, G., Marzetti, L., Ilmoniemi, R.J.: Source-based artifact-rejection techniques for TMS–EEG. J. Neurosci. Methods **382**, 109693 (2022)
4. Shulga, A., Lioumis, P., Kirveskari, E., Savolainen, S., Mäkelä, J.P.: A novel paired associative stimulation protocol with a high-frequency peripheral component: a review on results in spinal cord injury rehabilitation. Eur. J. Neurosci. **53**(9), 3242–3257 (2021)
5. Zubarev, I., Zetter, R., Halme, H.-L., Parkkonen, L.: Adaptive neural network classifier for decoding MEG signals. Neuroimage **197**, 425–434 (2019)

Early Prediction of Memory Disorders Using Primary Healthcare and Social Services Data

Heba Sourkatti[1], Tunc Asuroglu[2(✉)], Aino-Lotta Ilona Alahäivälä[3], Anna-Maija Tolppanen[4], and Jouni Ihalainen[3]

[1] VTT Technical Research Centre of Finland, Tekniikantie 1, 02150 Espoo, Finland
[2] VTT Technical Research Centre of Finland, Visiokatu 4, 33720 Tampere, Finland
tunc.asuroglu@vtt.fi
[3] VTT Technical Research Centre of Finland, Microkatu 1, 70211 Kuopio, Finland
[4] School of Pharmacy, University of Eastern Finland, Yliopistonrinne 3, 70211 Kuopio, Finland

Abstract. This study focuses on early identification of memory disorders among elderly individuals, utilizing data from social and healthcare services in Kuopio. A cohort of 26,000 citizens aged over 65 as of 2015 was utilized. Through a case-control study, individuals diagnosed with Alzheimer's disease (AD) and controls were identified. ANOVA and Mutual Information (MI) methods identified significant features including International Classification of Primary Care (ICPC) and International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes, onset age, number of patient visits, and types of services. Logistic regression and SHAP-guided gradient-boosting classifiers demonstrated promising predictive performance, suggesting potential for proactive interventions and targeted monitoring in individuals at risk of memory disorders.

Keywords: Machine Learning · Feature Selection · Predictive Modelling · Memory Disorders · Electronic Health Records

1 Introduction

We aimed to identify the risk of memory disorders at an earlier stage, enabling targeted interventions to monitor patient status and reduce disease risk through lifestyle counseling or cognitive testing. The study utilized a cohort of 26,000 Kuopio citizens aged over 65 as of the year 2015 who used social and healthcare services between 2010 and 2020.

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 500–502, 2024.
https://doi.org/10.1007/978-3-031-59091-7

2 Materials and Methods

We conducted a case-control study with individuals diagnosed with Alzheimer's disease (AD; N = 1,524) and those without any memory-related ailments or medications as controls (N = 8,005). Both groups are composed of 66% females. Features were retrieved as 6-month data frames extending from 6 months preceding the diagnosis (i.e., excluding the initial half a year) to 4 years before diagnosis (or matching date of the controls). These data frames encompass early detection across distinct time windows for disease prediction.

To identify the most crucial features differentiating AD patients from controls, we employed ANalysis of Variance (ANOVA) [1] and Mutual Information (MI) [2] feature ranking methods. Based on ANOVA and MI, various feature sets were constructed according to feature importance ranks. These feature sets were then benchmarked using machine learning models such as Logistic regression, Random Forest, Gaussian Naïve Bayes, and Multi-layer perceptron. Additionally, we explored an alternative feature-selection approach for gradient-boosting classifiers. Classifiers were trained across all data frames. Top features were identified with SHAP (SHapley Additive exPlanations) [3] and we subsequently retrained the models using these selected features.

3 Results

ANOVA and MI methods revealed that International Classification of Primary Care (ICPC) and International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes associated with memory disorders, onset age, number of patient visits, and the types of services were among the top-ranking features. The high-ranking features, identified by ANOVA in conjunction with Logistic regression, achieved a prediction performance of 0.846 Area Under Curve (AUC) and surpassed the performance of other models and feature sets. The alternative strategy of retraining gradient-boosting classifiers using the top SHAP features resulted in enhanced classifier performance, yielding an AUC of 0.86. Based on SHAP values, the most impactful features influencing prediction were mild cognitive impairment (MCI), onset age, and the number of memory tests. These offer a comprehensive patient profile in early-stage memory disorders, as supported by clinical consultation.

4 Discussion

In conclusion, this study investigates the early identification of memory disorders among elderly individuals, leveraging data from social and healthcare services in Kuopio. Utilizing ANOVA, MI feature ranking, and SHAP techniques, we identified onset age, memory disorder-related codes, and high service utilization as risk features from 6 months to 4 years prior to the actual disease onset. Logistic regression and SHAP-guided gradient-boosting classifiers demonstrated promising predictive performance, highlighting the potential for proactive intervention and targeted monitoring in individuals at risk of memory disorders.

References

1. Pathan, M.S., Nag, A., Pathan, M.M., Dev, S.: Analyzing the impact of feature selection on the accuracy of heart disease prediction. Healthc Analytics **2**(1), 100060 (2022). https://doi.org/10.1016/j.health.2022.100060
2. Mian Qaisar, S., Hussain, S.F. An effective arrhythmia classification via ECG signal subsampling and mutual information based subbands statistical features selection. J. Ambient Intell. Human Comput. **14**, 1473–1487 (2023). https://doi.org/10.1007/s12652-021-03275-w
3. Lundberg, S.M., Erion, G., Chen, H., et al.: From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. **2**, 56–67 (2020). https://doi.org/10.1038/s42256-019-0138-9

Developing a Patient Safety CDSS Platform Using Healthcare MyData

Wona Choi[1,2], Sung goo You[1,3], Dai-Jin Kim[4,5], and In Young Choi[1(✉)]

[1] Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul 06591, Republic of Korea
iychoi@catholic.ac.kr
[2] Department of Nursing, Graduate School, The Catholic University of Korea, Seoul 06591, Republic of Korea
[3] Department of Biomedicine and Health Sciences, The Catholic University of Korea, Seoul 06591, Republic of Korea
[4] Department of Psychiatry, College of Medicine, Seoul St. Mary's Hospital, Catholic University of Korea, Seoul 06591, Republic of Korea
[5] Department of Medical Informatics, College of Medicine, Seoul St. Mary's Hospital, Catholic University of Korea, Seoul 06591, Republic of Korea

Abstract. This study explores the development of a Clinical Decision Support Systems (CDSS) platform powered by healthcare MyData, aiming to enhance patient safety in medication and surgical interventions. Leveraging digital technology, particularly AI and CDSS, the research focuses on creating algorithms to manage hyperglycemia, predict Acute Kidney Injury (AKI), manage antithrombotic therapy, and assess postoperative infection risks. Through a structured methodology involving algorithm design, data simulation, expert review, and integration planning, the study lays the groundwork for a CDSS platform that utilizes patient-controlled health information. Preliminary assessments based on simulated data and expert feedback indicate promising potential for these algorithms to significantly improve medication and surgical safety. The integration of personalized patient data through MyData is expected to provide tailored and proactive safety alerts, marking a significant advancement in patient-centered care. This abstract summarizes our initial steps towards empirical testing and real-world application, highlighting the critical role of innovative digital tools in addressing healthcare safety challenges.

Keywords: MyData · Personal health record · Patient safety · Clinical decision support system

1 Background

In the wake of global initiatives to safeguard individual rights in personal data usage, governments are increasingly focusing on empowering citizens with control over their information. In South Korea, this movement is epitomized by 'MyData', a concept gaining traction worldwide under various designations. This paradigm shift from disease treatment to health prevention necessitates the integration of digital healthcare

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 503–505, 2024.
https://doi.org/10.1007/978-3-031-59091-7

tools like Clinical Decision Support Systems (CDSS). Research indicates that CDSS can significantly reduce medication errors, a leading cause of patient safety incidents, and enhance clinical outcomes.

2 Objective

Our research aims to leverage patient-controlled health information within the framework of 'MyData' to develop and validate Artificial Intelligence (AI) and CDSS technologies for patient safety management.

3 Methods

Our study proposes a theoretical framework and development strategy for a CDSS platform using healthcare MyData, aimed at improving medication and surgical safety. The methodology involves a structured process:

- Algorithm Design: Develop algorithms for managing hyperglycemia, predicting AKI, managing antithrombotic therapy, and assessing postoperative infection risks, based on clinical guidelines and risk factors.
- Data Simulation: Generate simulated patient datasets to test algorithm efficacy across diverse clinical scenarios.
- Expert Review: Engage healthcare professionals to evaluate the clinical relevance and impact of these algorithms.
- Integration Planning: Outline a framework for incorporating the algorithms into the MyData platform, detailing data flow, interface, and security protocols.
- This foundation prepares for empirical testing and real-world application.

4 Results

- Hyperglycemia Management Algorithm: Utilizes patient data and medication profiles to predict and alert healthcare providers about potential high blood sugar risks.
- AKI Prediction Algorithm: Analyzes patient-specific factors and current medications to forecast AKI risks, emphasizing age-related susceptibilities.
- Antithrombotic Therapy Management Algorithm: Assesses patient surgery and procedure histories to alert on antithrombotic prescriptions, considering bleeding risks.
- Postoperative Infection Prediction Algorithm: Leverages surgical history and patient demographics to predict infection risks post-lower limb surgeries.

This study will involve developing a platform that integrates these four algorithms with patients' MyData, to be piloted across four hospitals to 400 patients. As the platform is currently under development, empirical results are not available. However, preliminary assessments based on simulated data and expert feedback have been promising. The algorithm design phase has successfully identified key parameters and

risk factors associated with each health concern. Simulated testing scenarios have demonstrated the algorithms' potential to accurately flag risks, with initial accuracy estimates based on expert feedback suggesting a high degree of potential clinical utility. These findings, while preliminary, indicate that the developed algorithms could significantly contribute to medication and surgical safety once fully implemented and tested in real-world settings. The expert review process has also highlighted areas for further refinement, ensuring that the algorithms are both clinically relevant and aligned with current medical practices.

