

**International Journal on
Advances in Life Sciences**



The *International Journal on Advances in Life Sciences* is published by IARIA.

ISSN: 1942-2660

journals site: <http://www.iariajournals.org>

contact: petre@iaria.org

Responsibility for the contents rests upon the authors and not upon IARIA, nor on IARIA volunteers, staff, or contractors.

IARIA is the owner of the publication and of editorial aspects. IARIA reserves the right to update the content for quality improvements.

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy or print, providing the reference is mentioned and that the resulting material is made available at no cost.

Reference should mention:

International Journal on Advances in Life Sciences, issn 1942-2660
vol. 16, no. 1 & 2, year 2024, http://www.iariajournals.org/life_sciences/

The copyright for each included paper belongs to the authors. Republishing of same material, by authors or persons or organizations, is not allowed. Reprint rights can be granted by IARIA or by the authors, and must include proper reference.

Reference to an article in the journal is as follows:

<Author list>, "<Article title>"
International Journal on Advances in Life Sciences, issn 1942-2660
vol. 16, no. 1 & 2, year 2024, <start page>:<end page> , http://www.iariajournals.org/life_sciences/

IARIA journals are made available for free, proving the appropriate references are made when their content is used.

Sponsored by IARIA

www.iaria.org

Copyright © 2024 IARIA

Editor-in-Chief

Les Sztandera, Thomas Jefferson University, USA

Editorial Board

Ganesharam Balagopal, Ontario Ministry of the Environment Conservation and Parks | Technical Assessment and Standards Development Branch, Canada

Kazi S. Bennoor, National Institute of Diseases of Chest & Hospital - Mohakhali, Bangladesh

Razvan Bocu, Transilvania University of Brasov, Romania

Karin Brodén, Karlstad University, Sweden

Ozgu Can, Ege University, Turkiye

Young (Yang) Cao, Virginia Tech, USA

Jitender Deogun, University of Nebraska-Lincoln, USA

Duarte Duque, ALGORITMI Research Centre | LASI, University of Minho / 2Ai - School of Technology | IPCA, Portugal

Hassan Ghazal, Mohammed VI University of Health Sciences, Morocco

Piero Giacomelli, Fidia Farmaceutici SpA, Vicenza, Italy

Malina Jordanova, Space Research & Technology Institute | Bulgarian Academy of Sciences, Bulgaria

Hassan M. Khachfe, Lebanese International University, Lebanon

Peter Kieseberg, St. Pölten University of Applied Sciences, Austria

Evgeniy Krastev, Sofia University St. Kliment Ohridsk, Bulgaria

Ljerka Luic, University North, Croatia

Jose Manuel Molina Lopez, Universidad Carlos III de Madrid, Spain

Stefano Mariani, Politecnico di Milano, Italy

Julio César Mello Román, National University of Asuncion (UNA), Paraguay

Helena Pereira de Melo, NOVA School of Law, Portugal

Vitor Pinheiro de Almeida, Pontifícia Universidade do Rio de Janeiro (PUC-Rio), Brazil

Tamara Powell, Kennesaw State University, USA

Addisson Salazar, Universitat Politècnica de València, Spain

Les Sztandera, Thomas Jefferson University, USA

Paulo Teixeira, Polytechnic University of Cávado and Ave, Portugal

Genny Villa, Université de Montréal, Canada

Vivian Vimarlund, Linköping University, Sweden

CONTENTS

pages: 1 - 10

Enhancing User Engagement in DailyExp: A Tool for Collecting Cognitive Performance and Physiological Data with Engaging Behavioral Design

Xianyin Hu, Graduated School of Frontier Sciences, The University of Tokyo, Japan

Yuki Ban, Graduated School of Frontier Sciences, The University of Tokyo, Japan

Shin'ichi Warisawa, Graduated School of Frontier Sciences, The University of Tokyo, Japan

pages: 11 - 20

Assessment of Pharmacology Costs in Diabetes Treatment Using OMOP CDM: A Nationally Representative Study

Evgeniy Krastev, Sofia University St. Kliment Ohridski, Bulgaria

Petko Kovachev, Sofia University St. Kliment Ohridski, Bulgaria

Simeon Abanos, Sofia University St. Kliment Ohridski, Bulgaria

Ralitsa Krasteva, Specialized Hospital for Active Treatment of Children's Diseases, Bulgaria

Dimitar Tcharaktchiev, Medical University Sofia, Bulgaria

pages: 21 - 32

Building a Smarter Care Home: A Deep Dive Into HealthSonar's Architecture for Assisted Living

Adamantios Ntanis, PD Neurotechnology Ltd., Ioannina, Greece, Greece

Spyridon Kontaxis, PD Neurotechnology Ltd., Ioannina, Greece, Greece

George Rigas, PD Neurotechnology Ltd., Ioannina, Greece, Greece

Anastasia Pentari, Institute of Computer Science, Foundation for Research and Technology—Hellas, Heraklion, Greece, Greece

Kostas Tsiouris, Biomedical Research Institute, Foundation for Research and Technology—Hellas, Ioannina, Greece, Greece

Efstathios Kontogiannis, PD Neurotechnology Ltd., Ioannina, Greece, Greece

Eleftherios Kostoulas, PD Neurotechnology Ltd., Ioannina, Greece, Greece

Ilias Tsimperis, PD Neurotechnology Ltd., Ioannina, Greece, Greece

Theodoros Vlioras, PD Neurotechnology Ltd., Ioannina, Greece, Greece

Aristotelis Bousis, PD Neurotechnology Ltd., Ioannina, Greece, Greece

Styliani Zelilidou, Biomedical Research Institute, Foundation for Research and Technology—Hellas, Ioannina, Greece, Greece

Kalpso Tasiou, Biomedical Research Institute, Foundation for Research and Technology—Hellas, Ioannina, Greece, Greece

Manolis Tsiknakis, Institute of Computer Science, Foundation for Research and Technology—Hellas, Heraklion, Greece, Greece

Dimitrios Fotiadis, Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, Ioannina, Greece, Greece

pages: 33 - 43

Approaches to Improving Medication Adherence Prediction in Chronic Disease Patients

Ben Malin, Brunel University London, United Kingdom

Ejike Nwokoro, HealthNet Homecare, United Kingdom

Tatiana Kalganova, Brunel University London, United Kingdom

Joshua Hinton, HealthNet Homecare, United Kingdom

pages: 44 - 55

Decoding Key Variables Contributing to Right Ventricular Involvement in Ischaemic and Non-ischaemic Cardiomyopathy

Carlos Barroso-Moreno, Universidad Europea de Madrid, Spain
Hector Espinos Morato, Universidad Europea de Valencia, Spain
Enrique Puertas, Universidad Europea de Madrid, Spain
Juan José Beunza Nuin, Universidad Europea de Madrid, Spain
José Vicente Monmeneu, ASCIRES Biomedical Group, Spain
María P. López-Lereu, ASCIRES Biomedical Group, Spain
David Moratal, Universitat Politècnica de València, Spain

pages: 56 - 66

Developing a Process for Gathering and Managing Multi-site User Feedback for a Data Collection Project in a Rare Disease Setting. Using Design Science to Explicate a Knowledge Contribution

Sinéad Impey, ADAPT Centre, Trinity College Dublin, Ireland
Jonathan Turner, Technological University Dublin, Ireland
Frances Gibbons, ADAPT Centre, Trinity College Dublin, Ireland
Anthony Bolger, ADAPT Centre, Trinity College Dublin, Ireland
Gaye Stephens Stephens, ADAPT Centre, Trinity College Dublin, Ireland
Lucy Hederman, ADAPT Centre, Trinity College Dublin, Ireland
Ciara O'Meara, ADAPT Centre, Trinity College Dublin, Ireland
Ferran De La Varga, ADAPT Centre, Trinity College Dublin, Ireland
John Kommala, ADAPT Centre, Trinity College Dublin, Ireland
Matthew Nicholson, ADAPT Centre, Trinity College Dublin, Ireland
Daniel Farrell, ADAPT Centre, Trinity College Dublin, Ireland
Emmet Morrin, ADAPT Centre, Trinity College Dublin, Ireland
Miriam Galvin, ADAPT Centre, Trinity College Dublin, Ireland
Mark Heverin, Academic Unit of Neurology, Trinity College Dublin, Ireland
Éanna Mac Domhnaill, Academic Unit of Neurology, Trinity College Dublin, Ireland
Robert McFarlane, Academic Unit of Neurology, Trinity College Dublin, Ireland
Dara Meldrum, Academic Unit of Neurology, Trinity College Dublin, Ireland
Deirdre Murray, Academic Unit of Neurology, Trinity College Dublin, Ireland
Orla Hardiman, Academic Unit of Neurology, Trinity College Dublin, Ireland

pages: 67 - 79

AI-Based Digital Therapeutics Platform for Obesity Management

Sarfraz Khokhar, Rasimo Systems, US
John Holden, Rockford College of Medicine, IL., US
Catherine Toomer, Health Wellness and WL Centers, Aiken, SC, US
Linda Whitby, Rasimo Systems, US

pages: 80 - 91

Considerations of Application Examples of Gait Analysis Using MediaPipe

Yasutaka Uchida, Teikyo Univbarsity of Science, Japan
Tomoko Funayama, Teikyo University of Science, Japan
Eiichi Ohkubo, Teikyo University of Science, Japan
Yoshiaki Kogure, Teikyo University of Science, Japan
Yasunori Fujimori, Seirei Yokohama Hospital, Japan
Ryota Kimura, Seirei Yokohama Hospital, Japan
Kohei Kimura, Seirei Yokohama Hospital, Japan

Enhancing User Engagement in DailyExp: A Tool for Collecting Cognitive Performance and Physiological Data with Engaging Behavioral Design

Xianyin Hu

*Graduated School of Frontier Sciences
The University of Tokyo
Chiba, Japan*

Email: huxy214@gmail.com

Yuki Ban

*Graduated School of Frontier Sciences
The University of Tokyo
Chiba, Japan*

Email: ban@edu.k.u-tokyo.ac.jp

Shin'ichi Warisawa

*Graduated School of Frontier Sciences
The University of Tokyo
Chiba, Japan*

Email: warisawa@edu.k.u-tokyo.ac.jp

Abstract—Experiments in cognitive science that rely on laboratory-based settings are not only costly and time-consuming but also make it difficult to investigate individuals' cognitive states as they naturally fluctuate over time in daily life. In our initial study [1], we presented a practical tool implemented as a smartphone application DailyExp that aims to conveniently collect cognitive performance data in daily life settings. This application is accessible from major mobile platforms (iOS and Android), tied with a Fitbit account to collect physiological data at the same time. We employed engaging behavioral design to overcome problems faced experimenting in the wild, intended to improve data quality as well as data collection efficiency and evaluated them in a one-month-long experiment involving 10 participants. For this extended study, we implemented new features based on feedback from the preliminary study and tested them on 41 individuals. This paper provides implementation details of each cognitive task that was not covered in the initial paper, and included a comprehensive analysis of post-study questionnaires as well as a quantitative comparison of objective metrics evaluating user engagement and persistence. Our results demonstrated a statistically significant improvement in user engagement as well as persistence by adding new features.

Index Terms—data collecting tool; engaging design; physiological data; cognitive performance;

I. INTRODUCTION

Cognitive science has traditionally centered on comprehending the mechanisms of human cognition at an aggregate level. However, the exploration of individual differences has emerged as a progressively significant subject within the field. Recently, researchers have adopted a perspective that views individuals' cognition as a dynamic system that fluctuates. It is also suggested that the fluctuation in cognition is related to fluctuation in physiology from an embodied cognition perspective [2]. In this study, Our primary interest lies in facilitating studies that investigate individual differences and intra-individual variations within established cognitive mechanisms. We advocate for these investigations to be carried out in meticulously designed real-life settings, as opposed to laboratory settings, since laboratory environments may induce high arousal levels that shift individuals' cognitive and physiological states due to nervousness and unfamiliarity. Experiments carried out with authentic context ensure the accumulation of extensive data encompassing both population-wide variations and intrapersonal dynamics. To support such endeavors, we offered a practical tool DailyExp tailored to assist researchers in gathering data on cognitive performance and physiological signals within daily life settings.

Previous attempts have been made to adopt smartphones and smartwatches as assessment tools, including iVitality [3], DelApp [4], Cognition Kit [5] and UbiCAT [6]. These studies showed a good correspondence between data obtained from the mobile-based tools and that from the laboratory, indicating that mobile-based tools are feasible for evaluating cognitive function. However, challenges such as a lack of user engagement throughout a prolonged experiment still exist in an experiment conducted in the wild that depends largely on participants' voluntary behaviors.

This study provided a practical implementation of a tool for data collection of both cognitive performance, as well as physiological data in daily life settings with engaging behavioral design. An alpha version of the smartphone application DailyExp is readily available that can conduct various classical paradigms in cognitive science including N-back, Stroop, and Raven's Advanced Progressive Matrices (RAPM) test. The application was linked with a widely used commercial smartwatch (Fitbit) to collect physiological data at the same time. This application is accessible from major mobile platforms (iOS and Android). We employed multiple practices of engaging behavioral designs to overcome several challenges facing experimenting in the wild, including immediate reward and feedback, trackable performance to bolster user self-efficacy, transparent monetary rewards for enhanced psychological safety, and a ranking system to ignite social competition.

Building upon the features previously evaluated with 10 participants in our initial conference paper [1], we made two key modifications to DailyExp and tested on 41 individuals to address user feedback collected in the preliminary study. Firstly, we revised the ranking screen to display only the top 15 most active participants, concealing information about less-engaged users to prevent demotivation among others. Secondly, we implemented a daily limit on task completion to foster a sense of achievable goals and discourage procrastination. This adjustment was based on user comments highlighting the negative impact of unrestricted task completion on motivation. Also, this paper provided implementation details of each cognitive task that was not covered in the initial paper, and included a comprehensive analysis of post-study questionnaires as well as a quantitative comparison of objective metrics evaluating user engagement and persistence. Our results demonstrated a statistically significant improvement in

user engagement as well as persistence by adding new features.

The paper is structured as follows: In Section II, we reviewed existing literature on the utilization of smartphones and smartwatches as assessment tools, explored the feasibility of using mobile devices for data collection in daily life and highlighted the challenges associated with this approach. In Section III, we provide a detailed description of the implementation that covers the technical aspects of developing the tool. In Section IV, we evaluated the system through a preliminary user study involving ten participants over one month. We compiled several points for improvement based on the user interview in the preliminary study and subsequently enhanced the DailyExp. In Section V, we recruited 41 participants to use the improved version of DailyExp for one month. We compared the objective metrics including user engagement and persistence with those in the preliminary study and revealed the effectiveness of the enhancement. Both experiments in this study were approved by the Research Ethics Committee of the University of Tokyo (No. 22-399) and performed under written informed consent from all the participants. Finally, we present a summary in Section VI that highlights the achievements of this paper and discusses the potential applications of DailyExp in different domains of cognitive science, showcasing its versatility.

II. RELATED WORKS

In this section, we reviewed existing literature on the utilization of mobile devices as assessment tools, explored the feasibility of using mobile devices for data collection in daily life and highlighted the challenges associated with this approach.

Jongstra's team [3] developed a smartphone-based app iVitality to evaluate five cognitive tests (Memory-Word, Trail Making, Stroop, Reaction Time, and Letter N-back) in 151 healthy adults over 6 months. This study concluded that repeated smartphone-assisted cognitive testing is feasible with reasonable adherence and moderate validity for the Stroop and the Trail Making tests compared with conventional neuropsychological tests. They also addressed that smartphone-based cognitive testing seems promising for large-scale data collection in population studies.

Tieges' team developed the DelApp [4], a smartphone application aimed at objectively detecting attentional deficits in delirium patients. By leveraging mobile technology, this study addressed the need for accessible and accurate cognitive assessment tools in clinical settings, contributing to improved diagnosis and management of delirium. Moreover, this study showed the use of smartphone-based cognitive tasks allows for a broader range of participants, extending beyond healthy individuals to include those with varying cognitive conditions.

Dingler introduced a mobile toolkit designed to capture fluctuations in alertness and cognitive performance throughout the day [5]. The tool kit consists of three tasks, they are Psychomotoric Vigilance Task, Go/No-Go task, and Multiple Object Tracking task. To investigate the feasibility of using

smartphone-based cognitive tasks to assess diurnal fluctuations in arousal level, the researchers conducted an in-the-wild experiment with 12 participants who completed the test batteries multiple times a day over 1-2 weeks. The results demonstrated that circadian rhythms in alertness could be effectively captured in naturalistic settings, without the need for controlled laboratory conditions. This study provided insights into the utilization of smartphone-based cognitive test batteries to capture temporal dynamics of cognitive function, showing their potential applications in personalized healthcare and productivity management.

Hafiz [6] presented the UbiCAT, a smartwatch-based cognitive assessment tool. This tool implemented three cognitive tests — an Arrow test, a Letter test, and a Color test—adapted from the two-choice reaction-time, N-back, and Stroop tests, respectively. They evaluated the UbiCAT test measures against standard computer-based tests with 21 healthy adults. The results showed a strong correlation between the UbiCAT and standard computer-based tests, indicating its effectiveness. Usability ratings were high, and participants reported low discomfort while using the smartwatch. Despite some participants preferring computer-based tests due to familiarity, the UbiCAT offered the advantage of in-the-wild assessment, leveraging the ubiquity of wearable devices. However, it is limited by the less diverse interaction methods available on smartwatches compared to smartphones.

In summary, these studies demonstrated the feasibility of leveraging mobile devices for collecting cognitive performance data in a real-world setting. While smartwatches provide convenient assessment, smartphones offer a wider range of interaction, making them suitable for assessing a large variety of cognitive functions. However, a common challenge of these studies is maintaining participant engagement, as the amount and quality of data depend largely on voluntary participation. In this study, we provided a practical implementation of a tool for collecting not only cognitive performance data but also physiological signals linked with the widely used commercial smartwatch Fitbit. These physiological signals can be used to predict cognitive function and explore the underlying mechanisms of cognition and physiology. Moreover, instead of validating how well mobile-based cognitive tests align with results from laboratory settings, which is done in the precedent studies, we addressed the practice of engaging behavior design, aiming to enhance the quality and consistency of data collection in naturalistic settings.

III. IMPLEMENTATION

This section covered the technical aspects of developing the DailyExp. First, we provide a systematic overview. Then we described the implementation of the three cognitive tasks—Spatial 2-back, Stroop, and FluidIQ in detail, which was not covered in the initial conference paper due to page constraints. Finally, we demonstrated the features designed to handle unexpected user behavior and those implementing best practices for engaging behavioral design.

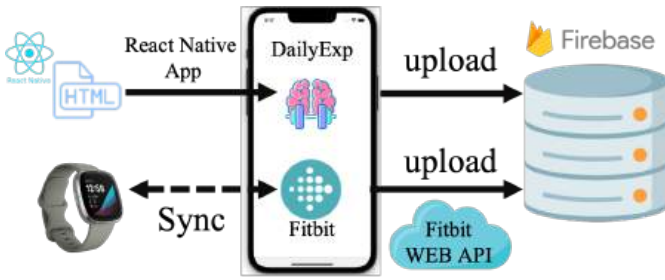


Fig. 1: Overview of the system design.

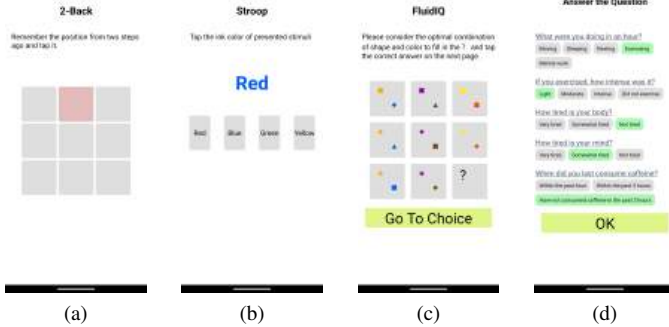


Fig. 2: Screenshots of DailyExp. (a) The 2-back task. (b) The Stroop task. (c) The FluidIQ task. (d) The post-task questionnaire.

A. System Overview

Fig. 1 showed the overview of the system design. The mobile client of DailyExp was developed using React Native, a web-based open-source framework for mobile application development. React Native was chosen to ensure compatibility across multiple platforms for iOS and Android devices. For the server side, Firebase's data storage service was utilized to store data, including users' daily summary data, cognitive performance data for various tasks, and physiological data grabbed from the Fitbit server using Fitbit web API.

B. Cognitive Tasks

In the alpha version of DailyExp, three well-established cognitive tasks were administrated to study working memory (N-back), attention and executive function (Stroop), and fluid intelligence (FluidIQ or Raven's Progressive Matrices) as shown in Fig. 2. (a)-(c). These tasks were selected due to their robustness and potential to have individual differences and intrapersonal fluctuations in the corresponding cognitive ability to be evaluated. Cognitive Performance data with the trial information, problem context, and users' responses will be recorded and uploaded to the Firebase server. The details of the implementation of the three tasks were described as follows.

1) *The Spatial Continuous 2-back Task*: The n-back task paradigm is commonly used as an assessment in psychology and cognitive neuroscience to measure a part of working memory and working memory capacity. In this study, we employed

a variation of the N-back task called the continuous N-back task. As suggested in previous work [7], the performance of continuous 2-back tasks reflected the level of mental fatigue and can be estimated from physiological signals using a deep learning model. Since the purpose of this study is to collect data that can be used to investigate intra-day fluctuations in individuals' cognitive states, we considered the continuous version instead of the original one a better candidate.

In the continuous N-back task, instead of responding only at the matched trials (when the current stimulus matches the stimulus presented in N trials before), but continuously respond with the stimulus presented in N trials before. Regarding the type of stimuli, we adopted spatial position, which aligns best with the touch-based input method on smartphones. The stimuli sequence was randomly generated. The task lasted for 5 minutes, and the stimuli were presented every 2 seconds. Thus there will be a total of 150 trials in each conduction.

Fig. 3 illustrated an example of four trials in the spatial continuous 2-back task. For explanatory purposes, numbers are assigned to each position in Fig. 3a, although these numbers were not displayed in the actual task. The presented position was highlighted in green to distinguish it from other positions. Participants were required to respond by tapping the corresponding position on the smartphone screen. The tapped position turned red to provide feedback. In this example, a sequence of positions (2, 4, 8, 5) were presented. On the third trial, when position 8 was presented, a correct response would be tapping position 2, which had been presented two trials before. Similarly, on the fourth trial, it was desired to respond by tapping position 4. Additionally, to assist users in developing an intuitive sense of the presentation intervals, the green color highlighting the current stimulus gradually faded over the 2-second interval.

Cognitive performance data collected in the spatial continuous 2-back task were listed in Table I. Specifically, we recorded stimulus up to three trials back to facilitate the analysis of correctness as well as error patterns. Regarding user response, both the response position and reaction time were recorded to enable a multidimensional analysis of human cognitive behavior.

2) *The Stroop Task*: The Stroop task was originally devised by John Ridley Stroop in 1935 to investigate the interference in serial verbal reactions [8], has been widely used in cognitive science and experimental psychology to study selective attention [9]. In the Stroop task, participants are presented with

TABLE I: Cognitive performance data recorded in spatial continuous 2-back task

| Trial information | trial ID |
|-------------------|--------------------|
| | trial start time |
| Problem context | current stimulus |
| | 1-back stimulus |
| | 2-back stimulus |
| | 3-back stimulus |
| User response | responded position |
| | reaction time |

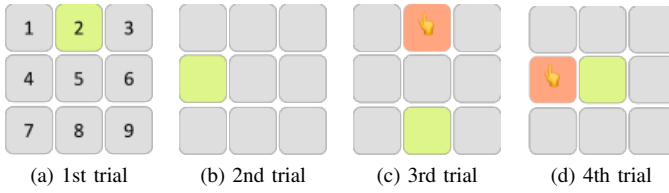


Fig. 3: 2-back detail explained. 2-back detail explained. 2-back detail explained. 2-back detail explained. 2-back detail explained.

TABLE II: Cognitive performance data recorded in Stroop task

| | |
|-------------------|-------------------------------------|
| Trial information | trial ID trial start time |
| Problem context | stimulus word stimulus ink color |
| User response | responded color reaction time |

words printed in specific colors. In the congruent condition, the color of the word matches the meaning of the word. In the incongruent condition, they do not match. Participants are asked to either say aloud the word or to report the color of the word. As a control condition, participants read words printed in black or report the color of non-word stimuli.

The Stroop task was implemented in DailyExp as shown in Fig. 2b. As mentioned in the previous section, this study aims to investigate the fluctuations in individual cognitive states, rather than the already validated Stroop effect. Therefore, we only presented the incongruent condition where the ink color of the stimulus was different from the meaning of the word, leading to a fairly difficult level for most of the users.

In each trial, two different colors were randomly selected to be presented as word or ink colors from a set of four colors: red (#ff4b00), blue (#005aff), green (#03af7a), and yellow (#fff100). Those colors are recommended in Color Universal Design, making them accessible to people with diverse color vision characteristics. Participants were instructed to tap the button corresponding to the ink color of the presented stimulus, and the presentation lasted for 3 seconds. To indicate there is a transition to the next trial, a 2-second retention interval was included after each presentation. During the retention interval, nothing was displayed. Consequently, the interval for each trial was 5 seconds. Cognitive performance data collected in the Stroop task were listed in Table II.

3) *The FluidIQ Task*: The RPM (Raven's Progressive Matrices) task, also known as the FluidIQ task, presents an incomplete geometric pattern and requires participants to select the missing piece from a set of options. The RPM task is a non-verbal test that does not require any specific knowledge or language skills. It is used worldwide to measure general human intelligence and abstract reasoning abilities and gained recognition as an IQ test since it was adopted as a part of the Monsa intelligence assessment. Intelligence is generally classified into two categories which are crystallized and fluid intelligence. Crystallized intelligence represents indi-

vidual abilities acquired over years of experience, education, and learning, while fluid intelligence represents insight and abilities independent of language or experiential knowledge. The non-verbal nature and the independence from specific knowledge of the task make it suitable for evaluating fluid intelligence. For this reason, it is also referred to as the FluidIQ task. Since the latter term more intuitively captures the essence, we will adopt the term FluidIQ in this study.

The FluidIQ task used in this study was implemented as shown in Fig. 2c, and a set of features as described in Table III were generated for each trial. The generation of the whole figure followed a certain set of rules. There are two possible rules or different-along the row and the column. Rules were randomly assigned to row and column despite a case when rule same was assigned to both since it would make the task too easy to be solved intuitively without involving the desired cognitive process. In summary, there were three possible combinations of row and column rules:

- Same along the row, different along the column.
- Different along the row, same along the column.
- Different along the row, different along the column.

If all four features follow the rule different along row/column and the rule same along the other axis, the total number of features needed to be generated would be 4, also denoted as $n^{\text{different}} = 4$. If all four features follow the rule different along both row and column, then $n^{\text{different}} = 8$, which is the most difficult. For example, in Fig. 2c, for Feature 1 (shape of the top-left figure), the rule along the row is same, and rule along the column is different. For Feature 2 (color of the top-left figure), the rule along the row is different, and rule along the column is same. For Feature 3 (shape of the bottom-right figure), both row and column rules are different. For Feature 4 (color of the bottom-right figure), the row pattern is different, and the column pattern is the same. In summary, for this trial, the number of different rules is $n^{\text{different}} = 5$. In this task, the systematic factor that determines the difficulty level is the number of different rules $n^{\text{different}}$, with possible values of $n^{\text{different}} \in [4, 5, 6, 7, 8]$.

To control individual execution strategies, we implemented the FluidIQ task to not display the problem and options at the same time. Instead, participants were required to observe an incomplete figure to distinguish the rules in the problem view and generate their answers. Participants then tap the "Go to Choice" button below to move to the option view. Once moved on to the option view, participants could not go back to the problem view. This implementation prevented participants from using a strategy of applying each option to the incomplete figure one by one, thereby achieved to force users to adopt a generative strategy.

There was no specific time limit for this task, and participants were allowed to perform at their own pace. The FluidIQ task consisted of 12 trials, taking around 5 minutes to complete. The cognitive performance data recorded was shown in Table IV.

TABLE III: Features and corresponding possible values in the FluidIQ task

| Feature | Description | Possible values |
|----------|----------------------------------|---|
| feature1 | shape of the top-left figure | circle, star, triangle, diamond, square |
| feature2 | color of the top-left figure | red(#ff4b00), blue(#005aff), green(#03af7a), orange(#f6aa00), purple(#990099), brown(#804000), yellow(#ffff100) |
| feature3 | shape of the bottom-right figure | same as feature1 |
| feature4 | color of the bottom-right figure | same as feature2 |

TABLE IV: Cognitive performance data recorded in FluidIQ task

| Trial information | trial ID trial start timestamp |
|-------------------------------|--|
| Problem context | stimuli combination in position 1 |
| | stimuli combination in position 2 |
| | ... |
| | stimuli combination in position 8 |
| User response | stimuli combination in answer |
| | stimuli combination in alternative choice1 |
| | stimuli combination in alternative choice2 |
| responded stimuli combination | |
| reaction time | |

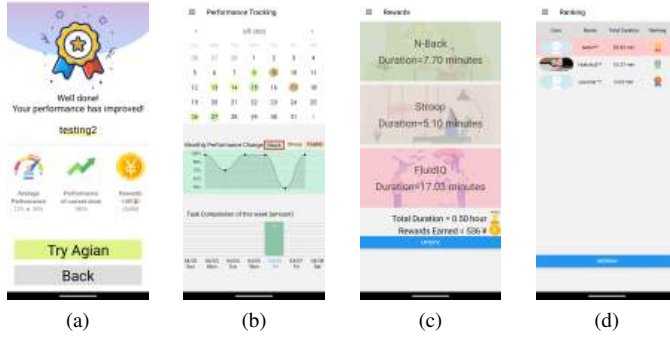


Fig. 4: Screenshots of DailyExp. (a) The encouragement view. (b) The performance tracking view. (c) The rewards view. (d) The ranking view.

C. Dealing with Unexpected User Behavior

An issue facing the experiment that relies on users' voluntary behavior is that users do not always behave in a desired manner. Predictable behaviors include forgetting to wear the smartwatch, responding randomly without involving the target cognitive process to be assessed, and an improper understanding of the procedures for the tasks. These behaviors will lead to a lack of data and noisy meaningless data. To address these issues, our application implemented several features to reduce unexpected behaviors.

Firstly, we provided reminders to wear the smartwatch before the start of any task. Secondly, practice mode with feedback will help users familiarize themselves with the task procedure. Moreover, user performance is recorded to monitor if it falls below a preset threshold. Users will be notified of invalid conduction due to their suboptimal performance.

Another concern facing an experiment in the wild is that it is difficult to control factors that are not the target of interest, such as physical activities and caffeine intake, which have a great influence on the cognitive and physiological state. As a solution, we provide a self-report questionnaire (Fig. 2. (d)) after completing the task. This enables data to be nicely categorized and analyzed afterward.

D. Engaging Behavioral Design

We leveraged multiple practices of engaging behavioral design aiming to improve the efficiency of data collection, which

is largely determined by user engagement. Fig. 4. (a) showed the encouragement view that popped up immediately after each task conduction, notifying users about how well they performed this time compared to the past. We expect this action-reward link to lead to the habitation of voluntary conduction of cognitive tasks. Fig. 4. (b) showed a performance tracking view from a long-term perspective. The calendar and line chart displayed the monthly task executions and performance fluctuations, while the bar chart showed the task executions for the current week. This feature took advantage of human's tendency to make more effort towards specific goals when they feel in control of their actions. We expect this feature to satisfy users' desire for autonomy and increase intrinsic motivation. Fig. 4. (c) illustrated the monetary rewards earned, along with detailed information, such as the amount of time spent on each task and the corresponding rewards. In addition to providing an external motivation, this feature also promotes transparency of the experiment and is expected to enhance the psychological safety of participants. Finally, a ranking view (Fig. 4. (d)) was implemented as a social motivation leveraging the competitive mindset by allowing users to see others' task executions.

IV. PRELIMINARY USER STUDY

To evaluate the effectiveness of collecting data of the alpha version of DailyExp as well as gain insights into application design, we conducted a preliminary study involving 10 users to use DailyExp in their daily life for one month. Both qualitative and quantitative analysis was performed.

A. Participants

We recruited 10 participants (6 males and 4 females, aged 21-27) as app users, who are graduated students from the University of Tokyo.

B. Procedure

All participants were scheduled to attend an orientation in our lab. At the beginning of the orientation, we explained the purpose and the procedure of the experiment. After that, the participants signed with informed consent if they agreed to participate in the experiment. Then, they were guided to install Fitbit and DailyExp on the smartphones that they use in daily life. The Fitbit application was necessary to synchronize with the Fitbit device for physiological data collection. The researcher distributed pre-assigned Fitbit accounts to the

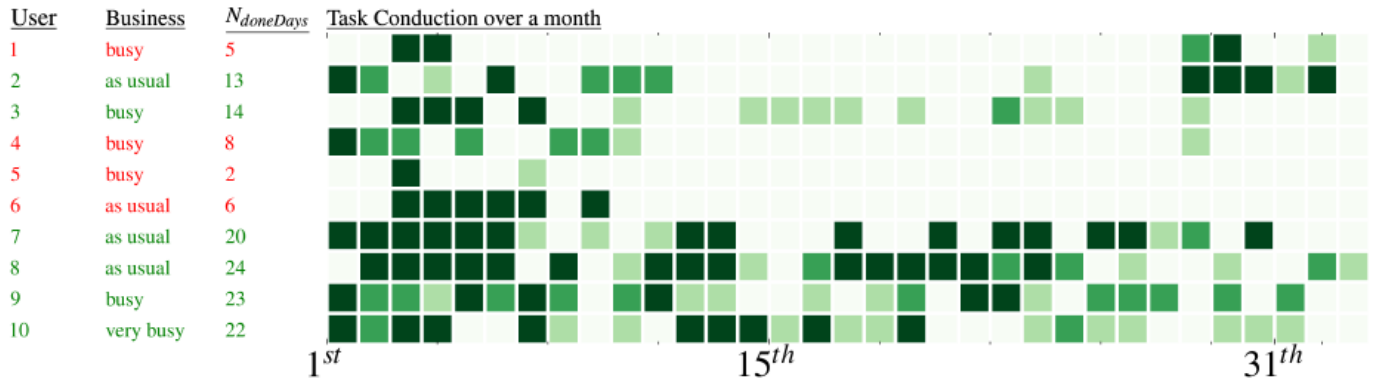


Fig. 5: User's busyness and task conduction in the preliminary user study. $N_{doneDays}$ denotes the number of days with task conducted. The cell color gradient indicated the number of task types performed (etc., the darkest grey indicated a completion of all three different tasks). Users printed in green ink are those who conducted tasks for more than 10 days and were considered active.

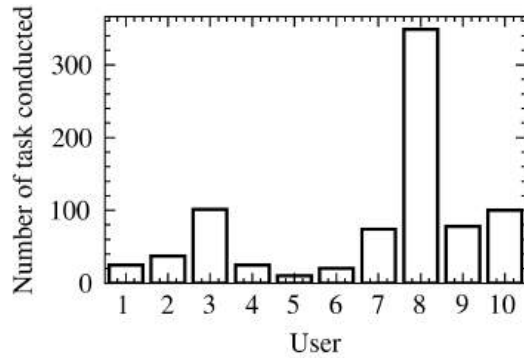


Fig. 6: Number of tasks conducted by each user in the preliminary user study.

participants and assisted them in logging in to DailyExp using Fitbit account. Subsequently, a document containing screenshots (Fig. 2) and descriptions of how to use DailyExp was presented. The participants were allowed to practice the cognitive tasks multiple times until they fully understood how to perform the three tasks correctly. Participants were informed that a monetary reward of 100 Japanese yen was given for each valid task completion. Besides the basic usage of conducting cognitive tasks, the researcher also provided explanations on other functionalities of DailyExp (Fig. 4) to the participants, and have them give a try on them. It has to be addressed that, participants were informed to conduct the cognitive tasks and use other functionality within the DailyExp at their own pace, without any restriction on the number of tasks to be completed each day or any limitations regarding task completion. Finally, instructions about the usage of the Fitbit device, the Fitbit application, and the synchronization process between them were provided. All the participants were requested to return the Fitbit devices at the end of the experiment and fill out a brief questionnaire. The questions and options for answers were listed in Table V.

C. Quantitative Evaluation

1) *Data Collected*: Throughout the preliminary user study, we obtained a total of 847 rounds of cognitive performance data, with 235 rounds for the spatial continuous 2-back task, 290 rounds for the Stroop task, and 322 rounds for the FluidIQ task.

2) *User Engagement*: We evaluated user engagement using two metrics. The first metric was the ratio of active users. A user is considered active if he/she conducted task for more than ten days throughout one month. Fig. 5 visualized the ten participants' task completion during the experiment period. It highlighted active users in green and non-active users in red, with their number of days when task was conducted at the left, denoted as $N_{doneDay}$, along with their business reported in the after-study questionnaire. As a result, four out of the ten users (users 7, 8, 9, and 10) actively engaged with DailyExp. Consequently, the ratio of active users in the preliminary user study was 60%. Notably, among the four non-active users, three reported themselves being busier than usual.

The other metric was the number of tasks conducted per user. As shown in Fig. 6, User 8 conducted 349 tasks and contributed to almost half of the total completions. Therefore, we treated User 8 as an outlier and calculated the remaining users' averaged completions, which is around 55 per user.

D. Qualitative Evaluation

1) *Reason of losing motivation*: In the preliminary user study, the top reasons cited for failure to sustain task execution were being busy during the period (50%) and tasks lasting too long (40%).

2) *Functionality that increases motivation*: Besides monetary rewards, The Ranking screen was highly rated as contributing to increased motivation for task execution (90%).

3) *Possible improvements*: The top suggested improvements were shorter task duration (80%), push notifications as reminders (60%) as well as more interesting tasks (60%).

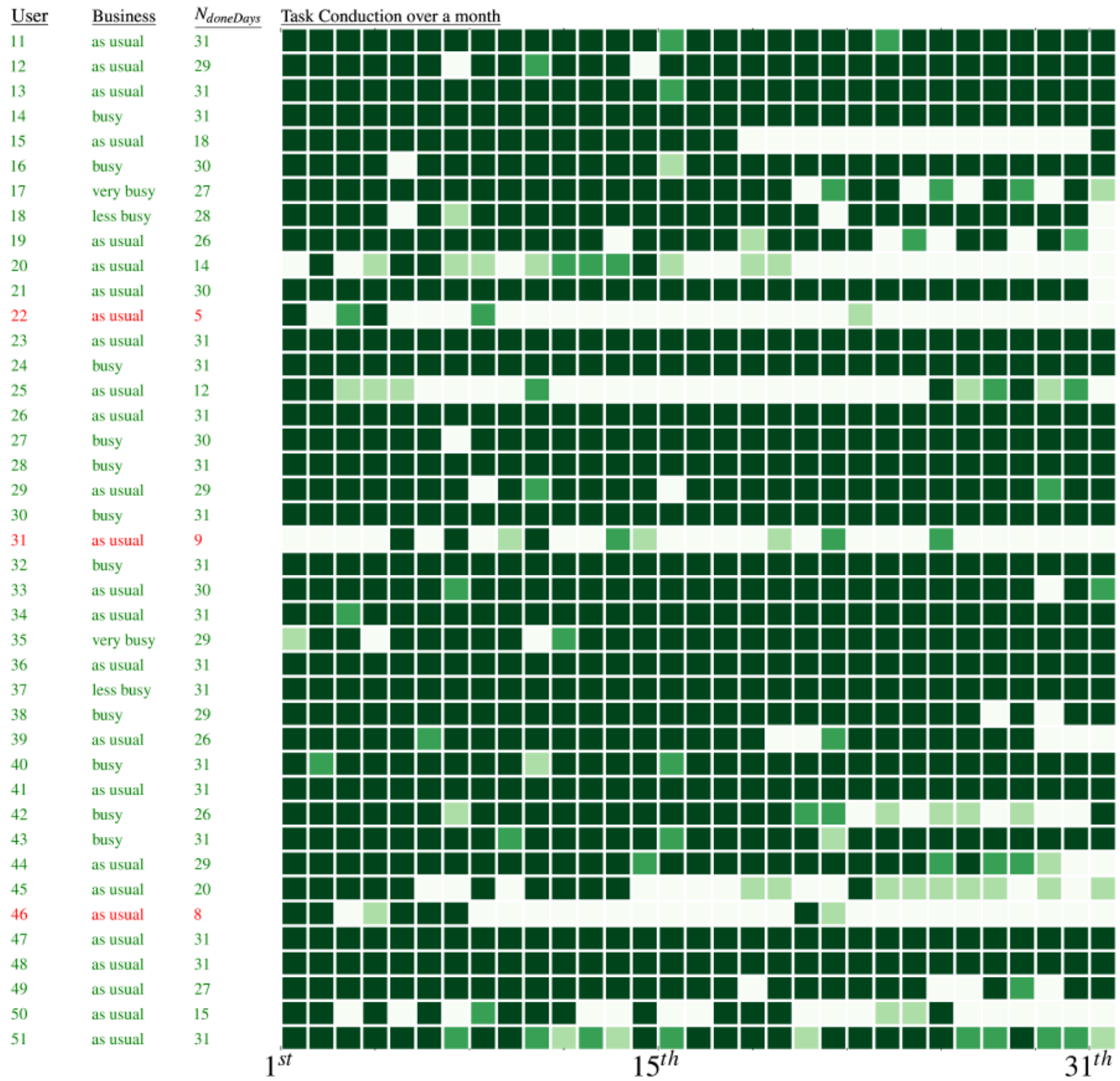


Fig. 7: User's busyness and task conduction in the user study. $N_{doneDays}$ denotes the number of days with task conducted. The cell color gradient indicated the number of task types performed (etc., the darkest grey indicated a completion of all three different tasks). Users printed in green ink are those who conducted tasks for more than 10 days and were considered active.

4) *Free comments from users:* Two users provided comments about a disadvantage of DailyExp to sustaining motivation as follows:

User 1: Since all users' task completion status can be seen on the ranking screen, I have a feeling that it was okay not to rush, after observing those who were not executing tasks much. This led to a decrease in my motivation for task execution.

User 9: My motivation was to earn rewards, and

since there are no restrictions on task execution each day, I believed a large amount of tasks could be completed in the last few days, so I procrastinated.