5 Conclusion

This study presents the initial steps towards the development of a MyData-powered CDSS platform aimed at improving patient safety through enhanced medication and surgical interventions. While actual implementation and testing have yet to be conducted, the theoretical groundwork and preliminary feedback suggest that such a platform has the potential to address critical safety issues in healthcare. The integration of personalized patient data through MyData enhances the platform's ability to deliver tailored and proactive safety alerts, embodying a significant advance in patient-centered care. Moving forward, the focus will be on completing the development of the platform, conducting rigorous empirical testing, and refining the algorithms based on real-world data and outcomes. This research underscores the importance of innovative digital tools in advancing patient safety and sets the stage for transformative changes in healthcare delivery.

Acknowledgement. This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HI23C0476).

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. US. Food and Drug Administration. Clinical decision support software: guidance for industry and Food and Drug Administration staff (2022)
2. Choi, W., Chun, J.W., Lee, S.J., Chang, S.H., Kim, D.J., Choi, I.Y.: Development of a MyData platform based on the personal health record data sharing system in Korea. Appl. Sci. **11**(17), 8208 (2021)
3. WHO Patient, World Health Organization: Patient safety curriculum guide: multi-professional edition (2011)

Navigating the Intensive Care Dilemma: Striking a Harmonious Balance in AI Automation

Irina Atkova[1(✉)], Miia Jansson[2,3], Petri Ahokangas[1], Henna Härkönen[2], and Gillian Vesty[3]

[1] Oulu Business School, University of Oulu, Oulu, Finland
irina.atkova@oulu.fi
[2] Research Group of Health Sciences and Technology, University of Oulu, Oulu, Finland
[3] RMIT University, Melbourne, Australia

Keywords: AI · Business Model · Automation · Augmentation · ICU

1 Background

The unprecedented possibilities brought about by the combination of Artificial Intelligence (AI) and the exponential growth in available data have ushered in a new era of task automation and augmentation. This technological advancement enables the automation of routine operations and facilitates the integration of AI-based solutions into managerial tasks, thereby augmenting the decision-making process. However, this transformative landscape is not without its challenges, with issues such as unemployment, competence deterioration, and algorithmic discrimination surfacing. Existing research contends that the nature of a task is a critical determinant in deciding whether it should be automated, augmented, or left to human execution. Yet, this body of knowledge falls short in addressing the crucial aspect of striking a balance between human and machine involvement in decision processes and subsequent task execution. Within this context, the medical field emerges as a particularly promising domain for augmentation, given the longstanding use of AI applications to assist clinicians in diagnoses.

2 Materials and Methods

Adopting a business model approach and drawing insights from a single case study within the intensive care context, our research challenges the notion that the nature of a task is the sole factor influencing the choice of AI application.

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 506–507, 2024.
https://doi.org/10.1007/978-3-031-59091-7

3 Results

We contend that, in addition to the nature of the task, factors such as the sequence of tasks, the organizational level of a task, and timing should be equally considered when deciding on task automation, augmentation, or human execution.

4 Conclusions

Theoretical implications of this study suggest approaching the automation-augmentation duality as a continuum, recognizing the dynamic interplay between the two. From a practical standpoint, our study offers a simple rule that empowers organizations to sense and respond to opportunities in the emerging and continuously evolving AI-enabled landscape.

Automatic Measurement of Body Composition from Computed Tomography Scans of Colon Cancer Patients

Helinä Heino[1(✉)], Reetta Häivälä[2], Egor Panfilov[1], Tero Rautio[1,2], Anne Tuomisto[1,2], Päivi Sirniö[1,2], Juha Saarnio[1,2], Sanna Meriläinen[2], and Aleksei Tiulpin[1,2]

[1] University of Oulu, Oulu, Finland
helina.heino@oulu.fi
[2] Oulu University Hospital, Oulu, Finland
reetta.haivala@pohde.fi

Abstract. Sarcopenia (i.e. involuntary muscle loss) has significant association with adverse health outcomes in tumor patients and elderly populations. On the other hand, Computed Tomography (CT) imaging is a routine part of patient care pathway in gastrointestinal malignancies, which leads to availability of such data for predictive modeling. Body composition (BC) metrics can be derived from CT data, but it laborious when done manually. Thus, automatic extraction of BC metric could provide value in clinical practice by allowing for rapid and repeatable assessment of tissue content from already existing CT data. Our study performed an external validation of two automatic BC measurement methods based on Deep Learning by leveraging data from preoperative venous-phase CT scans from Oulu University Hospital colon cancer patients. As a result, both methods showed promising results in analyzing colon cancer patients CT data when compared to a manual BC measurement workflow.

Keywords: Deep Learning · Computed Tomography · Body Composition

1 Introduction

Body composition (BC) metrics reflect the amount of muscle, subcutaneous fat, and visceral fat, and they are associated with multiple health conditions [1, 2]. In case of sarcopenia, i.e. involuntary muscle loss, BC is known to be connected to adverse health outcomes [3]. BC metric for sarcopenia assessment can be derived using different methods, where Computed Tomography (CT) imaging can be seen as a gold standard [4]. For colon cancer patients, several CT scans are typically acquired during their patient care pathway. Thus, use of automatic BC extraction methods for such patients would be beneficial, as necessary data already exist.

H. Heino, R. Häivälä, S. Meriläinen and A. Tiulpin — Equal contribution

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 508–510, 2024.
https://doi.org/10.1007/978-3-031-59091-7

2 Materials and Methods

Our study performed an external validation of two existing methods for automatic tissue segmentations to derive BC metrics. Analyzed data comprised preoperative venous-phase CT scans from Oulu University Hospital colon cancer patients from years 2019–2022 (n = 109, pixel spacing: [0.66–0.98 mm], slice thickness [0.60–1.25 mm]). From every CT scan, a single axial image from the third lumbar level (L3) was selected to assess muscle and fat distribution. Ground truth (GT) BC measurements (muscle: n = 107, subcutaneous fat: n = 88, visceral fat: n = 109) were done by an experienced research nurse, who traced tissues manually from axial (middle L3) CT images by using predetermined Hounsfield units (HU) windows (muscle: −29 to 150 HU, fat: −190 to -30 HU). Both studied automatic methods, namely Comp2Comp [5] and TotalSegmentator [6], relied on Deep Learning, however, they had different models for segmenting tissues. Comp2Comp method comprised model to detect middle axial image from the L3 level, model to perform tissue segmentation, and post-processing pipeline to obtain the BC metrics from the predictions. TotalSegmentator method provided a model to segment necessary tissues, but it did not contain a subsequent routine to derive the BC metrics. Thus, middle L3 slices were identified by Comp2Comp, segmented by TotalSegmentator, and obtained segmentations were processed by post-processing pipeline similar to Comp2Comp. Correlation (Pearson's *r*) and Bland-Altman analysis were applied to quantify how well the two studied methods performed in comparison to GT.

3 Results

The evaluation for BC metrics derived from both studied methods is presented in Table 1. The results contain metrics in case of muscle (n = 107), subcutaneous fat (n = 88), and visceral fat (n = 109). Comp2Comp received more agreement in muscle, and TotalSegmentator in visceral fat when analyzing results from Bland-Altman analysis. TotalSegmentator obtained better correlation when analyzing results from visceral fat, and it achieved same result in case of subcutaneous fat when compared to Comp2Comp.

Table 1. Automatic models' performance evaluated via correlation (Pearson's *r*) and Bland-Altman analysis (values are mean difference [limits of agreement], difference = GT-prediction).

Analyzed tissue	Method	*r*	Agreement, cm^2
Muscle	Comp2Comp	0.97	−2.27 [−18.12, 13.58]
	TotalSegmentator	0.91	11.50 [−13.31, 36.31]
Subcutaneous fat	Comp2Comp	0.98	−6.26 [−37.41, 24.89]
	TotalSegmentator	0.98	−14.96 [−47.82, 17.89]
Visceral fat	Comp2Comp	0.93	3.71 [−85.86, 93.27]
	TotalSegmentator	0.98	5.38 [−39.50, 50.27]

4 Discussion

Our study performed an external validation of two automatic BC extraction methods, namely Comp2Comp and TotalSegmentator. Both methods achieved excellent correlation (Pearson's *r*) when compared to GT. When considering results from Bland-Altman analysis, Comp2Comp is in more agreement in muscle, while TotalSegmentator - in visceral fat. As a conclusion, there could be a possibility to get a better method by getting the best of two, e.g., by retraining on the combined data, or just fusing on two methods. In general, both validated Deep Learning-based methods showed promising results when CT images from colon cancer patients from Oulu University Hospital were analyzed.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Brown, J.C., et al.: Body composition and cardiovascular events in patients with colorectal cancer: a population-based retrospective cohort study. JAMA Oncol. **5**(7), 967–972 (2019). https://doi.org/10.1001/jamaoncol.2019.0695
2. Al-Sofiani, M.E., et al.: Body composition changes in diabetes and aging. J. Diabetes Complications **33**(6), 451–459 (2019). https://doi.org/10.1016/j.jdiacomp.2019.03.007
3. Boutin, R.D., et al.: Sarcopenia: current concepts and imaging implications. AJR Am. J. Roentgenol. **205**(3), 255–266 (2015). https://doi.org/10.2214/AJR.15.14635
4. Tagliafico, A.S., et al.: Sarcopenia: how to measure, when and why. Radiol. Med. (Torino) **127**(3), 228–237 (2022). https://doi.org/10.1007/s11547-022-01450-3
5. Blankemeier, L., et al.: Comp2Comp: open-source body composition assessment on computed tomography. arXiv preprint, (2023). arXiv:2302.06568
6. Wasserthal, J., et al.: TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. Radiol. Artif. Intell. **5**(5) (2023). https://doi.org/10.1148/ryai.230024

Parental Views on the Digital Solution for Children's Day Surgery Pathway

Arja Rantala[1(✉)], Miia Jansson[2,3], Otto Helve[4], Pekka Lahdenne[5], Minna Pikkarainen[6], and Tarja Pölkki[2,3]

[1] Oulu University of Applied Sciences, Oulu, Finland
arja.rantala@oamk.fi
[2] Research Unit of Health Sciences and Technology, University of Oulu, Oulu, Finland
[3] MRC Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland
[4] Pediatric Research Center, Department of Pediatrics, Helsinki University Hospital, University of Helsinki, Helsinki, Finland
[5] Department of Pediatrics, Helsinki University Hospital, Helsinki, Finland
[6] Faculty of Health Sciences /Centre and Faculty of Technology, Art and Design (TKD), Oslo Metropolitan University, Oslo, Norway

Keywords: Children · Day Surgery · Digital Solution · Gamification · Nursing · Technology

1 Background

Day surgery is 12–24 h cost-effective pathway [1, 2] including various of nurse-delivered counselling throughout the pathway, such information of children's fasting time [3] and post-operative pain counselling [4]. Those elements could be delivered via digital solutions [5, 6]. However, there is not enough information about parental views on digital solutions of the information needed in solutions. The aim of the study was to describe parental views on the digital solution for children's day surgery pathway.