V. USER STUDY WITH IMPROVED DAILYEXP

To address the issues identified in the preliminary user study conducted with the alpha version of DailyExp, the following changes were made in the improved version:

TABLE V: Summary of Participant Responses in Post-Study Questionnaire. The last two columns indicated the number and percentage of participants who chose each answer option in the preliminary user study and user study using improved DailyExp, respectively. The top two most chosen answers for each question are highlighted in bold.

| question | answer | Preliminary user study | User study |
|--|---|------------------------|-----------------|
| During the experiment period, did your research or work become busier than you usually are? | not busy at all | 0 (0%) | 0 (0%) |
| | less busy | 0 (0%) | 2 (5%) |
| | as usual | 4 (40%) | 26 (63%) |
| | busy | 5 (50%) | 11 (27%) |
| | very busy | 1 (10%) | 2 (5%) |
| Please select the possible reason(s) listed below that contribute to the failure of sustaining task execution. (multiple choices possible) | I persisted | 6 (60%) | 35 (85%) |
| | I am busy during the period | 5 (50%) | 3 (7%) |
| | The tasks were too boring | 3 (30%) | 0 (0%) |
| | The tasks last too long | 4 (40%) | 3 (7%) |
| | It was bothersome to wear Fitbit device in daily life | 2 (20%) | 0 (0%) |
| | It is bothersome to answer the after task questionnaire | 0 (0%) | 0 (0%) |
| | I forget to open the app | 3 (30%) | 5 (12%) |
| From the following functionality of DailyExp, please select the one(s) that contributed to increasing your motivation for task execution. (multiple choices possible) | Performance Tracking | 2 (20%) | 6 (15%) |
| | Reward | 7 (70%) | 21 (51%) |
| | Ranking | 9 (90%) | 32 (78%) |
| | Encouragement right after task conduction | 2 (20%) | 15 (37%) |
| From the following possible improvements, please select the one(s) you believe would enhance the engagement and persistence for your task execution. (multiple choices possible) | Push notifications to remind me | 6 (60%) | 10 (24%) |
| | more monetary rewards | 4 (40%) | 34 (83%) |
| | apple watch instead of Fitbit | 2 (20%) | 1 (2%) |
| | the task become more interesting | 6 (60%) | 29 (71%) |
| | shorter duration | 8 (80%) | 18 (44%) |

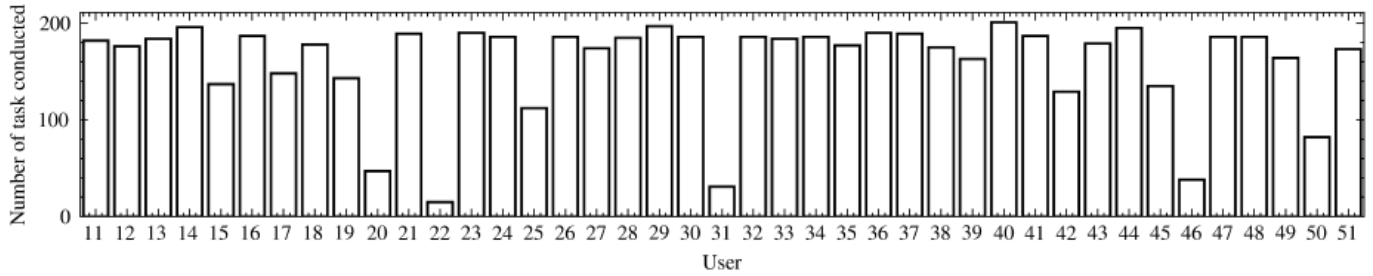


Fig. 8: Number of tasks conducted by each user in the user study.

- A limit of two valid completions per day was imposed for each task. After the limit is reached, task entry from the home screen will be disabled.
- The ranking screen was modified to display only the top 15 participants based on the number of task completions.

We conducted a user study involving 41 users to used the improved version of DailyExp in their daily life for one month. Both qualitative and quantitative analysis was performed.

A. Participants

In the study using an improved version of DailyExp, we recruited 41 participants (17 males and 24 females, aged 22-30) as app users, who are graduated students from the University of Tokyo. None of the participants from the preliminary study were permitted to take part in this subsequent study.

B. Procedure

The procedure of the orientation and instructions were almost the same as those conducted in the preliminary user study, despite that, we informed participants this time that there is a daily limit of two valid task completions for each type of task.

C. Quantitative Evaluation

1) *Data Collected*: We obtained a total of 6461 rounds of cognitive performance data, with 2115 rounds for the spatial continuous 2-back task, 2329 rounds for the Stroop task, and 2017 rounds for the FluidIQ task.

2) *User Engagement*: For the ratio of active users, as shown in Fig. 7, 38 out of the 41 users (other than user 22, 31, and 46) actively engaged with DailyExp. Consequently, the ratio of active users in the user study was 93%, showing a large improvement from that in the preliminary user study (60%).

Considering the number of tasks conducted, the average completion is approximately 157 tasks per user, which is three times the preliminary study's average of 55 tasks per user. For details, please refer to Fig. 8. We performed Mann-Whitney U test to compare the two groups of the number of tasks conducted. As results shown in Fig. 9, the group in the user study utilizing improved DailyExp showed a significant increase ($p=0.0008$) in the number of tasks conducted than that in the preliminary user study using the alpha version of DailyExp.

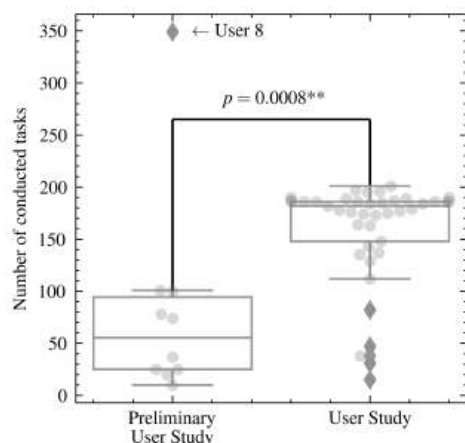


Fig. 9: A comparison of the number of conducted tasks between preliminary user study and user study using improved DailyExp

D. Qualitative Evaluation

1) *Reason of losing motivation:* In the user study, while the majority of the users sustained in task execution (85%), several users reported that they forgot to open the app and the experiment itself (12%).

2) *Functionality that increases motivation:* Similar to the result in the preliminary study, the Ranking screen was highly rated as contributing to increased motivation for task execution (78%). It is implied that a social competition atmosphere is a robust engaging feature and had a notable impact on the participants' motivation.

3) *Possible improvements:* The top suggested improvements were more monetary rewards (83%) and more interesting tasks (71%).

VI. DISCUSSION

The results of our study offer several insights into enhancing user engagement and motivation in mobile-based cognitive tasks. Firstly, the significant improvement in user engagement, from 60% to 93% of active user ratio and three times more task completion per user, indicated that the two modifications made to the app were effective in enhancing user motivation.

On the one hand, the implementation of a limit of two valid completions per day for each task appears to align with Locke's goal-setting theory [10]. In Locke's research, the effect of goal setting on motivation was emphasized, and it was confirmed that setting clear goals, which the individual accepts, leads to better performance compared to ambiguous goals.

On the other hand, the ranking screen revealed to be the most effective feature in the preliminary study, was enhanced by concealing information about inactive users. This enhancement likely contributed to the increased motivation by creating a more competitive environment and providing a clearer sense of progress and achievement for active users. The mechanism

behind this effect can be attributed to the psychological principle of social comparison, especially when their performance is visible and comparable. Thus, the concealment of inactive users may have heightened the perceived competition among active users, driving greater engagement and motivation.

VII. LIMITATIONS AND FUTURE WORK

It had to be addressed that we did not evaluate the two new features individually, meaning that the observed improvement in user engagement is likely an overall effect of both the introduction of achievable goals and the concealment of inactive users in the Ranking screen. This suggests that future studies should consider evaluating new features separately to better understand their impact on user motivation.

In future works, we plan to expand the coverage of cognitive aspects by administering more cognitive batteries and implement a web-based dashboard for experimenters, which would allow them to easily adjust system factors and design their experiments. We expect DailyExp to be a useful tool for creating a large-scale real-world cognitive performance and physiology database. This database has great potential to contribute to the field of cognitive science by providing valuable information for understanding individual differences, intra-personal fluctuations, and the embodied nature of cognitive processes. Potential research questions include studying the impact of human rhythms on cognitive processes across different timescales (e.g., daily circadian rhythm, monthly menstrual cycle) and identifying biomarkers of cognitive processes correlated with physiological features.

VIII. CONCLUSION

In this paper, we presented DailyExp as a comprehensive tool for collecting cognitive performance and physiological data in everyday life settings. Building upon the alpha version published in [1], we implemented two key improvements. Firstly, we enhanced the sense of achievable goals by limiting users to two valid completions per day for each task. Secondly, we improved the Ranking screen by concealing information about inactive users.

We conducted a one-month user study with 41 individuals, the results indicated that these updates effectively enhanced user engagement. Our study demonstrated the app's effectiveness as a practical smartphone application for conveniently collecting data in daily life settings, showing consistent usage by a significant portion of users and successful data collection across multiple tasks.

ACKNOWLEDGMENT

This work was supported by JST SPRING, Grant Number JPMJSP2108.

REFERENCES

- [1] Xianyin Hu, Yuki Ban, and Shin'ichi Warisawa. "DailyExp: A Tool for Collecting Cognitive Performance and Physiological Data in Daily Life with Engaging Behavioral Design." The Fifteenth International Conference on Advanced Cognitive Technologies and Applications, June 2023.

- [2] Wolfgang Tschacher, and Jean-Pierre Dauwalder, eds. "Dynamical Systems Approach To Cognition, The: Concepts And Empirical Paradigms Based On Self-organization, Embodiment, And Coordination Dynamics". Vol. 10. World Scientific, 2003.
- [3] Susan Jongstra, Liselotte Willemijn Wijsman, Ricardo Cachucho, Marieke Peternella Hoevenaars-Blom, Simon Pieter Mooijaart, and Edo Richard. "Cognitive testing in people at increased risk of dementia using a smartphone app: the iVitality proof-of-principle study." *JMIR mHealth and uHealth* 5, no. 5 (2017): e6939.
- [4] Zoë Tieghe, Antaine Strobhairt, Katie Scott, Klaudia Suchorab, Alexander Weir, Stuart Parks, Susan Shenkin, and Alasdair MacLulich. "Development of a smartphone application for the objective detection of attentional deficits in delirium." *International psychogeriatrics* 27, no. 8 (2015): 1251-1262.
- [5] Tilman Dingler, Albrecht Schmidt, and Tonja Machulla. "Building cognition-aware systems: A mobile toolkit for extracting time-of-day fluctuations of cognitive performance." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, no. 3 (2017): 1-15.
- [6] Pegah Hafiz, and Jakob Eyvind Bardram. "The ubiquitous cognitive assessment tool for smartwatches: design, implementation, and evaluation study." *JMIR mHealth and uHealth* 8, no. 6 (2020): e17506.
- [7] Xianyin Hu, Shinji Nakatsuru, Yuki Ban, Rui Fukui, and Shin'ichi Warisawa. "A physiology-based approach for estimation of mental fatigue levels with both high time resolution and high level of granularity." *Informatics in Medicine Unlocked* 24 (2021): 100594.
- [8] J. Ridley Stroop. "Studies of interference in serial verbal reactions." *Journal of experimental psychology* 18, no. 6 (1935): 643.
- [9] Edith Lavy, and Marcel Van den Hout. "Selective attention evidenced by pictorial and linguistic Stroop tasks." *Behavior Therapy* 24, no. 4 (1993): 645-657.
- [10] Edwin Locke, and Gary Latham. "Goal-setting theory." In *Organizational Behavior* 1, pp. 159-183. Routledge, 2015.

Assessment of Pharmacology Costs in Diabetes Treatment Using OMOP CDM: A Nationally Representative Study

Evgeniy Krastev

Faculty of Math. and Informatics
Sofia University St. Kliment Ohridski
Sofia, Bulgaria
eck@fmi.uni-sofia.bg

Petko Kovachev

Faculty of Math. and Informatics
Sofia University St. Kliment Ohridski
Sofia, Bulgaria
az@petko.info

Simeon Abanos

Faculty of Math. and Informatics
Sofia University St. Kliment Ohridski
Sofia, Bulgaria
simeonabanos@gmail.com

Ralitsa Krasteva

Specialized Hospital for Active Treatment
of Children's Diseases
Sofia, Bulgaria
ralitsakrasteva@outlook.com

Dimitar Tcharaktchiev

Department of Medical Informatics
Medical University
Sofia, Bulgaria
dimitardt@gmail.com

Abstract—Public diabetes reports rarely use large volume of observational health data because these data are usually heterogenous in terms of structure and semantic presentation of clinical concepts. In this paper we employ observational health data transformed to OMOP CDM database in a nationally representative study with the goal to estimate the pharmacology costs for diabetes treatment in Bulgaria during 2018. The OMOP CDM database contains health data from pseudonymized outpatient records of 501,065 patients with diabetes (45.3% male and 54.7% female). The mean age of the patients with Type 1 (42,249 patients) and Type 2 (458,816 patients) gives mean age 57.04 (CI 95%, [56.87, 57.22]) years for Type 1 and 66.38 (CI 95%, [66.35, 66.41]) years for Type 2. Drug costs are evaluated with respect to the major classes of drugs prescribed for diabetes treatment and diabetes comorbidities treatment. The annual average cost of drugs per patient with diabetes is estimated to 750 Euros. The obtained results are new and help to understand the trends and effects in using different classes of drugs for diabetes treatment. Novel drug diabetes therapies are found to be evolving in 2018, while the Metformin prescriptions prevail significantly. The costs in this study we evaluate both at patient-centric level by age groups and gender specifics and at high level in terms of cost distributions among the drug classes in each group. The results are graphically visualized, discussed, and compared in relation to existing public sources.

Keywords- *diabetes mellitus; diabetes register; nationally representative study; registry; database; OMOP CDM; Common Data Model; pharmacology; cost analysis.*

I. INTRODUCTION

Observational health data (OHD) are valuable resource for the assessment of cost effectiveness and management of healthcare services. Typical examples of OHD include values of plasma glucose, glycated hemoglobin (HbA1c) levels, measurements of blood pressure as well as prescriptions for drug therapy in accordance with a medical diagnose. Such clinical data are collected routinely during health procedures under real-world conditions during the execution of most of the business processes in the healthcare system. Therefore, OHD is indispensable in the evaluation of typical key

performance indicators for cost effectiveness like budget, satisfaction of healthcare standards and quality assurance. These indicators support decisions making for figuring out the reimbursement status of innovative drugs or introducing improvements in existing treatment protocols.

This paper extends the study in a related research work [1] with a deeper analysis of the drug prescriptions collected in a nationally representative study aiming to obtain accurate estimates of the pharmacology cost for treatment of patients suffering from diabetes mellitus (DM) Type 1 and Type 2. DM is a chronic metabolic illness with steadily increasing prevalence and associated to high risk for comorbidities such as blindness, kidney failure, heart attacks, stroke, and lower limb amputation [2] [3]. A large number of the patients with DM depend on life-saving insulin to control daily the glucose blood levels. Often, they need to take additionally one or more medications to avoid or treat the illness complications such as high blood pressure [4]. Therefore, DM treatment incurs significant expenses for life-long regular purchases of multiple medication products [5]. Part of the medication costs are reimbursed by health insurance funds, while the greatest part of these expenses harshly affects directly or indirectly the finances of the individuals, their families, and the society.

Recent research study [6] reports a total estimated cost of \$412.9 billion in 2022 (compared to \$327 billion in 2017 [7]) for treatment of diabetes in the USA, where a patient with diabetes spends on the average \$19,736 annually for medical treatment. Major sources for indirect costs include reduced or total loss of productivity, respectively, due to disability (\$28.3 billion) or premature death (\$32.4 billion). Thus, the medical expenditures of a patient with DM are 2.3 times higher than what expenditures would be in the absence of diabetes [8]. Cost effectiveness of diabetes treatment turns out to be even more important to consider in resource constrained low-income and middle-income countries [9] [10]. Most of the existing sources focus on the study of outpatient costs for treatment of DM Type 2 where the annual medication costs in 2018 are estimated in a wide range between \$15 and \$500 (median is \$177) depending on the country GDP [11]. It is noteworthy, that many of these studies aggregate and process data from statistical national-level surveys instead of using

original clinical documents. Therefore, the numerical results in these studies are obtained as statistical estimates from data extrapolated to the entire population of a country [12]. The thus obtained values of the numerical indicators in recent systematic reviews are suitable and sufficiently well describe the substantial financial burden imposed on the society by the diabetes illness. Apparently, management of healthcare services and improvement of medication treatment protocols of diabetes require a more accurate assessment of the pharmacology costs employing evidence-based data.

The limited availability of nationally representative data that provides patient-centric evidence for the medical treatment of diabetes appears to be the major obstacle to obtaining accurate evidence-based data for estimation of the pharmacology costs. Although huge volumes of OHD are generated in the healthcare system of each country, it is often difficult to overcome severe interoperability, methodological and organizational problems in any attempt to extend local evidence-based data processing at national level. Therefore, the number of national diabetes registers maintaining data about the health status of diabetics remains rather limited [13] [14]. The primary reason is that health data is persisted on heterogeneous platforms at multiple remote locations and besides, it is managed by disparate information models and technologies.

In practice, healthcare providers use Electronic Health Records (EHR) to persist systematically observational data as well as other medical information like prescribed medications, allergies, laboratory test results and demographics data about the patient in digital format [15]. Unlike the EHR, an Electronic Medical Record (EMR) such as the Outpatient record (OpR) provides a more detailed description of the patient's medical history than the EHR because it is maintained by a single healthcare provider (HP) [16]. Similarly to the EHR, the OpR captures rich observational data about the health status of a patient and allows the HP to follow it while prescribing treatment activities and procedures across time. In the general case, an EHR comprises the patient's EMRs from potentially different HPs. Thus, the EHR enables sharing of knowledge, skills and experience through communication between the actors in the healthcare system, serve the basis for research and education, satisfy organizational and legal requirements [17]. Nowadays, a lot of these opportunities for utilizing EHRs cannot be fully exploited. The reason is the lack of interoperability among the heterogeneous and proprietary nature of the software applications used by multiple HPs. Such interoperability problems stem from the primary distinction between EHRs and EMRs. EHRs are introduced for the purpose of sharing health data among organizations while EMRs serve the needs of a single HP. Therefore, the EMRs and in particular, the OpRs of a patient cannot be seamlessly integrated in the EHR of that patient.

Considerable research efforts have been made in the last twenty years to resolve the interoperability issues in the exchange of clinical data [18]. Data exchange schemas and standards for reference models have been introduced for sharing EHR data across clinicians, patients and communities [19] [20] [21]. This approach allows disparate health

information systems to effectively communicate, exchange data and process the exchanged data within and across the organizational boundaries. Services for accessing and sharing EHRs may accommodate their requirements with respect to three distinct levels of interoperability-foundational, structural and semantic interoperability [22]. Foundational interoperability is limited to the availability of information technology, allowing EHR data exchange. Structural interoperability upgrades foundational interoperability with requirements for representing the exchanged data in predefined syntax and thus, allowing interpretation of data at individual data field level. Most often interoperability at that level is used for exchange of observational data represented in terms of a Common Data Model (CDM) where the physical implementation could be a relational database or an XML Schema [23] [24]. The semantic interoperability level employs standard terminologies, classifications and vocabularies to encode EHR clinical data so that the receiving information systems can correctly interpret the clinical meaning of such data without human intervention [25] [26]. It is noteworthy that the clinical meaning is inferred not from the individual data values themselves rather from the way in, which such data are linked together as compound clinical concepts, hierarchically structured terms, problems or associated with preceding healthcare events. This interoperability level preserves the semantic context of the exchanged clinical data by representing clinical concepts in terms of standard reference models such as ISO/EN 13606 and HL7 FHIR. Therefore, the exchange of EHR extracts usually implements such semantic interoperability standards.

In this paper we consider a pharmacology case study that illustrates the potential of CDM to facilitate access to observational data and enhance population-based statistical research. It is motivated by the need for accumulating evidence on cost effectiveness and budget impact through Health technology assessment (HTA) [27]. The objective is to assess the burden of pharmacology costs spent for treatment of diabetes in a nationally representative dataset. The data source for this study is the Bulgarian Database Register (BDR) that is an Observational Medical Outcomes Partnership (OMOP) CDM standardized database publicly available at the EHDEN Portal [28] [29]. This database contains observational data (observation period 01.01.2018–31.12.2018) of all the outpatient records (6,887,876) issued in Bulgaria to patients with diabetes (501,065). The outpatient records are compiled by the general practitioners (GPs) and the specialists from ambulatory care for every patient encounter. In this case study the CDM appears to be the optimal solution for imposing structural interoperability in dealing with disparate data sources such as the variety of software applications employed to produce the outpatient records. Thus, the dataset of the BDR can be accessed remotely and return aggregated results by executing analytical code locally in the secure environment of the data custodian.

This paper is divided into sections as follows. In the following section, we make a brief overview of the existing CDM that enhance big medical data analytics [30] [31] [32] and elaborate on the OMOP CDM of the BDR. In Section III, we present aggregated results obtained by executing the

analytical code. In Section IV, we discuss the obtained results and compare them with existing research work [33]. Section V makes a conclusion and provides remarks on future work.

II. METHODS AND MATERIALS

This paper considers a case study where the original data sources are outpatient records created by a large number of GPs and specialists from ambulatory care using heterogeneous databases and client applications with disparate programming interface for data access, management and analysis. It entails problems caused by poor data interoperability such as patient-matching with observational data, pseudonymization of records, satisfying requirements for integrity and consistency of clinical data. The development of software tools for analysis and assessment of data in distributed dataset environment is rather complicated and inefficient as well. The need for imposing some kind of unification of these disparate data sources focused our attention on using CDM in this research.

The literature review provides convincing evidence that CDM are the preferred solution in cases of poor data interoperability when simultaneous analysis of disparate data sources is required [24] [34]. There are three most widely used CDMs for observational data research, namely, the OMOP CDM, the Sentinel and the Patient Centered Outcomes Research Institute (PCORNet). Each one of these CDMs has its strengths and weaknesses.

The PCORNet CDM [30] introduces its own standard organization and representation of EHR data for a distributed network of nine population-based Clinical Research Networks of data contributors (more than 14 billion diagnoses, 2.6 billion medication orders and 9.8 billion laboratory results) [35]. Nowadays, PCORNet persists data from the healthcare encounters of more than 30 million patients across the USA. The information model of the PCORNet CDM enables users to execute to query against the health data of a large number of patients and promptly receive the result in a standardized format. Special attention is dedicated to data quality by applying a two-stage for screening PCORNet-accessible data. During the first stage data is examined for conformance (adherence of EHRs to PCORnet CDM), completeness (diagnosis codes are not missing and correctly recorded), plausibility (ensure that the values are meaningful), and persistence (ensure that source data is preserved upon running queries). At the second stage we consider data quality issues, relevant to the specific research problem. A major weakness of this CDM is the missing support for clinical outcome measures as well as data linkage, for example, queries cannot “*de-duplicate*” patients appearing in multiple networks.

The Sentinel CDM was introduced in 2007 by the Federal Drug Agency (FDA) to monitor drug safety and includes EHR and register data in the following core subject areas utilization, registration, pharmacy, demographics, lab, death and vital signs (more than 463.3 million unique patient identifiers spanning the time period from 2000 to 2023, 19.7 billion drug dispensations, 20.2 billion unique medical encounters, 67.3 million patients with at least one laboratory test results) [31] [36]. Sentinel uses a distributed data approach in, which Data Partners maintain physical and operational control locally

over their electronic health data. The Data Partners are the organizations that collect routinely OHD on every patient encounter and transform it in the Sentinel CDM (SCDM) information model.

The SCDM is a data structure that standardizes administrative and clinical information across Data Partners. It is extensible to any data source because data is represented as detailed as possible avoiding embedded terminology mapping to maximize transparency and analytic flexibility. This way the network of Sentinel Data Partners allows FDA to generate larger datasets and study adverse events or drug dispensations even in small populations. Thus, the Sentinel CDM is flexible about demands for running data queries in any type of analysis. Data quality assurance practices are aligned with the FDA guidelines where three levels of quality checks are introduced to assess pre-defined data quality measures and characteristics [37]. Queries are processed in a distributed pattern as follows. Firstly, query requests are distributed to the data partners where the queries run locally. Next, query results with direct identifiers removed are returned to the central server for aggregation and final processing. It entails keeping copies of large amounts of data and time-consuming data synchronization even for simple queries. A critical system limitation is the inability to identify enough medical conditions of interest in OHD to a satisfactory level of accuracy [38]. Other weaknesses include limited data mapping, extensions of the CDM affect data usability, data granularity entails loss of information and local knowledge and finally, ongoing model refinements are driven entirely by the FDA.

The OMOP CDM was introduced about the same time as the Sentinel CDM for the purpose of studying the effects of medicinal products. Currently, it is extensively used in the US and Europe where it is underpinned by the Observational Health Data Sciences and Informatics (OHDSI) network and the EHDEN project of the EU (more than 153 EHDEN data partners, more than 1.12 billion unique patient identifiers) [39]. Similarly to Sentinel and PCORNet, the OMOP CDM maps disparate data sources to a “*patient-centric*” relational database with predefined tables linked directly or indirectly to patients. The tables correspond to the CDM core subject areas such as *PERSON*, *VISIT_OCCURRENCE*, *DRUG_EXPOSURE*, *MEASUREMENT*, *OBSERVATION*, *DEATH*. There are also tables describing *DEVICE_EXPOSURE*, *PROCEDURE_EXPOSURE* as well as standardized vocabularies for normalizing the meaning of data within the CDM. Thus, the OMOP CDM has the potential to meet the requirements of HTA.

The OHDSI OMOP CDM is well supported by software tools assisting the Extract-Transform-Load (ETL) process and ensuring data quality during the mapping steps. This allowed us to map to OMOP CDM health data from 6,887,876 outpatient records collected by the National Health Insurance Fund (NHIF) in Bulgaria from GPs or HPs upon the encounters of 501,065 patients with diabetes during 2018 [29] [40]. Meta data of the thus obtained OMOP CDM (v.5.3.1

[41]) of the Bulgarian Diabetes Register are published in the EHDEN Portal (Figure 1).

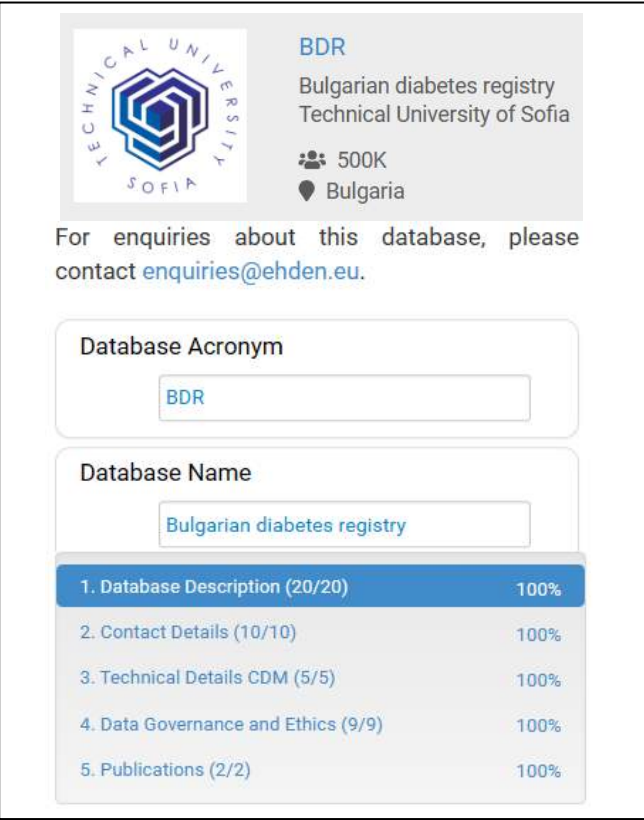


Figure 1. Link to the OMOP CDM of BDR inside the EHDEN Portal.

The distribution of diabetics (Type 1 and Type 2) relative to the population of the corresponding administrative region is displayed in Figure 2.



Figure 2. Distribution of patients with diabetes in Bulgaria in 2018.

This figure shows that most of the people living in the northern part of the country and especially, in the north-west part, have diabetes. These are the least populated regions of the country. It motivates us to explore the burden of costs spent for reimbursement of drugs for treatment of diabetes and

its related comorbidities (cardiovascular drugs, drugs for disorders of the eyes or the nervous and urological system), for the purpose of comparing it with related research work.

The original pseudonymized outpatient records were provided by the NHIF in XML format that needed data processing for making them valid against a single XML schema. For convenience, we loaded the adapted XML instances of outpatient records in a relational database that served as a source for the ETL process (Figure 3).

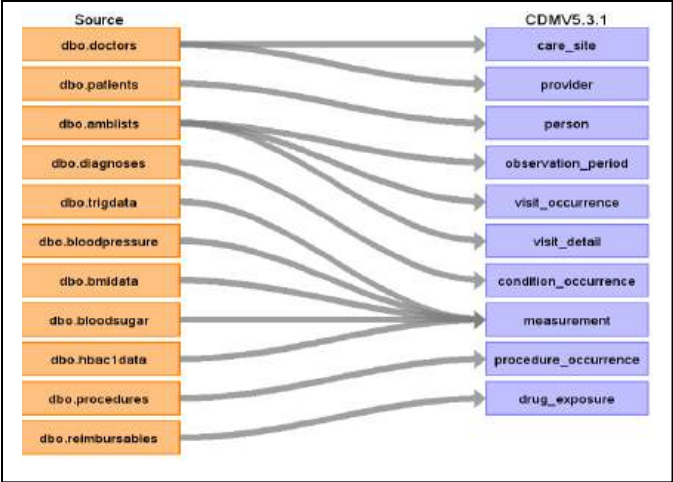


Figure 3. Mapping of outpatient records to OMOP CDM.

These records contain administrative data and coded clinical data describing health status or procedures such as:

- ✓ Date and time of the visit occurrence.
- ✓ Administrative data.
- ✓ Personal data, age, gender.
- ✓ Patient visit-related information.
- ✓ Diagnoses in ICD-10.
- ✓ ATC codes for medications reimbursed by the NHIF.
- ✓ Encodings for examinations and procedures.
- ✓ Codes describing specialized health care.
- ✓ Codes describing hospitalization need.
- ✓ Codes for planned consultations.
- ✓ Laboratory tests and medical imaging.

Observational data like patient *status*, *height*, *weight*, *Body-Mass-Index* or *blood pressure* were provided in the outpatient records as unstructured data in native language (Bulgarian text).

Special interests in this study represent the fields in the OMOP CDM table *drug_exposure* shown in Figure 4 where the field *drug_concept_id* encodes the drugs prescribed to diabetics and reimbursed by the NHIF [42]. It is noteworthy, that the Bulgarian national drug codes are represented in the ATC hierarchical classification system. Therefore, the standard vocabularies of the BDR are linked to ATC drug codes through *drug_concept_id*.

All other medicinal products represent no interest to the NHIF, and such products are recorded as unstructured free native text in the source dataset of outpatient records using their International Non-proprietary Names for Pharmaceutical Substances (INN). The extraction of INN of individual medications and mapping the INN to ATC codes requires a lot

of efforts and currently table DRUG_EXPOSURE does not include codes of non-reimbursed medicinal products.

Therefore, in this study the assessment of the pharmacology costs is limited to evaluating only the costs of medications that are reimbursed. In addition to table DRUG_EXPOSURE, the analytical code in this study makes use of tables PERSON, CONDITION_OCCURRENCE, VISIT_OCCURRENCE and OBSERVATION_PERIOD of the OMOP CDM. These are the core entities in the group of tables in the OMOP CDM database referred to as “Standardized clinical data”.

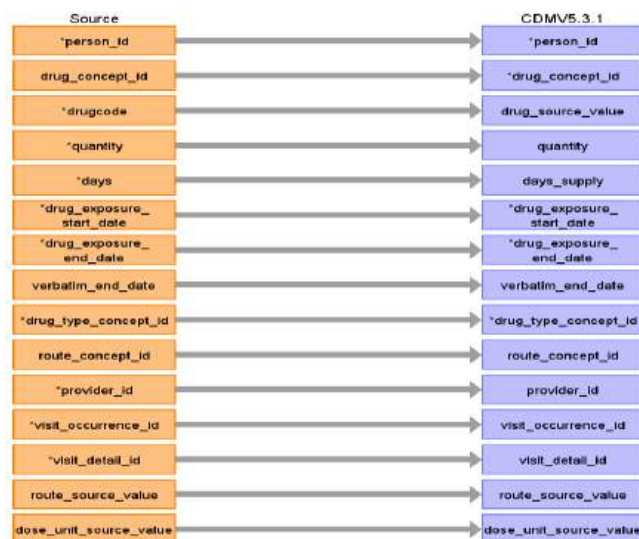


Figure 4. Mapping to table drug_exposure of the OMOP CDM.

Table 1. Drug classes for treatment of Diabetes.

| Code | Drug class | International Nonproprietary Name (INN) |
|------|--|--|
| T1 | Insulin | Insulin unique analogues and combination regimens |
| T2 | Sulfonylureas | Glyburide, Glipizide, Glimperide, Gliclazide, Tolbutamide, Chlorpropamide, Tolazamide, Vildagliptin |
| T3 | Biguanides | Metformin |
| T4 | Alpha-Glucosidase Inhibitors | Acarbose, Miglitol, Voglibose |
| T5 | Thiazolidinediones | Troglitazone, Rosiglitazone, Pioglitazone |
| T6 | Incretin-Dependent Therapies | Incretin, Exenatide, Liraglutide, Dulaglutide, Albiglutide, Lixisenatide, Semaglutide, Sitagliptin, Saxagliptin, Linagliptin, Alogliptin |
| T7 | Meglitinides | Nateglinide, Repaglinide |
| T8 | Sodium-Glucose Cotransporter Type 2 Inhibitors | Canagliflozin, Apagliflozin, Empagliflozin, Ertugliflozin, Dapagliflozin |
| T9 | Statin-Dependent therapies | Simvastatin, Lovastatin, Ravastatin, Fluvastatin, Atorvastatin, Cerivastatin, Rosuvastatin, Pitavastatin |

The existing literature distinguishes several distinct classes among the drugs for diabetes treatment [33] [43]. These classes are shown in Table 1, where the custom Code is introduced for the purpose of referencing the obtained results in the following section.

It is noteworthy, that currently, the drug class denoted as T8 in Table 1 is considered in the literature as the most modern and promising for DM treatment [33]. This is another reason to find out what is the share of prescriptions of these drugs. Similar interest represents the distribution of drugs prescribed for treatment of diabetes comorbidities.

The drug encodings displayed in Table 2 are introduced as shortcuts for referencing the drugs used for treatment of the most frequently encountered comorbidities of diabetes. By means of the respective drug Code in Table 2, it will be easier to quote these classes of drugs in the obtained results.

Table 2. Drug classes for Diabetes comorbidity treatment.

| Code | Drug class for comorbidity treatment | ATC code prefix |
|------|--------------------------------------|-----------------|
| A | Cardiovascular drugs | C01-C10 |
| A1 | Antithrombotic agents | B01 |
| N | Nervous system disorders | N01-N07 |
| G | Urological disorders | G04 |
| S | Ophthalmological disorders | S01 |
| L | Endocrine disorders | L02 |
| M | Treatment of bone diseases | M05 |
| R | Asthma drug categories | R03 |

Let's consider now the methods for evaluating the costs of the drug classes outlined in Table 1 and Table 2. The specification of the OMOP CDM v. 5.3.1. dedicates table COST to persist the cost of any medical case recorded in one of the OMOP tables such as DRUG_EXPOSURE, PROCEDURE_OCCURRENCE, VISIT_OCCURRENCE, VISIT_DETAIL or MEASUREMENT (Figure 4). The COST table belongs to a group of two tables entitled as “Standardized health economics” and it is not related referentially to any of the tables displayed in Figure 4. Once a payer (patient) completes a payment for a medicinal product or health service it triggers an event to record the payment details in the COST table. Such a scenario assumes the interaction of the OMOP CDM database with a kind of “claims” database and it is a use case that is feasible in a hospital environment. The outpatient records comprising the source dataset used in this study don't contain any payment details. Besides, ePrescriptions didn't exist in 2018. In view of the circumstances, instead of retrieving payment data from the COST table, we explored the publicly available price list for drugs reimbursed in 2018 by the NHIF [42]. By means of a custom developed Python script each of the prescribed drug codes was related to its price extracted from that price list.

Without loss of generality, in this study we assume that all the drugs prescribed to a patient are purchased in the current year. Based on this assumption, the “Standardized clinical data” set of tables in the OMOP CDM database allows to

group drug prescriptions and find totals per each prescribed drug that is reimbursed by the NHIF for patients suffering DM. Moreover, the “patient-centric” architecture of the OMOP CDM allows us to tally these numbers by type of diabetes, gender, age or any other property of the dataset. The results of executing these tasks are presented in the following section.

III. RESULTS

The BDR contains huge amounts of data that can provide rich information for treatment of diabetes. First of all, we get an accurate estimate for the diabetes prevalence (9.77%) in Bulgaria in 2018 (501,065 patients, 45.3% male and 54.7% female). Unlike other public data [12] [43], the diabetes prevalence is computed accurately taking into consideration the total number of individual patients with encounters registered by GPs or HPs and not by statistical estimates or extrapolation over the total population of the country.

Once we know the diabetes prevalence, it is important to learn what is the cost for diabetes treatment. The available data in the BDR allows to get detailed information on this issue from different perspectives. For shortness, here we present summary results that demonstrate the potential of HTA by limiting the scope of our research to drugs that are reimbursed by the National Health Insurance Fund as they are described in Table 1 and Table 2. The total cost (TC) of all the drugs prescribed for treatment of a patient with DM diagnose in Bulgaria in 2018 is 178,537,010 euros, where 96,201,239 euros is the amount for diabetes treatment with prescribed drugs from Table 1 that are reimbursed by the NHIF. These prescriptions cost the NHIF on the average about 356 euros annually per diabetic patient (Figure 5). Accordingly, 53.88% of the TC are for drugs prescribed for diabetes treatment (Table 1), where 61.55% is the share of the insulin class of drugs.

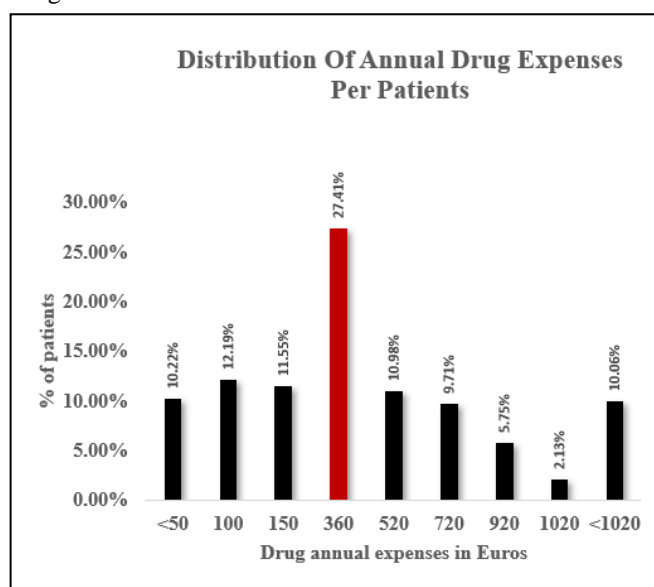


Figure 5. Distribution of annual pharmacology expenses for DM treatment.

A related research study [44] shows that the diabetes illness reaches its peak in the years after the age of 40. The expenses of patients with DM (Type 1 and Type 2) increase on the average up to 442 euros annually for this age group. Moreover, patients with diabetes undergo therapy for other illnesses like Hypertensive heart disease (ICD-10 code I11.0) or Tachycardia (ICD-10 code I48). Treatment of rare diseases or disorders caused by immune deficiency is even more expensive. These cases represent a huge burden in the overall amount reimbursed to patients for treatment of diabetes equivalent to 187,821,407 euros. Thus, the total average amount per patient that is paid annually by the NHIF for drug treatment of diabetes reaches 750 euros.

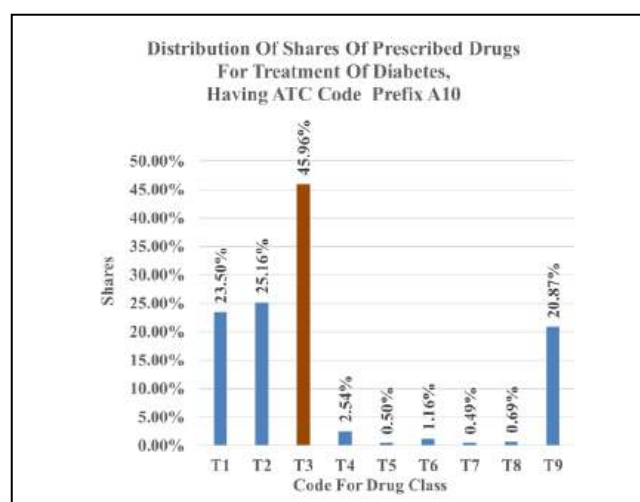


Figure 6. Shares of prescriptions for diabetes treatment.

Once we estimated the average cost for diabetes treatment, let's explore what is the share of modern drugs for diabetes treatment among all the prescribed drugs for diabetes treatment. Such are, for example, the drugs encoded as T8 in Table 1. Figure 6 shows that these drugs are rarely prescribed for diabetes treatment in Bulgaria during 2018 (0.69% of all the prescribed drugs from Table 1). Metformin drugs are the most frequently prescribed (T3 in Table 1). These kinds of drugs are usually prescribed for initial treatment of Type 2 diabetes and besides, the number patients with Type 2 diabetes (458,516; Male 45%, Female 55%) prevails significantly (91.5%) over the patients with Type 1 diabetes (42,249; Male 51%, Female 49%). This explains the peak value in the prescriptions for Metformin drugs (T3 in Table 1).

In terms of costs of the shares of the drugs in Table 1 are distributed as it is displayed in Figure 7. We notice that the largest expenses are attributed to the insulin class of drugs (T1 in Table 1) although it is the third most prescribed class of drugs in Figure 6. Drugs of that class are used for treatment of both DM with Type 1 and Type 2. For instance, the expenses for insulin drugs used for treatment of patients with DM Type 1 reach 99.13% of the total cost of drugs from Table 1 prescribed to these patients. Note, that the average price in Bulgaria for the insulin drug class has been about 60 euros against 16 euros for the Metformin drug class in 2018.

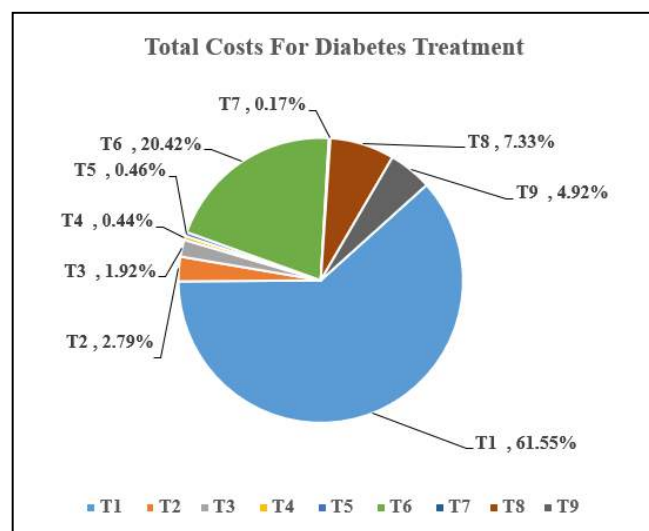


Figure 7. Total costs of drugs for diabetes treatment.

The above results provide evidence that the treatment of comorbidities accompanying the diabetes illness is almost as expensive as the treatment of the diabetes itself. Therefore, it is important to understand what the costs for treatment are of the most often encountered comorbidities.

In the existing literature there is enough evidence that the cardiovascular diseases, the disorders of the nervous system and the ophthalmological disorders are some of the most frequent comorbidities of diabetes. At the same time, little is known about the relative shares of these disorders with respect to the overall expenses for treatment diabetes comorbidities.

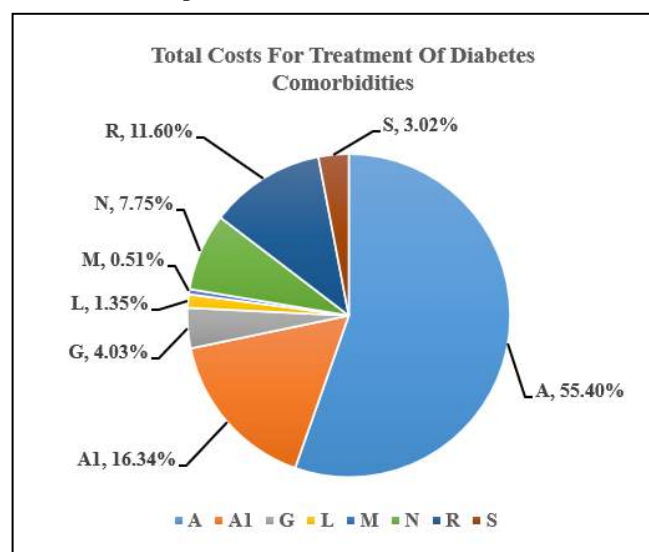


Figure 8. Total costs of drugs for treatment of diabetes comorbidities.

Figure 8 confirms that drugs for cardiovascular disorders and drugs with antithrombotic agents (code A and A1 in Table 2) have the greatest weight (71.74%) in the TC for treatment of comorbidities. The drugs for treatment of asthma of (code R in Table 2) are at the second place (11.60%) in the TC with average price of about 51 euros in 2018, where most of the

prescriptions are for medical products costing above the average value.

For comparison, the drugs for treatment of disorders of the nervous system (code N in Table 2) are at the third place with 7.75% share in the TC with average price of about 130 euros. Unlike the drugs prescribed for asthma treatment, most of these prescriptions are for medical products with prices significantly below the average for all the products with code N in Table 2. Such an increase in the costs for drugs prescribed to diabetics for treatment of accompanying asthma disorders is observed for the first time and it should be taken in consideration in regulatory decision making.

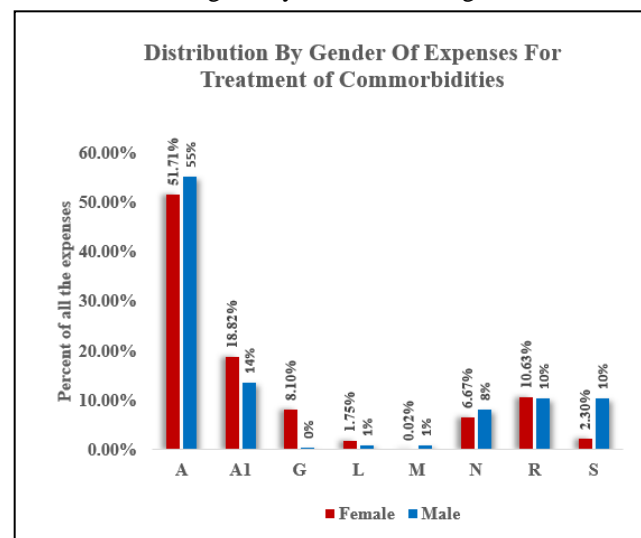


Figure 9. Distribution by gender of costs for Type 2 diabetes treatment.

The “patient-centric” architecture of the OHDSI CDM allows to investigate in greater detail the distribution of the pharmacology cost in terms of attributes gender and age in table PERSON (Figure 3).

For example, it is interesting to compare these costs for treatment of DM comorbidities subject to the gender of the patient. In Figure 9, we observe some notable differences in the distribution of these costs depending on the patient’s gender. Most significant differences are discovered in the expenses for treatment of comorbidities of classes A (cardiovascular), A1(antithrombotic), G (urological) and S(asthma) disorders. The expenses for treatment of male patients prevail classes A and S and that is a signal that comorbidities of DM are more specific to male patients. Similarly, we can conclude that comorbidities of classes A1 and G are more typical for female patients.

The analysis of the mean age of patients with Type 1 (42.249 patients) and Type 2 (458.816 patients) gives mean age 57.04 (CI 95%, [56.87, 57.22]) years for Type1 and 66.38 (CI 95%, [66.35, 66.41]) years for Type 2. The distribution in Figure 10 of the total costs of prescriptions per age groups of patients with Type 1 and Type 2 shows that the largest share of costs belongs to the age group 65-70 years. It coincides with the average age of patients with diabetes Type 2 that is the largest groups of patients receiving reimbursement by the NHIF for drugs used for treatment of diabetes.

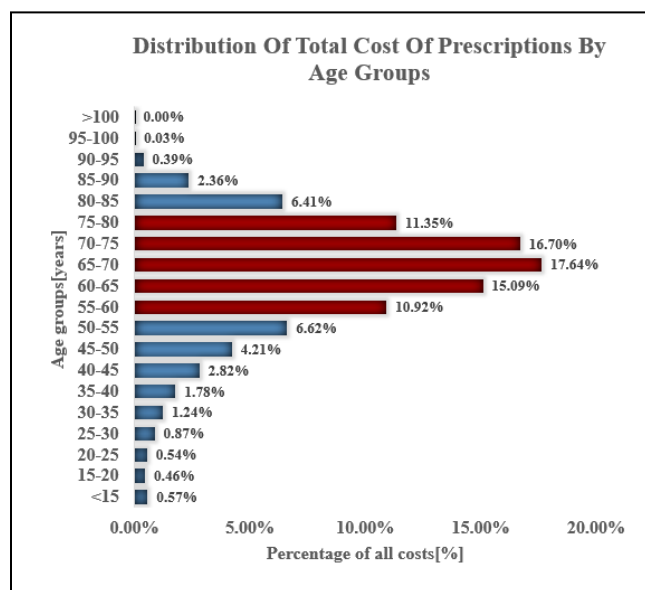


Figure 10. Distribution of costs by age.

IV. DISCUSSION

This paper reports results that are obtained by processing nationally representative data mapped to an OMOP CDM. The BDR is a physical implementation of that CDM with meta data published on the EHDEN Portal. It allows transparency in accessing data and verifying the integrity and consistency of these results. The BDR contains huge amount of pseudonymized observational data that allows to investigate diabetes treatment from different views through health assessment technologies.

A distinct feature of this study is that the obtained numerical results are obtained from evidence based OHD. The drug prescriptions are related to a large number of individual patients with a DM diagnose recorded in original clinical documents, the outpatient records issued by GPs and HPs on every patient encounter. Another important feature of the study is the nationally representative scope of the dataset of pseudonymized outpatient records used to extract data about the prescribed drugs. Unlike most statistical reviews, OHD for individual patients is not duplicated or extrapolated in this study. Data about diabetes prevalence coincide with recent reports. The here reported accurate value 9.7% of diabetes prevalence in Bulgaria in 2018 matches the statistically extrapolated prevalence data of the illness in Cyprus and Finland in 2021 [45]. A similar share of Diabetes Type 2 cases (91.5%) is established in more than 183 countries and territories [12].

The here considered pharmacology case study is just one example of the potential for exploring OHD mapped to an OMOP CDM. Without a restriction, data exploration could be extended to provide details with different level of granularity about the prescription of selected drugs or to group drug prescription by age and gender. In this regard, we must outline the following limitations that have to be taken in consideration.

First, it is rather difficult to find public literature with numeric data from population-based studies evaluating the burden of costs in diabetes treatment. In one such rare publication [43] we found evidence that matches close with our findings. Although this publication refers to data from 2014 and involves 312,223 patients from Italy, we established close correlation at several issues. For example, the share of costs on insulin drugs (T1 in Table 1) reported in that publication is 58.90% against the above quoted percentage 61.55%. Another match is established in the reported share of class A drug costs with respect to all drug costs 21.80% against 21.25 % found in our study with respect to the total of costs for drugs, prescribed for treatment of DM diagnose. There is, however, a significant difference in the average cost per diabetic patient, 1066 euros against 750 euros established from data in the BDR. This difference could be attributed to the known differences in the standard of life (and price levels) between both countries at that time.

Another issue that must be taken in consideration is that the NHIF does not reimburse always the full costs for prescribed drugs, while the amounts above quoted refer to the full drug costs. Moreover, in this study the analysis of costs considers data for patients that have health insurance.

Since the finance reports of NHIF are public [46], we managed to calculate the amounts really reimbursed by the NHIF for diabetic drugs (Table 1) to be 67,208,241 euros in 2018. As expected, this amount is about 30% less than the amount reported in the above section (96,201,239 euros). Here we must take in consideration that only a fraction of all the prescribed drugs in 2018 are dispensed to patients in the same year. Besides, the quantities of the prescribed drugs are usually greater than the quantities of the reimbursed drugs. Thus, we can conclude that the results reported in this paper are consistent with the real-life practice.

V. CONCLUSION AND FUTURE WORK

This paper demonstrates the potential of the OMOP CDM to facilitate access to observational data accumulated from heterogenous datasets and extract knowledge using standard statistical tools. The assessment of the burden caused by the pharmacology costs on the healthcare system is important for regulatory decision making as well as for drug suppliers in planning their market strategies. Even though the assessment of the pharmacology costs in this study considers only the drugs reimbursed by the NHIF and data for patients with health insurance, the study produces an evidence-based estimate of the financial burden of DM on the society from different points of view including gender and age group distributions. The obtained results help to understand the trends and effects in using different classes of drugs for diabetes treatment and especially, the trends in applying novel drug therapies for diabetes treatment. Public diabetes surveillance reports with such results are rather rare to find in the existing literature primarily because most often the datasets are heterogenous in terms of structure and lack of interoperability of the data sources. Unlike most regularly published reports in the public space, this paper reports results obtained from a population-based study rather than applying aggregated statistical estimates.

The BDR implements an open-source OMOP CDM that allows overcoming poor interoperability among heterogeneous and often, incompatible data providers. It contains the latest and complete dataset of outpatient records issued to 501,065 distinct patients with diabetes in Bulgaria at every encounter to GP or HP in 2018. Among other CDM briefly reviewed in this paper the OMOP CDM proves the best potential for applying health assessment technology in obtaining reliable, transparent and verifiable results though analysis of observational data.

The pharmacology case study makes public lot of new results that help understand better the burden of costs generated in the process of prescribing drugs for diabetes treatment. Two major groups of drugs are considered, drugs for treatment the diabetes and drugs for treatment of diabetes comorbidities. This study presents numerical evidence that the cost for treatment of diabetes comorbidities is as expensive as treatment of diabetes itself. It emphasizes the need for developing better strategies and policies for prophylactic and control of the diabetes illness in order to minimize deterioration of the patient health status and respectively, the minimize the cost burden on the society.

Numerical evidence shows that novel drug therapies of diabetes in this country are just beginning to evolve in 2018, while the prescriptions of Metformin drugs prevail significantly among all the rest. Contrary to the expectations, the costs of prescribed drugs for treatment of comorbidities in diabetes caused by asthma surmount the costs of prescribed drugs for therapy of the nervous system or urological disorders. The costs are evaluated both at patient-centric level as well as at high level in terms of cost distributions among the drug classes in each one of the two groups. The results are graphically visualized, discussed and compared in relation to existing public sources.

In our future work we focus on exploring the trends in using novel drug therapies for diabetes in Bulgaria. Preliminary results based on new public data sources during 2018-2021 show a significant and rapid increase in prescriptions of novel drug class therapies (T8 in Table 1), decrease in other prescriptions (T7 in Table 1) and stable interest in other (T3 in Table 1). Moreover, we work on updating the BDR with fresh data once it becomes available.

ACKNOWLEDGMENT

This research is supported by grant 4.80-10-59/09.04.2024 of the Scientific Research Fund of Sofia university St. Kliment Ohridski.

REFERENCES

- [1] E. Krastev, P. Kovachev, S. Abanos, and D. Tcharaktchiev, "Common Data Model for Interoperability of Observational Health Data: Bulgarian Diabetes Register Pharmacology Case Study", in GLOBAL HEALTH 2023: The Twelfth International Conference on Global Health Challenges, pp. 25-31, 2023. Available: https://www.thinkmind.org/articles/global_health_2023_1_50_70084.pdf. [Accessed 15 March 2024].
- [2] WHO, "Diabetes" 10 November 2021. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/diabetes>. [Accessed 10 March 2024].
- [3] International Diabetes Federation, "IDF Diabetes Atlas" 2021. [Online]. Available: <https://diabetesatlas.org/>. [Accessed 1 March 2024].
- [4] Centers for Disease Control and Prevention, "National Diabetes Statistics Report website. Estimates of Diabetes and Its Burden in the United States" 15 March 2023. [Online]. Available: <https://www.cdc.gov/diabetes/data/statistics-report/index.html>. [Accessed 15 March 2024].
- [5] R. Turner, C. Cull, V. Frighi, and R. Holman, "Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus: progressive requirement for multiple therapies (UKPDS 49). UK Prospective Diabetes Study (UKPDS) Group" JAMA, vol. 281, no. 21, pp. 2005-2012, 1999.
- [6] E. D. L. Parker, Janice, T. Mahoney, N. Ume, G. Yang, R. A. Gabbay, N. A. ElSayed, and R. R. Bannuru, "Economic Costs of Diabetes in the U.S. in 2022" Diabetes Care, vol. 47, no. 1, pp. 26-43, 2024.
- [7] R. Bosetti, L. Tabatabai, G. Naufal, T. Menser, and B. Kash, "Comprehensive cost-effectiveness of diabetes management for the underserved in the United States: A systematic review" PLoS one, vol. 16, no. 11, pp. e0260139, 2021.
- [8] American Diabetes Association, "Economic Costs of Diabetes in the U.S. in 2017" Diabetes Care, vol. 41, no. 5, pp. 917-928, 2018.
- [9] M. Kamusheva, M. Vassileva, A. Savova, M. Manova, and G. Petrova, "An Overview of the Reimbursement Decision-Making Processes in Bulgaria as a Reference Country for the Middle-Income European Countries" Frontiers in Public Health, vol. 6, 2018.
- [10] D. Flood, J. Seiglie, M. Dunn, S. Tschida, M. Theilmann, M. Marcus, et al., "The state of diabetes treatment coverage in 55 low-income and middle-income countries: a cross-sectional study of nationally representative, individual-level data in 680 102 adults" Lancet Healthy Longev, vol. 2, no. 6, pp. e340-e351, 2021.
- [11] C. Moucheraud, C. Lenz, M. Latkovic, and V. J. Wirtz, "The costs of diabetes treatment in low-and middle-income countries: a systematic review" BMJ Glob Health., vol. 4, no. 1, p. e001258, 2019.
- [12] GBD 2021 Diabetes Collaborators, "Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021" Lancet, vol. 402, pp. 203-234, 2023.
- [13] J. C. G. Bak, E. H. Serné, M. H. H. Kramer, M. Nieuwdorp and C. L. Verheugt, "National diabetes registries: do they make a difference?" Acta diabetologica, vol. 58, no. 3, pp. 267-278, 2021.
- [14] University Specialized Hospital for Active Treatment in Endocrinology, "Diabetes Register" USBALE, 2024. [Online]. Available: <https://usbale.org/bg/registar-zaharen-diabet/>. [Accessed 11 January 2024].
- [15] K. Dipak, "Electronic Health Record Standards, "Yearbook of medical informatics, vol. 45, no. 01, pp. 136-144, 2006.
- [16] L. Spooner and K. Pesaturo, "The Medical record" in Fundamental Skills for Patient Care in Pharmacy Practice, Jones & Bartlett Learning, 2014.
- [17] B. Blobel, "Interoperable EHR Systems—Challenges, Standards and Solutions" European Journal for Biomedical Informatics, vol. 14, no. 2, pp. 10-19, 2018.
- [18] HIMSS, Dictionary of Healthcare Information Technology Terms, Acronyms and Organizations, 5th ed., CRC Press, 2019, p. 75.
- [19] HL7, "HL7 FHIR release 5" HL7, 26 March 2023. [Online]. Available: <https://www.hl7.org/fhir/>. [Accessed 11 March 2024].

- [20] ISO/TC 215, "ISO 13606-1:2019 Health Informatics-Electronic health record communication-Part 1: Reference model" ISO, 2019. [Online]. Available: <https://www.iso.org/standard/67868.html>. [Accessed 6 March 2023].
- [21] openEHR, "The EHR Information Model" The openEHR Foundation, 17 November 2022. [Online]. Available: https://specifications.openehr.org/releases/RM/latest/ehr.html#_the_ehr_information_model. [Accessed 14 March 2024].
- [22] HIMSS, "Interoperability in Healthcare" Healthcare Information and Management Systems Society, 2023. [Online]. Available: <https://www.himss.org/resources/interoperability-healthcare>. [Accessed 11 March 2024].
- [23] J. Weeks and R. Pardee, "Learning to Share Health Care Data: A Brief Timeline of Influential Common Data Models and Distributed Health Data Networks in U.S. Health Care Research" eGEMs (Generating Evidence & Methods to improve patient outcomes), vol. 7, no. 1:4, pp. 1-7, 2019.
- [24] S. Kent, E. Burn, D. Dawoud, P. Jonsson, J. T. Østby, N. Hughes, P. Rijnbeek and J. C. Bouvy, "Common Problems, Common Data Model Solutions: Evidence Generation for Health Technology Assessment" Pharmacoeconomics, vol. 39, no. 3, pp. 275-285, 2021.
- [25] SNOMED International, "SNOMED CT" 2023. [Online]. Available: <https://www.snomed.org/>. [Accessed 14 March 2024].
- [26] WHO, "Anatomical Therapeutic Chemical Classification System" WHO Collaborating Centre for Drug Statistics Methodology, May 2023. [Online]. Available: <https://www.whocc.no/>. [Accessed 11 June 2023].
- [27] R. Claire, J. Elvidge and D. Dawoud, "EHDEN-D2.7-Workshop for HTA community report" Zenodo (<https://doi.org/10.5281/zenodo.7550919>), 2023.
- [28] EHDEN, "Bulgarian Database Register (BDR)" EHDEN Portal, 2023. [Online]. Available: <https://portal.ehden.eu/>. [Accessed 5 March 2024].
- [29] E. Krastev and E. Markov, "Bulgarian Diabetes Register-mapping onto OMOP CDM (CC#10)" January 2023. [Online]. Available: EHDEN Community Calls, <https://www.ehden.eu/bulgarian-diabetes-register-mapping-onto-omop-cdm-cc10/>. [Accessed 15 March 2024].
- [30] PCORnet, "PCORnet Common Data Model" The National Patient-Centered Clinical Research Network, 22 October 2020. [Online]. Available: <https://pcornet.org/news/pcornet-common-data-model/>. [Accessed 10 March 2024].
- [31] Sentinel, "Sentinel Common Data Model" Sentinel Operations Center, 3 June 2022. [Online]. Available: <https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model/sentinel-common-data-model>. [Accessed 11 March 2024].
- [32] Observational Health Data Sciences and Informatics, "OMOP Common Data Model" ohdsi.org, 2022. [Online]. Available: <https://ohdsi.github.io/CommonDataModel/>. [Accessed 10 March 2024].
- [33] A. D. Dahlén, G. Dashi, I. Maslov, M. M. Attwood, J. Jonsson, V. Trukhan, et al., "Trends in Antidiabetic Drug Discovery: FDA Approved Drugs, New Drugs in Clinical Trials and Global Sales" Front Pharmacol., vol. 12:80754, pp. 1-16, 2022.
- [34] H. Liyanage, S. Liaw, J. Jonnagaddala, W. Hinton and S. de Lusignan, "Common Data Models (CDMs) to Enhance International Big Data Analytics: A Diabetes Use Case to Compare Three CDMs" in Volume 255: Decision Support Systems and Education, Studies in Health Technology and Informatics, IOS Press, 2018, pp. 60-64.
- [35] C. B. Forrest, K. M. McTigue, A. F. Hernandez, L. W. Cohen, H. Cruz, K. Haynes, et al., "PCORnet@ 2020: current state, accomplishments, and future directions" Journal of clinical epidemiology, vol. 129, p. 60-67, 2021.
- [36] European Medicines Agency, "Common Data Model for Europe? -Why? Which? How?" 11-12 December 2018. [Online]. Available: https://www.ema.europa.eu/en/documents/report/common-data-model-europe-why-which-how-workshop-report_en.pdf. [Accessed 10 March 2024].
- [37] Sentinel Operations Center, "Sentinel Data Quality Assurance Practices" 27 February 2017. [Online]. Available: https://www.sentinelinitiative.org/sites/default/files/data/distributed-database/Sentinel_DataQAPractices_Memo.pdf. [Accessed 22 March 2024].
- [38] J. Brown, J. Maro, M. Nguyen and R. Ball, "Using and improving distributed data networks to generate actionable evidence: the case of real-world outcomes in the Food and Drug Administration's Sentinel system" JAMIA, vol. 27, no. 5, pp. 793-797, 2020.
- [39] Observational Health Data Sciences and Informatics, The Book of OHDSI, <https://ohdsi.github.io/TheBookOfOhdsi/>, 2021.
- [40] National Scientific Program eHealth, "National Scientific Program "Electronic Healthcare in Bulgaria" (e-health)" Sofia University. Faculty of Mathematics and Informatics, 2024. [Online]. Available: <https://ehealth.fmi.uni-sofia.bg/>. [Accessed 10 January 2024].
- [41] Observational Health Data Sciences and Informatics, "OMOP Common Data Model v5.3" 2021. [Online]. Available: <https://ohdsi.github.io/CommonDataModel/cdm53.html>. [Accessed 22 March 2024].
- [42] National Council on Prices and Reimbursement of Medical Products, "Appendix 1. Medicinal products intended for the treatment of diseases, which are paid under the Health Insurance Act (in Bulgarian)" 2 January 2019. [Online]. Available: https://www.nhif.bg/upload/12053/Izmenenie_Pril1_PLS-NZOK_01_01_2019.xls. [Accessed 22 March 2024].
- [43] L. Scalone, G. Cesana, G. Furneri, R. Ciampichini, P. Beck-Peccoz, V. Chiodini, et al, "Burden of Diabetes Mellitus Estimated with a Longitudinal Population-Based Study Using Administrative Databases" PLoS ONE, vol. 9, no. 12, pp. 1-11, 2014.
- [44] E. Krastev, S. Abanos, P. Kovachev and D. Tcharaktchiev, "Diabetes Prevalence and Duration Data Extracted from Outpatient Records Representative for the Bulgarian Population" in Stud Health Technol Inform., IOS Press, 2023, pp. 230-233.
- [45] statista, "Prevalence of diabetes in adult population in Europe 2021, by country" 2021. [Online]. Available: <https://www.statista.com/statistics/1081006/prevalence-of-diabetes-in-europe/>. [Accessed 22 March 2024].
- [46] NHIF, "Report on quarterly expenses for 2018. National Health Insurance Fund (in Bulgarian)" 4 November 2019. [Online]. Available: https://www.nhif.bg/bg/medicine_food/quarter-payments/2022. [Accessed 12 March 2024].

Building a Smarter Care Home: A Deep Dive Into HealthSonar's Architecture for Assisted Living

Adamantios Ntanis^{✉*}, Spyridon Kontaxis*, George Rigas*, Anastasia Pentari[†], Kostas Tsiouris[‡], Efstathios Kontogiannis*, Eleftherios Kostoulas*, Ilias Tsimperis*, Theodoros Vlioras*, Aristotelis Bousis*, Styliani Zelilidou[‡], Kalypso Tasiou[‡], Manolis Tsiknakis^{‡§}, and Dimitrios Fotiadis[¶]

*PD Neurotechnology Ltd., Ioannina, Greece

Email: {a.ntanis, s.kontaxis, g.rigas, s.kontogiannis, e.kostoulas, i.tsimperis, t.vlioras}@pdneurotechnology.com

[†]Institute of Computer Science, Foundation for Research and Technology—Hellas, Heraklion, Greece

Email: anpentari@ics.forth.gr, tsiknaki@ics.forth.gr

[§]Department of Electrical and Computer Engineering, Hellenic Mediterranean University, Heraklion, Greece

[‡]Biomedical Research Institute, Foundation for Research and Technology—Hellas, Ioannina, Greece

Email: tsiourisk@gmail.com, stylianizelil@gmail.com, kelkeltasi@gmail.com

[¶]Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, Ioannina, Greece

Email: fotiadis@cs.uoi.gr

Corresponding author: Adamantios Ntanis, a.ntanis@pdneurotechnology.com

Abstract—This study introduces HealthSonar, an innovative, unobtrusive and privacy-preserving health monitoring system, designed for the continuous monitoring of elderly individuals and patients with movement disorders, such as Parkinson's disease. HealthSonar offers a suite of features, including sleep quality tracking, sleep apnea event identification, mobility assessment, real-time fall detection and fall notification. The system itself comprises an impulse-radio, ultra-wideband radar device, a web portal, a dashboard, and a mobile application. HealthSonar is particularly well-suited for use in potentially hazardous areas of care homes and hospitals (such as bathrooms), where continuous monitoring of their residents/inpatients is essential, especially during nighttime. To demonstrate the system's utility, a care home deployment is presented, detailing the installation of the radar-based device, the monitoring procedure of the residents as well as the management of the system. Through this application, it is effectively demonstrated that HealthSonar can be seamlessly integrated into clinical settings improving the quality of life of elders/patients, preventing injuries and reducing the workload of their personnel as well as of the overall healthcare system in general.

Index Terms—ultra-wideband radar; health monitoring; sleep monitoring; gait analysis; fall detection; care home monitoring; assisted living.

I. INTRODUCTION

THIS work builds upon the foundation laid by a preceding study [1] which introduced a novel health and wellness monitoring system, the HealthSonar, based on ultra-wideband (UWB) ambient sensing radar technology, perfectly suited for monitoring elders and patients with sleep, neurological and movement disorders.

The world's older population is rising, along with the number of patients suffering from sleep, neurological and movement disorders [2]–[5]. More often than not, these two groups share similar health problems, usually revolving around sleep and mobility. Impaired mobility, in particular, can result in falls, which are common, yet serious, incidents that can lead to life-threatening injuries, or permanent disability. Falls and their associated costs are expected to surge significantly over the next two decades, posing a growing burden on the healthcare systems over time [6]. Current technologies fall short of addressing the growing need for practical wellness and health monitoring systems, particularly for the expanding elderly and patient populations. Ideal systems would monitor sleep, evaluate mobility, and identify falls – functionalities crucial for this demographic. A practical solution, targeting a holistic approach to wellness, could be of even greater interest, specifically for care homes and clinical environments, where residents are primarily elders and patients, in need of constant monitoring and care, for various health emergencies.

As of now, to the best of our knowledge, there is no single unified solution offering sleep monitoring, mobility monitoring and fall detection. Existing approaches, usually offer only a fraction of those features and lie in the use of either medical-grade diagnostic equipment (such as polysomnography), camera-based systems, or wearable solutions (such as actigraphy) [7]–[10]. Those systems can be expensive, cumbersome, privacy-invading, or solely geared towards clinical use. There are commercially available devices, such as smartwatches and fitness trackers, offering similar functional-

ity, but their wearable form significantly limits their usability. This is especially true for sleep tracking purposes, as a wearable solution would be cumbersome leading to user discomfort. Furthermore, current commercial devices often rely on complex on-device interfaces, alongside web and mobile applications with intricate functionalities. The primary challenge lies not with the applications themselves, but rather with the nature of the employed user interfaces (UIs). These UIs frequently neglect to incorporate design principles that cater to the specific requirements of elderly users and patients, thereby creating substantial barriers to adoption and use within these populations. For elderly users, unfamiliar with technology, clear and well-documented user interfaces (UIs) are essential. Patients with movement disorders, on the other hand, benefit from larger icons and minimal interactive elements to accommodate dexterity limitations.

Lately, progress in ultra-wideband (UWB) radar technology has resulted in the development of affordable, practical and accurate radar sensors [11]. These sensors can serve as a platform for building wellness and health monitoring solutions capable of addressing the needs of both elders and patients. Specifically, because of their high accuracy, penetration capabilities and reliability, UWB radars can track micro and macro motions, even through different weather conditions and obstacles (such as walls and furniture), making them ideal for a diverse set of applications. Among these are respiratory and heart rate extraction, presence detection, fall detection, people counting, gesture recognition, baby monitoring, assisted living of elderly people, as well as mobility monitoring. What is more, UWB radar technology provides an unobtrusive, contactless, low-consumption and privacy-focused approach, perfectly suited for devices meant to be used for long periods of time, inside the most private areas of people's homes, such as bedrooms, bathrooms etc. Due to their nature, radar-based devices are a suitable choice for applications pertaining to sleep, where a wearable solution would inevitably cause discomfort, while a camera-based solution would be privacy-intrusive [12]. Moreover, radars offer unparalleled opportunities for nighttime activity/movement monitoring as their penetrating abilities result in motion evaluation and posture recognition even in low temperatures where people would be covered with blankets during sleep [12]. Given radar-based devices are constantly emitting energy in the form of radiation, a logical concern would be that of safety regarding their continual use. Thankfully, UWB radars are safe to use, due to their low emitted power levels of non-ionizing radiation that are harmless for human health [11]. The contactless nature of radar-based devices unlocks new possibilities for continuous monitoring in assisted living and general wellbeing applications. Unlike wearable sensors that can be forgotten, misplaced, or uncomfortable for long-term use, radar-based devices offer a passive monitoring solution that seamlessly integrates into daily routines. This facilitates unobtrusive tracking of activity levels, sleep patterns, and even subtle changes in movement, empowering individuals in both assisted living and wellness contexts to gain valuable insights into their wellbeing.

UWB radars have been used extensively for wellness and

health monitoring in research settings for various applications. In the context of this work, we are mostly interested in the state-of-the-art research about UWB radar technology applied to vital signs extraction, sleep monitoring, presence detection, fall detection and mobility monitoring. The extraction of the respiratory and heart rate using devices based on UWB technology [13], [14], takes advantage of the ability of radar sensors to capture micro motions, in this case the oscillating motion of the chest resulting from the movement of the lungs and heart. By analyzing the subtle changes in the reflected radar signal caused by these micro motions, sophisticated algorithms can accurately extract respiratory and heart rate information. This contactless approach eliminates the need for uncomfortable or intrusive wearable sensors, offering a more natural and user-friendly way to monitor vital signs. Sleep monitors based on UWB radars, have been developed either as medical-grade diagnostic equipment with the goal of substituting the gold standard, yet cumbersome and expensive, polysomnography (PSG) or as general wellness tools for evaluating the quality of one's sleep. In either case, vital signs are commonly paired with additional features obtained from radar data, such as body posture or night movement [15]–[17]. This allows for applications such as detection of obstructive sleep apnea events [13], [18], [19] (the most frequent breathing disorder occurring during sleep [20]), differentiation between sleep and wake states, or classification of sleep stages [12], [16], [21], among others. UWB radars are gaining traction as ambient sensors in assisted living environments. They excel at human presence detection and fall identification, promoting safety and well-being for residents [22]–[24]. Radar-based devices are privacy-oriented and are ideal for installation in areas such as bedrooms and bathrooms, where fall incidents are relatively common for elders and patients with movement disorders. Last but not least, UWB technology has been applied to human mobility evaluation [25]–[27], a task that up to now primarily relied on wearables, such as devices based on inertial measurement units (IMUs) and instrumented insoles, or highly specific laboratory equipment, such as pressure sensitive mats and walkways. In applications like human presence detection, mobility monitoring, and fall detection, UWB radar sensors leverage their ability to remotely and precisely capture large-scale human movements (macro-motions), even through objects. This distinctive capability translates into rich spatiotemporal data, providing valuable insights into a person's movement patterns and location over time.

Within this work, we outline the different elements composing the HealthSonar system, its structure, and its diverse functions. Additionally, we present a detailed application of the system in a care home catering to elders or patients with movement disorders, such as Parkinson's disease. A care home application was specifically chosen because both residents and employees can greatly benefit from a system like HealthSonar, which was specifically designed and built with such an environment in mind. The system can be easily deployed in a plethora of other scenarios, even as a general wellness tool for improving the quality of life of both healthy individuals and patients, but its applicability in clinical environments and care homes is particularly noteworthy.

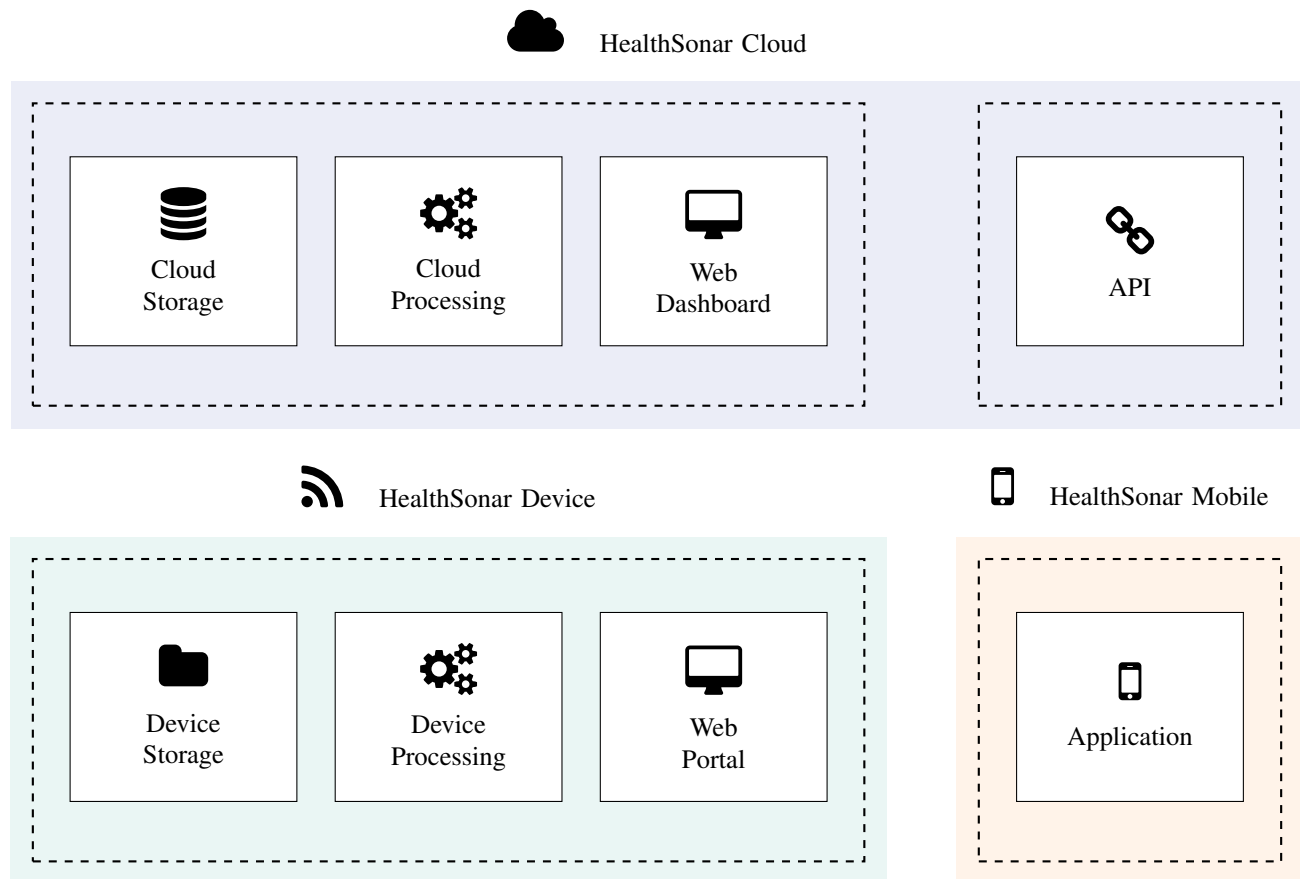


Fig. 1: The architecture of the HealthSonar system.

II. THE HEALTHSONAR SYSTEM

The HealthSonar system was developed to fill the gap in continuous, accurate, unobtrusive, privacy-focused and contactless health monitoring, designed to support elders (focusing on those with deteriorated mobility) and patients with sleep, or movement disorders. It achieves this by tracking the quality of their sleep, identifying sleep apnea events, evaluating their mobility and detecting fall incidents. The development of the system was propelled by recent advances in IR-UWB radar technology, especially the small size of available sensors and their affordable price, rendering them commercially viable options to serve as platforms for building health monitoring devices. The HealthSonar system consists of the following “components” (see Figure 1):

- 1) a radar-based monitoring device,
- 2) a web dashboard application,
- 3) a web portal to the radar device,
- 4) a mobile application (or app),
- 5) a Cloud data processing service.
- 6) a Cloud data storage service, and
- 7) an API for communication purposes.

The system was developed as an ecosystem catering to different users serving different purposes, thus it was designed for use by (1) elders and patients, (2) caregivers (for example,

family members or nurses and healthcare professionals of a care home), as well as by (3) attending physicians (for example, neurologists or sleep experts of hospitals and clinics). It was also designed to provide different administrative and monitoring utilities to each one of the aforementioned groups based on their roles. The HealthSonar system is meant to be straightforward and easy to use, while still being customizable for advanced, mainly research, use cases. Prioritizing the needs of the target audience, uncompromising usability was a core principle throughout the design process.

It is important to note that HealthSonar was built as a centralized, scalable system. As a result, multiple devices can be connected and administrated at once in order to cover the needs of large organizations serving many residents or inpatients. On the other hand, the system is modular enough to be usable even as a single unit, ideal for home use applications, or research purposes.

A. The radar-based monitoring device

The cornerstone of the HealthSonar system is the radar-based monitoring device, which was designed, developed and produced, in-house, by PD Neurotechnology Ltd. You can see a HealthSonar prototype in Figure 2. The case for the prototype was produced using 3D computer-aided design (CAD)

TABLE I: The specifications of the Aria Sensing LT102 radar module, the heart of the HealthSonar system.

| General specifications | Values |
|-----------------------------|--------------------|
| Radar's operating frequency | 6.5 GHz to 8.5 GHz |
| Temperature operating range | −40 °C to 85 °C |
| Radar module's dimensions | 36 mm×68 mm |
| Maximum power consumption | 220 mW at 5 V |
| Integrated antenna aperture | ±60° by ±60° |
| Typical detection range | 12 m |

and additive manufacturing techniques. The device comprises an Aria Sensing LT102 IR-UWB radar sensor, and a Raspberry Pi 4 Model B board. The radar module connects to the Raspberry Pi board via 1 external USB cable. The radar sensor identifies micro motions (such as the oscillations of the chest) and macro motions (such as the human gait), translating them into data, while the Raspberry Pi board is the main processing unit of the device. The specifications of the Aria Sensing radar module can be found in Table I. The HealthSonar device is powered through a wall socket instead of a battery, allowing it to operate indefinitely without requiring any user interaction for charging. The intended placement of the device is:

- 1) Next to a bed, on top of a nightstand facing the sleeping individual, enabling nighttime sleep monitoring.
- 2) Mounted on a bathroom wall, maximizing the coverage of the area, enabling human presence and fall detection.
- 3) Mounted on a tripod, with its line of sight facing a strip of the room, for general mobility evaluation.

B. The web dashboard application

The web dashboard (see Figure 3a) is a cloud-based web application accessible via the internet. It serves as the central hub for managing multiple (or, more aptly, at least one) connected HealthSonar devices. Tailored for healthcare professionals across care homes, hospitals, and other healthcare settings, the web dashboard offers a comprehensive suite of tools for patient data monitoring, device management, and overall system administration. Its functions include:

- 1) Manually initiating and terminating recording sessions.
- 2) Manually selecting data for upload to Cloud storage.
- 3) Accessing telemetry data (logs) for connected devices.
- 4) Accessing information (metadata) for each device.
- 5) Assigning the system's users to specific devices.
- 6) Viewing the connection status of each available device.
- 7) Managing the recorded data of all connected devices.
- 8) Viewing sleep/mobility reports and notifications.

C. The web portal to the radar

The web portal to the radar (see Figure 3b) is a web application hosted on the HealthSonar device itself, accessible solely via the local network the device is connected to. It facilitates direct communication with, and access to, a specific HealthSonar device, serving as the primary tool for configuring its settings. While the web portal shares a similar purpose with

the mobile application (see Section II-D), it uniquely offers configuration capabilities for the radar, which the mobile app does not.

Contrary to the web dashboard and mobile application, the web portal is designed for a more technical audience. Trained healthcare professionals or researchers in various healthcare and research environments, including care homes, hospitals, and medical research facilities, utilize the web portal as a tool for setting up the radar device, quickly scheduling and implementing sleep or mobility evaluation scenarios, gathering health and mobility data, as well as easily accessing stored data. Specifically, the functionality of the web portal is:

- 1) Connecting the device through Wi-Fi.
- 2) Configuring the settings of a device.
- 3) Initiating and terminating a recording.
- 4) Setting up a scheduler for a recording.
- 5) Accessing previously-stored radar data.

The web portal is intended for advanced, trained users who understand the consequences of directly accessing the HealthSonar device's settings and storage, particularly the effects of changing the internal parameters of the radar sensor. A clinical trial, or generally a similar research endeavor, would be the perfect use case for the web portal, enabling highly specialized configuration of the radar based on the needs of specific scenarios, as well as for accessing the generated data.

It should be noted that, under normal circumstances, access to the web portal should not be necessary (it would even be ill-advised), as the HealthSonar device is delivered to a user preconfigured with optimal settings. The full functionality of the system should be accessed through the web dashboard and mobile application. Throughout the design phase of the companion applications, an austere design language was purposely chosen for the web portal's user interface in order to discourage routine use. On the contrary, the web and mobile applications offer better usability through a "polished", "clean" and "friendly" user interface, encouraging their use.