2 Materials and Methods

The participants (N = 31) were parents whose children (under 16 years old) were admitted to the hospital for the day surgical treatments or magnetic resonance imaging in one university hospital in Finland. The inclusion criteria were as follows: parent or custodian of a child who was receiving a day surgical treatment at the selected hospital, ability to understand and write in Finnish, and access to a laptop or a mobile app for answering the questionnaire. The data were collected through an unstructured, open-ended questionnaire, and an inductive content analysis was conducted to analyze the qualitative data.

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 511–512, 2024.
https://doi.org/10.1007/978-3-031-59091-7

3 Results

The parental views on a digital solution were identified as the digital gaming solution for children and families to support care. This main category included three generic categories: 1) preparing children and families for the day surgery enabling virtual familiarization with care environment and waiting time via solution 2) gamification in the solution to support of care and overcome hospital anxiety and fear and 3) connecting people through the solution including interaction between medical staff and families as well as children's peer support via solution.

3.1 Conclusions

Families need relevant information about children's day surgery via a digital solution. Parents are ready and are open to digital gaming solutions that provide support and guidance and engage children in the day surgery pathways. A digital gaming solution may be a relevant tool to support communication between families and healthcare professionals and to provide versatile information about day surgeries.

References

1. Bailey, C.R., Ahuja, M., Bartholomew, K., Bew, S., Forbes, L., Lipp, A., et al.: Guidelines for day-case surgery. Anaesthesia **6**, 778–792 (2019). https://doi.org/10.1111/anae.14639
2. Rauta, S., Reponen, E.: Päiväkirurgian tavoitteet ja potilasvalinta. Duodecim (2023)
3. Rantala, A., Pikkarainen, M., Pölkki, T.: Health specialists' views on the needs for developing a digital gaming solution for paediatric day surgery: a qualitative study. J. Clin. Nurs. **29**, 3541–3552 (2020). https://doi.org/10.1111/jocn.15393
4. Hui, W.J., et al.: Parental experiences while waiting for children undergoing surgery in Singapore. J. Pediatr. Nurs. **52**, e42–e50 (2020). https://doi.org/10.1016/j.pedn.2020.01.004
5. Kerimaa, H., et al.: Effectiveness of a mobile app intervention for preparing preschool children and parents for day surgery: randomized controlled trial. J. Med. Internet Res. **25**, e46989 (2023). https://doi.org/10.2196/46989
6. Rantala, A., Pikkarainen, M., Miettunen, J., He, H., Pölkki, T.: The effectiveness of web-based mobile health interventions in paediatric outpatient surgery: a systematic review and meta-analysis of randomized controlled trials. J. Adv. Nurs. **76**, 1949–1960 (2020). https://doi.org/10.1111/jan.14381

Preliminary Findings from Community Health Workers' Experiences of a Digital Solution for Community-Based Rehabilitation in Primary Health Care

Michael Oduor[1(✉)], Eeva Aartolahti[1], Juliette Gasana[2], Hilmi Zadah Faidullah[3], David Tumusiime[2], and Katariina Korniloff[1]

[1] Institute of Rehabilitation, Jamk University of Applied Sciences, Jyväskylä, Finland
michael.oduor@jamk.fi
[2] College of Medicine and Health Sciences, University of Rwanda, Kigali, Rwanda
[3] Universitas Aisyiyah Yogyakarta (UNISA), Yogyakarta, Indonesia

Abstract. This study investigated community health workers' (CHWs) experiences of Physitrack's Inclusion App in Indonesia and Rwanda. The study is part of a global project to enhance rehabilitation access in low- and middle-income countries through a digital-first approach in primary health care. Real-world pilot testing of the application begun in May 2023. Local rehabilitation professionals pre-installed exercise programs for the most common conditions among service users in the application. CHWs, with limited prior knowledge of digital rehabilitation technologies, found the application user-friendly and beneficial. However, they requested more training on the exercise programs and identifying rehabilitation needs in the community. Implementing digital rehabilitation services into primary health care requires further research, CHW competency development, public outreach, and supportive policies from governmental institutions.

Keywords: Attitudes Towards Technology · Community Health Workers · Digital Rehabilitation · Primary Health Care · Usability

1 Introduction

At least one in every three people worldwide will need rehabilitation at some point in the course of their illness or injury [1]. The demand for rehabilitation services already exceeds the resources, leaving a large unmet need, especially in low- and middle-income countries (LMICs) with the highest burden of disease worldwide [2, 3]. As countries move towards integrated person-centered care, it is imperative that quality rehabilitation is embedded in service delivery models [4].

To enhance access to rehabilitation services in low-resource settings, health systems should be strengthened and reformed. Innovative technology, interprofessional

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 513–515, 2024.
https://doi.org/10.1007/978-3-031-59091-7

teamwork, and task shifting are recommended as catalysts for the reform [5]. This study aimed to investigate Rwandan and Indonesian community health workers' computer proficiency, attitudes towards health technology and their experiences of using a digital rehabilitation application, Physitrack's Inclusion App. The study is part of a global project that aims to increase access to rehabilitation in low- and middle-income countries by proposing a digital-first approach to providing rehabilitation services in primary health care.

2 Materials and Methods

Real-world pilot testing of the application begun in May 2023. Prior to this, local rehabilitation professionals in Rwanda and Indonesia identified the most common conditions among rehabilitation service users, and exercise programs for these conditions were pre-installed in the application. Community health workers (CHWs) received basic training and advice on sharing these exercise programs.

A purposively selected sample of CHWs (n = 53) responded to questionnaires about their computer proficiency, the application's usability, attitude towards information technology for health and participated in interviews and focus group discussions (FGD) six months later. None of the respondents had previous knowledge of digital rehabilitation technologies and many had limited experience with information technology.

3 Results

Overall, CHWs had a positive experience of using the Inclusion App. It was user-friendly and enhanced their interaction with end users. However, they specifically requested additional training on the exercise programs in the application, sharing these programs and identifying community members in need of rehabilitation.

4 Discussion

More research and digital competency development of CHWs, public outreach programs, and supportive health policies from governmental institutions are required to implement digital rehabilitation services in primary health care.

References

1. Cieza, A., Causey, K., Kamenov, K., Hanson, S.W., Chatterji, S., Vos, T.: Global estimates of the need for rehabilitation based on the Global Burden of Disease Study 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. **396**, 2006–2017 (2020). https://doi.org/10.1016/S0140-6736(20)32340-0

2. Chimatiro, G.L., Rhoda, A.J.: Scoping review of acute stroke care management and rehabilitation in low and middle-income countries. BMC Health Serv. Res. **19**, 789 (2019). https://doi.org/10.1186/s12913-019-4654-4
3. Kamenov, K., Mills, J.-A., Chatterji, S., Cieza, A.: Needs and unmet needs for rehabilitation services: a scoping review. Disabil. Rehabil. **41**, 1227–1237 (2019). https://doi.org/10.1080/09638288.2017.1422036
4. Stucki, G.: Advancing the rehabilitation sciences. Front. Rehabil. Sci. **1**, 617749 (2021). https://doi.org/10.3389/fresc.2020.617749
5. Snyman, S., et al.: The ICanFunction mHealth Solution (mICF): a project bringing equity to health and social care within a person-centered approach. J. Interprof. Workforce Res. Dev. **2**, 1–17 (2019)

Patient and Healthcare Professional Experiences on the Barriers, Facilitators, and Solutions of Digital Counseling Materials for Patients with Cerebrovascular Diseases

Kirsi Myllykangas[1,2(✉)], Henna Härkönen[1,2], Maria Kääriäinen[1,2], Mikko Kärppä[2,3], and Miia Jansson[1,2]

[1] Research Unit of Health Sciences and Technology, University of Oulu, Oulu, Finland
kirsi.myllykangas@oulu.fi
[2] Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland
[3] Neurocenter, Neurology, Oulu University Hospital and Research Unit of Clinical Medicine, Neurology, University of Oulu, Oulu, Finland

Abstract. The leading risk-factors of stroke, (e.g. hypertension), are related to people's lifestyle and habits, such as exercise, diet, smoking, and alcohol consumption. Patient counseling aims to positively affect the daily choices people make, but there is still only limited evidence on the effectiveness of different counseling solutions. Patient and healthcare professional experiences are needed in developing evidence based digital solutions to reduce the burden of stroke. This study is a part of a larger research project that co-creates and validates stroke prevention and diagnostic solutions together with different stakeholders.

Keywords: Nursing · Cerebrovascular Diseases · Stroke · Counseling · Patient Education · Patient Experience · eHealth · Digital Counseling

1 Background

The leading risk-factors of stroke, (e.g. hypertension), are related to people's lifestyle and habits, such as exercise, diet, smoking, and alcohol consumption. Patient counseling aims to positively affect the daily choices people make, but there is still only limited evidence on the effectiveness of different counseling solutions. Patient and healthcare professional experiences are needed in developing evidence based digital solutions to reduce the burden of stroke. This study is a part of a larger research project that co-creates and validates stroke prevention and diagnostic solutions together with different stakeholders. The aim of the study was to describe patient and staff experiences on the barriers, facilitators, and solutions of digital counseling materials for patients with cerebrovascular diseases.

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 516–517, 2024.
https://doi.org/10.1007/978-3-031-59091-7

2 Materials and Methods

We conducted semi-structured face-to-face interviews for 22 patients with cerebrovascular diseases (CVD) and 26 healthcare professionals on an acute stroke ward and on a CVD diagnostic and rehabilitation ward in a single university hospital in Finland. Data were analyzed deductively.

3 Results

According to the participants, digital materials were rarely used in patient counseling. The healthcare professionals reported lack of knowledge of digital counseling materials and challenges in sharing digital content to patients. Both patients and healthcare professionals wished for new high quality and interest awakening digital materials with multimedia content, plain language, and moderate length of information. The participants also wished for new counseling software and applications with easy-to-use search functions, two-way communication possibilities and reminders.

4 Conclusions

Both patients and healthcare professionals feel there is a need for the development of new digital counseling materials for patients with CVD. In the development, attention should be paid to visual appearance and accessibility to increase patient motivation to engage with the materials. Availability, usability and shareability of digital counseling materials should be supported for both the patients and professionals.

Legitimacy Challenges of MyData Approach to Personal Data Management Within the Context of GDPR Implementation

Aigerim Dairabekova(✉)

University of Oulu, 90570 Oulu, Finland
aigerim.dairabekova@oulu.fi

Abstract. Due to the increasing amount of data, paired with global digitalization, it is becoming more and more important to utilize this data for commercial and non-commercial purposes. Considering that much of this data is personal, it enables the remote tracking of people's lifestyles and monitoring of patients' health inside and outside of health organizations using different utilities. As such, this calls for an imperative shift towards human-centric data management, with the main idea of individuals controlling their personal data. This paper discusses the concept of MyData approach to personal data management and explores its convergence with the GDPR in Europe, which has created barriers to assessing personal data but has had a great impact on the regulation of the data and increased the transparency of the fair use of personal health data. This paper delves into the legitimacy challenges of implementing the MyData platforms within the framework of GDPR conformity. The paper also presents a framework for legitimation of a human-centric approach to data management and gives recommendations on improving GDPR norms.

Keywords: Human-centric personal data · Healthcare · MyData · GDPR

1 Introduction

With the entrance of big data into many commercial and non-commercial aspects of the world, it is becoming more and more crucial to study the construction and implementation of data-driven business models and learn to apply them [1]. Considering that much of this data appears to be personal, it can be used as a decision-making basis at a societal level. Unfortunately, practice shows that a great deal of this data is stored only with a small number of organizations, which limits the proper utilization of this data [2]. This problem could be solved by shifting from an organization-centered approach to management into human-centered personal data management and placing every individual in control of their own personal data [3]. The aim is to enable interoperability of personal data from different sectors of industry where fast-paced new digital business models are being introduced with new technological advancements. This, in turn, arises challenges of privacy that led European Union (EU) to passing the General Data Protection Regulation (GDPR) that has been effective since May 2018. The main

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 518–520, 2024.
https://doi.org/10.1007/978-3-031-59091-7

objective for introducing GDPR was to return control over the personal data to the citizens and residents and unify the existing regulations within the EU [4].

But how much and what type of data is appropriate for full utilization for certain purposes while providing personal data protection? This calls for a proper platform for intersectoral data interoperability with a predefined architecture that would give the premises for the successful creation of more personalized services. This was resolved by introducing the MyData model of balancing personal data usability and protection. MyData is a consent-based data management and control tool that is an infrastructure-level approach for ensuring data interoperability and portability. This novel procedural approach combines the digital rights of individuals with the needs of organizations and industries. The main pillars of the MyData model are human-centricity, data usability, and open business environment principles. These principles suggest that developing a human-centric approach to personal data management would always be the most practical and profitable [2]. While GDPR unifies EU data protection regulation, MyData provides GDPR-compliant architecture, tools, and practices. While GDPR aims at strengthening and clarifying practices for data security, the human-centric principles of MyData aim at enabling new services for personal data usage. Together they provide improved privacy and ensure trusted and fair utilization of data between organizations [5]. Hence this paper aims to analyze how GDPR is aligned with MyData concept, how they are legitimated and co-legitimated in the context of healthcare data management.

2 Materials and Methods

This study focuses on legitimizing, the process of moving from isolated incidents of a new practice to widespread acceptance. One of the principal approaches to assessing legitimacy is the normative approach, which sets out criteria for three types of legitimation: input, output, and throughput legitimation. This study reviews the literature on legitimation and applies input, output, and throughput legitimacy terms to approach the challenges of legitimizing MyData health platform within the context of GDPR implementation.

3 Results and Recommendations

The results of the analysis show that sharing by patients their personal (health) data through the MyData platform can be legitimized with GDPR norms in place, only if all input, output and throughput legitimacy is achieved. Throughput legitimacy in this case shows that a transparent way of policymaking through public negotiation and decision-making processes is needed. The study also proposes recommendations that are likely to improve the level of GDPR compliance of enterprises, which will, in turn, contribute to achieving the desired level of trust among the public. Raising trust will help gain full legitimation with the public, which will lead to better utilization of personal data. It is likely that non-compliance with GDPR of some enterprises will lead to poor usage of

personal data, which is fraught with misuse of third-party personal data, undercapacity usage of possible data, etc.

4 Discussion

The integration of citizen-centered data into a digital health platform increases the level of citizen involvement and their attitude towards technology and system use [6]. The MyData transition process requires the development of a mindset of engaging in self-care and preventive health for both health professionals and citizens. Maturation of existing platform technologies and mapping out how they will comply with regulations such as EU GDPR is a vital subject to explore in order for MyData platform to function properly. By giving individuals the power to determine how their data can be used, MyData approach enables the collection and use of personal data in ways that maximize the benefits gained and minimize the loss of privacy. GDPR and human-centric principles are complementary to each other. Hence, it is vital to ensure GDPR conformity of the enterprises involved and MyData platform operators to achieve full legitimacy. Citizens' personal health management awareness is enhanced by achieving full legitimacy.

Acknowledgments. This paper was written as part of the LNETN project, which 'has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (grant agreement No 860364). This communication reflects only the author's view and that the Agency is not responsible for any use that may be made of the information it contains.

Disclosure of Interests. None

References

1. Brownlow, J., Zaki, M., Neely, A. Urmetzer, F.: Data-driven business models: a blueprint for innovation (2015)
2. Poikola, A., Kuikkaniemi, K., Kuittinen O., Honko, H., Knuutila A., Lahteenoja V.: MyData – an introduction to human-centric use of personal data, 3rd edn. Helsinki, Finland: Finish Ministry of Transport and Communications (2020)
3. Iivari, M., Pikkarainen, M., Koivumäki, T.: How MyData is transforming the business models for health insurance companies. In: Camarinha-Matos, L.M., Afsarmanesh, H., Fornasiero, R. (eds.) PRO-VE 2017. IAICT, vol. 506, pp. 323–332. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-65151-4_30
4. Lindgren, P.: GDPR regulation impact on different business models and businesses. J. Multi Business Model Innov. Technol. **4**, 241–254 (2018). https://doi.org/10.13052/jmbmit2245-456X.434
5. Su, X., et al.: Privacy as a service: protecting the individual in healthcare data processing. Computer **49**(11), 49–59 (2016). https://doi.org/10.1109/MC.2016.337
6. Ahmed, I., Jeon, G., Chehri, A.: An IoT-enabled smart health care system for screening of COVID-19 with multi layers features fusion and selection. Computing **10**, 1–18 (2022)

Monte Carlo Simulation of Implanted Intracranial fNIRS

Netaniel Rein[1,2(✉)], Mordekhay Medvedovsky[1,2], and Michal Balberg[3]

[1] Department of Neurology and Agnes Ginges Center for Human Neurogenetics, Hadassah Medical Organization, Jerusalem, Israel
Netaniel.rein@mail.huji.ac.il
[2] The Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel
[3] Holon Institute of Technology, Holon, Israel

Abstract. Functional Near Infrared Spectroscopy (fNIRS) is limited to the external layers of the brain, and is impeded by motion artifacts and hemodynamic "noise" from scalp. Often, prior to epilepsy surgery, depth-electrodes are implanted, yet this approach is limited by "tunnel vision" effect. To overcome these limitations, we developed an implanted fNIRS (ifNIRS) sensor using optical fibers. To simulate the use of this sensor we employed a Monte Carlo (MC) method. When compared to standard scalp positioning (3 cm distance between emitter and detector), the simulated intracranial fNIRS resulted in a >20-fold increase in measured signal at 3 cm distance and a >4-fold increase at 5cm. Simulations with different scatter coefficients of the white matter demonstrated substantial differences in measurable signal for intracranial fNIRS. MC simulations highlighted notable differences between scalp and ifNIRS, emphasizing the advantages of intracranial positioning. We propose that ifNIRS can be combined with stereo EEG or responsive neurostimulation (RNS) to improve seizure detection.