D. The mobile application (mobile app)

The mobile application (see Figure 4) facilitates the initial setup of the system and provides reports based on data acquired by the HealthSonar device. It also enables elders/patients to evaluate their mobility and caregivers to receive notifications for fall incidents. Specifically, the mobile app was designed for use by elders/patients and their caregivers as a tool for:

- 1) Setting up and configuring the system before its first use.
- 2) Tracking sleep quality (e.g., sleep staging, duration).
- 3) Identifying sleep-related events (such as sleep apnea).
- 4) Evaluating mobility using gait assessment techniques.
- 5) Tracking gait quality, including fall risk indicators.
- 6) Receiving real-time notifications in the event of a fall.

Evaluating mobility is conducted using a Timed-Up and Go (TUG) test (see Section III-C), which is manually scheduled through the mobile application. Unlike feature-rich mobile apps accompanying commercial sleep and fitness trackers, the HealthSonar app prioritizes user-friendliness taking into account, first and foremost, the needs and experience of their



Fig. 2: A prototype of the HealthSonar device. On the left picture, the front side of the device can be seen, while on the right, the back side. The bottom part of the device houses the Raspberry Pi 4 and the top part contains the Aria Sensing LT102 radar module. Various connection ports are located on the top, back, and sides of the device. The radar connects to the Raspberry Pi board via an external USB cable.

users. The mobile application caters to users with two main limitations: unfamiliarity with technology, common among elders, and impaired dexterity experienced by patients with movement disorders like Parkinson's disease. Recognizing these limitations, the HealthSonar app is designed towards providing a user experience that is both concise and focused. This is achieved by intentionally designing it primarily as a reporting tool.

E. System data storage and processing

The radar data acquired during the operation of the HealthSonar system are stored and processed either locally on the monitoring device or on the Cloud after being uploaded. The decision depends on the time-critical nature of the resulting analytics for the health of the user. Local, on-device, storage and processing delivers near real-time outputs, crucial for situations where immediate action might be necessary. On the contrary, uploading to the Cloud, while valuable for long-term analysis, can introduce delays.

The fall detection and, to some degree, the mobility evaluation pipeline, leverage on-device storage and processing, producing their outputs as close to real-time as possible. Fall detection is a cornerstone of any assisted-living solution, but its effectiveness hinges on delivering real-time alerts immediately following the detection of a fall. Falls present a significant risk of serious injury, particularly for elderly individuals and patients with movement disorders. These populations are more vulnerable due to a combination of factors: increased fragility and the possibility of falls going unnoticed if they are unable to move and call for help afterwards.

Real-time fall notifications can be a matter of life and death, providing precious time for caregivers to attend to the needs

of the faller, potentially saving their life. Similarly, mobility evaluation (TUG test) benefits from producing results with minimal latency, mainly for two reasons. First, physicians conducting mobility evaluations require the resulting information as soon as the evaluation concludes, ideally while the patient is still present. This allows physicians to tailor treatment plans based on the patient's current condition and discuss them directly. Revisiting a patient's case at a later time would be impractical. Second, the results of a TUG test indicate a high risk of future falls. While this information is not strictly time-sensitive, early awareness allows for timely medical interventions and preventative measures to be taken, potentially avoiding serious future injuries. However, recognizing the potential value for research and custom analysis, the system allows users to upload the raw TUG test radar data to the Cloud for future use. Uploading is easily managed through the web dashboard application.

On the other hand, the sleep monitoring pipeline rely on Cloud storage and processing since time-critical feedback to the users is not required. Nighttime monitors typically work silently in the background, collecting data throughout your sleep. Upon waking, they automatically stop recording and process the data to generate a sleep report accessible through a mobile or web application. HealthSonar follows the same approach, automatically collecting sleep data throughout the night and concluding recording upon waking. It then generates a sleep report accessible via its mobile app and web dashboard.

Last but not least, presence detection acts as a prerequisite step for both sleep monitoring and fall detection, necessitating its real-time execution on the HealthSonar device. Unlike sleep monitoring and mobility evaluation, presence detection does not generate data for long-term analysis. Instead, it acts as a

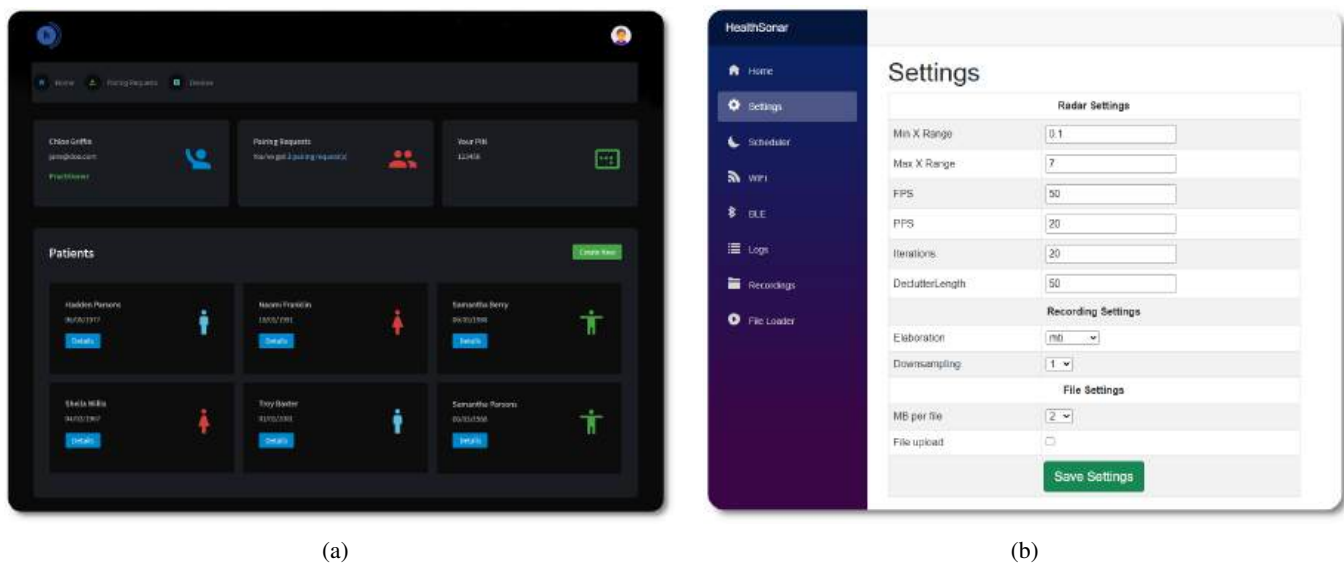


Fig. 3: (3a) A screenshot depicting the web dashboard's homepage layout, optimized for user-friendliness and clear presentation of health monitoring data. Designed to meet the demands of healthcare professionals in diverse settings like care homes and hospitals, the web dashboard provides a powerful toolkit for patient data monitoring, device management, and overall system administration. (3b) A screenshot depicting the Settings page of the radar's web portal. This page caters to trained users with advanced technical knowledge, allowing them to configure specialized radar parameters. These configurations are primarily intended for researchers conducting specialized studies. To ensure optimal performance out of the box, the HealthSonar device is preconfigured for typical use, eliminating the need to access the web portal under normal circumstances. To discourage routine use, a utilitarian, bare-bones user interface was chosen.

trigger, signaling the system to launch the appropriate pipeline (sleep or fall) when a user's presence is detected.

F. Application Programming Interface (API)

The communication between the Cloud services and the "local" parts of the HealthSonar system is enabled by an application programming interface (API). The API acts as the middleman for data access and integration between databases, web services and the device itself as well as the mobile application. More specifically, the API provides functionality enabling secure and stable communication between the HealthSonar device and the Cloud, safeguarding sensitive user data and guaranteeing data integrity. The functionality offered by the API serves a pivotal role in the use of the HealthSonar system and enables building, maintaining and scaling its overall ecosystem. The API opens the door for future innovation. It allows third-party developers to create applications that connect with the HealthSonar system. This not only empowers users with a wider range of tools to manage their health data, but also fosters a vibrant ecosystem of applications that leverage HealthSonar's features for personalized health insights and experiences.

G. HealthSonar: Customizable, Scalable

The HealthSonar system offers high customizability and scalability. This flexibility allows it to function effectively as a single unit or as a network of radar-based devices. Notably, the term "network" refers to a collection of HealthSonar devices managed by a central service, not a cluster of interconnected

devices themselves. Each device's radar parameters can be precisely adjusted to suit the specific application. Centralized management of all active devices within an organization is conveniently facilitated through the web dashboard application.

The HealthSonar system provides extensive customization options, allowing users to tailor its functionality to their specific needs. Users can choose to utilize at least one radar-based device along with the accompanying software components (including the web portal, web dashboard, and mobile app) based on their requirements. For research purposes, the web portal offers comprehensive functionality and grants elevated access to the embedded radar parameters. However, administrators should possess knowledge regarding the theory and operation of IR-UWB radars.

Furthermore, the HealthSonar ecosystem as a whole boasts extensibility and can seamlessly integrate with one or multiple radar-based devices as needed. This adaptability is particularly crucial for care homes and clinical environments, where requirements evolve based on the number of residents or inpatients. By effortlessly expanding, or reducing, the connected network of HealthSonar devices, organizations can efficiently manage their human resources according to demand. Routine monitoring tasks can be swiftly automated by the HealthSonar system, freeing up personnel to focus on other critical responsibilities.

III. SLEEP, MOBILITY, PRESENCE AND FALLS

The HealthSonar system leverages the IR-UWB radar module to capture rich micro and macro body movements. These

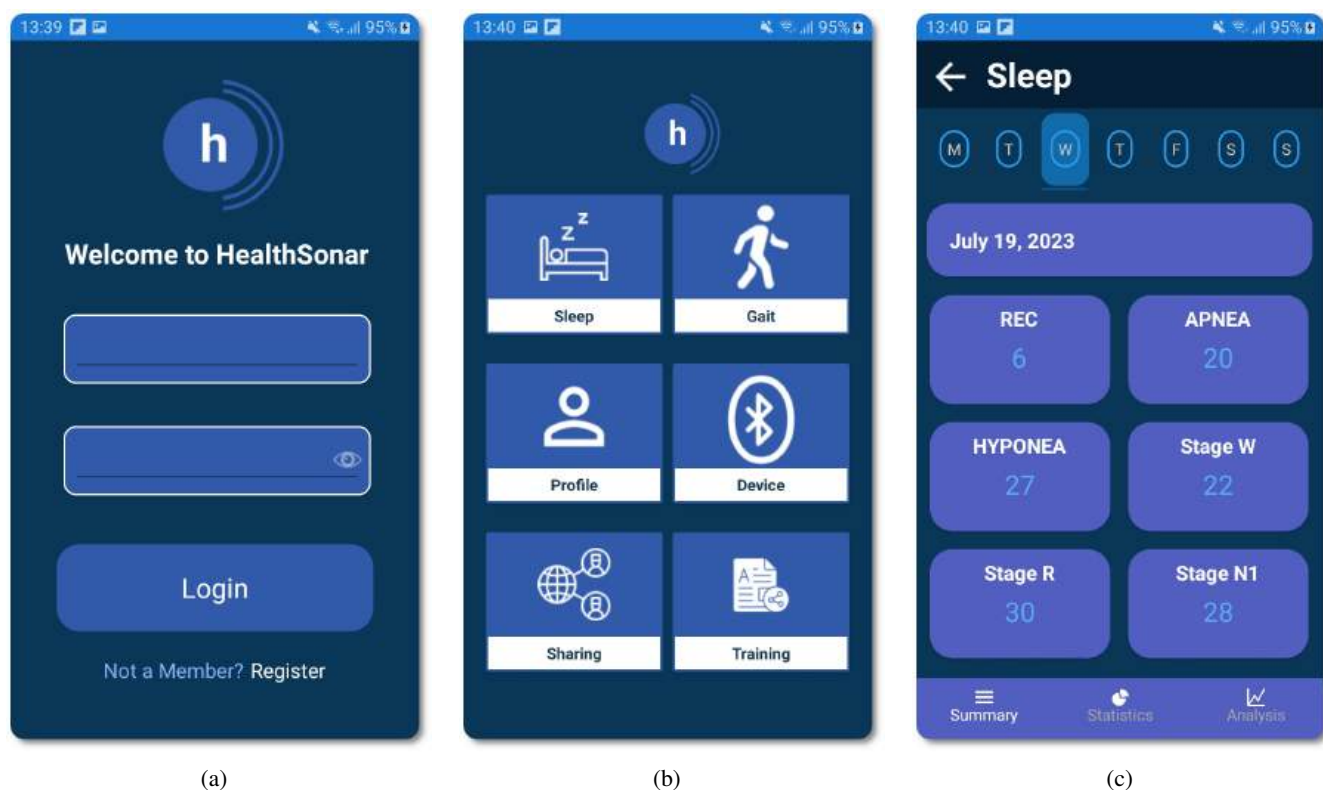


Fig. 4: Various screens of the HealthSonar mobile application. (4a) The login screen. (4b) The main menu. (4c) The sleep analytics. The app provides a user-friendly interface to the HealthSonar system, assisting users in configuring the device before its first use, as well as providing sleep metrics, sleep-related events, mobility metrics and real-time fall notifications.

captured movements provide valuable data for health and well-being monitoring, offering insights into mobility, posture, and potential sleep disturbances. HealthSonar offers the following functionalities:

- 1) Human presence detection within a space.
- 2) Nighttime sleep monitoring (e.g., quality, duration).
- 3) Sleep-related events identification (e.g., sleep apnea).
- 4) Mobility evaluation (e.g., gait, balance, fall risk).
- 5) Fall detection in high-risk environments.
- 6) Real-time fall notifications enabling timely action.

A comprehensive description regarding the features of the HealthSonar system (i.e., presence detection, sleep monitoring, mobility evaluation and fall detection), as well as the necessary setup, including suggestions, for its proper use per application can be found in Sections III-A, III-B, III-C and III-D.

The performance of the HealthSonar system was evaluated in the sleep lab of Evangelismos General Hospital (with patients suffering from sleep disorders), in the neurology clinic of the University Hospital of Ioannina (with patients suffering from), as well as in-house (with healthy individuals), through experimental setups similar to the ones described in the aforementioned sections (those are depicted in Figure 5). The results of the performance evaluation of the HealthSonar system have been published in peer-reviewed journals and conference proceedings [26], [28]–[31].

A. Presence detection

Presence detection serves as the essential foundation for both sleep monitoring and fall detection within the HealthSonar system. This crucial process runs continuously and locally on the device itself, acting as a prerequisite before activating either the sleep or fall detection pipelines.

By detecting human presence, HealthSonar can determine whether someone is lying in bed for sleep. If so, it automatically initiates the sleep monitoring pipeline to capture sleep data. Similarly, presence detection allows the bathroom-mounted HealthSonar to identify when someone is using the bathroom, triggering the fall detection pipeline as a precautionary measure.

B. Sleep monitoring

Sleep monitoring is a core function of HealthSonar. When the presence detection pipeline identifies someone lying in bed, it automatically initiates sleep monitoring. This continuous process gathers detailed data throughout the sleep period. Once the user leaves the bed, the monitoring stops, and the locally stored data is uploaded to the Cloud for in-depth analysis.

Unlike fall detection or mobility evaluation, sleep data processing leverages the Cloud infrastructure of the HealthSonar system. This is because real-time sleep metrics aren't crucial during the night. Instead, HealthSonar extracts valuable sleep insights upon data upload, including sleep stage classification (identifying wakefulness and sleep periods) and sleep apnea

event detection. Users can then access these metrics through the web dashboard or mobile application.

For accurate sleep staging and apnea detection, HealthSonar calculates respiratory rate and heartbeat during sleep monitoring. These metrics serve as essential preprocessing steps for an advanced and thorough sleep analysis (such as, sleep staging and sleep apnea detection) conducted in the Cloud.

C. Mobility evaluation

The HealthSonar system evaluates the mobility of an individual based on the Timed Up and Go Test (TUG) [32], a well-established, standardized test, used for assessing various aspects of one's mobility, such as gait, balance and risk of falling [33], [34]. Performing a TUG test requires minimal instrumentation, namely

- 1) an armchair (the arms are crucial for the test),
- 2) a reference marking the 3 m walking distance and
- 3) a stopwatch, (with variations being the iTUG tests [33]).

A participant begins the test seated in the armchair. Following instructions, they stand up unaided, walk for 3 meters, turn around 180 degrees at the reference marking the turning point, walk back to the chair, and sit down to complete the test. The performance of the participant is timed with the stopwatch, with the total time serving as the final score of the test.

HealthSonar utilizes a modified version of the Timed Up and Go (TUG) test to assess gait. The key difference lies in the walking path length. Instead of the standard 3 meters, HealthSonar utilizes a variable walking distance ranging from 3 to 5 meters. This extended path allows for the collection of a richer set of gait data, such as the total walking duration, the turning duration and the average gait speed [26], resulting in a more comprehensive representation of the participant's mobility. Note that due to the use of the radar-based device, this modified TUG test is considered instrumented (iTUG).

Mobility evaluation in HealthSonar, unlike presence detection and sleep monitoring, is user-initiated through the mobile app (or the web portal to the radar for advanced or custom use cases). The instrumented Timed Up and Go (iTUG) test, is typically infrequent and periodic, often conducted following a physician's recommendation. This test requires a controlled setting, specific procedures, and the HealthSonar device itself as the instrumentation properly setup for accurate results. Following completion of the TUG test, the collected raw gait data is stored and processed locally on the HealthSonar device. The extracted mobility evaluation metrics can then be presented to users through the mobile app or web dashboard. The web dashboard provides an additional option for users to manually upload the generated iTUG test data to Cloud storage for potential future use or analysis. This can be beneficial for identifying long-term trends or sharing data with healthcare professionals.

A typical setup for a HealthSonar iTUG test scenario can be seen in Figure 5. The placement of the HealthSonar device should allow for the effective monitoring of the path in front of the armchair during the testing process.

D. Fall detection

HealthSonar's fall detection functionality builds upon the presence detection capabilities of the HealthSonar system, ideally installed on a bathroom wall. This targeted approach focuses on the bathroom, a high-risk area for falls among elderly and patient populations [35], [36]. Due to the critical need for real-time monitoring in high-risk areas, fall detection runs continuously on the HealthSonar device itself. This local processing ensures timely response in case of a fall event.

Upon detecting a fall, the system automatically sends notifications to both the web dashboard and mobile app, alerting caregivers, physicians or other relevant healthcare professionals and enabling them to swiftly address the fallen individual's needs. Fall detection is a crucial feature in assisted living environments for elders and patients. Falls can lead to serious or even fatal injuries. Real-time identification allows for prompt intervention, potentially preventing further harm to the fallen individual.

IV. THE CARE HOME APPLICATION

The HealthSonar system's diverse and rich feature set allow it to be used by various user groups for a plethora of use cases. However, it is particularly well-suited for clinical environments and assisted-living scenarios as they take advantage of all its features and functionality. One of those scenarios, is a care home, housing elders or patients with movement disorders. The specific needs of a care home are similar to those of any clinical environment, such as a hospital or a clinic. Consequently, the presented application can be seamlessly implemented in any clinical environment with minimal modifications. For the purposes of this work, we consider a care home, with several rooms, each one resembling the layout depicted in Figure 5. The specific arrangement of each room is not crucial, while the only requirements are usually already met by a typical room (a nightstand and an armchair).

A. Device installation

For full coverage of a room, providing comprehensive monitoring, two HealthSonar devices are recommended. The first device is placed on a nightstand, primarily for sleep monitoring. When an iTUG test needs to be conducted, this same device can be conveniently repositioned to effectively monitor the subject's walking path (covering a 3-to-5 meter straight line walking distance). The second device is mounted on a bathroom wall, strategically positioned for optimal coverage of the area for fall detection. The targeted placement of the device prioritizes the bathroom since it is regarded as the most hazardous space for elders and patients in the context of falls associated with an elevated likelihood of resulted serious injuries [35], [36]. Refer to Figure 5 for a visualization of the suggested device placements for various functionalities: sleep monitoring, mobility evaluation (iTUG test), and fall detection.

B. Monitoring procedure

After installing the HealthSonar device, healthcare professionals can set it up for use through the mobile app with

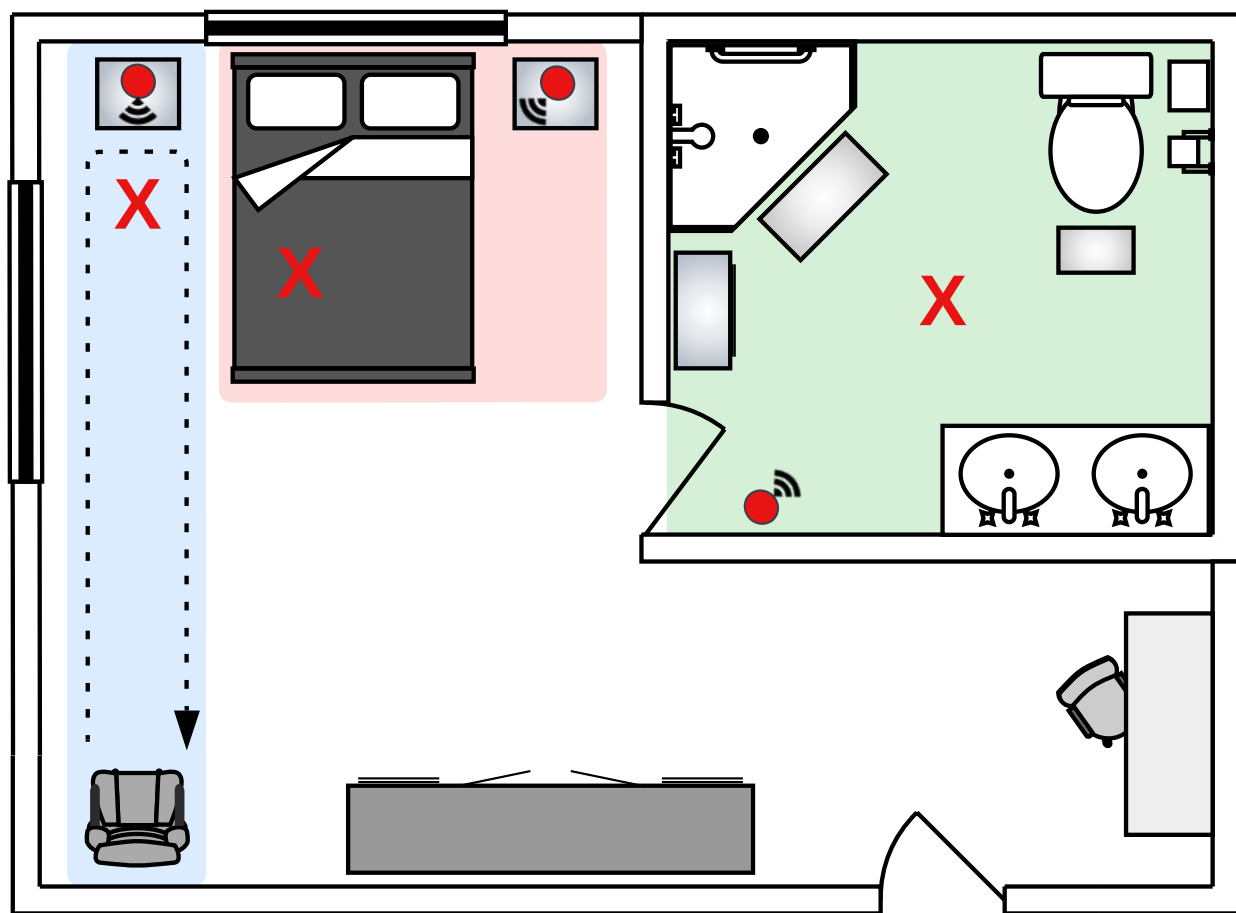


Fig. 5: Three illustrative scenarios showcasing the HealthSonar system's application within a typical bedroom of a care home or a similar clinical environment. **Green area:** A falling detection scenario. **Red area:** A sleep monitoring scenario. **Blue area:** A suggested Timed Up and Go (TUG) test scenario.

minimal effort. This involves registering an account for the resident and linking it to the device. Once configured, the device is ready for use.

Residents themselves don't need to interact with the device physically, as it's operational is contactless and unobtrusive. However, they can optionally access the HealthSonar system through a mobile app to view their sleep and mobility reports.

Sleep monitoring and fall detection run continuously in the background. Presence detection allows the system to automatically initiate these monitoring sessions. Manual session initiation is also available if needed.

C. Managing the system

The HealthSonar system is designed to be straightforward and user-friendly for busy care home personnel. It integrates seamlessly into their existing routines, requiring minimal additional time or effort.

Healthcare professionals leverage the web dashboard for comprehensive monitoring and resident data management. The dashboard provides functions for:

- 1) **Device Overview:** Gaining a quick view of all installed HealthSonar devices, including activity status, data upload status, and user registration.
- 2) **Device Management:** Assigning devices to residents, handling recorded data, accessing system information, and manually setting up recording sessions.
- 3) **Resident Analytics:** Reviewing detailed reports on resident sleep patterns (sleep stages, sleep apnea events) and gait metrics (total walking duration, turning duration, average gait speed). Leverage historical data to personalize care plans by evaluating residents' condition progression.
- 4) **Fall Notifications:** Receive real-time notifications regarding fall incidents taking place in residents' bathrooms, allowing for prompt intervention and injury prevention.

V. DISCUSSION

The rising number of elders and patients with sleep disorders (like insomnia and restless leg syndrome), neurological conditions (like Parkinson's disease), and movement disorders (like gait abnormalities) highlights the need for improved monitoring solutions. These groups face similar challenges, primarily related to sleep quality and mobility. Continuous, unobtrusive, and privacy-focused monitoring can offer valuable insights into sleep quality (e.g., sleep duration, fragmentation) and mobility state (e.g., gait patterns, fall risk assessment),

potentially leading to earlier interventions and improved health outcomes.

This work presents HealthSonar, a complete and unified assisted living solution designed and built from the ground up. It goes beyond a proof-of-concept, functioning as a comprehensive, multifunctional system for sleep monitoring, mobility evaluation, and fall detection. HealthSonar ensures streamlined operation and reliable data collection through its fully integrated elements, unlike systems relying on disparate hardware and software components. HealthSonar is comprised of

- 1) A prototype device, fabricated using additive manufacturing techniques (see Section II-A).
- 2) A suite of custom-developed software applications (see Sections II-B, II-C and II-D).
- 3) A set of novel algorithms implementing the system's functionality (see Section III).

The system is designed to be operated with minimal to no interaction from the user. Following installation and initial setup (see Section IV-A for device installation details), the radar-based device seamlessly runs in the background, unobtrusively monitoring sleep and detecting falls. Users can entirely bypass the software components and remain unaware of the system's operation. In such cases, healthcare professionals can access sleep reports and receive fall notifications. HealthSonar is entirely maintenance-free. Eliminating the need for batteries, the device can be permanently plugged into a wall socket, functioning continuously atop a nightstand or mounted on a wall. The system has been tested to function as intended under realistic conditions (see Section III).

Highlighting the versatility of the HealthSonar system, this work demonstrated its effectiveness in a care home setting. This adaptable scenario readily integrates into diverse clinical settings, requiring minimal adjustments to deliver comprehensive health/wellness monitoring.

Our presented approach for sleep monitoring, mobility monitoring and fall detection within a care home offers a comprehensive solution for clinical environments. However, the HealthSonar system's design allows for further customizations to address evolving healthcare requirements and potentially even entirely new applications. As an example, our use case prioritizes continuous fall detection in bathrooms, recognizing them as high-risk zones for falls in elderly and patient populations. This targeted approach maximizes effectiveness while minimizing intrusiveness in residents' living spaces. If necessary, continuous fall detection can be extended to bedrooms or other areas based on individual needs and risk factors.

As of now, the functionality of the HealthSonar system is significant, yet future development could extend it in various directions.

- 1) **Advanced Mobility Evaluation:** Enhance the mobility evaluation to identify freezing of gait events in Parkinson's patients. This could significantly improve patient care and early intervention strategies.
- 2) **Bedroom Fall Detection:** Based on user feedback, explore implementing continuous fall detection within

the bedroom area to provide a more comprehensive monitoring solution.

- 3) **Refined Heartbeat Extraction:** Incrementally improve the heartbeat extraction algorithm, a crucial preprocessing step for sleep monitoring, to ensure the accuracy of sleep data analysis.

While the software for HealthSonar (web/mobile app, radar portal) is considered feature-rich and fully-realized, delivered as a robust wellness monitoring application suite, the current prototype prioritizes functionality over design aesthetics. However, future iterations can explore a more refined design, potentially incorporating elements such as a user-friendly screen. This display could show time, weather, or news headlines, transforming the nightstand device into a versatile dashboard with additional functionalities.

VI. CONCLUSIONS

The increasing number of elders and patients with sleep and neurological disorders necessitates improved health monitoring solutions, particularly for sleep and mobility. This work presented HealthSonar, a novel contactless and unobtrusive system designed to enhance the well-being of elderly individuals and patients with sleep and movement disorders in assisted living settings. HealthSonar prioritizes user privacy while providing continuous sleep and mobility monitoring, as explored through a care home application scenario.

The presented care home scenario exemplifies HealthSonar's centralized management, enabling healthcare professionals to simultaneously monitor the health status and analytics of multiple residents. This seamless integration significantly improves workflow efficiency. For residents, HealthSonar provides a hassle-free experience with completely unobtrusive monitoring, eliminating the need for physical or digital interaction.

Continuous, privacy-preserving monitoring solutions like HealthSonar have the potential to revolutionize the way clinical organizations (care homes, clinics, hospitals, etc.) monitor their residents. By providing a continuous stream of actionable health data and analytics, HealthSonar empowers these organizations to optimize care delivery while alleviating the burden of repetitive and time-consuming tasks like continuous health monitoring.

ACKNOWLEDGMENT

This research was funded by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship, and Innovation, under the call RESEARCH-CREATE-INNOVATE (project name: HealthSonar, project code: T2EDK-04366)

REFERENCES

- [1] A. Ntani, S. Kontaxis, G. Rigas, A. Pentari, K. Tsiouris, E. Kontogiannis, E. Kostoulas, I. Tsimperis, T. Vlioras, A. Bousis, S. Zelilidou, K. Tasiou, M. Tsiknakis, and D. Fotiadis, "Healthsonar: A system for unobtrusive monitoring of elders and patients with movement disorders," in *Proceedings of the HEALTHINFO 2023, The Eighth International Conference on Informatics and Assistive Technologies*

- for Health-Care, Medical Support and Wellbeing. IARIA, 2023, pp. 60–65. [Online]. Available: https://www.thinkmind.org/index.php?view=article&articleid=healthinfo_2023_1_100_80044
- [2] D. T. Kasai, "Preparing for population ageing in the Western Pacific Region," *The Lancet Regional Health - Western Pacific*, vol. 6, p. 100069, Jan. 2021. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2666606520300699>
 - [3] J. Acquavella, R. Mehra, M. Bron, J. M.-H. Suomi, and G. P. Hess, "Prevalence of narcolepsy and other sleep disorders and frequency of diagnostic tests from 2013–2016 in insured patients actively seeking care," *Journal of Clinical Sleep Medicine*, vol. 16, no. 8, pp. 1255–1263, Aug. 2020. [Online]. Available: <http://jcsn.aasm.org/doi/10.5664/jcsn.8482>
 - [4] N. P. Gordon, J. H. Yao, L. A. Brickner, and J. C. Lo, "Prevalence of sleep-related problems and risks in a community-dwelling older adult population: a cross-sectional survey-based study," *BMC Public Health*, vol. 22, no. 1, p. 2045, Nov. 2022. [Online]. Available: <https://bmcpublishing.biomedcentral.com/articles/10.1186/s12889-022-14443-8>
 - [5] C. Ding, Y. Wu, X. Chen, Y. Chen, Z. Wu, Z. Lin, D. Kang, W. Fang, and F. Chen, "Global, regional, and national burden and attributable risk factors of neurological disorders: The Global Burden of Disease study 1990–2019," *Frontiers in Public Health*, vol. 10, p. 952161, Nov. 2022. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fpubh.2022.952161/full>
 - [6] C. S. Florence, G. Bergen, A. Atherly, E. Burns, J. Stevens, and C. Drake, "Medical costs of fatal and nonfatal falls in older adults," *Journal of the American Geriatrics Society*, vol. 66, no. 4, pp. 693–698, 2018.
 - [7] J. V. Rundo and R. Downey III, "Polysomnography," *Handbook of clinical neurology*, vol. 160, pp. 381–392, 2019.
 - [8] D. Fekedulegn, M. E. Andrew, M. Shi, J. M. Violanti, S. Knox, and K. E. Innes, "Actigraphy-Based Assessment of Sleep Parameters," *Annals of Work Exposures and Health*, vol. 64, no. 4, pp. 350–367, Apr. 2020. [Online]. Available: <https://academic.oup.com/annweh/article/64/4/350/5735350>
 - [9] N. Molinaro, E. Schena, S. Silvestri, F. Bonotti, D. Aguzzi, E. Viola, F. Buccolini, and C. Massaroni, "Contactless Vital Signs Monitoring From Videos Recorded With Digital Cameras: An Overview," *Frontiers in Physiology*, vol. 13, p. 801709, Feb. 2022. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fphys.2022.801709/full>
 - [10] V. Selvaraju, N. Spicher, J. Wang, N. Ganapathy, J. M. Warnecke, S. Leonhardt, R. Swaminathan, and T. M. Deserno, "Continuous Monitoring of Vital Signs Using Cameras: A Systematic Review," *Sensors*, vol. 22, no. 11, p. 4097, May 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/11/4097>
 - [11] G. Tiberi and M. Ghavami, "Ultra-Wideband (UWB) Systems in Biomedical Sensing," *Sensors*, vol. 22, no. 12, p. 4403, Jun. 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/12/4403>
 - [12] D. K.-H. Lai, Z.-H. Yu, T. Y.-N. Leung, H.-J. Lim, A. Y.-C. Tam, B. P.-H. So, Y.-J. Mao, D. S. K. Cheung, D. W.-C. Wong, and J. C.-W. Cheung, "Vision Transformers (ViT) for Blanket-Penetrating Sleep Posture Recognition Using a Triple Ultra-Wideband (UWB) Radar System," *Sensors*, vol. 23, no. 5, p. 2475, Feb. 2023. [Online]. Available: <https://www.mdpi.com/1424-8220/23/5/2475>
 - [13] M. Husaini, L. M. Kamarudin, A. Zakaria, I. K. Kamarudin, M. A. Ibrahim, H. Nishizaki, M. Toyoura, and X. Mao, "Non-Contact Breathing Monitoring Using Sleep Breathing Detection Algorithm (SBDA) Based on UWB Radar Sensors," *Sensors*, vol. 22, no. 14, p. 5249, Jul. 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/14/5249>
 - [14] Y. Lee, J.-Y. Park, Y.-W. Choi, H.-K. Park, S.-H. Cho, S. H. Cho, and Y.-H. Lim, "A Novel Non-contact Heart Rate Monitor Using Impulse-Radio Ultra-Wideband (IR-UWB) Radar Technology," *Scientific Reports*, vol. 8, no. 1, p. 13053, Aug. 2018. [Online]. Available: <https://www.nature.com/articles/s41598-018-31411-8>
 - [15] W. H. Lee, S. H. Kim, J. Y. Na, Y.-H. Lim, S. H. Cho, S. H. Cho, and H.-K. Park, "Non-contact Sleep/Wake Monitoring Using Impulse-Radio Ultrawideband Radar in Neonates," *Frontiers in Pediatrics*, vol. 9, p. 782623, Dec. 2021. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fped.2021.782623/full>
 - [16] H. B. Kwon, S. H. Choi, D. Lee, D. Son, H. Yoon, M. H. Lee, Y. J. Lee, and K. S. Park, "Attention-Based LSTM for Non-Contact Sleep Stage Classification Using IR-UWB Radar," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3844–3853, Oct. 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9403900/>
 - [17] R. de Goederen, S. Pu, M. Silos Viu, D. Doan, S. Overeem, W. Serdijn, K. Joosten, X. Long, and J. Dudink, "Radar-based sleep stage classification in children undergoing polysomnography: a pilot-study," *Sleep Medicine*, vol. 82, pp. 1–8, Jun. 2021. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1389945721001933>
 - [18] Y. Zhou, D. Shu, H. Xu, Y. Qiu, P. Zhou, W. Ruan, G. Qin, J. Jin, H. Zhu, K. Ying, W. Zhang, and E. Chen, "Validation of novel automatic ultra-wideband radar for sleep apnea detection," *Journal of Thoracic Disease*, vol. 12, no. 4, pp. 1286–1295, Apr. 2020. [Online]. Available: <http://jtd.amegroups.com/article/view/38166/html>
 - [19] Z. Wei, J. Xu, W. Li, X. Wang, Z. Qin, J. Zhou, and W. Wang, "Evaluation of a non-contact ultra-wideband bio-radar sleep monitoring device for screening of sleep breathing disease," *Sleep and Breathing*, vol. 26, no. 2, pp. 689–696, Jun. 2022. [Online]. Available: <https://link.springer.com/10.1007/s11325-021-02424-x>
 - [20] A. Kabir, S. Ifteqar, and A. Bhat, "Obstructive Sleep Apnea in Adults," *Hospital Practice*, vol. 41, no. 4, pp. 57–65, Oct. 2013. [Online]. Available: <http://www.tandfonline.com/doi/full/10.3810/hp.2013.10.1081>
 - [21] S. Toften, S. Pallesen, M. Hrozanova, F. Moen, and J. Grønli, "Validation of sleep stage classification using non-contact radar technology and machine learning (Somnofy®)," *Sleep Medicine*, vol. 75, pp. 54–61, Nov. 2020. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1389945720301027>
 - [22] H. Li, A. Mehul, J. Le Kerneec, S. Z. Gurbuz, and F. Fioranelli, "Sequential Human Gait Classification With Distributed Radar Sensor Fusion," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 7590–7603, Mar. 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9306810/>
 - [23] J. Maitre, K. Bouchard, and S. Gaboury, "Fall Detection With UWB Radars and CNN-LSTM Architecture," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 1273–1283, Apr. 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9212552/>
 - [24] J. C.-W. Cheung, E. W.-C. Tam, A. H.-Y. Mak, T. T.-C. Chan, W. P.-Y. Lai, and Y.-P. Zheng, "Night-Time Monitoring System (eNightLog) for Elderly Wandering Behavior," *Sensors*, vol. 21, no. 3, p. 704, Jan. 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/3/704>
 - [25] J. C. Ayena, L. Chioukh, M. J.-D. Otis, and D. Deslandes, "Risk of Falling in a Timed Up and Go Test Using an UWB Radar and an Instrumented Insole," *Sensors*, vol. 21, no. 3, p. 722, Jan. 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/3/722>
 - [26] A. Ntanas, N. Kostikis, I. Tsimperis, K. Tsiouris, G. Rigas, and D. Fotiadis, "Evaluating Parameters of the TUG Test Based on Data from IMU and UWB Sensors," in *2022 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. Thessaloniki, Greece: IEEE, Oct. 2022, pp. 142–147. [Online]. Available: <https://ieeexplore.ieee.org/document/9941700/>
 - [27] A.-K. Seifert, M. G. Amin, and A. M. Zoubir, "Toward Unobtrusive In-Home Gait Analysis Based on Radar Micro-Doppler Signatures," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 9, pp. 2629–2640, Sep. 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8613848/>
 - [28] S. Kontaxis, A. Ntanas, I. Tsimperis, E. Kostoulas, C. Tsitsou, T. V. Gortsas, E. Dermatas, D. Polyzos, and G. Rigas, "On the detection of indoor falls using an ultra-wideband radar," in *2023 IEEE Conference on Antenna Measurements and Applications (CAMA)*. IEEE, 2023, pp. 915–919.
 - [29] K. M. Tsiouris, G. Rigas, S. Zelilidou, E. Florou, F. S. Kanellos, E. Kosmas, I. Tsimperis, E. Vagiakis, and D. I. Fotiadis, "Multi-cohort evaluation of an automated sleep stage detection methodology using ecg and respiration signals," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2023, pp. 4154–4158.
 - [30] A. Pentari, G. Rigas, A. Ntanas, T. Kassiotis, D. Manousos, E. Florou, E. Vagiakis, D. Fotiadis, and M. Tsiknakis, "Can the ir-uwb radar sensor substitute the psg-based primary vital signs?" in *Proceedings of the HEALTHINFO 2023, The Eighth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing*. Valencia, Spain: IARIA, 2023, pp. 66–71. [Online]. Available: https://www.thinkmind.org/library/HEALTHINFO/HEALTHINFO_2023/healthinfo_2023_1_110_80057.html
 - [31] A. Pentari, D. Manousos, T. Kassiotis, G. Rigas, and M. Tsiknakis, "Respiration and heartbeat rates estimation using ir-uwb non-contact radar sensor recordings: A pre-clinical study," in *Workshop Proceedings of the EDBT/ICDT 2023 Joint Conference*, 2023. [Online]. Available: https://eur-ws.org/Vol-3379/HeDAI_2023_paper404.pdf
 - [32] D. Podsiadlo and S. Richardson, "The Timed 'Up & Go': A Test of Basic Functional Mobility for Frail Elderly Persons," *Journal of*

- the American Geriatrics Society*, vol. 39, no. 2, pp. 142–148, Feb. 1991. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1111/j.1532-5415.1991.tb01616.x>
- [33] P. Ortega-Bastidas, B. Gómez, P. Aqueveque, S. Luarte-Martínez, and R. Cano-de-la Cuerda, “Instrumented Timed Up and Go Test (iTUG)—More Than Assessing Time to Predict Falls: A Systematic Review,” *Sensors*, vol. 23, no. 7, p. 3426, Mar. 2023. [Online]. Available: <https://www.mdpi.com/1424-8220/23/7/3426>
- [34] B. M. Kear, T. P. Guck, and A. L. McGaha, “Timed Up and Go (TUG) Test: Normative Reference Values for Ages 20 to 59 Years and Relationships With Physical and Mental Health Risk Factors,” *Journal of Primary Care & Community Health*, vol. 8, no. 1, pp. 9–13, Jan. 2017. [Online]. Available: <http://journals.sagepub.com/doi/10.1177/2150131916659282>
- [35] R. Blanchet and N. Edwards, “A need to improve the assessment of environmental hazards for falls on stairs and in bathrooms: results of a scoping review,” *BMC Geriatrics*, vol. 18, no. 1, p. 272, Dec. 2018. [Online]. Available: <https://bmcgeriatr.biomedcentral.com/articles/10.1186/s12877-018-0958-1>
- [36] J. A. Stevens, E. N. Haas, and T. Haileyesus, “Nonfatal bathroom injuries among persons aged ≥ 15 years—United States, 2008,” *Journal of Safety Research*, vol. 42, no. 4, pp. 311–315, Aug. 2011. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0022437511000582>

Approaches to Improving Medication Adherence Prediction in Chronic Disease Patients

Ben Malin

Dept. Electronic and Electrical Engineering
Brunel University London
London, United Kingdom
e-mail: ben.malin@brunel.ac.uk

Ejike Nwokoro

Patient Insights and Data Strategy Unit
HealthNet Homecare
London, United Kingdom
email: ejike.nwokoro@healthnethomecare.co.uk

Tatiana Kalganova

Dept. Electronic and Electrical Engineering
Brunel University London
London, United Kingdom
email: tatiana.kalganova@brunel.ac.uk

Joshua Hinton

Patient Insights and Data Strategy Unit
HealthNet Homecare
London, United Kingdom
email: joshua.hinton@healthnethomecare.co.uk

Abstract— This study aims to identify the impact of a patient's treatment/ support service duration (LOS) on the ability of a machine learning model to predict their medication adherence. The insight generated from this study can support the adaptation of patient support interventions, based on the evolution of predicted adherence at different treatment or service durations. For adherence prediction, we use medication delivery data, driven by the patient's prescription, to calculate a patient's stock level at any given time during their participation in a homecare support service, whilst allowing for medication stockpiling. This data is visualized and inputted into a Convolutional Neural Network (CNN). To evaluate adherence for a range of LOS values, every patient's first x months on service are extracted, with the final month used as the target variable. To define nonadherence, we use Proportion of Days Covered (PDC) of 100% for this period, where if a patient does not have any medication during this month they will be classed as nonadherent. Using this approach, we found that as LOS changes, there is a variation in both the proportion of the population that are adherent as well as the prediction model's performance. Across the studied timeframe of 4-12 months LOS, proportion of the study population that are adherent varies between 54.6% and 66.1%, with an Area Under the Curve (AUC) varying from 88.1% to 98.6% for our best-performing model. We also found that additional variables linked with adherence such as: communications with the service provider, demographic data, socioeconomic data and diagnosis-specific average PDC, improve model performance. The model in this study achieves its highest AUC and adherence prediction accuracy of 98.6% and 92.8% respectively, at an LOS of 9 months. Additional evaluation was performed to identify variation across therapies offered through a Homecare service. The results from this evaluation show diversity across both adherence to these therapies as well as the accuracy to be expected from the adherence predictions. We conclude that this diversity is linked to medication delivery/prescription frequency, volume of medication stock prescribed as well as therapy-specific diagnosis differences.

Keywords- medication adherence; CNN; healthcare; homecare; adherence prediction; length of service

I. INTRODUCTION

This study builds upon our previous research on adherence prediction, with aims of optimizing the network, whilst providing further analysis on the impact service duration has on predictive performance and adherence [1]. Tailored interventions that are deployed proactively or at treatment initiation have been championed by many as an impactful approach to tackling poor medication adherence [2][3]. The ability to identify/predict patients who are likely to become nonadherent at the beginning of their treatment journey and intervening early, when they are relatively more receptive to targeted interventions, should increase the chances of preventing the deterioration of poor medication adherence behavioral patterns later. Importantly, research has shown that the dynamic prediction of nonadherence risk can allow for the efficient deployment of interventions that are known to be effective in improving adherence [2][3].

Reducing nonadherence is directly linked to more favorable health outcomes as well as reduced financial burden [4][5]. Interventions influence adherent behavior differently, based upon when the intervention takes place [6]. Putting a priority on early identification of poor adherence, in addition to adaptation of these interventions can lead to greater overall adherence and more favorable health outcomes.

This also supports a more proactive and sustainable approach to healthcare delivery. To this end, and with a view to expanding on the findings of our previous study [1], the objective of this current study was to explore the extent to which several methodological and design approaches impact on the performance of a Machine Learning (ML) model in predicting poor medication adherence in patients who are newly enrolled in a Homecare Patient Support Program. The aforementioned methodological and design approaches include: (1) exploring the dynamics of patient adherence across differing treatment duration timeframes (2) increased granularity in the novel visualization of medication delivery data (3) Patient data selection that is based on the initial period on treatment (4) incorporating additional data to map

patient characteristics, including demographic and socioeconomic data (5) implementation of therapy-specific models.

Furthermore, we examined to what extent the area under the characteristic curve (AUC) and prediction accuracy can change as more data accrues and patients spend more time in a Homecare Patient Support Program. This will allow for more rigorous evaluation of the performance level that can be expected across patients who have been on the service for varying lengths of time. This is crucial as the dynamic interaction between the many factors that drive poor medication adherence can change over time, even with the same medication and the same patient [6]-[9]. Therefore, the continuous monitoring of the changing risk of poor medication adherence in a patient is key in implementing proactive interventions that are designed to tackle negative adherence behaviors. To our knowledge, this has not been studied in detail in existing literature - although some studies have demonstrated connections between treatment duration and adherence prediction, these studies have had a somewhat limited scope [1][8][10][11].

This paper is an extension to our previously conducted study on adherence prediction [1] and is structured into the following sections. In the “Literature review” section, existing literature is reviewed for variables that have benefited adherence prediction as well as the treatment duration timeframe applied for each study. “Data and preprocessing” discusses design decisions as well as the details of our study dataset. The “Methodology” section will explain our implementation of various strategies used to achieve the study objectives. In “Results”, we will present our findings based on the discussed methodology. The “Discussion” section will provide our analysis of the results. Finally, the “Conclusion” will summarize these insights, as well as provide suggestions for future research.

II. LITERATURE REVIEW

Adherence prediction using ML has been extensively studied [12][13]. In our previous work, we focused on adherence prediction model design considerations, particularly with respect to the choice of adherence metrics and the network architecture [1]. The scope of this paper is to extend upon this previous work, by exploring strategies implemented in other studies, to identify which of these strategies can further improve adherence prediction performance and accuracy. Furthermore, we sought to investigate the effectiveness of adherence prediction across a range of treatment durations for patients participating in a Homecare Support Program.

As an initial step in achieving these objectives, we reviewed published scientific literature with the aim of identifying a range of variables (e.g., demographics, treatment duration, disease burden etc.) that other studies within this field have utilized for their adherence prediction. Additionally, we reviewed the links between adherence metrics and the patient demographics they have been applied

to. The most prevalent adherence metric is the Proportion of Days Covered by medication (PDC). The number following PDC denotes the percentage of days that the patient is required to have medication stock for to be deemed adherent.

Table I contains an overview of various relevant studies, showing the data that is utilized by the adherence prediction networks in each study. A patient’s Length of time on Service (LOS), which is an indication of treatment duration, is considered a crucial measure in the context of our study. Additionally, a patient’s LOS also influences the maximum quantity of data that is provided to the network for the patient. In other words, the longer the LOS, the more data is available. Adherence prediction is commonly applied to patients who are relatively new to their treatment, often starting from treatment naivety [8][11][14]. These studies also typically have a single LOS requirement for their patients, from which all testing is performed. As a result of this, little research has been conducted into the dynamics of patient adherence across a timeframe, or the impact of treatment duration on predictive adherence. Although, it has been shown that providing longer timeframes for naïve patient data to networks can improve performance, though this not been studied for longer treatment duration ranges [8][11].

Several studies have found that the inclusion of other types of variables other than medication delivery data, does improve predictive capabilities [14]-[16]. Findings from our previous study also support this point [1]. Most studies analyzed have included some form of patient demographics, most commonly age and gender, with performance gains or correlations with adherence being associated [13][15][17]. Similarly, a number of studies also incorporate a variable that represents the burden associated with specific diagnoses - often comorbidities, drug complexity or average PDC, with both variables having been shown to improve prediction performance or links with adherence [12][14][18][19].

Our previous work focused on predicting adherence for a patient’s most recent month of service, for patients with 3-4 months LOS as well as for all patients regardless of LOS [1]. This approach allows for varied levels of patient experience with their medication from which adherence was predicted. However, this did not evaluate in detail the impact of LOS on adherence and adherence prediction. Due to the research gaps identified across our previous work as well as other studies, we are extending the scope of our previous study to address the relationship between treatment duration and adherence, as well as the further optimization of our network [1].

III. DATA AND PREPROCESSING

Patients included in our study are those diagnosed with long-term conditions and who have been receiving direct to home delivery of their medication as well as nurse support for medication self-administration at home from a clinical homecare provider (HealthNet Homecare Ltd). The study dataset contains, but is not limited to, demography, LOS, primary diagnosis, medication delivery confirmation, delivery communications and communication medium, and

Table I. Patient treatment duration and variable evaluation for adherence prediction

| Author, year | Therapy Area | Patient treatment duration | Adherence metric duration | Input Variables | |
|--------------------------------|--|--|--------------------------------------|----------------------|------------------------------|
| | | | | Patient demographics | Diagnosis/ medication burden |
| Franklin et al., 2015 [8] | Cardiovascular disease | Naïve to 3 months | PDC80 30 day | ✓ | ✓ |
| Lucas et al., 2017 [11] | Cardiovascular disease | Naïve to first prescription | PDC80 5 year | ✓ | ✓ |
| Kumamaru et al., 2018 [14] | Cardiovascular disease | Naïve to first prescription | PDC80 1 year | ✓ | ✓ |
| Haas et al., 2019 [20] | Fibromyalgia | Diverse | Self-reported adherence | ✓ | ✓ |
| Kim et al., 2019 [10] | Smoking addiction | Naïve to 4/16 weeks | Daily consumption | ✓ | |
| Galozy et al., 2020 [16] | Hypertension | 2 years | PDC80 1 year | ✓ | ✓ |
| Gao et al., 2020 [15] | Hypertension | Various | PDC80 1 year | ✓ | |
| Koesmahargyo et al., 2020 [21] | Diverse – predominantly mental diagnoses | 1-2 weeks | PDC80, 1 day and 1 week | ✓ | ✓ |
| Wang et al., 2020 [22] | Crohn's disease | 6 months minimum, 36 months average | Self-reported adherence | ✓ | ✓ |
| Wu et al., 2020 [23] | Type 2 Diabetes | Diverse | PDC80 1 year | ✓ | |
| Gu et al., 2021 [24] | Diverse diagnoses | N/A - 1 Week of medication data used | Next medication consumption | ✓ | |
| Kharrazi et al., 2021 [18] | Diverse diagnoses | 30-day prescription fill rate. 2 year PDC value provided | Hospitalizations same year/next year | ✓ | ✓ |
| Hsu et al., 2022 [25] | Cardiovascular disease | 2 years observations | 5 years PDC80 | ✓ | ✓ |
| Malin et al., 2023 [1] | Diverse diagnoses – asthma, dermatitis, psoriasis and more | Up to 12 months | 31 days PDC100/ PDC80 | | |
| <i>(Proposed work)</i> | | | 31 days PDC100 | ✓ | ✓ |

whether the patient receives enhanced nurse support (where enhanced nurse support is used to aid medication adherence).

The calculation of every patient's medication stock follows the methodology outlined in our previous study, with every medication delivery being converted into the number of days' worth of medication it provides [1]. Whilst taking into account, and allowing for, the stockpiling of this medication before it is fully depleted, a medication stock timeline is generated for every patient from when they joined the service until the current date. Using this medication stock

history for patients, time periods can be extracted and processed into visualized time-series data. This preprocessing step follows the structure laid out in our previous work [1].

Most of the design decisions selected for this study have stayed consistent with the previous work, with the continued use of a PDC requirement of 100% across one month to define adherence. This month period is kept separate from the visualized time-series data and is the target variable, representing whether the patient did or did not possess

medication every day in that month. The network architecture remains a CNN with visualized medication stock data [1].

IV. METHODOLOGY

To achieve our objectives, as well as to cover areas of existing research gaps, the impact of varying treatment durations (i.e. LOS) on predicting adherence will be focused on in this study. Additionally, due to the desire for optimizing performance from the previous work, the implementation of patient demographics and disease burden will be evaluated, as these data points have shown utility or links to adherence within other studies [14]-[16][26].

A. Medication stock granularity

As discussed in detail in our previous study, the concept of representing numerical/time-series data into the visual data domain is a common preprocessing technique that is used in signal processing for the improvement of performance, and was inspiration for this work [27][28]. In our previous study, we apply similar techniques for the novel application of visualizing delivery data. However, in this study we only utilized four colors to represent patient medication delivery and possession information within a given period - this approach will be referred to as the block medication stock approach [1]. One drawback of this approach is the resulting loss of information with respect to the specific quantity of medication that each patient has at a given time. In view of this, we sought to test a more granular approach, by using a gradient of colors to represent the range of medication stock that each patient will have. For this purpose, a gradient of colors between red and green was created which is mapped to 0-31 days of medication stock i.e. 0 days is red, 31+ days is green, and every quantity in between has a unique shade which reflects the exact quantity of medication stock. Like with the previous approach, this new approach will use the color white to represent the period before a patient receives their first delivery. This allows for heterogeneous data to be inputted into the network – in other words, allowing the model to accommodate patients with varying treatment durations.

This new strategy will be referred to as the gradient medication stock approach. Figure 1 shows the new approach in comparison to the previous approach, for two patients. The theory behind this changed approach is that the additional granularity of data in the images should provide more information to the CNN from which to learn features that constitute adherent behavior.

B. Patient data selection

In our previous study the prediction model was tested using the most recent 12 months of medication stock data for every patient in the dataset [1]. Considering our objective of predicting adherence in the patients' first 3 – 4 months on treatment, the approach from our previous study has a risk of not being reflective of patient behavioral patterns that can occur during their first few months on a treatment. Moreover,

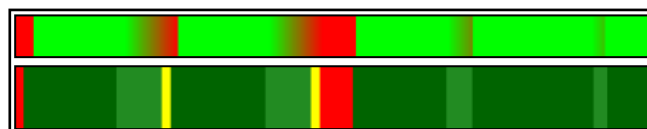


Figure 1a) Patient A.
Top: Gradient medication stock, Bottom: Block medication stock

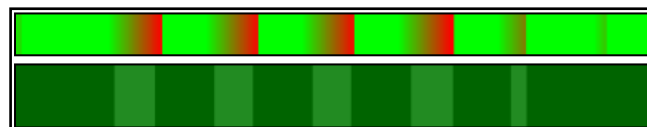


Figure 1b) Patient B.
Top: Gradient medication stock, Bottom: Block medication stock

Figure 1. Medication stock image representation comparison

training the network on the most recent 12 months of medication stock could be weighted towards longer LOS. To reduce this risk, we sought to evaluate whether the use of all patients' first 12 months on service would improve the prediction model's performance. Importantly, this new approach does not alter the testing dataset that is used, thus ensuring comparability across tests and studies. Furthermore, this change allows for features of naïve patients to be learned from when training the CNN model.

C. Treatment Duration (Length of time on service)

It is common within medication adherence studies for adherence prediction to be applied to patients who are relatively new to their treatment, often starting from a point of treatment naivety [8][11][14]. Crucially however, most studies in this area typically have a single LOS or treatment duration requirement for their study participants, from which all prediction model testing is performed [8][11][14]. To our knowledge therefore, little research has been conducted into the dynamics of patient adherence across varying treatment duration timeframe, or the impact of treatment duration on the prediction of adherence. Notably, it has been shown that providing longer timeframes of patient data, beginning from naivety, into an adherence prediction network can improve performance. However, this has not been studied for longer treatment duration ranges [8][11].

In our previous study, we focused on predicting adherence in two different patient scenarios: (1) for a patient's most recent month on a Homecare Support Program and (2) for patients with LOS of 3-4 months [1]. Although this approach allows for varied levels of patient experience with their medication to be taken into consideration when predicting adherence, it did not evaluate in detail the impact of LOS (and varying treatment durations) on adherence and adherence prediction. This is what we intend to evaluate in this study.

As previously stated, one of the major objectives of this study is to evaluate adherence across varying LOS. To achieve this, every patient has their full medication stock timeline extracted. The specific LOS (or treatment duration) that is being evaluated dictates where the patient medication stock data is cut-off. For example, if a patient has 15 months of medication data and we are evaluating adherence for an

LOS of <6 months then their first 6 months are taken and used for training; this process will occur for every patient.

This strategy ensures that every available patient is utilized, and that their data is processed in a way to capture their medication delivery and possession timeline from when they were at a specific LOS. This allows for more rigorous testing as the effective dataset size is far larger. It also allows for the use of cross-validation testing, as the whole dataset is now homogeneous with the evaluation criteria.

D. Incorporating additional patient variables

As demonstrated in our previous study, adherence prediction performance improved when additional relevant data such as enhanced service status and delivery communications are incorporated into the model [1]. Against this backdrop, we elected to identify further data variables that could potentially enhance the model performance even more. additional variables were selected on the basis of their potential to reflect patient behavior, and provided that they have already been collected as a part of our study dataset. It is important to note that the inclusion of a variable in the dataset means that such variables can either be routinely collected, in whole or in part, by virtue of a patient participating in a Homecare Support Program or that such variable can be collected/calculated from available and verifiable healthcare or population datasets.

In this study, two additional variables have been derived from data that is collected. These variables are average PDC and IMD (Index of Multiple Deprivation). The average PDC variable has been created for every diagnosis that is supported by the HealthNet Homecare Patient Support Program. For our study, diagnosis-specific average PDC has been used in our study as a proxy for disease severity. Other studies have shown correlation between the severity of disease and adherence, as well as direct links between an average PDC variable and predictive adherence performance [14][26][29].

IMD is used as another proxy for patient behavior, as several studies have shown that there is a correlation between deprivation/economic status and medication adherence [30][31].

To calculate an IMD value for each patient, the open-source IMD percentiles for the UK can be used, which correlate to specific Lower Layer Super Output Areas (LSOA) [32]. This can be mapped to partial postcode data, and the median IMD for that region is used. This gives a rough measure of deprivation for any given patient within the dataset. In addition to these derived variables, age, gender, enhanced service status and delivery communications with the patient are also evaluated due to noted benefits within other studies [1][14][15][16].

Our previous CNN tests that have implemented similar variables have visually appended this data to the images that are used for training and testing [1]. However, this approach is less feasible with more complex data, such as the average PDC and IMD, which are numeric data. To implement this data into a CNN the architecture must be modified to allow

for numeric and categorical data in addition to the image-based data that has previously been utilized. To input this data, for each patient their numeric and categorical variables are processed and concatenated into the network before the final dense layer – fusing heterogeneous data into a single network, with the aim of enhancing the level of information provided by each data point. This strategy has proven effective in other domains [33][34]. However, delivery communications continue to be visually appended to the medication stock timeline image.

E. Therapy specific training

There are four separate therapies (with different indications) represented in our study dataset. The therapy areas covered by the therapies included in our study are respiratory disease, dermatology, rheumatology, and gastroenterology. The included therapies have different features associated with them, with varying quantities of medication delivered as well as differing levels of disease severity and drug complexity. Table II shows the quartiles of how much medication is delivered to a patient in one delivery cycle.

It has been shown that a patient who is supplied with more medication, more regularly is likely to have different adherence patterns to a patient who receives larger deliveries less frequently [16][35]. As the four therapies within our study dataset cover a diverse range of diagnoses as well as varied prescription size, it is likely that there will be different behaviors associated with these therapies, which we wish to identify.

Due to this level of variability between patients within the study dataset, it is worth evaluating whether the network is over-generalising to the adherent behaviours for patients of specific therapies where there is a higher proportion of samples.

To evaluate the influence of this therapy imbalance, the use of a patient's therapy to create sub-groups for training specialised models was tested. If the overall performance improves then this would indicate that there is relevant information gained through using smaller, therapy specific networks. These results will be compared to models trained on all therapies, and both methodologies will be evaluated for their therapy-specific performance as well.

Table II. Quantity of medication delivered per therapy

| Therapy | Q1 Medication Stock Days | Median Medication Stock Days | Q3 Medication Stock Days | Percentage of total patients |
|---------|--------------------------|------------------------------|--------------------------|------------------------------|
| A | 56 | 56 | 56 | 43.4 |
| B | 28 | 28 | 84 | 35.9 |
| C | 56 | 56 | 56 | 13.5 |
| D | 56 | 84 | 84 | 7.2 |

V. RESULTS

A. Medication stock granularity

Following the methodology discussed previously, the use of gradient medication stock images have been evaluated against block medication stock images. Examples of these images can be seen in Figure 1. The results of this testing can be seen in Table III, where all results have been averaged across five runs.

The use of the new gradient methodology improves the capability of the network to generalize across both adherence and nonadherence. This is shown by a 0.9% AUC increase, and a 20.7% increase in overall accuracy. Due to this performance increase, this methodology will be utilized for all future testing and optimization.

B. Patient data selection

Initial results compare the performance of training using all patient's first 12 months of data, against the use of their most recent 12 months on the service, whilst testing on a separate 3-4 month test set, as detailed in our previous work [1]. These results can be seen in Table IV. Through using the first months of data for every patient, the AUC increases by 2.17%, and the accuracy increases by 2.70%. These results validate the use of the first months of a patient's medication stock data instead of their most recent months, when training a network to predict adherence for low LOS patients. This is likely due to the greater compatibility between patient behaviors, resulting from similar levels of naivety across the training and testing sets.

C. Treatment duration (length of time on service)

Table V shows the results of 5-fold cross-validation testing across various LOS categories for all patients. Following the 4th month, all patients should have received two deliveries, and this is when performance improves, which is corroborated in the literature where adherence prediction performance has increased following a prescription refill [8][11].

Table III. Medication supply visualization comparison

| Experiment | Accuracy (%) | AUC (%) |
|---------------------------|--------------|---------|
| Block medication image | 54.19 | 82.12 |
| Gradient medication image | 74.92 | 82.99 |

Table IV. Evaluation of patient medication period used for training

| Experiment | Accuracy (%) | AUC (%) |
|--------------------------------|--------------|---------|
| Naïve - 12 months patient data | 77.62 | 85.16 |
| Latest 12 months patient data | 74.92 | 82.99 |

Table V. Predictive adherence comparison across a range of target months

| Adherence prediction month | Accuracy (%) | AUC (%) | Adherent Population (%) |
|----------------------------|--------------|--------------|-------------------------|
| 4 | 73.30 | 85.25 | 54.6 |
| 5 | 92.19 | 98.11 | 66.0 |
| 6 | 84.54 | 94.28 | 65.5 |
| 7 | 89.96 | 96.82 | 66.0 |
| 8 | 88.63 | 96.24 | 64.9 |
| 9 | 92.63 | 98.36 | 66.1 |
| 10 | 88.24 | 96.07 | 61.1 |
| 11 | 90.68 | 96.70 | 62.3 |
| 12 | 89.56 | 97.52 | 60.9 |
| Average | 87.75 | 95.48 | 63.0 |

Additionally, it is worth noting that AUC does not linearly increase as more patient medication delivery data is provided. Instead, it appears to rise and fall cyclically. This can be explained through the medication delivery cycles that correspond to specific LOS months. This is shown in Figure 2 where the mean medication delivery and possession data is plotted across the first year of treatment for all patients. The weakest performance is seen in months 4 and 6 - it is during these months that a large portion of patients receive a medication delivery. This is indicated by the rise in mean medication stock days in Figure 2, as well as from Table II showing that all patients are expected to receive a delivery during month 6 (the 3rd delivery for therapies A and B and the 2nd delivery for therapies C and D). These factors likely contribute to weaker performance, as before an expected delivery the medication stock in a patient's possession will be lowest, leading to greater uncertainty for the network.

Disparities between the balance of adherent and nonadherent patients across the LOS months can also be linked to these performance variations, as the worst AUC was seen in the model with the least adherent population, and the best AUC was reached by the model with the most adherent population.

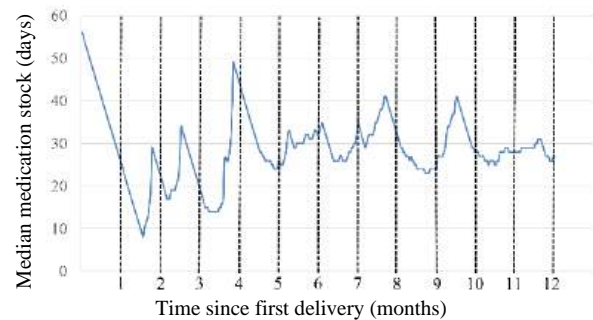


Figure 2. Median medication stock across every patient's first year of treatment

Generally, AUC improves as more patient data is supplied. This can be hypothesized to be due to patient behaviors becoming more stable as their treatment duration increases, as well as the network receiving more data with which to identify adherence patterns for patients. The best performing network was predicting the adherence for the 9th month across all patients, with 92.6% of predictions correct, with an AUC of 98.4%.

D. Incorporating additional patient variables

As shown in our previous study, and other studies, including a wide range of heterogeneous patient data can improve performance [1][21][24]. As determined by the literature review conducted, we have incorporated additional patient demographic data into the network following the trends shown in these studies. The additional data variables incorporated includes: age, gender, IMD and whether the patient receives enhanced nurse support. These variables have been grouped together under the column demographic data. The average PDC for each diagnosis is also trialed, attempting to capture the level of disease burden for each patient.

Additionally, medication delivery communications with patients have been visually appended to the medication delivery and possession images as they were previously shown to improve performance [1]. The results of this experimental testing can be seen below in Table VI. When delivery communications are incorporated there is an overall increase in AUC across most LOS months, as the networks with this variable had the highest average AUCs. This corroborates the findings in our previous study. However, most network configurations attain comparable performance, suggesting that the influence of the additional variables are not fundamental to adherence prediction, but do provide some benefit. This is most notable at 4 months, which should be the hardest predictive task due to limited prescription data [8][11], where integration of delivery communications increases both accuracy and AUC by 2-3%. As more patient data is included into the network, these performance differences diminish, but still convey utility.

E. Therapy specific training and evaluation

Correlation has been found between disease severity, drug complexity and prescription duration with respect to adherence [14][26][29][35]. All these features are directly linked to the therapy that a patient is on. As a result, we have performed tests on networks that have been trained and tested exclusively on specific therapies, using five-fold cross-validation testing. This will be referred to as the therapy specific approach, whilst the previous methodology that has been used will be referred to as the therapy agnostic approach. In this new therapy specific approach, the entire study dataset is segmented by therapy to create four smaller, therapy specific datasets, from which cross-validation testing is performed. Table VII compares the performance achieved by both approaches across the entire dataset, with therapy specific results being averaged across therapies. Additionally, Table VIII has been created which breaks down the performance per therapy. For the therapy agnostic approach, these results have been segmented by therapy after cross-validation testing was performed. The performance of therapy specific models was weaker than that found with the therapy agnostic models, but performance was closest when training and testing on the therapies with the most samples (therapies A and B). The explanation for this weaker performance is likely due to fewer training samples, which can result in overfitting. A further explanation is the difference in patient characteristics between therapies are potentially less significant than hypothesized. The therapies within our study dataset all have comparable drug complexity, which could be limiting the utility of this approach. It is plausible however, that as more data accrues for each therapy, there comes a point where the therapy specific models have enough samples to reliably outperform the agnostic models.

Despite the therapy specific models being outperformed by the therapy agnostic models, the segmentation of performance by therapy provides additional insight into the variations associated with predicting adherence for specific therapies. Patients associated with therapy D have the highest

Table VI. Predictive adherence comparison across a range of target months and incorporated variables

| Variables | | | | Maximum LOS months | | | | | | | | | | | | | | | | Average | | | | | |
|-------------------|-----------------|------------------|-------------|--------------------|------|---------|------|--------------|------|---------|------|--------------|------|---------|------|--------------|------|---------|------|--------------|------|---------|------|--|--|
| | | | | 4 | | 5 | | 6 | | 7 | | 8 | | 9 | | 10 | | 11 | | | | 12 | | | |
| Medication supply | Delivery comms. | Disease severity | Demographic | Accuracy (%) | | AUC (%) | | Accuracy (%) | | AUC (%) | | Accuracy (%) | | AUC (%) | | Accuracy (%) | | AUC (%) | | Accuracy (%) | | AUC (%) | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | |
| ✓ | | | | 73.3 | 85.3 | 92.2 | 98.1 | 84.5 | 94.3 | 90.0 | 96.8 | 88.6 | 96.2 | 92.6 | 98.4 | 88.2 | 96.1 | 90.7 | 96.7 | 89.6 | 97.5 | 87.7 | 95.5 | | |
| ✓ | ✓ | | | 76.3 | 87.3 | 92.3 | 98.3 | 85.7 | 94.3 | 89.9 | 96.2 | 84.2 | 95.9 | 88.9 | 97.5 | 88.5 | 96.2 | 91.7 | 97.7 | 91.7 | 97.4 | 87.7 | 95.7 | | |
| ✓ | | ✓ | ✓ | 73.6 | 85.9 | 92.0 | 98.0 | 84.7 | 93.6 | 89.0 | 96.7 | 88.3 | 96.3 | 92.6 | 98.4 | 88.0 | 96.3 | 89.4 | 97.5 | 90.7 | 97.3 | 87.6 | 95.6 | | |
| ✓ | ✓ | ✓ | ✓ | 76.0 | 88.1 | 92.5 | 98.2 | 86.3 | 95.0 | 89.8 | 96.4 | 89.3 | 96.8 | 92.8 | 98.6 | 89.3 | 96.6 | 86.3 | 95.2 | 90.4 | 97.5 | 88.1 | 95.8 | | |

Table VII. Predictive adherence comparison across a range of target months for therapy-specific and agnostic models

| Experiment | Months | | | | | | | | | | | | | | | | | | Average | |
|------------------|--------------|---------|--------------|---------|--------------|---------|--------------|---------|--------------|---------|--------------|---------|--------------|---------|--------------|---------|--------------|---------|---------|------|
| | 4 | | 5 | | 6 | | 7 | | 8 | | 9 | | 10 | | 11 | | 12 | | | |
| | Accuracy (%) | AUC (%) | Accuracy (%) | AUC (%) | Accuracy (%) | AUC (%) | Accuracy (%) | AUC (%) | Accuracy (%) | AUC (%) | Accuracy (%) | AUC (%) | Accuracy (%) | AUC (%) | Accuracy (%) | AUC (%) | Accuracy (%) | AUC (%) | | |
| Therapy agnostic | 73.3 | 85.3 | 92.2 | 98.1 | 84.5 | 94.3 | 90.0 | 96.8 | 88.6 | 96.2 | 92.6 | 98.4 | 88.2 | 96.1 | 90.7 | 96.7 | 89.6 | 97.5 | 87.7 | 95.5 |
| Therapy specific | 68.5 | 76.9 | 88.2 | 94.5 | 78.0 | 87.5 | 84.0 | 90.9 | 85.1 | 90.6 | 87.1 | 87.9 | 82.1 | 86.0 | 87.6 | 93.6 | 87.0 | 93.7 | 83.1 | 89.1 |

Table VIII. Predictive adherence comparison across a range of target months and therapies

| Months | Adherence (%) | Therapy A | | | | Adherence (%) | Therapy B | | | | Adherence (%) | Therapy C | | | | Adherence (%) | Therapy D | | | |
|--------|---------------|------------------|---------|------------------|---------|---------------|------------------|---------|------------------|---------|---------------|------------------|---------|------------------|---------|---------------|------------------|---------|------------------|------|
| | | Therapy agnostic | | Therapy specific | | | Therapy agnostic | | Therapy specific | | | Therapy agnostic | | Therapy specific | | | Therapy agnostic | | Therapy specific | |
| | | Accuracy (%) | AUC (%) | Accuracy (%) | AUC (%) | | Accuracy (%) | AUC (%) | Accuracy (%) | AUC (%) | | Accuracy (%) | AUC (%) | Accuracy (%) | AUC (%) | | Accuracy (%) | AUC (%) | | |
| 4 | 75.9 | 86.5 | 71.7 | 81.5 | 68.9 | 81.5 | 62.6 | 75.6 | 71.2 | 78.6 | 75.5 | 85.3 | 92.7 | 98.1 | 72.4 | 79.0 | 75.9 | 86.5 | 71.7 | 81.5 |
| 5 | 92.8 | 98.1 | 92.0 | 97.7 | 90.9 | 97.2 | 85.9 | 96.1 | 93.1 | 97.7 | 85.7 | 85.3 | 97.3 | 99.4 | 88.7 | 83.2 | 92.8 | 98.1 | 92.0 | 97.7 |
| 6 | 79.5 | 90.5 | 76.1 | 91.2 | 92.7 | 98.1 | 83.1 | 92.7 | 75.8 | 80.5 | 71.7 | 64.9 | 70.3 | 81.6 | 68.7 | 62.2 | 79.5 | 90.5 | 76.1 | 91.2 |
| 7 | 93.0 | 98.5 | 85.7 | 94.1 | 85.7 | 92.6 | 85.1 | 93.4 | 93.0 | 97.4 | 89.2 | 89.8 | 96.0 | 99.4 | 93.4 | 86.9 | 93.0 | 98.5 | 85.7 | 94.1 |
| 8 | 86.6 | 95.2 | 87.2 | 95.4 | 91.2 | 97.4 | 86.5 | 93.7 | 80.8 | 84.6 | 73.0 | 73.7 | 95.4 | 98.9 | 82.0 | 74.9 | 86.6 | 95.2 | 87.2 | 95.4 |
| 9 | 91.5 | 98.0 | 92.0 | 97.9 | 94.4 | 98.7 | 86.6 | 90.6 | 94.1 | 97.4 | 84.5 | 84.8 | 84.8 | 92.6 | 79.1 | 85.9 | 91.5 | 98.0 | 92.0 | 97.9 |
| 10 | 89.1 | 96.9 | 82.8 | 89.4 | 87.1 | 95.1 | 85.4 | 94.3 | 85.6 | 91.7 | 73.4 | 83.1 | 94.1 | 97.7 | 86.2 | 69.7 | 89.1 | 96.9 | 82.8 | 89.4 |
| 11 | 89.1 | 96.0 | 89.5 | 97.1 | 93.0 | 97.6 | 92.0 | 97.3 | 87.4 | 91.1 | 79.4 | 63.3 | 88.4 | 92.4 | 78.9 | 66.2 | 89.1 | 96.0 | 89.5 | 97.1 |
| 12 | 89.2 | 97.2 | 85.9 | 92.0 | 90.4 | 97.3 | 86.3 | 90.3 | 87.0 | 95.7 | 90.4 | 95.8 | 89.0 | 97.1 | 86.0 | 92.5 | 89.2 | 97.2 | 85.9 | 92.0 |
| Avg. | 87.4 | 95.2 | 84.8 | 92.9 | 88.3 | 95.1 | 83.7 | 91.5 | 85.3 | 90.5 | 80.3 | 80.7 | 89.8 | 95.2 | 81.7 | 77.8 | 87.4 | 95.2 | 84.8 | 92.9 |

average accuracy of 89.8% and an AUC of 95.3% across all treatment duration timeframes. This is despite the relatively smaller number of training samples for this therapy.

Furthermore, Figure 3 has been created to demonstrate the variation in delivery cycles across therapies, providing further insight into these results. This shows that the 6th month correlates with expected deliveries for therapies A, C and D and a large portion of patients within these therapies

having low medication stock at the end of month 5. This can largely explain the drop in AUC seen in month 6.

The variation in the model performance between different treatment durations and therapies indicates that there is utility in this information. Due to these distinct results attained by different therapies, using the patient LOS and their therapy can lead to more precise performance estimates in a real-world environment.

VI. DISCUSSION

Compared to our previous study, the prediction performance for patients with a 3-4 month LOS has improved, demonstrated by the AUC which increased from 82.12% to 85.16%. This improvement was driven by the more thorough analysis in this study, varying the granularity of medication stock data as well as the timeframe of patient data used for training. In addition to this, the treatment duration timeframe that adherence is predicted for was systematically evaluated. This approach allowed for analysis into the expected performance that can be achieved for specific points in a patient’s length of service.

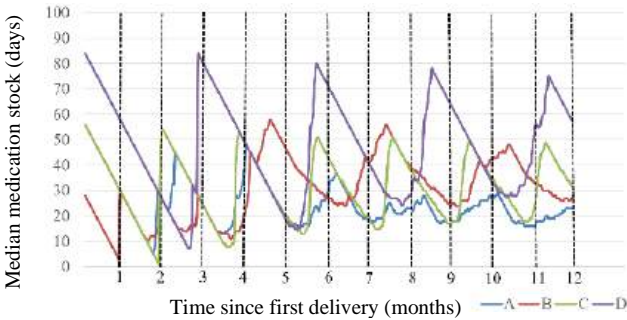


Figure 3. Median medication stock across every patient’s first year of treatment per therapy

We found that our model's performance is strongest at the 9 months LOS range, with an accuracy of 92.8% and an AUC of 98.6%. The variations in model performance observed across the different LOS's performance is likely due to the variation in medication delivery cycles, with there being significant trends in quantity of medication stock in a patient's possession with their time on service. This best-performing model utilized delivery communications, age, gender, IMD, diagnosis-specific average PDC, a variable detailing the presence of additional nurse support and the application of visualized medication possession time-series data. This corroborates with previous studies that have shown the benefit for adherence prediction when data relating to diagnosis, demographics, and communications are incorporated into the model [1][14][15][21][24]. As well as demonstrating how data can be enhanced through heterogeneity.

Additionally, through analysis into the predictions across the therapies included in this study, we found distinct differences in performance for specific therapies at certain months. This is linked to medication delivery cycle trends and can be used to further inform our level of confidence for classifications.

The results achieved in this study are crucial because tackling the issue of poor medication adherence requires the ability to accurately identify which patients have the greatest risk of poor adherence very early on in the patient's treatment journey and before negative adherence behaviors have set in or deteriorated. Needless to say, tailored interventions would be more impactful if implemented earlier, when patients are still relatively better engaged and subsequent interventions can also be amended accordingly as patients' poor adherence risks changes with the passage of time.

VII. CONCLUSION

This study set out to evaluate the difference in adherence prediction performance across varying patient treatment durations, as well as to optimize AUC achieved through our previous study [1]. The use of naïve patient data and more granular image data improved the AUC by 3.0% on a subset of patients with LOS between 3-4 months. Further analysis was conducted across a range of patient LOS values, finding the highest AUC reached when predicting the adherence during the 9th month, with an AUC of 98.6%. We find utility in the inclusion of delivery communications, improving AUC by approximately 0.2% when compared to comparable models without this variable. Likewise, the use of demographic data improves AUC by approximately 0.1%, with our best-performing model utilizing all of these variables. Though, there is scope for further modification of the CNN architecture to process these variables, as CNNs are novel within this domain and there are many approaches that can be taken.

When gathering all results attained through 5-fold cross-validation and averaging across treatment durations, we have identified disparity between the adherence of patients, as well

as the prediction results, across the therapies offered. When averaged across all studied months, there is a 5% difference in AUC between the best performing therapy and the worst, indicating therapy specific characteristics linked to adherence. Through our analysis across therapies and LOS, greater specificity can be attained with regards to expected performance for patients in real-world situations.

The ability to predict the risk of poor medication adherence offers immense value to healthcare providers and to patients. However, intervening accordingly (with the appropriate intervention delivered through the appropriate channel for each patient) in response to such predicted risk is of equal importance. To this end, an area of further study includes the prediction of how patients' preferences in terms of the type, format and channel of interventions could change over time.

ACKNOWLEDGMENT

This work was conducted as part of a predictive adherence project funded by HealthNet Homecare UK LTD.