Keywords: Functional Near Infrared Spectroscopy · fNIRS · Intracranial fNIRS · Stereo EEG · Monte-Carlo Simulation

1 Introduction

Functional Near Infrared Spectroscopy (fNIRS) has been used for decades to measure changes in cerebral hemodynamics during various tasks and conditions. However, due to physical and safety limitations of the measurements, the current practical penetration depth is limited to the external layers of the cortex at about 2 cm below the scalp. Patients with drug-resistant epilepsy sometimes undergo surgery as part of their assessment, in which depth-electrodes are implanted into the parenchyma, through the skull, for several days [1]. The electrodes are fixated to the skull using an anchor bolt. The purpose of this implantation is to localize the epileptogenic zone. Electrical fields that originate from the epileptogenic zone may be obscured by electric fields closer to the implanted electrode, resulting in a "tunnel vision" effect. Optical signals in the

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 521–523, 2024.
https://doi.org/10.1007/978-3-031-59091-7

brain, especially in the near-infrared range of the spectrum, are spatially diffuse by nature and therefore may provide information about epileptic activity-related changes in cerebral hemodynamics, further away from the electrode. Such sensing is conducted relatively deep within the brain and requires overcoming the depth limitation of current non-invasive fNIRS. To overcome this limitation, we have designed and built a novel implanted fNIRS (ifNIRS) sensor, comprising an anchor bolt and an electrode with embedded optical fibers, that can replace existing depth electrodes used in stereo-EEG (SEEG). The integration of optical fibers provides several benefits, including the elimination of scalp blood flow contamination, deeper penetration for detecting hemodynamic changes, and reduction of motion artifacts. Despite the clear benefits of ifNIRS, the optimal design of such a device depends on the optical properties of the brain, which have not been determined precisely yet, specifically for white matter (WM). We used Monte Carlo (MC) simulations with varying optical properties to assess the effect of this variability on the optimal configuration of intracerebral electrodes. Subsequent paragraphs, however, are indented.

2 Methods

Using MCXLAB, a five layered slab MC simulation of light propagation through the head was constructed. MCXLAB is a MATLAB (MathWorks, Natick, MA, USA) interfaced version of Monte Carlo eXtreme (MCX) open-source, GPU-accelerated simulation suite2 and employing a pseudo-random number generator. The five layers - scalp, skull, CSF, grey matter (GM) and WM - were of a thickness of 6, 8, 4, 10 and 72 mm respectively. The optical properties of the brain were initially set to the values of the MCXLAB [2] which are based on a study of excised human tissue [3]. Several locations for emitters and detectors were simulated. Firstly, the commonly used scalp positioning of both the emitter and detector was simulated at 1, 2, 3, 4 and 5 cm apart. Secondly, both the emitter and detector were placed near the interior surface of the skull at 2, 3, 4, 5 cm apart, simulating a bolt position. Finally, an emitter was placed 1, 2, 3, 4 and 5 cm intracerebrally to simulate the electrode, while the detector remained near the inner surface of the skull to simulate the bolt. Each of these simulations was run with three reduced scattering coefficients (μs' of 0.85, 2.565 and 6.015 mm^{-1}). The detector's aperture was a disk 0.6 mm in diameter directed internally. The emitter was pencil-shaped and directed intracranially when placed on the scalp and in the skull, and a 2 mm long line-shaped when placed intracranially. $5•10^7$ photons were launched in every simulation. In each simulation, the number of photons traversing each layer, and the average partial path length (APPL) through each layer were recorded. The standard for usable measured signal was set at the result of scalp positioning 3 cm apart.

3 Results

When compared to a simulation of scalp positioned source and detector at 3 cm distance, the number of photons detected in a simulation of a skull-bolt positioned emitter and detector resulted in a >20-fold increase in measured signal at 3 cm distance and a >4-fold increase at 5 cm. At scalp positioning 3 cm apart 100% and 77% of photons passed through the scalp and GM respectively. At bolt positioning at all distances, 30–39% and ~100% of photons passed through the scalp and GM respectively. The ratio of APPL through GM compared to APPL through scalp was 1.6–2.25 for all distances with scalp positioning, 4.1–6.5 with bolt positioning, and 4.1–8.9 with the electrode emitter and bolt detector.

The simulation with scalp positioning and with bolt positioning produced similar results with all scatter coefficients of WM. A substantial difference between simulations was found with the emitter simulated in the brain and the detector positioned in the skull. The usable signal was measured up to 3, 4, 5 cm deep for the different μs' values.

4 Conclusions

The MC simulation demonstrated differences between scalp and intracranial fNIRS in the measured signal, APPL ratio between GM and scalp and percentage of photons traversing GM and scalp. These findings suggest an advantage for positioning fNIRS optodes intracranially. Different scatter coefficients previously reported caused a drop in measured signal. This was previously demonstrated [4], but will pose a specific problem with depth electrode seeing as WM has the greatest variability in reported optical coefficients, soliciting the need for further research to optimize electrode design. In the future, ifNIRS combined with responsive neurostimulation could be implanted chronically and used to prevent seizures.)

Acknowledgements. This study was funded by the Israeli Innovation Authority.

References

1. Gonzalez-Martinez, J., et al.: Stereotactic placement of depth electrodes in medically intractable epilepsy: technical note. J. Neurosurg. **120**(3), 639–644 (2014)
2. Fang, Q., Boas, D.A.: Monte Carlo simulation of photon migration in 3D turbid media accelerated by graphics processing units. Opt. Express **17**(22), 20178–20190 (2009)
3. Yaroslavsky, A., Schulze, P., Yaroslavsky, I., Schober, R., Ulrich, F., Schwarzmaier, H.: Optical properties of selected native and coagulated human brain tissues in vitro in the visible and near infrared spectral range. Phys. Med. Biol. **47**(12), 2059–2073 (2002)
4. Russomanno, E., Kalyanov, A., Jiang, J., Ackermann, M., Wolf, M.: Effects of different optical properties of head tissues on near-infrared spectroscopy using Monte Carlo simulations. Adv. Exp. Med. Biol. **1395**, 39–43 (2022)

Enabling Rapid Multi-locus Transcranial Magnetic Stimulation with Pulse-Width Modulation

Heikki Sinisalo[1(✉)], Mikael Laine[1], Jaakko O. Nieminen[1], Victor H. Souza[1], Matti Stenroos[1], Renan H. Matsuda[1,2], Ana M. Soto[1], Elena Ukharova[1], Tuomas Mutanen[1], Lari M. Koponen[1], and Risto J. Ilmoniemi[1]

[1] Department of Neuroscience and Biomedical Engineering, Aalto University, 02150 Espoo, Finland
heikki.sinisalo@aalto.fi
[2] Department of Physics, Faculty of Philosophy Sciences and Letters of Ribeirão Preto, University of São Paulo, Ribeirão Preto, São Paulo 14040-901, Brazil

Abstract. This study explores the use of pulse-width modulation (PWM) to control stimulation strength of multi-locus transcranial magnetic stimulation (TMS) device for advanced brain stimulation protocols. Our findings from healthy volunteers show similarities in the motor responses elicited by PWM pulses and conventional TMS pulses, with some differences under certain experimental conditions. These findings suggest that coupling PWM technique with multi-locus TMS offers a promising alternative to conventional pulses, enabling more flexible cortical stimulation.

Keywords: Transcranial Magnetic Stimulation · Multi-locus · Pulse-width Modulation · Waveform · Pulse Sequence · Stimulation Strength

1 Introduction

Transcranial magnetic stimulation (TMS) is a non-invasive brain stimulation method that is used to diagnose and modulate brain activity, both in basic research and clinical setting [1]. In TMS, a coil placed on the scalp generates strong magnetic pulses that induce electric field (E-field) in the cortex, activating neurons. As part of the EU-funded ConnectToBrain project, our goal is to expand TMS methodology to enable stimulation of multiple spatially separate cortical targets within the timescale of ongoing neuronal activity.

Stimulating nearby cortical sites requires physically moving the TMS coil, which can take a few seconds. Thus, a multi-locus TMS (mTMS) device was constructed, comprising an array of overlapping coils to electronically manipulate the induced E-

H. Sinisalo and M. Laine — Equal contribution in first authorship.

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 524–527, 2024.
https://doi.org/10.1007/978-3-031-59091-7

field patterns [2]. This feature is crucial for expanding the spatial degrees of freedom with TMS, enabling new stimulation protocols and automated stimulation algorithms.

The ability to generate rapid pulse sequences with the mTMS device for cortical network stimulation has still been a challenge. Conventionally, the stimulation strength of each TMS pulse is controlled by adjusting the amount of electric charge stored in a high-voltage capacitor before releasing it to the stimulation coil. This approach severely limits the speed at which sequential pulses with different strengths can be delivered, as large charge adjustments can take several seconds. To overcome this limitation, we used pulse-width modulation (PWM) to enable stimulation strength adjustment by exerting specific temporal control over the released current waveform. It is difficult to predict how the waveform shape influences the evoked brain activity due to the complexity of the neuronal mechanisms. Thus, we compared PWM and conventional pulses by stimulating the primary motor cortex and measuring the motor responses evoked by each pulse type.

2 Materials and Methods

The study was approved by the Ethics Committee of the Hospital District of Helsinki and Uusimaa. We first developed a computer program that models the power electronics of the mTMS device and turns waveform of a conventional pulse into its PWM counterpart with minimum average coil current deviation. Then, we tested the PWM paradigm on six healthy volunteers. For each participant, we stimulated the primary motor cortex with each pulse type and measured the resting motor threshold (RMT) and motor evoked potential (MEP) amplitude with five stimulation strength levels ranging from 105% to 140% RMT. With each pulse type, the RMT was measured four times using an automated algorithm, and MEP amplitudes 20 times per stimulation strength level. The experiment was repeated with combinations utilizing 2, 3, and 5 coils of the mTMS coil array to measure if the number of overlapping waveforms affects the pulse type differences. During the experiments, the subjects rested on a chair and the motor activity of their hand was measured with surface electrodes attached to the skin over the abductor pollicis brevis muscle.

After the experiments, we measured the E-fields of all delivered pulses with an in-house search coil. Pulse stimulation strength was calculated as the peak of the E-field time integral over the pulse duration. With MEP amplitudes, the stimulation strengths were scaled relative to the average RMT of the conventional pulse. The RMT and MEP amplitude data were analyzed separately with linear mixed-effects models. We tested the effect of pulse type and its interaction with the coil combinations on the stimulation strengths of pulses at RMT intensity. Similarly, we tested the effect of pulse type and its interactions with the coil combinations and stimulation strength on log-transformed MEP amplitudes. Statistical significance was set at $p < 0.05$.

3 Results

Preliminary analysis indicated that the motor responses with the pulse types were similar, but some differences were present at specific experimental conditions. A statistically significant difference in RMT values was found with the 5-coil combination ($t_{123} = -4.55, p < 0.001$), where the PWM pulses had 8% higher stimulation strength than the conventional pulses. With the 5-coil combination, the average relative standard deviation of RMT values was 4% across the subjects and pulse types. Statistically significant difference was also found with MEP amplitudes at stimulation strength of 110% RMT with 2-coil ($t_{3261} = -3.12, p = 0.018$) and 3-coil combinations ($t_{3262} = -2.82, p = 0.045$), where the PWM pulses elicited 28% and 23% greater MEP amplitudes, respectively. The average relative standard deviations of the log-transformed MEP amplitudes with the 2-coil and 3-coil combinations were 24% and 22%, respectively, across the subjects, pulse types, and stimulation strengths of around 110% RMT.

4 Discussion

We demonstrated the use of PWM waveforms with the mTMS device to enable changing the stimulated cortical site with a sub-millisecond delay. Our preliminary results indicate that the motor responses elicited by the PWM pulses and conventional pulses were mostly similar. The RMT values were 8% higher with the PWM pulses when all five coils of the mTMS array were used. However, due to hardware restrictions, low strength PWM waveforms were excluded, potentially lowering E-field strength and inflating the RMT values. The PWM pulses lead to around 25% higher MEP amplitudes at 110% RMT stimulation strength when using coil combinations comprising two or three coils. The MEP amplitudes did not differ when using combination of all five coils. A comparative study with single TMS coil, however, found 8% lower RMT with PWM pulses but no dependency between MEP amplitude differences and stimulation strength [5]. In summary, the PWM pulse technique provides a promising alternative to the conventional pulses and, when used together with an mTMS coil array, enables stimulation of cortical networks in the time scale of ongoing neuronal activity.

References

1. Rossini, P.M., et al.: Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: basic principles and procedures for routine clinical and research application: an updated report from an I.F.C.N. Committee. Clini. Neurophysiol. **126** (6), 1071–1107 (2015)

2. Nieminen, J.O., et al.: Multi-locus transcranial magnetic stimulation system for electronically targeted brain stimulation. Brain Stimul. **15**(1), 116–124 (2022)
3. Memarian Sorkhabi, M., Wendt, K., O'Shea, J., Denison, T.J.: Pulse width modulation-based TMS: primary motor cortex responses compared to conventional monophasic stimuli. Brain Stimul. **15**(1), 980–983 (2022)

Impact of Digital Services on Healthcare Performance - An Umbrella Review

Henna Härkönen[1,2(✉)], Sanna Lakoma[3], Miia Jansson[1,2,4], and on behalf of the research group

[1] Research Unit of Health Sciences and Technology, University of Oulu, Oulu, Finland
henna.harkonen@oulu.fi
[2] MRC Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland
[3] University of Helsinki, Helsinki, Finland
[4] RMIT University, Melbourne, Australia

Keywords: Digital technology · Delivery of healthcare · Health impact assessment

1 Background

Digital services have increased rapidly during the last decade with the potential to address challenges relating to accessibility, availability, and costs of healthcare. Applicability of digital services is currently limited due to heterogenous and low-quality evidence of their impact. This umbrella review aimed to evaluate the impact of digital services on the four aspects of healthcare performance. The research question was: What is the impact of digital services on population health, service costs, and satisfaction of patients and healthcare professionals?

2 Materials and Methods

A search was performed to Centre for Reviews and Dissemination, Cochrane, Ovid Medline, Scopus, and Web of Science in June 2022 with total of 790 studies identified. The methodological quality was assessed. Digital services were identified using a pre-made definition. The impact of digital services was categorized as no evidence, no dominance, and mixed and positive.

3 Results

The review included 66 mostly (64%) high quality studies. The impact of digital services was mixed on population health, with mostly positive impact in some medical specialties. Impact on costs was mixed, with cost reduction reported in many of the

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 528–529, 2024.
https://doi.org/10.1007/978-3-031-59091-7

studies. Impact on patient satisfaction was positive and mixed on healthcare professionals.

4 Conclusions

Digital services can be viable options or additions to many healthcare service contexts. Mixed and potentially positive population health outcomes, high patient satisfaction and cost savings support wider adoption. Mixed healthcare professional satisfaction highlights the need study the implementation of digital services. Varied long-term research is needed to study digital services and their impact mechanisms in healthcare.

Effect of Digital Therapeutics Based on Cognitive Behavioral Treatment for Alcohol Use Disorder: Exploratory Clinical Study

Ji-Won Chun[1], Wona Choi[1,2], In Young Choi[1], and Dai-Jin Kim[3,4(✉)]

[1] Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul 06591, Republic of Korea
[2] Department of Nursing, Graduate School, The Catholic University of Korea, Seoul 06591, Republic of Korea
[3] Department of Psychiatry, Seoul St. Mary's Hospital, College of Medicine, Catholic University of Korea, Seoul 06591, Republic of Korea
kdj922@catholic.ac.kr
[4] Department of Medical Informatics, Seoul St. Mary's Hospital, College of Medicine, Catholic University of Korea, Seoul 06591, Republic of Korea

Abstract. This study assesses a 12-week CBT-based digital therapy (DTx) for Alcohol Use Disorder (AUD) in South Korea, emphasizing digital solutions due to the growing AUD problem and COVID-19 impact. The DTx, integrating mobile apps and virtual reality, focuses on improving awareness of alcohol use triggers and behavioral modification. It includes Motivational Enhancement Therapy to encourage behavior change. The trial compared DTx against traditional therapy in 30 AUD patients. Results showed DTx's superiority in abstinence and reduced alcohol consumption, highlighting its potential as an effective, accessible AUD treatment and a complement to existing methods.

Keywords: Digital Therapeutics · Alcohol Use Disorder · Cognitive Behavioral Treatment

1 Background

Recently, the growth of the digital healthcare service market is accelerating. The need to utilize digital software programs in the field of behavioral modification and chronic diseases is especially being emphasized (Digital Therapeutics Alliance, 2018). In South Korea, drinking is the number one disease burden factor. Also, the prevalence of alcohol use disorder (AUD) and the rate of drunk driving accidents are the highest among the OECD Countries (Korea Health Promotion Institute, 2022). In particular, due to the influence of COVID-19, alcohol-related deaths exceeded 10 per 100,000 people for the first time in 2020 which highlights that alcohol use (AU) is a serious social problem (Korea Health Promotion Institute, 2022). Cognitive behavioral therapy (CBT) is an effective evidence-based treatment for AUD, one of the representative chronic diseases (Magill et Al., 2019). To improve the AU problem, a CBT-based 12-

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 530–532, 2024.
https://doi.org/10.1007/978-3-031-59091-7

week digital program, combination of mobile application and virtual reality, was developed.

2 Objective

The safety and effectiveness of CBT-based DTx, targeted toward patients with AUD, were verified in this exploratory clinical study.

3 Methods

In this study, the CBT-based DTx was designed to improve awareness of cravings and behavioral processes by identifying the situations and emotions that trigger AU. Furthermore, there was included training content to correct dysfunctional and irrational thinking patterns for AU and to cope with trigger factors. In addition, Motivational Enhancement Therapy (MET) was used to enhance the therapeutic effect by promoting changes in AU behavior with consideration of user's motivation. From January to September 2022, an exploratory clinical study was conducted to provide 12 weeks of treatment to improve addiction in 30 patients diagnosed with AUD. The subjects who were qualified to the screening assessment were randomly assigned to either the 'digital therapy group (DTG)' or the 'basic therapy group (BTG)'. DTG had digital therapy that provides digital content education through mobile application and virtual reality. Meanwhile, BTG got basic treatment which provides both written and video educational materials. At the time of 12th week, abstinence from alcohol drinking cessation and average daily alcohol consumption were evaluated.

4 Results

In the 12th week, the effect of abstinence from alcohol drinking cessation was higher in the DTG, utilizing the digital therapeutics, compared to the BTG. The DTG showed a 40% abstinence rate in the 12th week, and the BTG showed a 20% abstinence rate. In addition, the average alcohol consumption per day was 2.7 drinks in the DTG and 4.3 drinks in the BTG which shows a low level of AU in the DTG.

5 Conclusion

Since the COVID-19 pandemic, in-patient treatment for AUD has been continuously decreasing, and psychosocial services for alcoholics are insufficient compared to other diseases. DTX is an alternative treatment that can overcome the practical limitations of existing addiction treatments. The results of this study suggest the potential of improving clinical effectiveness by supplementing drug and outpatient treatment. The CBT-based DTx will increase accessibility to alcoholism treatment and can contribute

to the growth of patient-centered participatory medicine. Also, healthcare workers could monitor the patient's daily condition and reflect DTx in the treatment plan.