REFERENCES

- [1] B. Malin, T. Kalganova, E. Nwokoro, and J. Hinton, "Medication Adherence Prediction for Homecare Patients, Using Medication Delivery Data," in *HEALTHINFO 2023, The Eighth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing*, Valencia, Spain: IARIA, 2023, pp. 30–38. Available: https://www.thinkmind.org/index.php?view=article&articleid=healthinfo_2023_1_60_80027
- [2] T. Patel, "Medication nonadherence: Time for a proactive approach by pharmacists," *Canadian Pharmacists Journal*, vol. 154, no. 5. SAGE Publications Ltd, pp. 292–296, Sep. 01, 2021. doi: 10.1177/17151635211034216.
- [3] A. G. G. Stuurman-Bieze, E. G. Hiddink, J. F. M. van Boven, and S. Vegter, "Proactive pharmaceutical care interventions decrease patients' nonadherence to osteoporosis medication," *Osteoporosis International*, vol. 25, no. 6, pp. 1807–1812, 2014, doi: 10.1007/s00198-014-2659-8.
- [4] P. M. Ho *et al.*, "Effect of Medication Nonadherence on Hospitalization and Mortality Among Patients With Diabetes Mellitus," *Arch Intern Med*, vol. 166, no. 17, pp. 1836–1841, Sep. 2006, doi: 10.1001/archinte.166.17.1836.
- [5] F. Kleinsinger, "The Unmet Challenge of Medication Nonadherence," *The Permanente Journal/Perm J*, vol. 22, pp. 18–033, 2018, doi: 10.7812/TPP/18-033.
- [6] E. Wiecek, F. S. Tonin, A. Torres-Robles, S. I. Benrimoj, F. Fernandez-Llimos, and V. Garcia-Cardenas, "Temporal effectiveness of interventions to improve medication adherence: A network meta-

- analysis,” *PLoS One*, vol. 14, no. 3, p. e0213432, 2019, doi: 10.1371/journal.pone.0213432.
- [7] E. Unni, O. O. Shiyanbola, and K. B. Farris, “Change in Medication Adherence and Beliefs in Medicines Over Time in Older Adults,” *Glob J Health Sci*, vol. 8, no. 5, pp. 39–47, Sep. 2015, doi: 10.5539/gjhs.v8n5p39.
- [8] J. M. Franklin, A. A. Krumme, W. H. Shrank, O. S. Matlin, T. A. Brennan, and N. K. Choudhry, “Predicting adherence trajectory using initial patterns of medication filling,” *American Journal of Managed Care*, vol. 21, no. 9, pp. 537–544, 2015.
- [9] M. T. Brown and J. K. Bussell, “Medication adherence: WHO cares?,” *Mayo Clin Proc*, vol. 86, no. 4, pp. 304–314, Apr. 2011, doi: 10.4065/mcp.2010.0575.
- [10] N. Kim *et al.*, “Predictors of adherence to nicotine replacement therapy: Machine learning evidence that perceived need predicts medication use HHS Public Access,” *Drug Alcohol Depend*, vol. 205, p. 107668, 2019, doi: 10.1016/j.drugalcdep.2019.107668.
- [11] J. E. Lucas, T. C. Bazemore, C. Alo, P. B. Monahan, and D. Voora, “An electronic health record based model predicts statin adherence, LDL cholesterol, and cardiovascular disease in the United States Military Health System,” 2017, doi: 10.1371/journal.pone.0187809.
- [12] A. Bohlmann, J. Mostafa, and M. Kumar, “Machine Learning and Medication Adherence: Scoping Review,” *JMIR Med*, vol. 2, no. 4, p. e26993, 2021, doi: 10.2196/26993.
- [13] B. Uchmanowicz, E. A. Jankowska, I. Uchmanowicz, and D. E. Morisky, “Self-Reported Medication Adherence Measured With Morisky Medication Adherence Scales and Its Determinants in Hypertensive Patients Aged ≥ 60 Years: A Systematic Review and Meta-Analysis,” *Front Pharmacol*, vol. 10, no. March, pp. 1–11, 2019, doi: 10.3389/fphar.2019.00168.
- [14] H. Kumamaru *et al.*, “Using Previous Medication Adherence to Predict Future Adherence,” 2018, doi: 10.18553/jmcp.2018.24.11.1146.
- [15] W. Gao *et al.*, “A Clinical Prediction Model of Medication Adherence in Hypertensive Patients in a Chinese Community Hospital in Beijing,” *Am J Hypertens*, vol. 33, no. 11, pp. 1038–1046, Nov. 2020, doi: 10.1093/AJH/HPAA111.
- [16] A. Galozy and S. Nowaczyk, “Prediction and pattern analysis of medication refill adherence through electronic health records and dispensation data ☆,” 2020, doi: 10.1016/j.yjbinx.2020.100075.
- [17] L. Thunander Sundbom and K. Bingefors, “Women and men report different behaviours in, and reasons for medication non-adherence: a nationwide Swedish survey,” *Pharm Pract (Granada)*, vol. 10, no. 4, pp. 207–221, Oct. 2012, doi: 10.4321/s1886-36552012000400005.
- [18] H. Kharrazi, X. Ma, H. Y. Chang, T. M. Richards, and C. Jung, “Comparing the Predictive Effects of Patient Medication Adherence Indices in Electronic Health Record and Claims-Based Risk Stratification Models,” *Popul Health Manag*, vol. 24, no. 5, pp. 601–609, 2021, doi: 10.1089/pop.2020.0306.
- [19] B. Jimmy and J. Jose, “Patient Medication Adherence: Measures in Daily Practice,” 2011.
- [20] K. Haas, Z. Ben Miled, and M. Mahoui, “Medication Adherence Prediction Through Online Social Forums: A Case Study of Fibromyalgia,” *JMIR Med Inform*, vol. 7, no. 2, Apr. 2019, doi: 10.2196/12561.
- [21] V. Koesmahargyo *et al.*, “Accuracy of machine learning-based prediction of medication adherence in clinical research,” *Psychiatry Res*, vol. 294, p. 113558, 2020, doi: 10.1016/j.psychres.2020.113558.
- [22] L. Wang *et al.*, “Applying machine learning models to predict medication nonadherence in crohn’s disease maintenance therapy,” *Patient Prefer Adherence*, vol. 14, pp. 917–926, 2020, doi: 10.2147/PPA.S253732.
- [23] X. W. Wu, H. B. Yang, R. Yuan, E. W. Long, and R. S. Tong, “Predictive models of medication non-adherence risks of patients with T2D based on multiple machine learning algorithms,” *BMJ Open Diabetes Res Care*, vol. 8, no. 1, Mar. 2020, doi: 10.1136/bmjdr-2019-001055.
- [24] Y. Gu *et al.*, “Predicting medication adherence using ensemble learning and deep learning models with large scale healthcare data,” *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-98387-w.
- [25] W. Hsu, J. R. Warren, and P. J. Riddle, “Medication adherence prediction through temporal modelling in cardiovascular disease management,” *BMC Med Inform Decis Mak*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12911-022-02052-9.
- [26] M. R. DiMatteo, K. B. Haskard, and S. L. Williams, “Health Beliefs, Disease Severity, and Patient Adherence: A Meta-Analysis,” *Med Care*, vol. 45, no. 6, pp. 521–528, Mar. 2007.
- [27] A. N. Sayed, Y. Himeur, and F. Bensaali, “From time-series to 2D images for building occupancy prediction using deep transfer learning,” *Eng Appl Artif Intell*, vol. 119, p. 105786, 2023, doi: <https://doi.org/10.1016/j.engappai.2022.105786>.
- [28] Z. Wang and T. Oates, “Imaging Time-Series to Improve Classification and Imputation,” Palo Alto, California USA, Jul. 2015, doi: <https://doi.org/10.48550/arXiv.1506.00327>.
- [29] K. A. Hommel, L. A. Denson, and R. N. Baldassano, “Oral medication adherence and disease severity in pediatric inflammatory bowel disease,” *Eur J*

- Gastroenterol Hepatol*, vol. 23, no. 3, 2011, doi: 10.1097/MEG.0b013e328344019c.
- [30] M. E. Wilder *et al.*, “The Impact of Social Determinants of Health on Medication Adherence: a Systematic Review and Meta-analysis,” *J Gen Intern Med*, vol. 36, no. 5, pp. 1359–1370, May 2021, doi: 10.1007/s11606-020-06447-0.
 - [31] E. A. Mamaghani, E. Hasanpoor, E. Maghsoodi, and F. Soleimani, “Barriers to Medication Adherence among Hypertensive Patients in Deprived Rural Areas,” *Ethiop J Health Sci*, vol. 30, no. 1, pp. 85–94, Jan. 2020, doi: 10.4314/ejhs.v30i1.11.
 - [32] MHCLG, OCSI, NISRA, and S. Government, “Index of Multiple Deprivation,” 2015. doi: 10.20390/enginddepriv2015.
 - [33] T. N. Wolf, S. Pölsterl, and C. Wachinger, “DAFT: A universal module to interweave tabular data and 3D images in CNNs,” *Neuroimage*, vol. 260, p. 119505, 2022, doi: <https://doi.org/10.1016/j.neuroimage.2022.119505>.
 - [34] S.-J. Heo *et al.*, “Deep Learning Algorithms with Demographic Information Help to Detect Tuberculosis in Chest Radiographs in Annual Workers’ Health Examination Data,” *Int J Environ Res Public Health*, vol. 16, no. 2, Jan. 2019, doi: 10.3390/ijerph16020250.
 - [35] S. King, C. Miani, J. Exley, J. Larkin, A. Kirtley, and R. A. Payne, “Impact of issuing longer- versus shorter-duration prescriptions: a systematic review,” *Br J Gen Pract*, vol. 68, no. 669, pp. e286–e292, Apr. 2018, doi: 10.3399/bjgp18X695501.

Decoding Key Variables Contributing to Right Ventricular Involvement in Ischaemic and Non-ischaemic Cardiomyopathy

Carlos Barroso-Moreno

*Faculty of Biomedical and Health Sciences
Universidad Europea de Madrid
Madrid, Spain
22014885@live.uem.es*

Hector Espinos Morato

*Faculty of Engineering
Universidad Europea de Valencia
Valencia, Spain
hector.espinos@universidadeuropea.es*

Enrique Puertas

*S. Architecture, Engineering and Design
Universidad Europea de Madrid
Madrid, Spain
enrique.puertas@universidadeuropea.es*

Juan José Beunza Nuin

*Faculty of Biomedical and Health Sciences
Universidad Europea de Madrid
Madrid, Spain
juanjose.beunza@universidadeuropea.es*

José Vicente Monmeneu

*Cardiac Magnetic Resonance
ASCIREs Biomedical Group
Valencia, Spain
jmonmeneu@ascires.com*

María P. López-Lereu

*Cardiac Magnetic Resonance
ASCIREs Biomedical Group
Valencia, Spain
mplopezl@ascires.com*

David Moratal

*Center for Biomaterials and Tissue Engineering
Universitat Politècnica de València
Valencia, Spain
dmoratal@eln.upv.es*

Abstract—Cardiomyopathy is a condition affecting the heart muscle, poses challenges to effective blood pumping by the heart. While prior research predominantly concentrated on the left ventricle, recent investigations underscore the significance of the right ventricle. This study aims to ascertain the clinical and cardiac parameters influencing right ventricular engagement in both ischaemic and non-ischaemic cardiomyopathy. A database comprising 56,447 subjects, collected between 2008 and 2020 by the ASCIREs Biomedical Group, forms the basis of this investigation. The methodology encompasses two main blocks: the clinical aspect utilizes decision trees for enhanced interpretability, while the technical aspect employs Machine Learning to achieve a higher degree of prediction accuracy. Power Business Intelligence and RapidMiner are the main software tools enabling data transformation and deep data analysis within the EU legal framework for health data. The outcomes reveal the pivotal influence of disparities in aortic artery beat volume, pulmonary vascular volume and aortic arch as key factors. Remarkably, the RapidMiner tool, employing the decision trees algorithm, attains an impressive Area Under the Curve (AUC) of 0.873 with XGBoost and decreases to AUC= 0.791 with SVM. In conclusion, the study underscores the ability to identify crucial clinical variables associated with right ventricular involvement, offering the potential to streamline diagnostic procedures and reduce associated timeframes in cardiomyopathy scenarios.

Keywords—Cardiomyopathy; Machine Learning Algorithms; right ventricular; Pulmonary Vascular Resistance.

I. INTRODUCTION

This paper is an extended and updated presentation of the research based on the publication and presentation at the HEALTHINFO 2023 conference in Valencia, Spain [1]. The extension allows to show details of the methodology used and a greater number of results.

Among cardiovascular diseases, ischaemic heart disease accounts for 16% of all deaths worldwide, rising from over 2 million deaths in 2000 to 8.9 million in 2019, and has become the disease attributed with the largest increase in deaths since 2000 [2]. Other cardiac conditions, such as non-ischaemic cardiomyopathy, arrhythmia, valvular heart disease, and heart failure are highly prevalent in developed countries and also cause high morbidity and mortality [3]. Due to the complexity and high prevalence of these diseases, a better understanding of the pathophysiology, as well as earlier diagnosis is of vital importance to increase the success rate of therapies, which is reflected in a reduced level of disability and lower mortality. To this end, for decades, all attention has been directed to the study of the left ventricle, making the right ventricle the "forgotten side of the heart". On the other hand, a direct extrapolation of the knowledge acquired about the physiology of the left side of the heart to the right side is not possible, as the normal right ventricle is anatomically and functionally different from the left ventricle. However,

in recent years, advances in non-invasive cardiac imaging techniques have made it possible to discover the importance of the right ventricle in different cardiac diseases [4]. Therefore, there is a need for a better understanding of those factors that influence right ventricular dysfunction, given the accumulating evidence of their clinical relevance from both a symptomatic or diagnostic and prognostic perspective.

Numerous researchers have demonstrated the feasibility of applying Machine Learning (ML) algorithms in health studies to predict strokes [5], ICU patients with Covid-19 [6], prostate cancer [7] or acute coronary diseases syndrome [7] and others. The joint use of this type of algorithms with visualisation tools, such as Power BI is used in numerous areas: restaurant marketing [8], rental car [9], tourism [10] or fashion industry [11]. However, the diagnostic use of decision-making tools is still rare in the health sector. In recent years, these types of diagnostic aid tools have become popular. For example, a success story is the application of this type of tools in the private health sector in Finland, which has based its health system's decisions on data, identifying key factors [12].

The RapidMiner tool is a data analytics platform that offers manipulation and predictive analytics with a graphical interface. The platform is used to analyse large datasets, especially using decision trees in the health domain. Beyond healthcare, RapidMiner showcases its adaptability in diverse domains, exemplified by its utilization in social network analysis as demonstrated by Anand *et al.* [13], tourism research highlighted by Seovcanac *et al.* [14], and the burgeoning field of electric vehicles, as explored by Fernandes *et al.* [15]. The tool is expansive utility underscores its significance as a multifaceted instrument contributing to advancements in analytics across various sectors [16].

Previous studies have identified influential variables in right ventricular compromise, such as: pulmonary arterial hypertension (PAH) associated with pressure overload [17]; diabetes, dyslipidemia [18], blood flow [19] and habits, such as smoking [20], family history of disease [21] and others. These multifaceted insights from previous research collectively contribute to a comprehensive understanding of the diverse variables influencing right ventricular health.

The main objective of this project is to determine those clinical and cardiac parameters that influence the involvement of the right ventricle in ischemic and non-ischemic cardiomyopathy using ML techniques. To this end, predictive models capable of identifying patients with right ventricular dysfunction will be developed and the key parameters used by the models will be studied from the point of view of their clinical implication.

This paper is organized as follows. Section II presents the details of dataset and it describes the methodology for Power BI and RapidMiner. Section III describes the results focusing on significant variables, distributions and decision trees. Finally, Section IV presents the conclusions and directions for future work.

II. MATERIALS AND METHODS

The employed methodology involves a comprehensive evaluation of the prevalent supervised classification algorithms in ML: Support Vector Machines (SVM) [22], decision trees [23], Random Forest [24], and neural networks [25]. The assessment of results encompasses the utilization of key metrics such as precision, sensitivity, specificity, and the Area Under the Curve (AUC). This last variable is a pivotal measure to evaluate the performance of binary classification models.

The dataset is sourced from the ASCIRES Biomedical Group database, comprising 56,447 records encompassing variables collected meticulously between 2008 and 2020.

A. Overall methodology flow

The flow of methodology in the research project is visually presented in Figure 1. The first step corresponds to the extraction of data from the ASCIRES servers. The application of the ETL process allows managing server extractions, customised transformations and loads for a first exploratory analysis [26]. The transformations seek a pre-processing of the data to adapt it to the legal framework with secure encryption for variables such as names, surnames, addresses or removal of the date of birth by the variable 'age', among others. The next point of the process focuses on improving the quality of the data, considering aspects such as the consistency of dates of birth, age, adaptation of formats or grouping of similar categories that are differentiated by orthographic criteria (such as, for example, Fusion or Fusion). Finally, validations of duplicate patients are performed using the unique patient ID. For the exploratory analysis, statistical mathematics such as mean, median or histograms, among others, are applied. Likewise, the multivariate algorithms of Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA) allow us to detect the variables that provide the most information.

The second step is the division of the resulting database into an 80% for training and a 20% for testing the algorithm. This division is not static since the algorithms randomly select patients for training and testing.

The third step corresponds to the development and integration of the ML algorithms, which are: Linear binary SVM, binary classification tree, neural networks, Naive Bayes and discriminant analysis. The choice of these algorithms is based on medical and technical criteria, with simple algorithms (SVM) seeking to explain the results, and complex neural network algorithms achieving the highest scores in the evaluation. This situation allows the creation of feature vectors.

Fourth to model cross-validation is a technique for evaluating ML models by training various subsets of available input data and evaluating them with other complementary subsets of data. The operation of cross-validation is divided into four steps: division into subsets of data, the default is 10 folds; 1 fold is reserved for validation and the remaining folds for training the model; this process is repeated as many times as folds are available, evaluating the accuracy; finally, a set of performance metrics and results are generated from all the data. This technique allows to assess the reliability of any

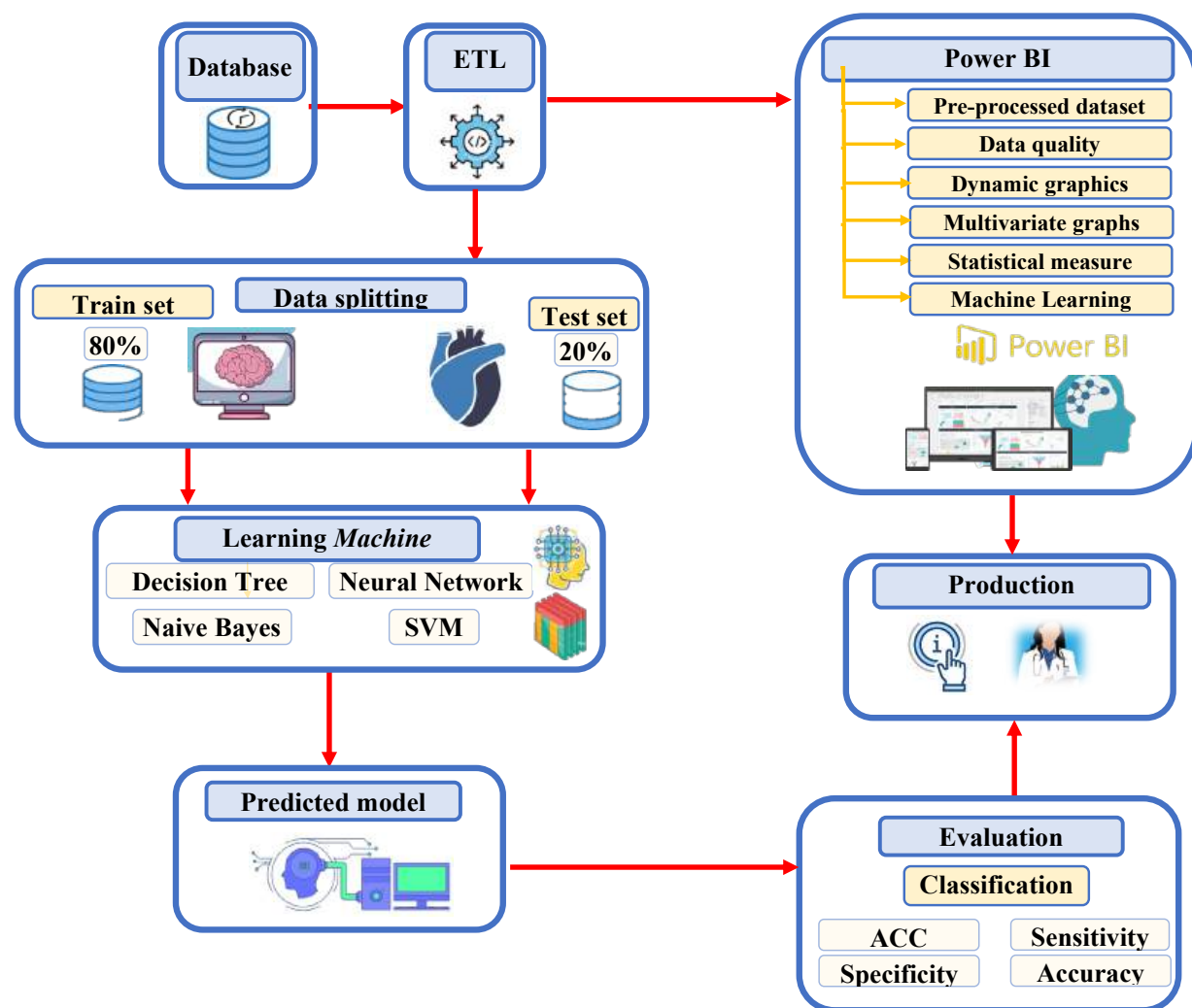


Figure 1. Overview of the flow of methodology in the research project.

trained model by testing the variability of the dataset, thus obtaining the final values for the evaluation. This requires a high computational cost due to random partitioning and use of a larger number of data, by training and evaluating the model several times [27].

The fifth step corresponds to the prediction model generated. The generation of the model requires a high computational cost of more than 1 hour of computation with the equipment used. The advantage lies in the speed of the result when introducing a new patient for prediction, in tenths of a second. Therefore, the models require a high computational load to be generated, but their medical application is instantaneous, fulfilling the medical requirement of minimising waiting times.

The sixth and seventh steps are the model scoring results and the model evaluation criteria, respectively. Each model has its own peculiarities; as a general rule, the higher the complexity of the algorithm, the higher the prediction score, but also the lower the control of the understanding of the learning process, becoming black boxes for the researcher

and clinician. The evaluation measures for the algorithms are mainly: ACC, SE, SP, NPV and AUC [28]. At this point it is necessary to take into account different results for each algorithm applied and also the different models of the database as input parameter. All this is done in RapidMiner.

The last step corresponds to the entry into production of the algorithms, providing health professionals with the results of the predictor in an informative way. The advantage of the methodological scheme used is the speed of the results, cost savings and the ability to diagnose ischaemic and non-ischaemic cardiomyopathy.

Another complementary methodology employed in the research is the Power BI business tool developed by Microsoft, which allows the integration of various data analysis and results gathering techniques. The first step is the loading of data into the Power BI tool with connectors already integrated in the tool. In this loading, data pre-processing is carried out, such as the formatting of variables, or the elimination of duplicates with the selection of options, among others. This

automation makes it possible to maintain high data quality. Subsequently, various types of graphs are created: column, linear, combined, circular, ring, card, scatter, bubble, meter, etc. All the graphs constructed are dynamic and interactive, an advantage of Power BI, allowing the researcher to filter the graphs according to their research needs. In addition, there are a series of ML algorithms implemented in Power BI that allow quantifying relationships of variables with exitus. Finally, all this information is accessible to healthcare professionals. In Figure 1 a dashed line is drawn to production because this analysis is investigative and time-consuming, unlike ML algorithms designed to provide real-time feedback on patient survival.

B. Software used

The research uses two software tools: Power BI and RapidMiner.

- Power BI is a data analytic service from Microsoft that provides interactive graphs focused on analytic intelligence to generate reports [29]. In the present research it allows easy visualisation of the database to automate the analysis.
- RapidMiner is a software for data analysis and data mining by chaining operations in a graphical environment [30]. Version 9.10.013 is used to obtain the results of the ML models.

C. Data preparation

At this point, clinical filters and patient labels are made according to age, gender, systole and diastole of the right ventricle. The volume of data cleaning by means of the filters means, that the initial database has 56,447 patients and 1815 variables; after applying the filters, these are reduced to 12,083 patients and 120 variables. Of the latter subgroup, 7153 are labelled as unaffected and the target group is 4944 with right ventricular involvement.

The process applies logical cleaning, such as: (i) Removal of inconsistent data, such as the presence of letters in numerical values. (ii) Elimination of erroneous data, heights < 1 metre and > 2.3 metre or ages < 0 and > 120 years. (iii) Checking whether the numeric value zero represents such a value or is a null value (NULL). (iv) For having the same content as other variables, but with a different name, e.g., Vol.eyec.Ao for vol.lat. (v) For having all data set to 0; (vi) Transformation into international units of certain variables, such as wood units. (vii) For containing inconsistent data from the clinical perspective (outlayer).

An additional step is the elimination of variables with higher correlations, in order to avoid multicollinearity in our database. These steps are the following:

- (i) $\text{Frac.reg.Ao.por.vol.lat}$ to $\text{Ao.reg.Vol.beat.vol.dif}$ with $\rho = 0.91$ (ii) $\text{Ao.reg.Vol.beat.vol.dif}$ to $\text{Vol.reg.Pulm.dif.vol.lat}$ with $\rho = -1$ (iii) $\text{Frac.reg.Ao.por.vol.lat}$ to $\text{Vol.reg.Pulm.dif.vol.lat}$ with $\rho = -0.92$ (iv) IMVI to MVI with $\rho = 0.93$ (v) NLVEDV to NLVESV with $\rho = 0.94$ (vi) Weight (kg) to S.Corp with $\rho = 0.95$

- (vii) NRVSV to RVSV with $\rho = 0.95$ (viii) NLVSV to LVSV with $\rho = 0.95$ (ix) NRVEDV to RVEDV with $\rho = 0.95$ (x) RWT...relative.wall.thickness to RWT.2...relative.wall.thickness.pwd.sd with $\rho = 0.90$ (xi) LVEDV to LVESV with $\rho = 0.93$ (xii) NLVEDV to LVEDV with $\rho = 0.95$ (xiii) NLVEDv to LVESV with $\rho = 0.90$ (xiv) IVTSVD to RVESV with $\rho = 0.97$ (xv) NLVESV to LVESV with $\rho = 0.97$

D. Data labels

Aligned with the main objective of the research, the database is labelled to assess whether the patient has right ventricular involvement. For the identification of these patients, the consensus tables specified by the European Society of Cardiology [31] establishes ranges of variables (age, gender, systoles, diastole and others) to identify RV involvement. These values are adapted in Table I, The normal values of RV systolic and diastolic parameters vary according to age and gender. The standard of normal values used for the recognition of impairment is used with a similar range in current papers, such as that of the researchers Petersen *et al.* (2019) [32].

The absolute values of end-systolic volume (ESV), end-diastolic volume (EDV) and body mass have been used, whose values are provided automatically within the framework of clinical tests. Clinics automatically provide based on clinical evidence frameworks.

Systolic volume (SV) is calculated by the difference between EDV and ESV; Additionally, ejection fraction (EF) is calculated as SV/VDE . Sex, body surface area (BSA), and age are independent predictors of several RV parameters, as suggested by previous well-established studies [33]. The standardised values of EDV/BSA and ESV/BSA are obtained from these variables.

The filtered database uses functions to apply binary labeling of patients: Normal or Abnormal (RV involvement). As an example, a 6 year old male patient with an EDV of 211 EDV (mL) is labeled as abnormal (RV impairment) because he is not within the range of (105.205) set in the parameters of Table I. If the same patient has an EDV of 204 EDV (mL), the patient is considered to have no RV involvement (normal), if the patient also meets the other variables in their corresponding ranges.

The database input and output variables are described in Table II. The \bar{X} is the average; SD is the standard deviation; Max is the maximum value and Min is the minimum value. The database contains 12,083 patients, of which 76.6% ($n = 9260$) are men and 23.4% ($n = 2823$) are women. The mean age is 62.49 years with a standard deviation of 14.1. The average body mass index (BMI) is 27.87 ($SD = 4.69$), the formula is $BMI = \text{Weight}(\text{kg})/[\text{Height}(\text{m})]^2$. According to the BSA, the average is 18.04 ($SD = 2.8$).

The output result corresponds to the classification of RV involvement with patients without RV involvement (normal) in a percentage of 59.1% ($n = 7139$) and with RV involvement (abnormal) in a percentage of 40.9% ($n = 4944$). Missing data is 0 because these records are removed in the preprocessing

TABLE I

RIGHT VENTRICLE LABELS. STANDARD RANGES BY RV VOLUMES, SYSTOLIC FUNCTION AND MASS BY AGE INTERVAL (95% CONFIDENCE INTERVAL).
ADAPTED FROM MACEIRA *et al.* (2006) [31].

| Right Ventricle labels | | | | | | |
|---|-----------|-----------|-----------|-----------|-----------|-----------|
| Age (years) | 20-29 | 30-39 | 40-49 | 50-59 | 60-69 | 70-79 |
| Males | | | | | | |
| Absolute values | | | | | | |
| 1-1 EDV (mL) SD 25.4 | (127.227) | (121.221) | (116.216) | (111.210) | (105.205) | (100.200) |
| ESV (mL) SD 15.2 | (38.98) | (34.94) | (29.89) | (25.85) | (20.80) | (16.76) |
| SV (mL) SD 17.4 | (74.143) | (74.142) | (73.141) | (72.140) | (71.139) | (70.138) |
| EF (%) SD 6.5 | (48.74) | (50.76) | (52.77) | (53.79) | (55.81) | (57.83) |
| Mass (g) SD 14.4 | (42.99) | (40.97) | (39.95) | (37.94) | (35.92) | (33.90) |
| Normalized to BSA | | | | | | |
| EDV/BSA (mL/m ²) SD 11.7 | (68.114) | (65.111) | (62.108) | (59.105) | (56.101) | (52.98) |
| ESV/BSA (mL/m ²) SD 7.4 | (21.50) | (18.47) | (16.45) | (13.42) | (11.40) | (8.37) |
| Females | | | | | | |
| Absolute values | | | | | | |
| 1-1 EDV (mL) SD 21.6 | (100.184) | (94.178) | (87.172) | (81.166) | (75.160) | (69.153) |
| ESV (mL) SD 13.3 | (29.82) | (25.77) | (20.72) | (15.68) | (11.63) | (6.58) |
| SV (mL) SD 13.1 | (61.112) | (59.111) | (58.109) | (56.108) | (55.106) | (53.105) |
| EF (%) SD 6 | (49.73) | (51.75) | (53.77) | (55.79) | (57.81) | (59.83) |
| Mass (g) SD 10.6 | (33.74) | (31.72) | (28.70) | (26.68) | (24.66) | (22.63) |
| Normalized to BSA | | | | | | |
| 1-1 EDV/BSA (mL/m ²) SD 9.4 | (65.102) | (61.98) | (57.94) | (53.90) | (49.86) | (45.82) |
| ESV/BSA (mL/m ²) SD 6.6 | (20.45) | (17.43) | (14.40) | (11.37) | (8.34) | (6.32) |

TABLE II

INPUT VARIABLES FOR OUTPUT VARIABLE LABEL.

| Input | | | | |
|----------------|------------|-------|------|---------|
| Variable | Categories | n | % | Missing |
| Gender | Males | 9260 | 76.6 | 0 |
| | Females | 2823 | 23.4 | 0 |
| | \bar{X} | SD | Min | Max |
| Age | 62.49 | 14.1 | 20 | 95 |
| BMI | 27.87 | 4.69 | 13.1 | 71.4 |
| BSA | 18.04 | 2.8 | 4.22 | 53.13 |
| EDV/BSA | 70.94 | 28.34 | 12.0 | 403.3 |
| ESV/BSA | 33.86 | 21.03 | 0.9 | 251.9 |
| Output | | | | |
| Variable | Categories | n | % | Missing |
| DV involvement | Normal | 7139 | 59.1 | 0 |
| | Abnormal | 4944 | 40.9 | 0 |

step so as not to distort the output of the ML algorithms in later steps.

In the first tests, the variables in table are eliminated due to their direct relationship in table calculations. These are discarded, as detailed in Table III: right ventricular systolic volume (RVSV), left ventricular systolic volume (LVSV), right ventricular end-diastolic volume (RVEDV), left ventricular end-diastolic volume (LVEDV), right ventricular end-systolic volume (RVESV), left ventricular end-systolic volume (LVESV). However, they are reintroduced in the prediction tests, as there is no clear correlation with the output variable and they are not detected as primary variables for predicting RV involvement.

TABLE III

VARIABLES ELIMINATED BY INDIRECT USE IN THE OUTPUT VARIABLE.

| Eliminated variables | | | | |
|----------------------|-----------|-------|-----|-----|
| | \bar{X} | SD | Min | Max |
| RV.DTD | 32.78 | 7.33 | 6 | 100 |
| RVSV | 70.73 | 27.24 | 4 | 250 |
| LVSV | 75.05 | 24.16 | 8 | 200 |
| RVEDV | 135.5 | 55.69 | 23 | 500 |
| LVEDV | 196.89 | 75.97 | 33 | 600 |
| RVESV | 64.68 | 40.56 | 2 | 437 |
| LVESV | 122.08 | 69.62 | 15 | 500 |

E. Legal framework and data security

The General Data Protection Regulation (GDPR) is a European regulation that regulates the protection of natural persons in the processing of personal data and the free movement of such data, which includes the law Regulation (EU) 2016/679 of the European Parliament and of the Council [34]. The Ley Orgánica de Protección de Datos Personales y Garantía de los Derechos Digitales (LOPD-GDD) is the Spanish Organic Law 3/2018, of 5 December, which aims to guarantee and protect the processing of personal data according to the guidelines set out in the RGPD [35]. In Spain, the public body in charge of ensuring compliance with this law is the Spanish Data Protection Agency (AEPD). Both European and national legislation are present during the research, and some highlights are mentioned below. The consent of the patients according to the legal basis for the treatment must be explicit for which the data are used, at the point that affects the university the research purpose is preserved, so the authorisation of the patients is not needed again. Furthermore, there is an action protocol for cases of loss or theft of data, which is developed

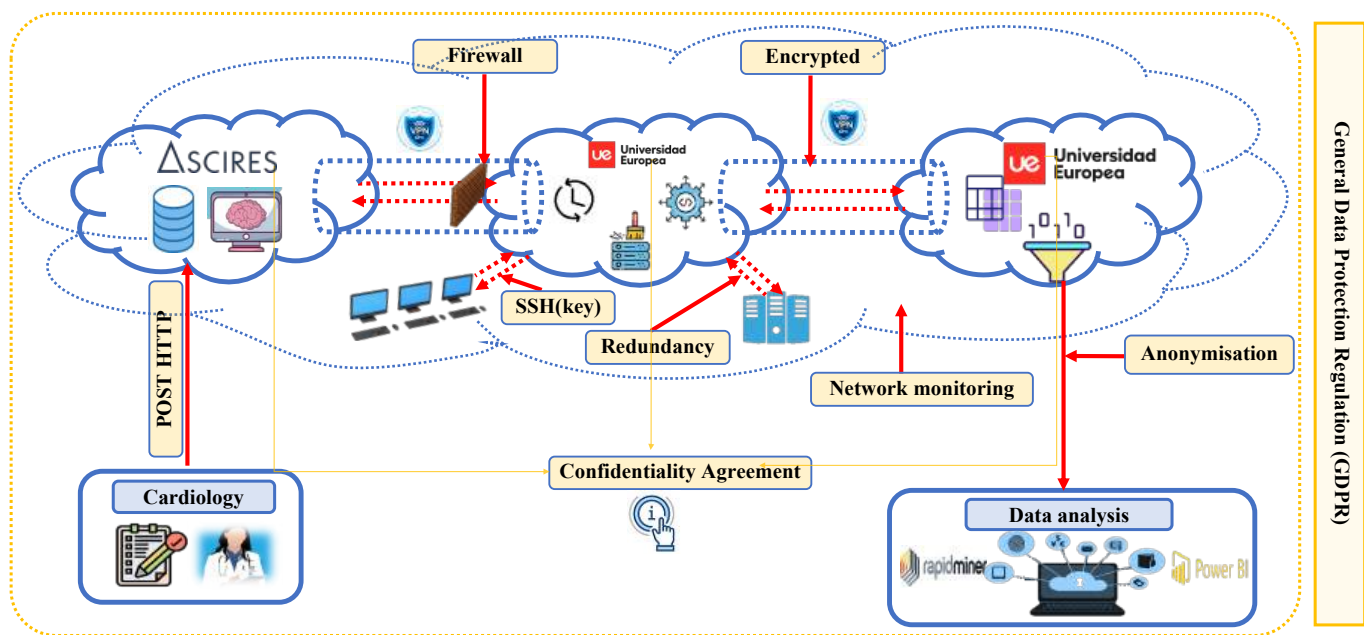


Figure 2. Overview of the legal framework and data security in research project.

by the Data Protection Delegate.

After mentioning the legislation in force and their respective protocols for action, Figure 2 compiles the techniques used both to comply with the legislation and the proactive part of security. The description in the figure follows the data flow in the process. This process starts after a new stroke patient is admitted and the physician enters the data in a form that is sent with the HTTPS protocol to the ASCIREs systems. This protocol prevents other users from intercepting the confidential information between the client and the server.

Once the data has been uploaded to the system, ASCIREs generates Virtual Private Network (VPN) to the EMU systems to upload new data on a regular basis. During this connection, a Firewall is set up to ensure that only the right person and files receive the data, and to block access to unauthorised users [36]. Researchers accessing such data need to have the SSH key, as this protocol establishes secure communications between two systems using client-server architecture to connect to the machine remotely. In addition, regular backups and checks allow for the creation of a backup system, the so-called redundancy. The company periodically carries out a vulnerability analysis to identify possible loopholes and to address them.

Different public and private institutions are involved in data traffic, which is why the confidentiality and non-disclosure agreement covers issues related to the database.

The data security with the various protocols ensures high security of digital privacy to prevent unauthorised access to the data, so it is worth mentioning that the description contains the protocols and techniques used, which can be detailed. A part of the process is omitted due to the confidentiality and privacy agreements of all parties.

III. RESULTS

After data preparation with a single model, the samples of the training set (n_{train}) and test sets (n_{test}) are 8458 and 3625, respectively. These patients are changed by cross-validation in the tests.

The results of the algorithms are shown in Table IV, which are Accuracy (ACC), Sensitivity (SE), Specificity (SP), Positive Predictive Value (PPV), Negative Predictive Value (NPV) and Area Under the Curve (AUC).

The algorithm with the highest AUC in RapidMiner is performed with XGBoost ($AUC = 0.87$), although the difference is small with the neural network algorithm ($AUC = 0.85$), determining a predominance of neural networks as predictors RV involvement. Although Random Forest algorithms have a ($AUC = 0.82$), and the lowest of the results is from Support Vector Machine (SVM) with a $AUC = 0.79$. The results are quite promising, as it means that we can predict almost with an accuracy of 9 out of 10 patients performing the clinical tests whether they have RV involvement.

Nevertheless, exercising caution is paramount when directly comparing the algorithms. While XGBoost yields the most favorable outcomes in this study, the inherent complexity of ML algorithms often renders them as "black boxes," challenging the interpretation of their inner workings. Conversely, decision trees, while achieving comparatively lower results, offer a distinct advantage in terms of interpretability, providing a valuable clinical perspective. The efficiency of the algorithms, coupled with the optimization techniques and the size of the database, ensures swift processing times, consistently under three minutes. These computations were conducted on a MacBook Pro 2.6 GHz Intel Core i7 6-core computer with 6 GB 2400 MHz DDR4 memory.

TABLE IV
COMPARISON AND EVALUATION OF DIFFERENT ALGORITHMS IN
RAPIDMINER.

| Algorithms | ACC | SE | SP | NPV | AUC |
|----------------|------|------|------|------|------|
| SVM | 0.76 | 0.76 | 0.76 | 0.77 | 0.79 |
| Random Forest | 0.78 | 0.78 | 0.79 | 0.78 | 0.82 |
| Neural Network | 0.78 | 0.78 | 0.79 | 0.77 | 0.85 |
| XGBoost | 0.82 | 0.78 | 0.80 | 0.77 | 0.87 |

Table V shows the main variables with the greatest weight detect in the ML algorithms, whose selection is automatic as key risk predictors. The descriptive analysis of the variables are gender in male with 76.6% ($n = 9260$) and female 23.4% ($n = 2823$), This is logical as it is a necessary variable for patient labelling. Dyslipidemia is present in 51.6% ($n = 6231$) of patients, hypertension has a 59.0% ($n = 7133$), diabetes has a low presence with a prevalence rate of 0.7% ($n = 85$), however, type II diabetes (low involvement) has a high incidence with 67.3% ($n = 8131$). The rest of the patients of each typology do not present the pathology. Smoking patients are high with 72.8% ($n = 8800$). Stent implantation is high with 82.2% ($n = 9937$) because the database is made up of patients who go to the cardiologist and suffer from some kind of affection or signs of affection of the heart. The stress study as a medical test is performed in 48.5% of patients ($n = 5857$), although the use of this aggressive test is decreasing every year.

In reference to the numerical variables, the most relevant is RVP [Wood] with $\bar{X} = 14.99$ ($SD = 1.27$). El aortic arch is $\bar{X} = 25.71$ ($SD = 3.68$), height is $\bar{X} = 167.86$ ($SD = 9.13$), Weight (kg) is $\bar{X} = 78.63$ ($SD = 14.96$). The Regurgitate Volume of the Aorta Artery (Reg.Vol.Ao.) is available for $\bar{X} = 4.31$ ($SD = 24.27$). The Diameter of the Aorta with Pulmonary Artery (DoA.PA) is $\bar{X} = 0.96$ ($SD = 0.25$). The Descending Thoracic Aorta (Desc.T.Ao) es de $\bar{X} = 24.43$ ($SD = 3.85$). Sinus pressure (Sinus.P) is $\bar{X} = 34.1$ ($SD = 4.77$). Posterior Wall in Diastole (PWD) has an average of $\bar{X} = 8.63$ ($SD = 2.75$). Finally, the ratio of diastolic volumes (diastolic RV) are of $\bar{X} = 1.59$ ($SD = 0.71$).

A. Pulmonary Vascular Resistance (PVR)

Pulmonary Vascular Resistance (PVR) is the mean pressure drop from the main pulmonary artery to the divided left atrium. The units of measurement are Wood's units, which arise from the equivalence one Wood's unit = $80 \cdot s \cdot cm^{-5}$.

PVR is defined by the Swan-Ganz catheter from a central vein [37], the formula is:

$$PVR = 80 * (PAP - CEP) * CO, \quad (1)$$

therefore depend on Pulmonary Arterial Pressure (PAP) in mmHg units; the Capillary Locking Pressure (CEP) in mmHg units; and Cardiac Output (CO) in l/min units. The normal value for a subject is 1-2 [Wood], but this increases with age,

TABLE V
ANALYSIS OF THE VARIABLES WITH THE GREATEST WEIGHT IN THE ML
ALGORITHM IN THE ASCIRES BIOMEDICAL GROUP DATABASE.

| Main variables | | | | |
|----------------------------|------------|-------|-------|-------|
| Variable | Categories | n | % | Miss. |
| Gender | Males | 9260 | 76.6 | 0 |
| | Females | 2823 | 23.4 | 0 |
| Dyslipidemia | Yes | 6231 | 51.6 | 0 |
| | No | 5852 | 48.4 | 0 |
| Hypertension | Yes | 7133 | 59.0 | 0 |
| | No | 4950 | 41.0 | 0 |
| Diabetes | Yes | 85 | 0.7 | 0 |
| | No | 11998 | 99.3 | 0 |
| Diabetes II | Yes | 8131 | 67.3 | 0 |
| | No | 3952 | 32.7 | 0 |
| Smoker | Yes | 8800 | 72.8 | 0 |
| | No | 3283 | 27.2 | 0 |
| Stent | Yes | 9937 | 82.2 | 0 |
| | No | 2146 | 17.8 | 0 |
| Stress study | Yes | 5857 | 48.5 | 0 |
| | No | 6226 | 51.5 | 0 |
| | \bar{X} | SD | Min | Max |
| PVR [wood] | 14.99 | 1.27 | 0.49 | 18.97 |
| Aortic.Arch | 25.71 | 3.68 | 2 | 60 |
| Height (cm) | 167.86 | 9.13 | 131 | 205 |
| Weight (kg) | 78.63 | 14.96 | 35 | 187 |
| Vol.reg.Ao. dif.vol.lat | 4.31 | 24.27 | -247 | 305 |
| DoA.PA | 0.96 | 0.25 | 0.01 | 5.01 |
| Desc.T.Ao | 24.43 | 3.85 | 2 | 63 |
| Sinus.P | 34.1 | 4.77 | 9 | 83 |
| PWD | 8.63 | 2.75 | 1 | 125 |
| diastolic RV | 1.59 | 0.71 | 0.17 | 10.88 |

as determined by the correlation table and Vizza *et al.* (2022) [38], increasing by 0.2 Wood in subjects over 50 years of age.

The database provided uses an estimation model to obtain the PVR value, described in equation 2 [39]:

$$PVR[Wood] = 19.38 - (4.62 * \ln(PAAV) - (0.08 * RVEF)) \quad (2)$$

where PAAV are in centimeter per second and RVEF in percentage.

B. Distributions

The original database model, which is not illustrated in the document, records an AUC value of 98% primarily due to the high proportion of patients in the database who do not show right ventricular (RV) involvement, causing imbalance in the data analysis. Consequently, the algorithms default to estimating a normal value, leading to a high accuracy rate (with sensitivity nearly at 0). However, through meticulous data filtering, the database achieves a balanced representation, mitigating this potential issue in the analysis of observations. Another notable aspect of the distributions involves excluding data during filtering if any of the variables are empty. This exclusion addresses the risk of bias introduced by eliminating patients who inherently did not complete data due to specific circumstances. Given the substantial volume of data, this

does not impede the analysis; instead, it enhances algorithmic performance, yielding more robust and accurate results.