Acknowledgement. This work was supported by the Technology Innovation Program (or Industrial Strategic Technology Development Program- 200009227, Development of CBT based Digital Therapeutics for Alcohol and Nicotine Addiction Disorder.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Korea Health Promotion Institute. 2022 Alcohol Harm Prevention Day: Forum for a Society without Alcohol Harm. In: Korea Health Promotion Institute (KHPI) (2022)
2. Magill, M., et al.: A meta-analysis of cognitive-behavioral therapy for alcohol or other drug use disorders: treatment efficacy by contrast condition. J. Consult. Clin. Psychol. **87**(12), 1093–1105 (2019)
3. Digital Therapeutics Alliance. Digital Therapeutics: Combining technology and evidence-based medicine to transform personalized patient care. Digital Therapeutics Alliance (2018). http://www.dtxalliance.org

Effects of a Digital Patient Journey Solution on Patient Outcomes and Health Care Utilization: A Pragmatic Randomized Controlled Trial

Miia Jansson[1,2,3(✉)] and The Icory-consortium

[1] Research Unit of Health Sciences and Technology, University of Oulu, Oulu, Finland
miia.jansson@oulu.fi
[2] MRC Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland
[3] RMIT University, Melbourne, Australia

Keywords: Patient Reported Outcome Measures · Randomized Controlled Trial · Mobile health applications

1 Background

Digital services (e.g., mobile health applications) have been proposed as a promising and safe alternative to usual care. However, the implementation of digital services has been hampered by conflicting results and only moderate- to low-quality evidence [1].

2 Materials and Methods

This rigorous, pragmatic, randomized controlled trial evaluated the short-term effects of a digital patient journey solution on patient outcomes and health care utilization in patients with total hip and knee arthroplasty [2]. Randomly assigned patients in the control group (n = 35) received usual care, while patients in the intervention group (n = 34) received the digital patient journey solution in addition to usual care. The primary outcome measure was the health-related quality of life. The secondary outcome measures included functional recovery, pain, self-efficacy, patient experience, adherence to fast-track protocol, and health care utilization. Self-efficacy was measured using an adapted version of the Healthcare Technology Self-Efficacy Scale. Patients were followed from a pre-operative surgical visit to a post-operative follow-up visit at 6 to 12 weeks after surgery.

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 533–534, 2024.
https://doi.org/10.1007/978-3-031-59091-7

3 Results

During the study, health-related quality of life, functional recovery, pain, adherence to the fast-track protocol, and healthcare utilization did not differ between the study groups. However, self-efficacy to use digital health services increased in the intervention group compared to the control group ($p = 0.027$).

4 Conclusions

Use of the digital patient journey solution was not superior to usual care in terms of patient-reported outcomes and health care utilization. However, the application improved patients' self-efficacy to use digital services, which may lead to greater demand for such services as patients become more familiar with mobile health applications. Future research should explore long-term effects and consider patient preferences in adopting such applications.

References

1. Härkönen, H., et al.: Impact of digital services in healthcare and social welfare - an umbrella review. Int. J. Nurs. Stud. **152**, 1–18 (2024)
2. Jansson, M., et al.: Short-term effects of a digital patient journey solution on patient-reported outcomes and health care utilization in arthroplasty: a pragmatic randomized con-trolled trial. Finnish J. EHealth EWelfare **15**(3), 305–321 (2023)

A Mobile Application Intervention Helps Alleviate Preschool Children's Fear in the Entire day Surgery Service Chain

Heli Kerimaa[1,3(✉)], Mervi Hakala[1,2,3], Marianne Haapea[3,5], Hannu Vähänikkilä[4], Willy Serlo[2], and Tarja Pölkki[1,3]

[1] Research Unit of Health Sciences and Technology, University of Oulu, Oulu, Finland
heli.kerimaa@oulu.fi
[2] Oulu University Hospital, Oulu, Finland
[3] Medical Research Center Oulu, University Hospital and University of Oulu, Oulu, Finland
[4] Northern Finland Birth Cohorts, Arctic Biobank, Infrastructure for Population Studies, Faculty of Medicine, University of Oulu, Oulu, Finland
[5] Research Service Unit, Oulu University Hospital, Oulu, Finland

Abstract. This study evaluates the effectiveness of a mobile application intervention on the fear and pain experienced by preschool children during their preparation for day surgery and describes how parents and nurses assess children's pain.

Keywords: Day Surgery · Fear · Mobile App · Pain · Preparation · Preschool Child · Randomized Controlled Trial

1 Introduction

Around half of all surgical procedures in Finland are conducted as day surgery (DS). This percentage is expected to increase, as well-planned DS is recommended and cost-effective for healthcare organizations and families [6, 9]. DS is a short procedure for children where they are admitted and discharged within a few hours [2]. High-quality preparation materials and cooperation with the parents are crucial for successful DS.

The number of day surgical procedures in preschool children has increased worldwide [7]. Due to their stage of development, preschool children tend to experience more fear related to surgery [8]. Children who experience preoperative fear may also experience emotional distress and increased pain during hospitalization. This can lead to more pain after surgery, requiring more medication and slower recovery [4]. Parents play a vital role in preparing and supporting their children during and after DS. They need to be informed about pain management, including painkillers, to avoid undermedication and fear of side effects [11]. Tailored pain assessment and treatment for each child, involving parents and educating them to assess their child's pain is crucial [5].

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 535–537, 2024.
https://doi.org/10.1007/978-3-031-59091-7

Digital preparation methods can provide an excellent opportunity for DS preparation [1, 4]. However, there is a need for more information on effective and client-oriented interventions that cater to preschool children and their parents. In the past, research on mobile application interventions aimed at reducing pain has primarily focused on either older children [3] or children with cancer [10]. The study aimed to (a) evaluate the effectiveness of a mobile application intervention on the fear and pain experienced by preschool children during their preparation for DS and (b) describe how parents and nurses assess children's pain.

2 Material and Methods

The Randomized Controlled Trial (RCT) was conducted between 2018 and 2020 in a Finnish university hospital's pediatric day surgery department. Preschool children (2–6 years old) who underwent elective DS were randomized into intervention (n = 36) and control (n = 34) groups when one was prepared using a mobile application while others received standard preparation. Data was collected using reliable and valid measures, including FAS (children's fear), PPPM (children's pain behavior), WBS and VAS (children's pain). The children's outcomes were measured at four different times: before and after surgery, measurements were taken at home (T1 and T4, child and parent), and at the hospital (T2 and T3, child, parent, and nurse).

3 Results

The mobile application intervention did not reduce pain levels in preschool children. However, the intervention group did experience a decrease in fear levels over time, while no such change was observed in the control group. Both groups experienced a decrease in fear levels after the surgery, but only the intervention group showed a significant decrease, as per the statistical analysis. Parental assessments revealed that both groups had similar levels of pain, which increased after the surgery. The change was statistically significant only in the control group. Nurses' assessments of pain levels showed no discernible difference between the two groups. However, both groups experienced a statistically significant increase in pain, as assessed by the nurses, which was consistent with parental assessments.

4 Discussion

In this study, neither the intervention group nor the control group of children experienced any significant pain before the surgery. This was confirmed by self-assessment of pain by the children, as well as by assessments provided by nurses and parents. However, after the surgery, both groups of children experienced an increase in pain, which was further confirmed by assessments provided by the children, parents, and nurses. This study demonstrated that preschool children can distinguish between feelings of pain and fear and can accurately evaluate their pain.

Digital interventions, such as mobile application intervention, can be an alternative to traditional preparation methods when preparing preschool children for DS. The intervention did not increase the child's pain; in fact, it reduced the fear experienced by the child throughout the entire day surgery service chain.

References

1. Agbayani, C.-J.G., Fortier, M.A., Kain, Z.N.: Non-pharmacological methods of reducing perioperative anxiety in children. BJA Educ. **20**(12), 424–430 (2020). https://doi.org/10.1016/j.bjae.2020.08.003
2. Bailey, C.R., et al.: Guidelines for day-case surgery 2019. Anaesthesia **74**(6), 778–792 (2019). https://doi.org/10.1111/anae.14639
3. Birnie, K.A., et al.: ICanCope PostOp: user-centered design of a smartphone-based app for self-management of postoperative pain in children and adolescents. J. Med. Internet Res. **21** (4) (2019). https://doi.org/10.2196/12028
4. Copanitsanou, P., Valkeapaa, K.: Effects of education of paediatric patients undergoing elective surgical procedures on their anxiety—a systematic review. J. Clin. Nurs. **23**(7–8), 940–954 (2015). https://doi.org/10.1111/jocn.12187
5. Cohen, L.L., Donati, M.R., Shih, S., Sil, S.: Topical review: state of the field of child self-report of acute pain. J. Pediatr. Psychol. jsz078 (2019). https://doi.org/10.1093/jpepsy/jsz078
6. de Luca, U., Mangia, G., Tesoro, S., Martino, A., Sammartino, M., Calisti, A.: Guidelines on pediatric day surgery of the Italian Societies of Pediatric Surgery (SICP) and Pediatric Anesthesiology (SARNePI). Ital. J. Pediatr. **44**(1), 35 (2018). https://doi.org/10.1186/s13052-018-0473-1
7. Erhaze, E.K., Dowling, M., Devane, D.: Parental presence at anaesthesia induction: a systematic review. Int. J. Nurs. Pract. **22**(4), 397–407 (2016). https://doi.org/10.1111/ijn.12449
8. Fortier, M.A., et al.: Web-based tailored intervention for preparation of parents and children for outpatient surgery (WebTIPS): formative evaluation and randomized controlled trial. Anesth. Analg. **120**(4), 915–922 (2015). https://doi.org/10.1213/ANE.0000000000000632
9. Kleye, I., Hedén, L., Karlsson, K., Sundler, A.J., Darcy, L.: Children's individual voices are required for adequate management of fear and pain during hospital care and treatment. Scand. J. Caring Sci. **35**(2), 530–537 (2021). https://doi.org/10.1111/scs.12865
10. Sarialioğlu, A., Kurudirek, F., Oluç, T.: The effect of storybook reading on children's preoperative fear and anxiety levels: a randomized controlled study. Child: Care, Health Dev. cch.13100 (2023). https://doi.org/10.1111/cch.13100
11. Stinson, J.N., et al.: Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J. Med. Internet Res. **15**(3), e51 (2013). https://doi.org/10.2196/jmir.2350
12. Tam, M.T., et al.: Barriers and facilitators to effective pain management by parents after pediatric outpatient surgery. J. Pediatr. Health Care **34**(6), 560–567 (2020). https://doi.org/10.1016/j.pedhc.2020.06.008

Neonatal Procedural Pain Management Using Digital Solutions – A Study Protocol of a Randomized Controlled Trial with Crossover Design

Anna-Kaija Palomaa[1,2(✉)], Eveliina Seppälä[1,2], Sadegh Moradi[3], Hany Ferdinando[1], Teemu Myllylä[1,2,3], and Tarja Pölkki[1,2]

[1] Research Unit of Oulu Health Sciences and Technology, University of Oulu, Oulu, Finland
apalomaa@student.oulu.fi
[2] MRC Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland
[3] Opto-Electronics and Measurement Technique Research Unit, University of Oulu, Oulu, Finland

Keywords: Digital Intervention · Maternal Presence · Neonate · Neonatal Intensive Care · Non-Pharmacological · Pain · Randomized Controlled Trial

1 Introduction

Neonates are exposed to repeated painful procedures as part of their medical care in neonatal intensive care unit (NICU) [1]. Recurrent and untreated procedural pain in the early life can cause immediate physiological instability and, in the long term, to impair neurological and cognitive development [2]. Therefore, pain assessment and management have a crucial role in neonatal care [3]. Neonatal pain research has focused on the use of non-pharmacological pain relief strategies from the viewpoint of nurses, but research on mother-driven non-pharmacological pain management is still lacking [4, 5]. Skin-to-skin contact (SSC), in which a naked, diaper-dressed neonate is placed on the mother's bare chest, has been shown to reduce neonatal pain intensity and behavioral distress during needle-related procedures [6]. In the SSC the mother herself is a mediator for pain relief, but not all mothers are always able to be present in the neonatal intensive care unit and to provide SSC [7]. In recent years, there has been increased interest in the use of digital solutions to involve mothers in neonatal pain relief, but the evidence of their effectiveness is mixed [8]. There has also been interest in the use of technology to assess pain in neonates, including near-infrared spectrometry (NIRS), alongside pain assessment scales [9].

The aim of this study is to compare the effectiveness of digital solution, so called digital maternal presence intervention, to skin-to-skin contact and oral glucose for pain relief in neonates undergoing a heel lance procedure, and to evaluate the use of technology in pain assessment.

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 538–540, 2024.
https://doi.org/10.1007/978-3-031-59091-7

2 Materials and Methods

This study is a randomized controlled trial with a crossover design, where all neonates will receive three different interventions in a randomized order during heel lance: 1) 30% oral glucose (standard care), 2) mother's recorded heart sounds and vibrations using a nucu™ -multisensory pad + 30% oral glucose (digital maternal presence intervention) and 3) skin-to-skin contact + 30% oral glucose (live maternal presence intervention). The study population will consist of neonates (n = 36) born between 32–42 gestational age (GA) and are under the care in the NICU of the Oulu University Hospital in 2023–2024.

We hypothesize that 1) digital maternal presence intervention will provide more effective pain management during heel lance compared to standard care and neonates will recover more quickly from the procedure; 2) skin-to-skin contact will provide more effective pain relief compare to digital maternal presence intervention or standard care, and the neonates will recover from the procedure faster and 3) PIPP-R (Premature Infant Pain Profile Revised) and NIAPAS (Neonatal Infant Acute Pain Assessment Scale) pain scores will correlate with changes in physiological variables.

The regional medical research ethics committee of the Wellbeing services county of North Ostrobothnia has pre-evaluated this research project and issued a statement on it.

3 Results

The primary outcomes are 1) infant's pain intensity measured following the heel lance using NIAPAS and PIPP-R; 2) pain-induced changes in physiological variables (heart rate (HR), oxygen saturation (SpO2), respiratory rate (RR) measured following the heel lance using patient monitor (Philips IntelliVue MX800) and 3) pain-induced concentration changes of regional cerebral oxygenated-hemoglobin (HbO) measured with fNIRS (Glymphometer). The secondary outcomes are 1) recovery measured after blood sampling and 2) correlation between pain scores measured by pain assessment scales (PIPP-R and NIAPAS) and changes in physiological variables (HR, SpO2, RR) and concentration changes of oxygenated-hemoglobin.

4 Discussion

This study provides new evidence on the use of digital solutions and parental involvement in neonatal pain management, as well as the utility of digital measurement methods in the assessment of neonatal pain. The results of the study can be used in nursing decision-making when considering the best pain relief method in terms of both efficiency and parental involvement.

References

1. Cruz, M.D., Fernandes, A.M., Oliveira, C.R.: Epidemiology of painful procedures performed in neonates: a systematic review of observational studies. Eur. J. Pain **20**(4), 489–498 (2016). https://doi.org/10.1002/ejp.757
2. Williams, M.D., Lascelles, B.D.X.: Early neonatal pain-a review of clinical and experimental implications on painful conditions later in life. Front. Pediatr. **8**, 30 (2020) DOI: https://doi.org/10.3389/fped.2020.00030
3. Campbell-Yeo, M., Eriksson, M., Benoit, B.: Assessment and management of pain in preterm infants: a practice update. Children **9**(2), 244 (2022). https://doi.org/10.3390/children9020244
4. Pillai Riddell, R.R, et al.: Non-pharmacological management of infant and young child procedural pain. Cochrane Database Syst. Rev. **6**(CD006275) (2023). https://doi.org/10.1002/14651858.CD006275.pub4
5. Ullsten, A., Andreasson, M., Eriksson, M.: State of the art in parent-delivered pain-relieving interventions in neonatal care: a scoping review. Front. Pediatr. **9**(651846) (2021). https://doi.org/10.3389/fped.2021.651846
6. Johnston, C., et al.: Skin-to-skin care for procedural pain in neonates. Cochrane Database Syst. Rev. **2**(2), CD008435 (2017). https://doi.org/10.1002/14651858.CD008435.pub3
7. Palomaa, A.K., Korhonen, A., Pölkki, T.: Factors influencing parental participation in neonatal pain alleviation. J. Pediatr. Nurs. **31**(5), 519–527 (2016). https://doi.org/10.1016/j.pedn.2016.05.004
8. Palomaa, A.K., Huhtala, S., Tuomikoski, A.M., Pölkki, T.: Effectiveness of technology-based interventions compared with other non-pharmacological interventions for relieving procedural pain in hospitalized neonates: a systematic review. JBI Evid Synth. **21**(8), 1549–1581 (2023). https://doi.org/10.11124/JBIES-22-00179. PMID: 37218335
9. Roué, J.M., Rioualen, S., Gendras, J., Misery, L., Gouillou, M., Sizun, J.: Multi-modal pain assessment: are near-infrared spectroscopy, skin conductance, salivary cortisol, physiologic parameters, and Neonatal Facial Coding System interrelated during venepuncture in healthy, term neonates? J. Pain Res. **11**, 2257–2267 (2018). https://doi.org/10.2147/JPR.S165810

Towards Remote 3D Pathological Diagnosis System Using LUCID, a Tissue-Clearing Reagent, and HandySPIM, a Novel Compact 3D Imaging Device

Hiroshi Onodera[1], Kenji Uehara[2], Takumi Murakami[3], Asami Tanaka[1], Yusuke Morishita[4], and Junji Yumoto[1,4(✉)]

[1] The University of Tokyo, Bunkyo-Ku, Tokyo 113-0033, Japan
yumoto@ipst.s.u-tokyo.ac.jp
[2] Miyuki Giken, Co. Ltd., Bunkyo-Ku, Tokyo 113-0033, Japan
[3] Nippon Electric Glass, Co., Ltd., Otsu 520-8639, Shiga, Japan
[4] PhotonTech Innovations, Co. Ltd., Bunkyo-Ku, Tokyo 113-0033, Japan

Abstract. We introduce LUCID, a tissue clearing reagent suitable for pathological diagnosis that we have developed, and HandySPIM, a selective plane illumination microscope developed with a new concept. These two technologies enable affordable and convenient three-dimensional imaging of biological information, and will also make remote three-dimensional pathological diagnosis techniques possible in the future.

Keywords: Tissue Clearing Reagent · Selective Plane Illumination Microscope · Remote 3D Pathology

1 Introduction

LUCID is a tissue-clearing reagent that enables three-dimensional (3D) imaging of biological specimens. Notably, it offers higher transparency compared to other reagents, and it allows for immunostaining and nuclear staining. Furthermore, biological samples cleared with LUCID can be preserved for up to 10 years, making it suitable for pathological diagnosis. However, 3D imaging requires expensive imaging equipment, making it impractical for cost-conscious pathological examinations.

2 Materials and Methods

To address this challenge, we developed a compact and user-friendly Selective Plane Illumination Microscopy (HandySPIM), which uses a 5–20 μm thick glass thin film as a waveguide to form a two-dimensional sheet of light, serving as the excitation light.

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 541–542, 2024.
https://doi.org/10.1007/978-3-031-59091-7

3 Results

The resolution of HandySPIM in the thickness direction depends on the thickness of the glass film; a thicker film allows for the verification of three-dimensional structures without moving the sample. Additionally, the cost of HandySPIM can be as low as 1/30th to 1/50th of that of commercially available SPIMs.

4 Discussion

By combining LUCID and HandySPIM, and further integrating network and information processing technologies, remote 3D pathological diagnosis can be achieved at a low cost. This makes it feasible for implementation even in small hospitals and clinics, where significant improvements in the accuracy of pathological diagnoses, time reduction, and cost savings are highly anticipated.

Author Index

M. Särestöniemi et al. (Eds.): NCDHWS 2024, CCIS 2084, pp. 543–547, 2024.
https://doi.org/10.1007/978-3-031-59091-7