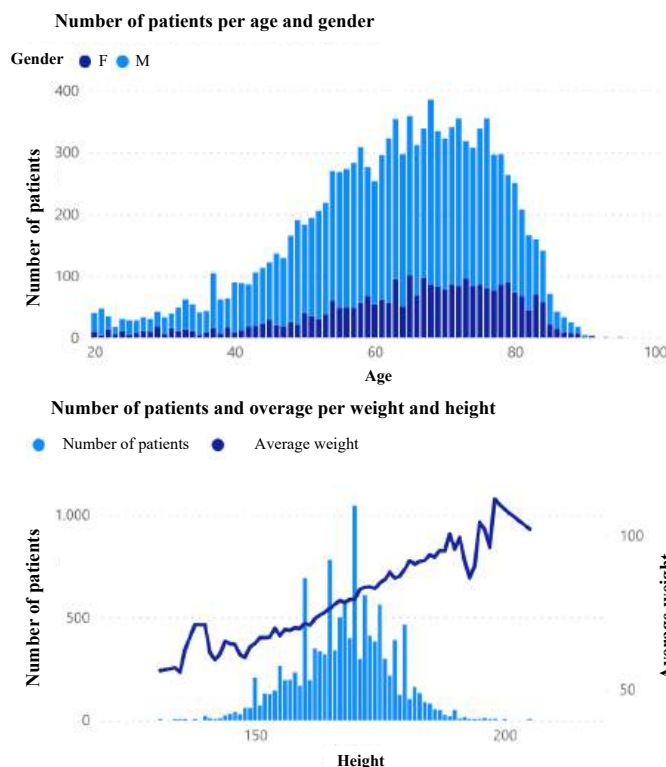


Figure 3. Multivariate histograms of number of patients, (top) age as a function of gender, (bottom) height as a function of mean weight.

The use of Power BI allows for quick visualisation of these and many other complexities, mitigating the risk of bias or inconsistencies in the data. Illustrated in figure 40, the histogram shows the distribution of patients by age and gender, highlighting the correlation between older age and higher incidence, especially among men. It should be noted that the age range between 60 and 80 years comprises almost half of the cases of right ventricular (RV) involvement. For height and weight, a discernible correlation emerges, prompting the use of body mass index (BMI) as a predictive variable in database labelling.

Figure 4 is the multivariate exploration of Right Ventricular End-Diastolic Volume (RVEDV) based on gender and patient count. The analysis distinctly reveals that the female anatomy tends to exhibit a smaller volume compared to male anatomy. This nuanced understanding is pivotal in appropriately labeling Right Ventricular (RV) involvement and aligns with findings from prior studies. The lower section of the figure depicts the scatter plot representing the relationship between Non-Right Ventricular End-Diastolic Volume (NRVEDV) and Indexed Cardiothoracic Surface Volume Difference (ICTSVD). This analysis serves to identify outliers, aiding in the exclusion of patients with significantly distorted values. The use of these graphs facilitates the establishment of acceptable ranges for each variable without adversely impacting the algorithms.

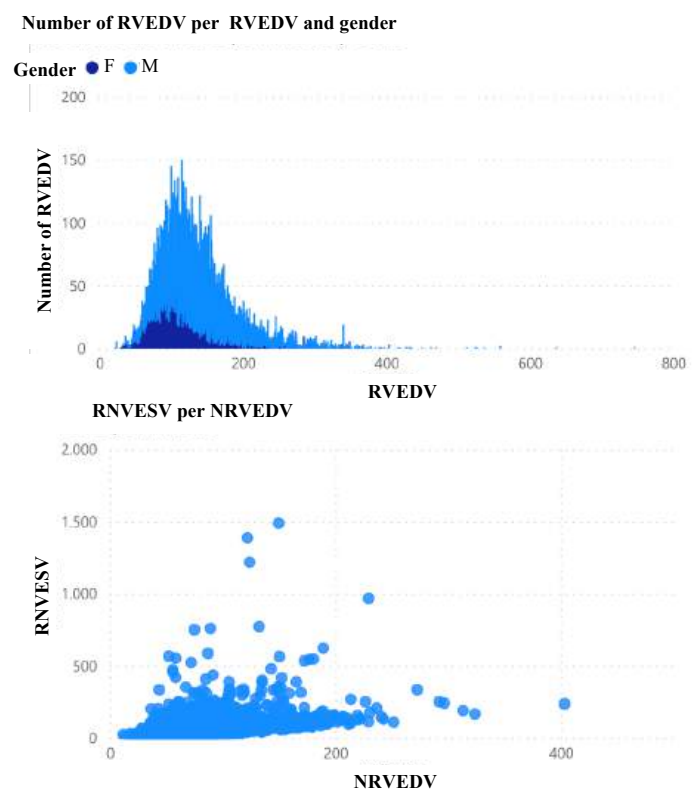


Figure 4. Multivariate histogram of the number of patients, (top) RVEDV by gender, and scatter plot of NRVEDV with RNVESV.

Furthermore, the interactive nature of the graphs streamlines the process of filtering out outliers, allowing for consensus-building with medical professionals regarding their exclusion based on clinical criteria and predefined ranges for each variable. This iterative procedure is replicated for all variables employed in the data labeling process.

In accordance with PVR [Wood], as illustrated in Figure 5, the data distributions are notably concentrated within the ranges of 12 to 18 [Wood]. This concentration validates PVR as a pertinent variable meriting consideration in the research. Such focus within a specific range facilitates the exclusion of outlier patients, particularly those with values exceeding 6. This deliberate exclusion is crucial for preventing algorithmic confusion during predictions and ensuring the reliability of the research findings.

C. Key influencer in Power BI

Power BI services have powerful features such as key influencer visualisation that allow you to detect and understand the factors that drive results from the database [40]. Figure 6 provides the relationships of the two most influential variables to understand the key concepts, in this case corresponding to the PVR variable in wood units and the Aortic Regurgitation Volume by Beat-Volume Difference (Ao.reg.Vol.beat.vol.diff) or also known as ARVBVD. Technically, Power BI uses decision trees that evaluate the impact of variables on the target metric to identify and rank key factors. Determining

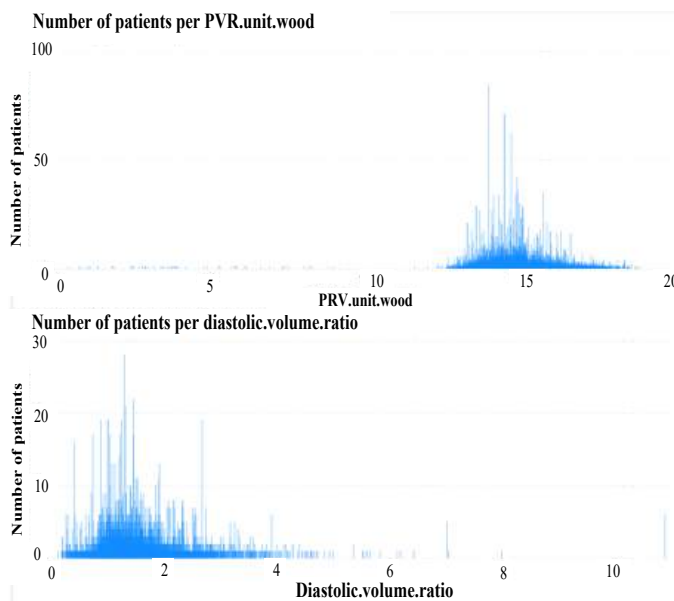


Figure 5. Multivariate histograms of the number of patients, (top) as a function of PVR [Wood], (bottom) as a function of diastolic volume ratio.

that the behaviour of a patient with Aortic Regurgitation Volume greater than 33 applies a factor of 0.59 on each PVR unit, if it is less than -34 it affects the same. Finally, Aortic Regurgitation Volume less than -46 applies a factor of 0.34 in each PVR unit. Both variables, when calculated in tree are similar to those determined in RapidMiner in Figure 9.

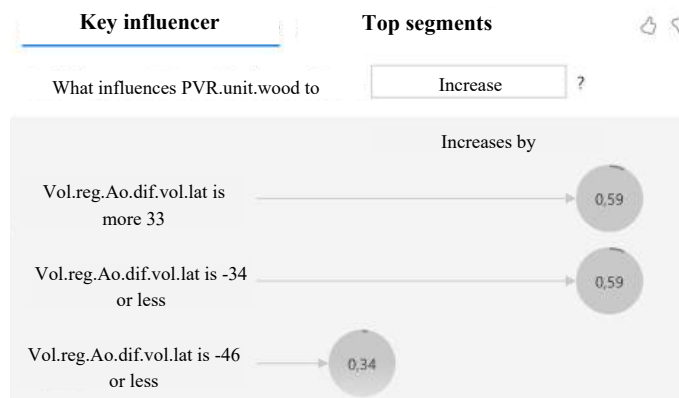


Figure 6. Influencing factors as a function of two key variables

Similarly, there is a cluster analysis tool that allows the identification of natural groups or clusters within the dataset providing a clearer structure of the internal organisation [41]. Figure 7 represents three clusters detected on the two most influential variables detected in Figure 9. The first cluster shows how ARVBVD values in the upper ranges $\pm 100 - 40$ with PVR values below 10 [Wood] are quite dispersed. If we decrease the ARVBVD range to $\pm 40 - 20$ and values between 10-30 [Wood] of PVR it starts to converge. The third cluster converges, finding the ARVBVD range at $\pm 20 - 0$ and values above 30 [Wood] PVR, this means a severity in patients with

such values because it means an elevated pulmonary pressure with a low volume difference.

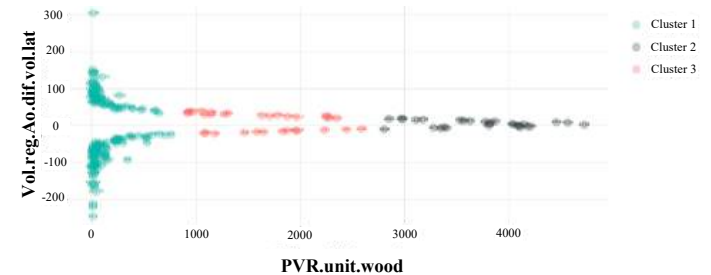


Figure 7. Automatic cluster detector

Another technique for pattern identification is Principal Component Analysis (PCA). The technique allows to reduce the dimensionality of the dataset by identifying the main directions of variability in the data. The 8 shows the two main components, which contribute more than half of the information, assuming $PC1 = 37.98\%$ and $PC2 = 26.10\%$ of the information contribution. These results indicate that the information is concentrated in a few variables, as determined above.

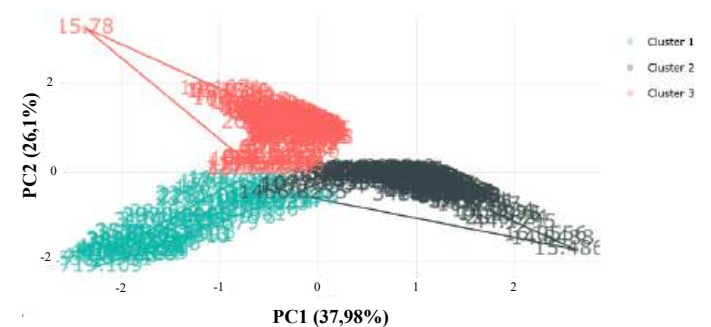


Figure 8. Cluster-based PCA analysis

D. Decision trees in RapidMiner

The decision tree has several advantages. First, as noted above, it is easy to interpret, which facilitates the decision-making process for clinicians. Second, it allows for easy manipulation of the data. Third, it excels in both speed of execution and efficiency of design. However, a notable drawback is its relatively low predictive power.

The solution to this problem is to perform pruning to optimise the best decision tree. However, care must be taken to avoid overfitting, as excessive pruning can lead to this problem. The hyperparameters are carefully adjusted to mitigate the risk of overfitting, and the analysis reveals no observable cases of overfitting in the decision tree model.

According to the algorithm, the patient can be classified with an AUC (91.2%) for RV involvement, based on the interpretation of the provided tree, Figure 9:

- RV involvement (Abnormal): If $PVR.en.units.wood > 15.221 + Height > 1.40m + DoA.PA > 0.087 + Aortic.arch > 14$.

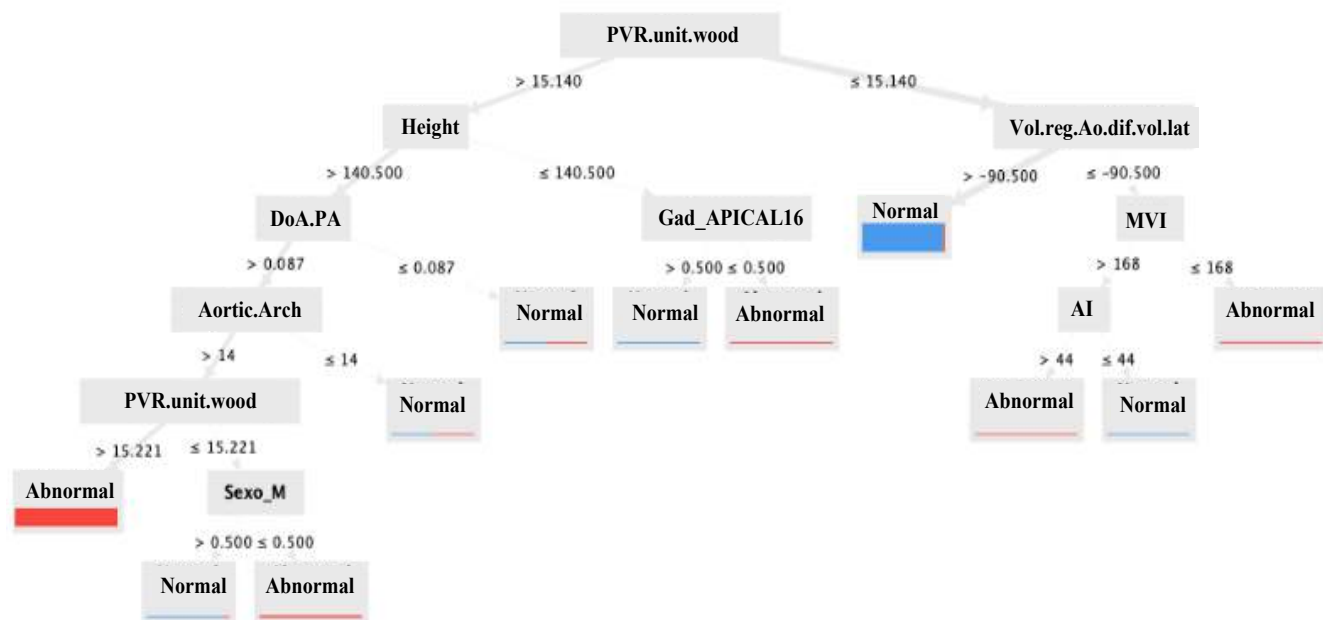


Figure 9. Decision tree pruned in RapidMiner.

- No RV involvement (normal): If $PVR.en.units.wood \leq 15.221$ and $Ao.reg.Vol.beat.vol.dif > -90.500$.

With this philosophy, decision trees are generated to generate visualisations that follow the branches to generate those visualisations on the medical side. On the technical side we use more complex but less visual algorithms to fit higher quality predictor data. Another important variable in the decision tree is Aortic Regurgitation Volume by Beat-Volume Difference ($Ao.reg.Vol.beat.vol.dif$).

The following tree would be a similar interpretation, if we remove the variable $PVR.in.wood.units$, the variable $reason.of.diastolic.volumes$ stands out, but the accuracy is reduced to 66.6% in the decision tree and 78.03% in the Gradient Boosted Trees.

If we include the variables gender and age, which were not initially included because they are used to classify the DV affectations, the results do not change. However, if we prune the tree, the variables age and gender are included, but they are not decisive, and a test with grouping by decades of age is carried out to ensure that they win.

The variables related to left ventricular involvement have been incorporated into the analysis. However, the algorithm does not recognize them as primary determinants for right ventricular involvement. It is essential to acknowledge the interplay and mutual influence between these variables, even though the algorithm does not designate them as key elements specifically for right ventricular involvement. This underscores the complexity of the interactions between left and right ventricular aspects, warranting further investigation into their collective impact on cardiac health.

Previous studies with ML algorithms in cardiology achieve

predictions greater than 90%, as they focus on achieving the best results, not on their interpretability [42]. It is interesting to see that our study achieves close values with interpretability by combining both tools (RapidMiner and Power BI). The results obtained are in line with recent research, which states that neural networks are the most predictive of cardiac parameters [43] [44].

This investigation has some relevant points. The main one is the creation of a tool to support clinical diagnosis that cardiologists can use for the prescription of new tests and a more detailed follow-up, similar to previous experiences already carried out [28] [45]. A second point is the creation of a visual interface, which allows dynamic monitoring, which facilitates dynamic interpretation, generating reports of high statistical value. It is worth exploring new algorithms to improve the interpretation of those key factors in the involvement of the right ventricle, thus improving the possible diagnosis. Tests could be carried out with different databases to create an algorithm that is robust enough to be able to limit any bias that the database used may contain. On the other hand, the authors aim to create a standardized protocol of measurements and tests that is carried out in daily clinical practice. The benefits of data analysis in cardiology using these types of techniques are evident, allowing them to increase the quality of diagnosis, prognosis and therapy.

IV. CONCLUSION

The ML algorithms and decision trees presented in this research show a remarkable ability to discern the most influential variables associated with right ventricular involvement, performing this task with a concise set of parameters. This

efficiency translates into the potential reduction of both the number of diagnostic tests and their corresponding durations, facilitating expeditious interventions in cases of ischaemic and non-ischaemic cardiomyopathy. Consequently, the primary objective of identifying clinical parameters that influence right ventricular involvement in these specific cardiomyopathies is successfully met.

The findings highlight the substantial influence of pulmonary vascular resistance and the difference in aortic artery beat volume, serving as pivotal factors. Notably, the XGBoost algorithm within decision trees attains an AUC of 87.3%. Additionally, variables such as height (BMI), DoA.PA, and Ao.reg.Vol.beat.vol.dif emerge as influential factors, contributing significantly to the predictive of the model.

Acknowledgements

The research is funded by the European University through the project "Identification of factors influencing right ventricular involvement in ischaemic and non-ischaemic cardiomyopathy" through the project 2022/UEM18. The authors of this study wish to express their gratitude to ASCIRES Biomedical Group for the confidentiality agreement reached with the European University; Silvia Ruiz-España (Universitat Politècnica de València) for the documentation compiled from the database; to the advice of the interdisciplinary working group of Machine Learning Health-UEM; as well as to many other researchers for being a source of inspiration in the convergence of technology applied to health.

REFERENCES

- [1] C. Barroso-Moreno, H. Espinos-Morato, E. Puertas, J. J. Beunza-Nuin, J. V. Monmeneu, D. Moratal, and M. P. López-Lereu, "Identifying key factors in right ventricular involvement in ischaemic and non-ischaemic cardiomyopathies," in *HEALTHINFO 2023, The Eighth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing*. IARIA, 2023, pp. 1–8.
- [2] O. W. Health, "The top 10 causes of death in the world." OWH, 2020. [Online]. Available: <https://www.who.int/es/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- [3] E. Boot, M. S. Ekker, J. Putaala, S. Kittner, F. E. De Leeuw, and A. M. Tuladhar, "Ischaemic stroke in young adults: a global perspective," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 91, no. 4, pp. 411–417, 2020.
- [4] L. Cavigli, M. Focardi, M. Cameli, G. E. Mandoli, S. Mondillo, and F. D'Ascenzi, "The right ventricle in "left-sided" cardiomyopathies: the dark side of the moon," *Trends in Cardiovascular Medicine*, vol. 31, no. 8, pp. 476–484, 2021.
- [5] S. Dev, H. Wang, C. S. Nwosu, N. Jain, B. Veeravalli, and D. John, "A predictive analytics approach for stroke prediction using machine learning and neural networks," *Healthcare Analytics*, vol. 2, p. 100032, 2022.
- [6] S. Martínez-Agüero, A. G. Marques, I. Mora-Jiménez, J. Álvarez-Rodríguez, and C. Soguero-Ruiz, "Data and network analytics for covid-19 icu patients: a case study for a spanish hospital," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 12, pp. 4340–4353, 2021.
- [7] S. L. Goldenberg, G. Nir, and S. E. Salcudean, "A new era: artificial intelligence and machine learning in prostate cancer," *Nature Reviews Urology*, vol. 16, no. 7, pp. 391–403, 2019.
- [8] K. K. Halim, S. Halim *et al.*, "Business intelligence for designing restaurant marketing strategy: A case study," *Procedia Computer Science*, vol. 161, pp. 615–622, 2019.
- [9] D. Andriansyah and L. Nulhakim, "The application of power business intelligence in analyzing the availability of rental units," in *Journal of Physics: Conference Series*, vol. 1641, no. 1. IOP Publishing, 2020, p. 012019.
- [10] M. Mariani, R. Baggio, M. Fuchs, and W. Höepken, "Business intelligence and big data in hospitality and tourism: a systematic literature review," *International Journal of Contemporary Hospitality Management*, vol. 30, no. 12, pp. 3514–3554, 2018.
- [11] A. Sánchez-Ferrer, H. Pérez-Mendoza, and P. Shiguihara-Juárez, "Data visualization in dashboards through virtual try-on technology in fashion industry," in *2019 IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI)*. IEEE, 2019, pp. 1–6.
- [12] M. Ratia, J. Myllärniemi, and N. Helander, "The new era of business intelligence: Big data potential in the private health care value creation," *Meditari Accountancy Research*, vol. 26, no. 3, pp. 531–546, 2018.
- [13] V. Anand and S. Ghosh, "Sentiment analysis of social network for information professionals: A case study of links discussion forum," *International Journal of Information Science and Management (IJISM)*, pp. 63–74, 2024.
- [14] M. Seočanac and D. Dimitrovski, "Factors influencing tourists' nightlife experience in belgrade," *Consumer Behavior in Tourism and Hospitality*, vol. 17, no. 4, pp. 436–452, 2022.
- [15] J. Fernandes Andry, J. Gunadi, G. Dwinoor Rembulan, and H. Tannady, "Big data implementation in tesla using classification with rapid miner," *International Journal of Nonlinear Analysis and Applications*, vol. 12, no. Special Issue, pp. 2057–2066, 2021.
- [16] B. R. Pagliaro, F. Cannata, G. G. Stefanini, and L. Bolognese, "Myocardial ischemia and coronary disease in heart failure," *Heart Failure Reviews*, vol. 25, no. 1, pp. 53–65, 2020.
- [17] C. R. Greyson, "Ventrículo derecho y circulación pulmonar: conceptos básicos," *Revista Española de Cardiología*, vol. 63, no. 1, pp. 81–95, 2010.
- [18] J. Sanz, D. Sánchez-Quintana, E. Bossone, H. J. Bogaard, and R. Naeije, "Anatomy, function, and dysfunction of the right ventricle: Jacc state-of-the-art review," *Journal of the American College of Cardiology*, vol. 73, no. 12, pp. 1463–1482, 2019.
- [19] S. Wang, H. Wang, M. Ng, Y. Tada, G. Pontone, J. Urmeneta, I. Saeed, H. Patel, C. Mariager, J. V. Monmeneu-Menadas *et al.*, "Quantification of myocardial blood flow using stress cardiac magnetic resonance for the detection of coronary artery disease," *European Heart Journal-Cardiovascular Imaging*, vol. 24, no. Supplement_1, pp. 119–375, 2023.
- [20] J. M. Oakes, J. Xu, T. M. Morris, N. D. Fried, C. S. Pearson, T. D. Lobell, N. W. Gilpin, E. Lazartigues, J. Gardner, and X. Yue, "Effects of chronic nicotine inhalation on systemic and pulmonary blood pressure and right ventricular remodeling in mice," *Hypertension*, vol. 75, no. 5, pp. 1305–1314, 2020.
- [21] J. Ingles, C. Burns, R. D. Bagnall, L. Lam, L. Yeates, T. Sarina, R. Puranik, T. Briffa, J. J. Atherton, T. Driscoll *et al.*, "Nonfamilial hypertrophic cardiomyopathy: prevalence, natural history, and clinical implications," *Circulation: Cardiovascular Genetics*, vol. 10, no. 2, p. e001620, 2017.
- [22] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [23] S. K. Murthy, "Automatic construction of decision trees from data: A multi-disciplinary survey," *Data Mining and Knowledge Discovery*, vol. 2, pp. 345–389, 1998.
- [24] T. Ho-Kam, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1. IEEE, 1995, pp. 278–282.
- [25] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data mining*, 2016, pp. 785–794.
- [26] V. Manickam and M. Rajasekaran Indra, "Dynamic multi-variant relational scheme-based intelligent etl framework for healthcare management," *Soft Computing*, vol. 27, no. 1, pp. 605–614, 2023.
- [27] M. W. Browne, "Cross-validation methods," *Journal of Mathematical Psychology*, vol. 44, no. 1, pp. 108–132, 2000.
- [28] J. J. Beunza, E. Puertas, E. García-Ovejero, F. Villalba, E. Condes, G. Koleva, C. Hurtado, and M. F. Landecho, "Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease)," *Journal of Biomedical Informatics*, vol. 97, p. 103257, 2019.
- [29] T. Lachev and E. Price, *Applied Microsoft Power BI Bring your data to life!* Prologika Press, 2018.

- [30] V. Kotu and B. Deshpande, *Predictive analytics and data mining: concepts and practice with rapidminer*. Morgan Kaufmann, 2014.
- [31] A. M. Maceira, S. K. Prasad, M. Khan, and D. J. Pennell, "Reference right ventricular systolic and diastolic function normalized to age, gender and body surface area from steady-state free precession cardiovascular magnetic resonance," *European Heart Journal*, vol. 27, no. 23, pp. 2879–2888, 2006.
- [32] S. E. Petersen, M. Y. Khanji, S. Plein, P. Lancellotti, and C. Bucciarelli-Ducci, "European association of cardiovascular imaging expert consensus paper: a comprehensive review of cardiovascular magnetic resonance normal values of cardiac chamber size and aortic root in adults and recommendations for grading severity," *European Heart Journal-Cardiovascular Imaging*, vol. 20, no. 12, pp. 1321–1331, 2019.
- [33] J. M. Gardin, M. K. Rohan, D. Davidson, A. Dabestani, M. Sklansky, R. Garcia, M. L. Knoll, D. White, S. K. Gardin, and W. L. Henry, "Doppler transmitral flow velocity parameters: relationship between age, body surface area, blood pressure and gender in normal subjects," *American journal of Noninvasive Cardiology*, vol. 1, no. 1, pp. 3–10, 1987.
- [34] E. Commission, "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016," European Commission, Tech. Rep., 2021.
- [35] J. Calvo-López and B. Varela-Adsuara, *La adaptación al nuevo marco de protección de datos tras el RGPD y la LOPDGDD*. Wolters Kluwer España ISBN: 978-84-9090-345-2, 2019.
- [36] P. Nyakomitta and S. Abeka, "Security investigation on remote access methods of virtual private network," *Global Journal of Computer Science and Technology*, 2020.
- [37] J. Swan-Harold, W. Ganz, J. Forrester, H. Marcus, G. Diamond, and D. Chonette, "Catheterization of the heart in man with use of a flow-directed balloon-tipped catheter," *New England Journal of Medicine*, vol. 283, no. 9, pp. 447–451, 1970.
- [38] C. D. Vizza, I. M. Lang, R. Badagliacca, R. L. Benza, S. Rosenkranz, R. J. White, Y. Adir, A. K. Andreassen, V. Balasubramanian, S. Bartolome *et al.*, "Aggressive afterload lowering to improve the right ventricle: a new target for medical therapy in pulmonary arterial hypertension?" *American journal of Respiratory and Critical Care Medicine*, vol. 205, no. 7, pp. 751–760, 2022.
- [39] A. García-Alvarez, L. Fernandez-Friera, J. G. Mirelis, S. Sawit, A. Nair, J. Kallman, V. Fuster, and J. Sanz, "Non-invasive estimation of pulmonary vascular resistance with cardiac magnetic resonance," *European Heart Journal*, vol. 32, no. 19, pp. 2438–2445, 2011.
- [40] M. Ehrenmueller-Jensen and M. Ehrenmueller-Jensen, "Discovering key influencers," *Self-Service AI with Power BI Desktop: Machine Learning Insights for Business*, pp. 53–80, 2020.
- [41] B. Powell, *Mastering Microsoft Power BI: expert techniques for effective data analytics and business intelligence*. Packt Publishing Ltd, 2018.
- [42] T. Smole, B. Žunković, M. Pičulin, E. Kokalj, M. Robnik-Šikonja, M. Kukar, D. I. Fotiadis, V. Pezoulas, N. S. Tachos, F. Barlocco *et al.*, "A machine learning-based risk stratification model for ventricular tachycardia and heart failure in hypertrophic cardiomyopathy," *Computers in Biology and Medicine*, vol. 135, p. 104648, 2021.
- [43] P. Revuelta-Zamorano, A. Sánchez, J. L. Rojo-Álvarez, J. Álvarez-Rodríguez, J. Ramos-López, and C. Soguero-Ruiz, "Prediction of healthcare associated infections in an intensive care unit using machine learning and big data tools," in *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016: MEDICON 2016, March 31st-April 2nd 2016, Paphos, Cyprus*. Springer, 2016, pp. 840–845.
- [44] R. García Carretero, L. Vigil-Medina, O. Barquero-Perez, I. Mora-Jimenez, C. Soguero-Ruiz, and J. Ramos-Lopez, "Machine learning approaches to constructing predictive models of vitamin d deficiency in a hypertensive population: a comparative study," *Informatics for Health and Social Care*, vol. 46, no. 4, pp. 355–369, 2021.
- [45] A. Hernández-Casillas, S. Del-Canto, S. Ruiz-España, M. P. López-Lereu, J. V. Monmeneu, and D. Moratal, "Detection and classification of myocardial infarction transmuralities using cardiac mr image analysis and machine learning algorithms," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 1686–1689.

Developing a Process for Gathering and Managing Multi-site User Feedback for a Data Collection Project in a Rare Disease Setting. Using Design Science to Explicate a Knowledge Contribution

Sinéad Impey, Jonathan Turner, Frances Gibbons,
Anthony Bolger, Gaye Stephens, Lucy Hederman,
Ciara O'Meara, Ferran De La Varga, John Kommala,
Matthew Nicholson, Daniel Farrell, Emmet Morrin,
Miriam Galvin
ADAPT Centre
Trinity College Dublin
Dublin, Ireland
e-mail: simpey@tcd.ie, jonathan.turner@tudublin.ie,
frances.gibbons@adaptcentre.ie,
anthony.bolger@adaptcentre.ie, gabe.stephens@tcd.ie,
hederman@tcd.ie, ciara.omeara@adaptcentre.ie,
ferran.delavarga@adaptcentre.ie,
john.kommala@adaptcentre.ie,
matthew.nicholson@adaptcentre.ie,
daniel.farrell@adaptcentre.ie, morrinem@tcd.ie,
galvinmi@tcd.ie

Mark Heverin, Éanna Mac Domhnaill, Robert
McFarlane, Dara Meldrum, Deirdre Murray, Orla
Hardiman
Academic Unit of Neurology
Trinity College Dublin
Dublin, Ireland
e-mail: mark.heverin@tcd.ie, amacdomh@tcd.ie,
mcfarlor@tcd.ie, meldrumd@tcd.ie, dmurray1@tcd.ie,
hardimao@tcd.ie

Abstract— Large datasets are required to understand disease progression, investigate treatment options and discover potential cures in rare neurological conditions such as Amyotrophic Lateral Sclerosis. Generating large datasets for such rare conditions requires the participation of multiple specialist clinical sites. The Precision ALS project is a partnership between multiple specialist Amyotrophic Lateral Sclerosis clinical sites (n=10) and industry partners across Europe that seeks to collect and analyse multi-modal data collected from participants with the disease. The project is managed by an expert group comprised of clinical, technical, business, legal and interdisciplinary specialists. Adopting a design science approach, the project created an artefact – a data collection tool. This tool is used by data collectors at each specialist clinical site to capture a range of patient information, including biological and socio-economic data. Applying an iterative approach, the initial user requirements were based on extensive collaborative projects undertaken by clinical sites, which formed the basis of a bespoke worksheet. Additional modifications were introduced through project group discussions and engaging members from one of the clinical specialist sites. The lessons learned from this initial work were formulated into a knowledge contribution in the form of a process for integrating pre-existing paper-based data collection processes, additional requirement gathering and managing user feedback. The process is a way to identify and manage the early digital representation of data gathering paper worksheets, along with the volume and variety of unstructured feedback generated during a diverse range of multi-site user engagements. A central component of the process is three themes that were adapted from the initial feedback and presented as three pillars: Digital Worksheet; Usability; and

Process/actors. These three pillars were embedded in a four-step information pipeline (capture, record, review and target), which provided a structure to ensure that the management of this feedback was transparent and auditable from source to decision. This paper describes the knowledge contribution of the user requirements gathering process explicated through a design science approach targeted towards the data collection tool resulting from this process.

Keywords- amyotrophic lateral sclerosis; design science; knowledge contributions; requirements gathering; data collection.

I. INTRODUCTION

This paper expands on work presented at HealthInfo23 on the development of a data collection tool (DCT) for Amyotrophic Lateral Sclerosis (ALS) [1]. ALS is an incurable progressive neurodegenerative disease responsible for up to 10,000 deaths per year in Europe; it is the most common form of the motor neurone diseases [2][3]. ALS is a complex genetic disease. Approximately 15% of patients have a family history of ALS or frontotemporal dementia. Of these, approximately 70% are associated with known genetic variants, the commonest being mutations in C9orf72, SOD1, FUS and TDP 43 [4][5]. Of the remaining 85%, the pathogenic mechanisms are poorly understood, with at least 30 at-risk genes associated with ALS [6]. To further elucidate that disease heterogeneity, and to identify subgroups of patients with shared clinical and pathogenic features that would be suited to targeted drug therapies,

collaboration between clinicians and data scientists is required in the collection, curation and analysis, including by machine learning methods [7]. However, generating such large datasets for a rare disease like ALS is challenging due to the low numbers of affected individuals. In response to this challenge, the Precision ALS (P-ALS) project [7][8] was initiated as a partnership between ten clinical sites in Europe, all part of the TRICALS [9] network of 48 specialised ALS sites, interested industry partners, and technical and clinical researchers.

A complexity of combining data from many sites is in ensuring consistency in data collection. A solution chosen for the data collection strand of the P-ALS project is to use a common data collection tool (DCT) at each site. The tool contains a common set of questions across all collection sites and, where appropriate, a defined list of valid responses. These agreed “core data elements” were developed in paper format at TRICALS sites in the context of extensive previous collaborations over a 15-year period. These were provided to the ADAPT Centre [10] as a structured paper ‘worksheet’ to be used as the basis for the development of the DCT. The research adopted a lightweight Agile approach [11].

This paper describes research undertaken as part of the initial development stages of the DCT in the P-ALS project. The purpose of this research was to explicate the processes required to refine and stress test the suitability of the worksheet, to evaluate usability of the DCT, and to support the TRICALS sites collect data.

Following this introduction, the paper is organised as follows. Section II presents background to the research. Section III describes the methodology – elaborated action design research (eADR). Section IV describes application of eADR activities in this work. Section V presents some of the feedback generated from user engagements. Section VI discusses our conclusions and plans for future work. The acknowledgment and references close the article.

II. BACKGROUND

Wicked problems are complex, multifaceted issues that lack clear definitions and have a high degree of ambiguity [12]. To address the wicked problem described in this paper, an elaborated action design research (eADR) approach [13] was adopted. This provided a structured approach to conducting research in information systems and related fields that not only produced a design artefact but also design knowledge. A core component of eADR is the combination of action research and design science [14]. This combination results in an ensemble of knowledge contributions.

eADR is a four-stage method – diagnosis, design, implementation, and evolution – that emerged from design science roots [14][15] and is illustrated in Figure 1. For this paper, only the first stage of eADR, diagnosis, is described.

According to Mullarkey and Hevner [14], diagnosis is problem-centered. Each stage of the eADR can take several cycles to complete, depending on research requirements. Each eADR stage comprises five activities: Problem Formulation/Action Planning (P); Artefact Creation (A); Evaluation (E); Reflection (R); and Formalisation of Learning (L). For each stage, an artefact is created to address

the problem formulated. Artefact creation is an important part of design science research [14].

The knowledge contributions of this work are a DCT artefact and a ‘Three Pillar Information Pipeline’ that embeds a process for gathering and managing feedback from multiple sources that could potentially be targeted towards user requirements. This paper reports on the development and evaluation of the information pipeline explicated through a design science approach and targeted towards the data collection tool resulting from this process.

III. METHODOLOGY

The P-ALS principal investigator (PI), the leading clinical academic at the National Specialist ALS Centre, is based in Ireland. The DCT project sub-group comprised a clinical subject expert group from Beaumont National Centre & Academic Neurology Trinity College Dublin and a technical development group from Trinity College Dublin. Some members were jointly members of both groups; to differentiate these members, they will be referred to as the ‘interdisciplinary’ group in this paper. In addition to the project group, and the individual sites, there was a data collectors’ group. Data collectors are responsible for using the DCT at each partner site to capture participant information. There is at least one designated data collector at each of the ten sites. The legal and financial expertise necessary to ensure the success of the project was provided by a separate group.

Development of the DCT was in two concurrent streams. The first stream was concerned with gathering technical requirements through review of the previously agreed TRICALS “master worksheet” along with additional discussions and technical workshops with the project group and one partner site. Acting as a test site, it was chosen as it was linked to the project PI and was geographically convenient for the majority of the group. These requirements were subsequently reviewed and refined at the other sites (n=9) during the project. Requirements refinement focused on the needs of clinicians, data collectors and analysts. These requirements informed the iterative development of the DCT by the technical team.

The second stream was concerned with the refinement of the previously generated worksheet to ensure data collected was accurate and capable of being collated into a single dataset. It was important to ensure data collected was capable of being collated into a single dataset, to ensure future cross-site analysis of collected data.

A number of workshops occurred throughout the initial stage of the project to gather additional requirements for both the DCT and the worksheet. These workshops were facilitated by and involved P-ALS team members from the ten data collection sites. Feedback was recorded and reviewed after each workshop.

An early challenge that emerged during the research was the volume of unstructured information that had been implicit in the originally generated TRICALS paper-based collection processes. This was reflected in the structural heterogeneity across sites involved in the project (n=10) along with the variety of stakeholders across a range of

TABLE I. APPLICATION OF EADR TO THIS RESEARCH

| Activity | Performed in this research |
|---|---|
| Problem formulation/ action planning (P) | Managing the potential high volume and variety of feedback captured from a range of stakeholders including ten different geographical sites that could translate into user requirements so that nothing is overlooked could be challenging. |
| Artefact creation (A) | An information pipeline to manage feedback so that it is auditable and ensures feedback reaches its target. |
| Evaluation (E) | The pipeline was developed based on group discussions and engagement with one site. To evaluate its utility, it was used as the additional sites. |
| Formalisation of learning (L) | Lessons learned were formulised into an illustration of the evaluated pipeline. |
| Reflection (R) | Reflection was continuous throughout the process during project group discussions. |

interests including design/development of the DCT and analysis of data captured. The challenge of gathering additional user requirements from ten individual sites, and with a range of stakeholders, displays the characteristics of a ‘wicked’ problem: it is ‘ill-formulated’, information is ‘confusing’, with many stakeholders often holding conflicting perspectives [14][15].

Design of a bespoke digital data collection tool, based on the TRICALS worksheet, was carried out in-house in the ADAPT Centre, with the development team based at Trinity College Dublin. This team met regularly with clinicians and data analysts to ensure that the developed tool met their needs. A tablet-based application approach was chosen to ensure portability and to enable operation without a working internet connection. This enabled data collection to be performed using the tablet at locations without wi-fi internet connection, which could include some participants’ homes. The data collection form structure implemented in the tablet follows the worksheet structure and content developed by TRICALS sites and is configured using a metadata-driven approach, where the user interface is described by metadata rather than hard-coded, allowing easy updates without the need to modify the application code itself. Development followed a lightweight Agile [11] approach with regular prototype releases to project stakeholders. The application is deployed via a mobile device management solution to minimise security and device management concerns. Using Android with Mobile Device Management software provides the mechanism to distribute private apps and client certificates and allows for restricted and secure access to the server. Android provides a more open and accessible development platform than Apple and iOS, which does not provide a distribution mechanism for the small scale required.

The four stages of eADR as applied to this research are shown in Figure 1. Table I describes each step of an eADR stage and describes how each step was carried out in the Diagnosis stage.

IV. APPLICATION OF EADR ACTIVITIES

For each stage of the eADR five activities are performed. This section discusses the activities performed during this stage (Diagnosis) and the outcome.

A. Problem Formulation (P)

The first activity surrounded problem formulation/action planning (P). Problems are formulated by reviewing the learning from previous stages [14]. As Diagnosis was the first stage performed in this research, the problem was developed based on early group discussions with the P-ALS group members and following initial project workshops. From this work, it was observed that due to the high number of stakeholders (including ten participating sites from different European countries), a potentially large volume of information could be generated and transformed into user requirements.

B. Artefact Creation (A)

Artefacts are created to address the problem formulated. Artefact creation is an important part of design science research [14]. Throughout the creation process, the abstract nature of the artefact created will develop. To manage the feedback generated during workshops and discussions, an artefact in the form of an information pipeline was proposed. The purpose of this artefact was to ensure that all feedback was recorded and reviewed, and to facilitate the efficient flow of information from the source to an appropriate target

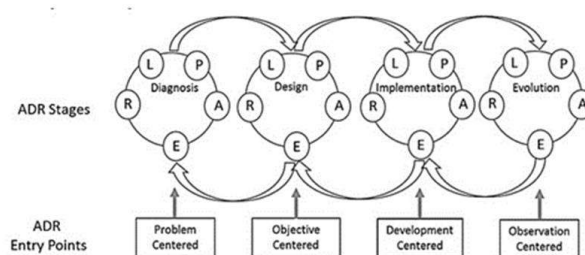


Figure 1. The four stages of elaborated action design research [14].

so that a decision could be made. To do this, a proposed route from information identification to categorisation to reaching an appropriate target was constructed by the project group. To categorise the information in the artefact, outputs from the early work completed (paper worksheets, workshops and group discussions) and feedback from one of the ten sites was analysed and three broad themes emerged – feedback relating to the original worksheet, to the usability of the technology, and to the data collection process including actors involved in those processes. These themes were reviewed by the project group and refined into ‘Three Pillars’: Digital Worksheet, Usability, and Process/actors. Each pillar represents a broad area that was explored to identify user requirements.

Pillar 1, “Digital Worksheet”, captured feedback related to the early digital format of the TRICALS master worksheet. These could be technical or clinical. Pillar 2, “Usability”, focused on user experience requirements. Pillar 3, “Process/actors”, investigated requirements relating to the flow of data into the collection tool from the various data sources that contribute to the final dataset, and the actors required to perform these tasks.

In the pipeline, all feedback received was given a unique reference code and assigned to a pillar, and the nature, provenance and outcome of the submission was recorded. This research uses the term ‘Four-step Information Pipeline’ to describe the pipeline in its entirety.

C. Evaluation (E)

The eADR process has roots in action design research (ADR). A central component of ADR is to build, implement and evaluate a relevant artefact [15]. Building and implementing in eADR are presented as individual stages (design and implementation, respectively), whereas evaluation is the activity of all stages. The purpose of the evaluation is to ensure artefacts created meet their intended purpose. The ‘Four-step Information Pipeline’ artefact was created based on project group discussions and workshops but only captured feedback from one of the ten sites. To ensure its suitability across all sites, the pipeline was used to manage feedback from the online and in person site engagements at all sites.

User feedback was identified in three ways. Firstly, prior to the in-person site visits, data collectors at each site were given an evaluation version of the DCT and instructed to enter dummy data. These data were reviewed by the project team for any anomalies. Secondly, lessons learned from previous TRICALS projects in which data were gathered using paper-based format were applied. These previously captured datasets (comprising clinical research from 22,000 patients) is referred to as extant data. Finally, in-person visits were arranged following online engagement meetings with each site. Through the initial online site engagement meetings, variations in the data collected or the interpretation of the question was captured and included in the information pipeline for review by the project group. These discussions also determined how ready each site was to ‘go live’, so that, if necessary, a suitable remedy or interim measure could be initiated so that the site could move toward collecting data in

digital format using the collection tool. A digital document was used to record relevant information using a set of headings that emerged as the project progressed. The final set of headings are shown in the following text:

Pillar 1 (Digital Worksheet) – highlighted any known challenges related to the question set (or worksheet) that needed to be addressed before a site could go live.

Pillar 2 (Usability) – highlighted any known challenges related to the usability of the DCT that needed to be addressed before a site could go live.

Pillar 3 (Process/Actors) - identified a data collection process for the site and actors required to collect data. Under this heading any challenges to either the process or actors that could be a barrier to go live were identified, addressed or a suitable interim step was proposed.

Ethics - under this heading it was established if ethics was completed and approved, in progress, or not started.

Governance - under this heading it was established if other legal agreements required (such as data sharing agreement, privacy statement) were completed and approved, in progress, or not started.

Collection Tool Provided - information captured under this heading related to whether a site had received a DCT and that all the necessary security measures were in place.

Data collector in situ/contact details - this noted if a data collector was in post at the site. Potential answers were Yes, in post; Post filled awaiting staff member; No. Reasons for ‘no’ could include that the job had to be advertised or the candidate had not taken up the position. It was important that the site visits included meeting with the data collector as a large part of the site visit surrounded providing training.

A ‘traffic-light’ system was employed to indicate site readiness. When the outcomes for all headings were positive (i.e., all ‘Yes’), the site was deemed ‘Green’ and an in-person visit to the site was arranged. If readiness was imminent the site was deemed ‘Amber’ and the project group focused on addressing known challenges. Some sites required extensive clarifications prior to formal ethics approval, so were deemed ‘Red’ as no data collection could be initiated. The Site Readiness for Go Live document was frequently reviewed to track progress for each site.

Using this Site Readiness for Go-Live document allowed The DCT project sub-group members to access a high-level overview of all sites and quickly identify challenges that individual sites were having.

The site visit following a ‘Green’ decision followed a protocol that was constructed to ensure a standardised approach across all sites. This protocol reflected the process used at the first site. This protocol included the length of time of the visit and who would be attending. Attending each site visit were two members of the project team, who met with, at minimum, a data collector from the site visited. Each site was encouraged to include as many stakeholders as they wished in these visits, but it was imperative that the data collectors were available. The visit took place over two days to ensure that the site was not over-burdened, being mindful that clinical areas are busy environments.

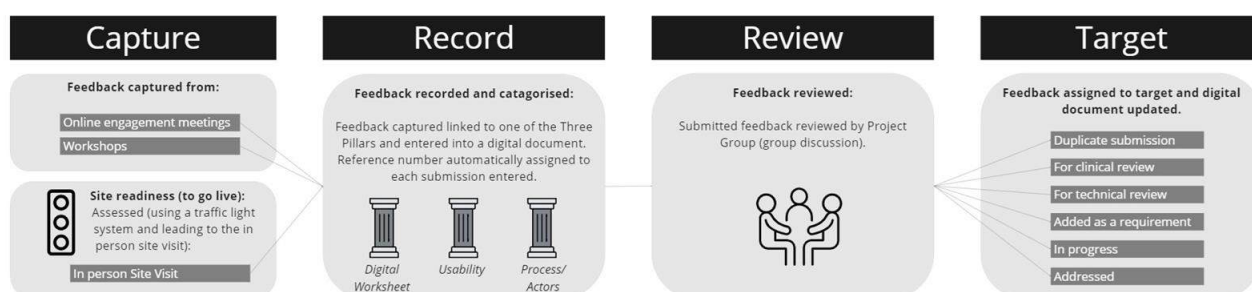


Figure 2. Four-step information pipeline.

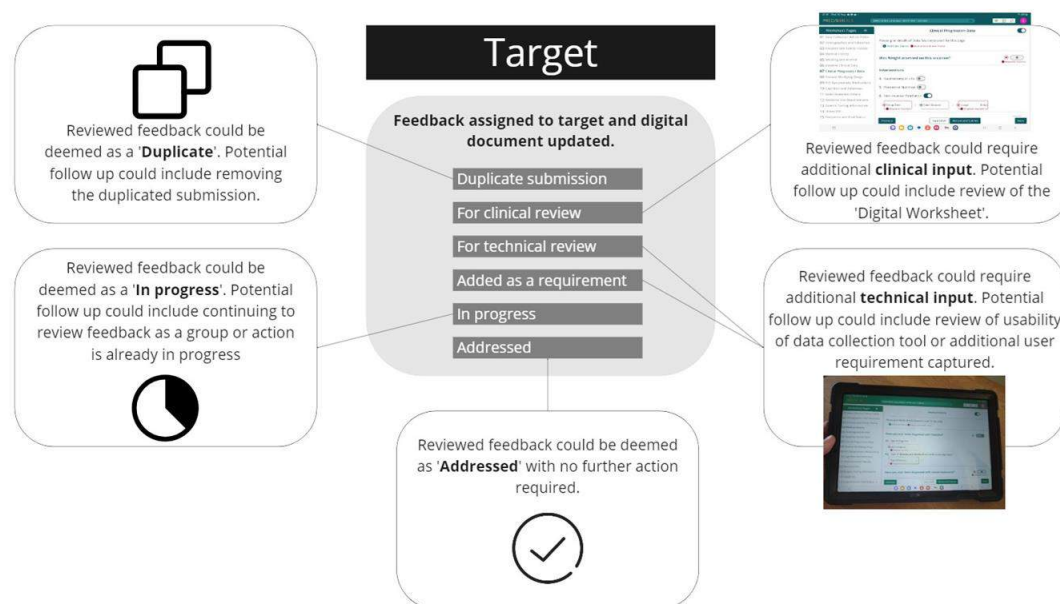


Figure 3. Potential follow ups once user feedback has been assigned to a target

It was envisaged that once each site visit had been completed satisfactorily, the site could 'go live' with participant data collection, so it was important that no barrier remained unaddressed. Feedback captured during the visit was entered into the 'Four-step Information Pipeline'.

A major focus of each site visit was a mock data collection interview. This interview was conducted by one researcher who acted as the patient participant and the data collector, while the other researcher took notes.

The purpose of this interview was fourfold:

1. Understanding: The researcher and data collector ensure all questions are fully understood so that the appropriate data could be collected accurately (related to pillar 1).
2. Training: As a training exercise to ensure the data collector can use the DCT (related to pillar 2) and is comfortable performing an interview with a real patient.

3. Discuss potential data collection process and actors required (related to pillar 3)

4. Generate new user requirements: The interview is observed by a second researcher who takes notes.

During the interview, the data collector at the study site used the DCT to capture the interview responses (from the researcher) and was encouraged to discuss their thoughts on the interpretation of the question, what type of answer they might expect and how the response could best be captured in the tool. For example, one aim of the mock interview was to ensure that the data collector is comfortable with the wording of the questions and understands the rationale and meaning behind them and to provide training if any queries arose. This will help to ensure the data collected is relevant and that the collector understood the rationale for the questions. Along with this training, the purpose of this exercise was to check the contents of the digital worksheet

to identify any site-specific issues that could limit the ability of the project to combine data from the ten individual sites.

The three pillars were used to structure the site visit discussion. While the three pillars are presented as separate entities, during the site visits it became apparent that there can be an overlap in discussion. For instance, during the mock interview the discussion might include the type of data (relates to pillar 1), how it could be captured in the data collection tool (related to pillar 2), the potential source of the data and how/who collected it (related to pillar 3).

The notes gathered during the site visits were first reviewed by the research group members involved in the mock data collection interview to ensure they accurately represented what was discussed. At the end of the visit, the team further reviewed all the feedback gathered to ensure all points generated were captured. This validated feedback was then submitted to a digital spreadsheet and further reviewed in a post-site-visit meeting with the project manager, clinical group and technical group for a decision (to include as a user requirement) or for further discussion.

In addition to discussions, during the mock data collection interview, the data collector was observed to see how they engaged with the DCT. This observation included noting when they reverted to use of paper to record information for later data entry.

D. Formalisation of learning (L)

According to Mullarkey and Hevner [14], the lessons learned during the development of an artefact should be formalised into design knowledge. The pipeline proposed has four parts to managing feedback. These are capture, record, review and assign to a target. Ways to capture feedback included workshops, online engagement meetings and in person site visits. The in-person visit adopted a mock data collection interview to review the usability of the tool and questions it contained, along with reviewing the data collection process used at each site. The feedback was recorded in a digital document and automatically assigned a reference number and from this it was reviewed by the project group and assigned to a target (for instance, 'for clinical review' or 'for technical review').

To date five out of ten site visits have been completed with ongoing online site engagement. No major changes were required to the pipeline, apart from refinements to wording. For example, initially the worksheet pillar was referred to as 'dataset/worksheet'. 'Dataset' was removed as it was thought to refer to any data collected rather than the question set. A further change was the identification of additional targets for the reviewed feedback. These describe who the feedback was directed to. Initially it was envisioned that the targets would align with the pillars. However, from reviewing feedback these categories did not fully capture the type of follow-up the feedback required. Therefore, each reviewed feedback remained assigned to one of the three pillars but was assigned a query type. These are:

Duplicate submission – this referred to any submission that was currently represented in the recorded feedback.

For clinical review – refers to any submission that required additional review by the clinical members of the group before a decision can be made.

For technical review - refers to any submission that required additional review by the technical members of the group before a decision can be made.

Added as a requirement – referred to any submission that was accepted as a user requirement during the review without any further exploratory or information required.

In progress – describes a submission that has been recorded but an answer is not possible during the current review session. This could be for a number of reasons, such as, additional expertise or information is required to reach a decision.

Addressed – This denotes any submissions that are addressed and an outcome recorded.

The need for these query types arose as some queries crossed more than one pillar. It is expected that some of these findings could be incorporated into a set of standard operating procedures to give operational guidance to each site. Based on the outcomes of the evaluation, an illustration, Figure 2, of the pipeline was produced to depict the process from start to finish across the four steps. The final artefact created is a pipeline that is proposed as a way to manage feedback generated in a multi-site healthcare technology development project. The aim of the pipeline is to manage user requirements from initial feedback to final decision while providing an auditable record of all decisions. Figure 3 shows the potential follow-up once user feedback has been assigned to a target.

E. Reflection (R)

In the eADR activities, Mullarkey and Hevner [14] propose a reflection activity that is performed in every stage. The purpose of reflection is to identify learnings that can be incorporated into the developing artefact. In this research, reflection took the form of project group discussions that occurred during research. Members of the project group were clinical, technical and interdisciplinary. An overview of how the activities were applied is shown in Table I.

V. FEEDBACK GENERATED

The next section describes the feedback gathered during online engagement meetings and site visits. It is proposed that this information could be useful for system designers developing data collection tools for a rare disease setting. The text is structured using the 3 pillars.

A. Pillar 1: Digital Worksheet

"Digital Worksheet" relates to the set of questions to be used in the collection tool. The questions used in the digital worksheet were based on successfully conducted collaborative clinical research by the TRICALS sites, funded by European grants (e.g., Joint Programme in Neurodegeneration (JPND), FP7 and Horizon) group.

The aim of this pillar, therefore, was to further review and refine the paper worksheet, to generate a digital version and to identify any potential harmonisation challenges. This work was undertaken to ensure consistency of data collection across the sites and presented an opportunity to ensure the consistency of meaning of the fields contained within the worksheet. Feedback from the site visits raised a number of common queries relating to understanding the digital worksheet and individual questions within the worksheet. These queries included wording issues, for example, whether a question relating to children was intended to capture the number of biological children only, or any and all dependent children including those adopted or from a new partner.

To address this finding, a document with answers to frequently asked questions relating to the digital worksheet was constructed. Once evaluated, the content could be integrated into the DCT. To access this content, data collectors can click on 'information buttons' associated with a question to quickly access information relating to a specific question if required.

This work also evaluated the consistency of meaning of the fields contained within the digital worksheet, to identify the potential range of answers and if standard application programming interfaces (APIs) could be incorporated into the data collection tool.

In addition, the work identified a range of harmonisation challenges.

1) *Numerical data challenges*: Units, particularly currencies; preference for numerical values being recorded as precise values or banding.

2) *Social/cultural challenges*: Societal mores may vary in how acceptable it is to reveal personal information, such as income, in a research study.

3) *Documentation practices*: Assumptions may be made at individual sites that their way of working is documented with sufficient precision, for example, when recording a patient's forced vital capacity (FVC), the maximum amount of air that the patient can exhale, the patient can be standing or sitting for the measurement. Different sites use different positions. The measurement is then recorded as 'FVC sitting', 'FVC lying' or simply 'FVC' (without the patient's position being recorded).

As each harmonisation challenge was identified, it was shared among all data collectors (across all sites) for feedback and to see if it was a site-specific challenge or global.

Feedback gathered during this part of the research included noting variations in how each variable was interpreted by each site, being mindful that this is a multi-national project; that each variable is relevant for data analysis or that there is an analysis impact of the omission of the variable; and what, if any, challenges exist in collating multisite data. Apart from identifying any data variable issues, reviewing the question set also helped ensure data collectors were comfortable with their understanding of the wording of the questions and the rationale behind them. It is envisioned that this will help

improve the quality and consistency of the data collected once routine data collection commences.

B. Pillar 2: Usability

The second pillar was concerned with usability of the DCT. Usability of the tool was focused on clear presentation of required data items and ease of data entry. Prior to the site visit, online site engagement meetings, with representatives from each site, were given a complete overview of project and walkthrough the DCT. Users then had the opportunity to explore the tool independently, with members of the development team available to answer any questions and to capture verbal feedback.

To record a baseline measure of usability of the DCT, 13 users from across all sites completed a user survey. The tool used was version 3 of the Post-Study System Usability Questionnaire (PSSUQ) [16]. The PSSUQ is a 16-point questionnaire that is used to measure users' satisfaction with a product. Questions include, for example 'Overall, I am satisfied with how easy it is to use the system' or 'It was simple to use the system', with users giving each statement a score ranging from 1 ('Strongly agree') to 7 ('Strongly disagree'). The PSSUQ produces an overall score together with three sub-scores for system usefulness, information quality and interface quality.

Results of the baseline PSSUQ are shown in Table II. Possible scores range from 1 (best) to 7 (worst). For each of the three sub-scale scores (system usefulness, interface quality and information quality) and for the overall score, the tool performed well, exceeding the mean scores calculated by Sauro and Lewis [17]. It is intended that the tool will be further refined over the period of its development and use, with changes in usability measurable by repeat use of the questionnaire.

An early usability issue noted was that dropdown menus containing a high number of potential answers, particularly medications, interfered with users' experience of the tool as the dropdown covered the majority of the screen.

C. Pillar 3: Processes (including actors)

Technologies are influenced by the environment within which they are situated and vice versa. This is referred to as reciprocal shaping by Sein et al. [15]. Understanding the context is therefore very important. A problem faced by the development team was that the collection tool was to be located in ten sites from different European countries or cities, all working in different healthcare systems. The final pillar allowed the tool to be observed being used at the sites, to gain a better understanding of how data collectors planned to capture participant data using the collection tool and who were the actors needed to produce, review or record these data. Understanding data collection processes and relevant existing workflows allowed the project to generate further user requirements.

In addition to processes and actors, under this pillar, current data collection systems and any other artefacts used at the site to collect data were identified and the impact on data collection reviewed. The purpose of this was to

understand how these could fit with the P-ALS DCT to minimise the impact of the data collection process on clinical work at the data collection sites. This pillar ensures that the data collection process required by the tool does not impose inefficient or impractical working practices on the data collection sites.

From this an understanding of how the tool would be used at the sites was developed. For example, initially it was thought that the primary source of patient information would be an interview. However, following the site visits it was noted that common to all sites was a collection process whereby a portion of the information would be transcribed from existing clinical documents (paper and electronic) into the tool. The remainder would be captured in an interview at a later time. This led to a requirement that data entered in one location be available to view when the data collector went to add additional information (from the interview).

The data collection tool contains 15 pages. Each page focuses on a particular division of data to be collected, for example 'Smoking and Alcohol' use, or 'Socio-Economic Details'. Table III lists the 15 pages and describes their contents. Table IV gives an example of the contents of one page in more detail. This allows for some pages to be skipped when not required in a particular data collection encounter, e.g., during a repeat data collection encounter when the focus of the data collector is on fields that may have changed since the last encounter, such as clinical progression or resource use. A change resulting from the feedback was the rearrangement of the pages. This resulted in pages containing questions that would be collected during an encounter to be grouped together. A sample data collection page is shown in Figure 4.

Another change identified was that, as not all questions would be asked in each participant encounter, users required a way to know what had been asked or what needs to be asked (during the current data collection encounter). A requirement around carrying over previous data is currently in development to address this.

Future iterations of the collection tool aim to extend data collection beyond the clinical encounter and patient interview to include data from wearables, carers, and other modalities. While no unified data collection process was found that suited all sites, the arrangement of the worksheet (now question set) was also amended to better match the flow of information. For example, information that is likely to be updated frequently, such as clinical progression, was grouped together, whereas socio-economic information is captured at specific intervals so was presented separately.

Along with the question set and usability review, existing data collection processes at the site were examined through discussions with the data collection. This included understanding the role of data collectors; registration of participants; clinical coding systems used; information systems used; and remote monitoring of participants. The data on participants are personal and sensitive, and each collection site was required to follow their local processes for ethics approval, data protection impact assessment and to

TABLE II. SCORES FROM PSSUQ

| Question number | Question text | Average Score |
|-----------------|--|---------------|
| 1 | Overall, I am satisfied with how easy it is to use the system | 2.2 |
| 2 | It was simple to use this system | 2.2 |
| 3 | I was able to complete the tasks and scenarios quickly using this system | 2.4 |
| 4 | I felt comfortable using this system | 2.6 |
| 5 | It was easy to learn to use this system | 2.3 |
| 6 | I believe I could become productive quickly using this system | 2.0 |
| 7 | The system gave error messages that clearly told me how to fix problems | 3.1 |
| 8 | Whenever I made a mistake using the system, I could recover quickly and easily | 2.7 |
| 9 | The information (such as SOP) provided with this system was clear | 2.3 |
| 10 | It was easy to find the information I needed | 2.0 |
| 11 | The information was effective in helping me complete the tasks and scenarios | 2.3 |
| 12 | The organization of information on the system was clear | 1.9 |
| 13 | The interface was pleasant | 2.0 |
| 14 | I like using the interface of this system | 2.2 |
| 15 | The system has all the functions and capabilities I expect it to have | 2.7 |
| 16 | Overall, I am satisfied with this system | 2.0 |
| | Overall PSSUQ score | 2.3 |
| | System usefulness | 2.3 |
| | Information quality | 2.4 |
| | Interface quality | 2.3 |

sign a data transfer agreement. It was for each site to ensure that they comply with their local data protection laws.

TABLE III. SECTIONS IN THE DCT

| Section number | Section name | Description |
|----------------|------------------------------|--|
| 00 | Index Capsule | Personal identifying information, such as name, address, date of birth, that should be kept private. Not available during routine use of the data collection tool. |
| 01 | Data Collection Admin Fields | Details of the data collection, such as the date of collection, data source (e.g., patient interview or healthcare record), data collector's ID. |
| 02 | Demographics and Education | Demographic information of the participant including details of their education history and number of children. |
| 03 | Ancestry and Family History | Ethnicity and geographic origin of participant, parents and grandparents; medical history of relatives. |
| 04 | Medical History | Including specifically history of diabetes, cholesterol, psychiatric conditions, and medications taken for these |
| 05 | Smoking and Alcohol | History of smoking and alcohol consumption |
| 06 | Baseline Clinical Data | Details of ALS onset, first symptoms and diagnosis |
| 07 | Clinical Progression Data | Data recorded at each clinic visit. Includes weight, whether gastrostomy is in situ, FVC score, use of NIV, ALSFRS-R score and sub-scores. |
| 08 | Disease Modifying Drugs | History of taking disease-modifying medication (primarily Riluzole, Edaravone or Tofersen) |
| 09 | ALS Symptomatic Medication | History of taking symptomatic medications or therapeutic interventions |
| 10 | Cognition and Behaviour | Dates and scores of cognitive screening (primarily ECAS), behavioural screening, dementia status |
| 11 | Socio-Economic Details | Including information on household income, employment history and occupation, and state benefits |
| 12 | Resource Use | Information on care received, including community-based care, social care, palliative care, counselling services, hospital care, aids and appliances received, and additional costs due to ALS |
| 13 | Genetic Testing Information | History of genetic testing, including which genes were tested for, dates of tests and results of tests. |
| 14 | Linked IDs | the key unique local IDs that allow the Precision ALS participants information to be linked to all other research information. |
| 15 | Endpoints and Vital Status | Patient vital status, date and place of death, dates of permanent mechanical ventilation and tracheostomy if applicable. |

TABLE IV. FIELDS WITHIN SECTION 10 OF THE DCT

| Question number | Question | Notes | Frequency |
|-----------------|---|--|-----------|
| 01 | Was cognitive screening performed? | 'Yes' or 'No' | Recurrent |
| 02 | Which assessment was used? | 'ECAS' or 'Other, please specify' | Recurrent |
| 03 | Date of cognitive screening | | Recurrent |
| 04 | ECAS ALS_SPECIFIC_Total Score | 0-100 | Recurrent |
| 05 | ECAS ALS Specific Classification | 'Normal' or 'Abnormal' | Recurrent |
| 06 | ECAS Total Score | 0-136 | Recurrent |
| 07 | Cognitive Screening Status | 'Normal' or 'Abnormal' | Recurrent |
| 08 | Was behavioural screening performed? | 'Yes' or 'No' | Recurrent |
| 09 | Which assessment was used? | ECAS, BBI, Apathy Scale, FTDQ. Other, please specify | Recurrent |
| 10 | Date of behavioural screening | | Recurrent |
| 11 | Behavioural status | Mild impairment, Moderate impairment, Severe impairment , No Impairment | Recurrent |
| 12 | Does the person have dementia? | 'Yes' or 'No' | Once-off |
| 13 | Source of Diagnostic Dementia Information | Pre-existing Dx before visiting clinic , Neurologists notes, ECAS , Full battery , neuropsychological testing, Other, Please Specify | Once-off |
| 14 | Date of diagnosis | | Once-off |
| 15 | Type of dementia | FTD (Frontotemporal Dementia), Amnesic Dementia, Other, Please Specify | Once-off |

Figure 4. An example page from the data collection tool

VI. CONCLUSIONS AND FUTURE WORK

User requirements for the development of a data collection tool were constructed from the feedback from different groups including ten clinical sites. The work described in this paper addresses the wicked problem of managing the potential high volume and variety of feedback captured from a range of stakeholders including ten different geographical sites that could translate into user requirements so that nothing is overlooked. To address this problem, a knowledge contribution in the form of a process for an auditable information pipeline was developed to manage this feedback so that it reached its target, the DCT. The initial pipeline was developed based on project group discussions and engagement with an initial test site.

The pipeline proposes four parts to managing feedback. These are capture, record, review and assign to a target. Ways to capture feedback included workshops, online engagement meetings and in-person site visits. The in-person visit adopted a mock data collection interview to review the usability of the tool and questions it contained, along with reviewing the data collection process used at each site. The feedback was recorded in a digital document and automatically assigned a reference number and from this it was reviewed by the project group and assigned to a target (for instance, for clinical review or for technical review).

Feedback is categorised into three pillars: digital worksheet, usability, and process/actors. Each pillar represents a broad area that was explored to identify user requirements. Pillar 1, “Digital Worksheet”, captured feedback related to the question set. These could be technical or clinical. Pillar 2, “Usability”, focused on user experience requirements. Pillar 3, “Process/actors”, investigated requirements relating to the flow of data into the collection tool from the various data sources that contribute to the final dataset, and the actors required to perform these tasks. The evaluation process did not lead to identification of further pillars, and we believe that the three-pillar model has proved to be both helpful and robust.

From this work, challenges were identified that could potentially limit the collected data to be combined from many sites into a single dataset. For example, under pillar 1 (digital worksheet) wording and harmonisation were identified. The latter included numerical, social/cultural and documentation challenges. Development and refinement of the information pipeline will continue at the remaining sites.

ACKNOWLEDGMENT

This work is funded by the Science Foundation Ireland grant SP20/SP/8953. OH’s work is also funded from grants 16/RC/3948 and 13/RC/2106_P2.

REFERENCES

- [1] S. Impey, J. Turner, F. Gibbons, A. Bolger, G. Stephens, et al. "Dataset, Usability and Process - Developing an Interdisciplinary, Multi-modal Data Collection Tool and Platform for a Rare Disease," HEALTHINFO 2023, The Eighth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing. Valencia, 13th-17th November 2023. ISBN: 978-1-68558-105-3. Available from https://www.thinkmind.org/articles/healthinfo_2023_1_50_80022.pdf
- [2] M. Ryan, M. Heverin, M. A. Doherty, N. Davis, E. M. Corr, et al. "Determining the incidence of familiarity in ALS: a study of temporal trends in Ireland from 1994 to 2016," *Neurology Genetics*. 2018;4:e239.
- [3] C. A. Johnston, B. R. Stanton, M. R. Turner, R. Gray, A. H. Blunt, et al. "Amyotrophic lateral sclerosis in an urban setting: a population-based study of inner city London". *Journal of Neurology*. 2006;253:1642-3.
- [4] S.A. Goutman, O. Hardiman, A. Al-Chalabi, A. Chió, M. G. Savellieff, et al. "Recent advances in the diagnosis and prognosis of amyotrophic lateral sclerosis". *The Lancet. Neurology*. 21 (5): 480-493, May 2022, doi:10.1016/S1474-4422(21)00465-8. PMC 9513753. PMID 35334233
- [5] MedlinePlus. *Amyotrophic lateral sclerosis*. [Online]. Available from: <https://medlineplus.gov/genetics/condition/amyotrophic-lateral-sclerosis> [Accessed 2024.05.31]
- [6] P. Masrori and P. Van Damme. "Amyotrophic lateral sclerosis: a clinical review," *European Journal of Neurology*, 2020 Oct;27(10):1918-1929. doi: 10.1111/ene.14393. Epub 2020 Jul 7. PMID: 32526057; PMCID: PMC7540334.
- [7] R. McFarlane, M. Galvin, M. Heverin, É. Mac Domhnaill, D. Murray, et al, "PRECISION ALS—an integrated pan European patient data platform for ALS," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*. 2023. 24:5-6, 389-393, DOI: 10.1080/21678421.2023.2215838
- [8] Precision ALS. *Precision ALS*. [Online]. Available from: <https://www.precisionals.ie> [Accessed 2024.05.31]
- [9] TRICALS. *TRICALS: The Highway Towards A Cure* [Online]. Available from <https://www.tricals.org> [Accessed 2024.05.31]
- [10] Adapt Research Centre. *ADAPT: The Global Centre of Excellence for Digital Content and Media Innovation*. [Online]. Available from <https://www.adaptcentre.ie/> [Accessed 2024.05.31]
- [11] Agile Alliance. *What is Agile?* [Online]. Available from: <https://www.agilealliance.org/agile101> [Accessed 2024.05.31]
- [12] C. Churchman. "Wicked problems." *Management Science*, 1967:14(4), B141-142
- [13] H.W.J. Rittel and M.M. Webber. "Dilemmas in a General Theory of Planning." *Policy Sciences*, 1973:4 (2): 155-169, doi:10.1007/bf01405730
- [14] M. Mullarkey, A. R. Hevner. "An elaborated action design research process model." *European Journal of Information Systems*, 2019:28(1):6-20. doi: 10.1080/0960085X.2018.1451811
- [15] M. K. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren. "Action Design Research." *MIS Quarterly* 35, no. 1 (2011): 37-56. <https://doi.org/10.2307/23043488>.
- [16] J.R. Lewis. "IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use". *International Journal of Human-Computer Interaction*. 1995. 7:1, 57-78, DOI: 10.1080/10447319509526110
- [17] J. Sauro and J.R. Lewis, *Quantifying the user experience: Practical statistics for user research*, 2nd ed. Cambridge, MA: Morgan-Kaufmann, 2016.

AI-Based Digital Therapeutics Platform for Obesity Management

Sarfraz Khokhar^{1*}, John Holden²

¹Rasimo Systems, Raleigh, NC, USA
khokhar@rasimo.com

²Rockford College of Medicine, IL, USA
jholden@uwhealth.org

Catharine H. Toomer³, Linda Whitby⁴

³Health Wellness and WL Centers, Aiken, SC, USA
wellness@total-weight-loss.com

⁴Rasimo Systems, Raleigh, NC, USA
drwhitby@rasimo.com

Abstract—Obesity is a global problem that has had a significant impact on society and the economy. The consequences are ominous, with serious health risks. Millions of people are dying every year from complications of obesity and comorbidities. Despite efforts by governments and health agencies, obesity continues to rise. Most of the approaches to management and treat obesity have not been successful because they did not shape people's lifestyle and the solutions that were provided for lifestyle modification are not multidisciplinary, they focus on only specific aspects. Obesity management mandates multidisciplinary approach with effective patient engagement, enhanced patient-healthcare provider communication, better adherence to therapy, minimize therapeutic inertia, motivation, more informed treatment decisions by the healthcare provider, and addressing psychosocial conditions. We designed and developed an AI (artificial intelligence) based digital therapeutics platform (named SureMediks) to the multidisciplinary mandate for obesity management and treatment. We tested the efficacy of our proposed platform (solution) with a 24-week field trial and achieved 13.9% weight loss of the initial weight.

Keywords- obesity; weight loss; digital therapeutics; artificial intelligence; expert systems.

I. INTRODUCTION

Obesity has matched epidemic proportions, with at least 2.88 million people dying every year [1] as a result of being overweight or obese and a whopping economic and social impact of \$1.7 trillion dollars [2]. The costs include \$1.24 trillion in lost productivity and \$480.7 billion in direct healthcare costs [3]. Once associated with high-income countries, obesity is now also prevalent in low and middle-income countries. Government agencies, non-governmental organizations, and the private sectors have been publishing their expert advice as good practices for a healthy lifestyle, in their research and field trials for decades and acknowledge that this pandemic is ever-increasing.

Despite ubiquitous information about nutrition and exercise, more fitness awareness, and more food and activity tracking devices, over 42% of the US adult population is living with obesity [4]. The world obesity rate grew proportionally as well [5]. The statistics show a significant increase from a decade ago, as depicted in Figure 1. The consequences are ominous; obesity is associated with serious

health risks including heart, liver, gallbladder, kidneys, joints, breathing disorders, sleep apnea, diabetes, and several types of cancer [6]. Obesity can disrupt the hormonal balance that regulates ovulation and menstrual cycles, leading to irregular or absent periods and reduced chances of conception in women and can impair the quality and quantity of sperm, as well as cause erectile dysfunction and reduced libido in men [7]. The medical community continues struggling to find successful ways to encourage weight loss and provide effective interventions.

Lifestyle intervention faces challenges like compliance issues making weight loss difficult. Despite this, it continues to be a crucial component of obesity treatment. Digital tools augment lifestyle interventions by offering personalized support catering to the need for continuous interaction and support beyond conventional primary care settings. However, there is a need for a more comprehensive approach in utilizing digital tools to address the multifaceted aspects of obesity treatment effectively.

Traditional digital health methods of lifestyle modification have limited effectiveness in managing obesity as they lack multidisciplinary approach and engagement of HealthCare Provider (HCP). The use of AI health coaching and predictive guidance for weight loss [8-11] is comparable with in-person HCP treatment; however, it lacks patient engagement and treatment adherence.

Studies combining approaches and technologies showed better results. A clinical trial conducted showed that the use of a mobile application that used AI algorithms and gamification techniques to provide personalized feedback led to a significant reduction in body weight, body mass index (BMI), and waist circumference [12]. However, there is a need for effective, holistic, adaptive, cost effective, user-friendly, and integrated digital solution to manage obesity.

In the 21st century, AI and health technological advancement have enabled the development of digital therapeutics. Digital Therapeutics (DTx) are defined as evidence-based therapeutic interventions for patients by means of qualified software programs and medical devices to prevent, manage, or treat medical conditions.

Digital therapeutics can be more flexible than other treatment methods to address patients' individual needs [13]. These technologies employ various techniques, such as

mobile applications, wearable devices, and online platforms, to improve the effectiveness of treatment interventions [14].

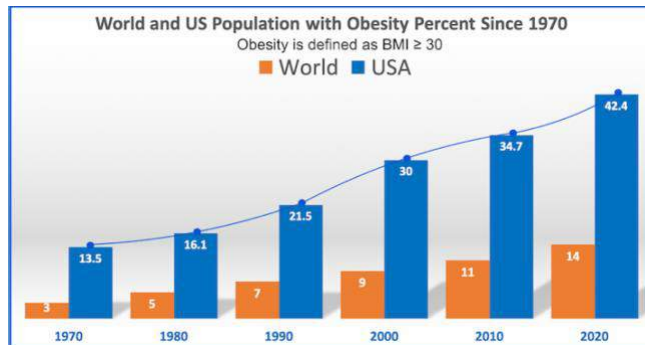


Figure 1: World and US obesity growth in the last six decades.

AI-based DTx would be an ideal complement to the pharmaceutical or even surgical weight loss offerings. AI along with related technologies offer a promising approach for the management of obesity, as they use Machine Learning (ML) algorithms and/or expert systems (ES) to personalize treatment plans for patients. We propose an AI-based DTx integrating all the approaches tried before but individually in a unified single platform, SureMediks. It includes the following approaches and features:

A. Remote monitoring of body weight

Remote monitoring of body weight can help improve weight management outcomes and reduce costs associated with in-person clinic visits. Patients can be provided with digital scales to monitor their weight, which transmits data to a remote monitoring platform such as a mobile app or web portal, allowing healthcare providers to track progress and make necessary adjustments to treatment plans. In a randomized controlled trial (RCT) study conducted among 230 adults, there was a reduction of over 3% BMI among participants who used telemedicine-enabled remote monitoring of body weight compared to the control group [15].

B. Virtual coaching

Virtual coaching can facilitate continuous support and motivation to patients throughout their weight management journey. Healthcare providers can engage with patients through regular phone calls or online messaging using secure platforms. A meta-analysis of 21 RCTs found that by engaging patients in virtual coaching, significant weight loss outcomes could be achieved in a three to six-month period compared to traditional care alone [16].

C. Short-term goals

The weight loss participants who reach their short-term goals have better long-term weight loss [17]. Long-term goals prepare them for a new lifestyle; however, long-term goals are far away and a patient may get lost somewhere in the path in pursuit of the goal far away. It is beneficial to lose

weight in a series of smaller short-term goals. The short-term goal achievement effects persist over time, and in fact, induce users to accomplish even more ambitious short-term goals in the future [18]. The successful weight loss, including bariatric Laparoscopic Roux-en-Y gastric bypass operation (LRYGB) conforms to exponential decay [19]. We investigated into mathematical modeling of a successful weight loss behavior and formulated the Khokhar WL Formula [20]. The Khokhar WL formula to generate short-term goals is depicted in the equation below:

$$W_{loss} = \frac{\Delta W}{1 - e^{-\frac{r}{10}}} \left(e^{-\frac{rn}{10}} - e^{-\frac{r\tau}{10}} \right); r, \tau \neq 0; \quad (1)$$

Here, W_{loss} , ΔW , τ , and r are weight to lose at each short-term goal, total weight loss, time to lose weight in weeks, and r is a special parameter respectively, we called, r , the curve tension. In the formula, n is the week number sequence for short-term weight loss. For example, for $n = 1, 2, 3$, it will determine the required weight loss for the first, second, and third weeks. Figure 2 depicts an example of the short-term weight loss goals for a period of 24 weeks to achieve a long-term goal of 15 Kg. In our implementation, the patient's weight loss curve is dynamically adjusted based on their performance to reinforce their continuous interest and motivation as they see they are achieving their short-term goals. If a patient is struggling with his or her short-term goals, the system detects this and reassigns a different set of short-term goals which are relatively easier to achieve.

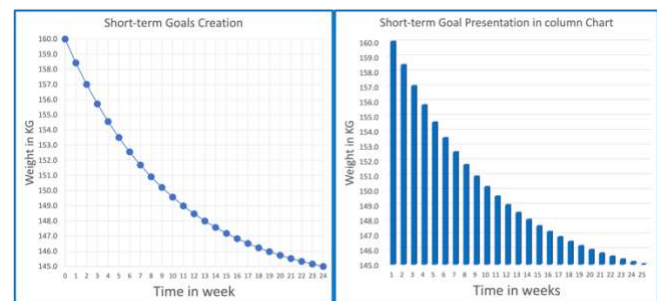


Figure 2: Weekly goals in a curve and columns presentation generated by Khokhar WL formula for 24 weeks period

Conversely, if the short-term goals are too easy for the patients, the system detects this phenomenon and raises the bar a little to keep them motivated and challenged. The SureMediks scale shows the patients their short-term goals on the scale screen along with their overall progress.

D. AI-generated feedback and predictive analyses

Having weight loss tools or platforms available to the patients and the providers to deliver real-time feedback and in-the-moment support may assist the patient with initiating and maintaining changes over the entirety of a day. Technology-based weight loss interventions offer the potential to deliver help at the exact moment necessary to support health behavior change in a way that was never possible before. Frequent and timely feedback to encourage

the patients and provide them with support and education shows very productive results and is a key predictor in weight loss [21]. Artificial intelligence offers sophisticated support in weight loss. The intensity of the feedback intervention can be continuously adjusted depending on the patient's weight loss, and it is at a reduced cost compared to a non-optimized intensive intervention [22].

SureMediks, the platform of this study, includes an ES. Expert System is a branch of AI that mimics the decision-making processes of human experts in specific domains. These systems are designed to provide guidance, advice, and recommendations to users based on their input and the knowledge (KB) rules programmed into the system [23]. These rules in KB can be updated as system learns new facts about the patients and their behavior. The integral components of an ES and its operation are depicted in Figure 9. The Knowledge Acquisition System of the ES extracts the expert knowledge and saves (learns) it in Knowledge Base (KB) as rules. Inference Engine (IE) activates these rules based on current and historical data and provides the guidance and education stored and learned in the KB. IE also updates the rules in KB dynamically. Explanatory Systems interprets patient's data and explain to the patients through charts and graphs in the mobile app.

In the context of patients' guidance and education, expert systems can provide personalized and interactive programs for managing and treating various health conditions, including obesity [24] and diabetes [25-26]. These systems can analyze patient data, such as medical history, symptoms, and lifestyle factors, and provide tailored recommendations and interventions to support patients in making informed decisions about their health.

AI feedback system was designed to address the primary barriers to successful weight loss, such as the complexity of dietary information, ineffective motivational strategies, and intermittent physical activity. By delivering real-time personalized feedback, SureMediks helps individuals remain on track, and offer corrective strategies when necessary. Additionally, it offers access to human expert guidance, which can further help individuals develop healthier behaviors that last longer.

E. Guidance through remote interaction and real-time and offline monitoring

The use of remote monitoring, focused on effective guidance, has proved to be a tool to support the care of overweight patients. During the Covid-19 epidemic, remote weight loss patient monitoring has been acceptable to large patients' population [27]. Remote consultation and monitoring, in general, were accepted as an attractive and convenient alternative during the Covid-19 pandemic, not just the weight loss patients, Weight loss behavioral interventions delivered remotely, without face-to-face contact between participants and weight-loss coaches, patients with obesity achieved and sustained clinically significant weight loss over a period of 24 months [28].

With remote guidance and remote monitoring, the weight loss provider imparts a sense of closeness with the patient, which keeps the patient accountable and motivated. This narrative is an integral part of our study and we wanted to have this factor integrated with the other factors for better weight loss results. Our remote guidance and motoring platform (dashboard) were designed to manage, communicate and monitor our participants (patients). The dashboard showed us patients' weight loss progress in real-time or offline. Along with weight loss progress, we could observe the changes in the patients' body metrics such as fats, muscles, metabolic rate, body water, and muscle protein. This gave us insight into how the participants' diet and physical activity are coming along with their weight loss effort. Our coaches then adjusted the relevant guidance for the participants as needed.

F. Motivational and moral support

Motivation is the state of being driven or encouraged to do something or act a certain way. In the context of weight loss, it involves the desire to achieve the desired result, which in our case, is weight loss. Patients' autonomous motivation to participate in a weight-loss program is positively related to their staying in the program and losing weight during the program [29]. Various motivational factors for losing weight may lead to successful weight loss and long-term weight maintenance [30]. Motivational support is such a strong factor in weight loss that a motivation-focused program offers an effective alternative to traditional skill-based programs [31].

The motivation was the center-staged of our program. Each week when the participants fulfilled the weekly target weight the feedback sent by the AI agent had an integrated rewarding and reinforcing motivational message which further strengthened their commitment and motivation. The dynamic and adaptive short-term goals algorithm kept their short-term goals achievable, giving them the sense that they could do it. For those participants who could not achieve their weekly short-term goals, the AI agent sent constructive and supporting motivational messages along with the feedback which held them up and motivated in the weight loss program.

G. Community support and accountability

Numerous studies have indicated that social support can significantly impact weight loss success. The participants who received social support had a greater likelihood of achieving their weight loss goals [32]. Furthermore, the study revealed that participants who received social support experienced a reduced risk of weight regain. In contemporary society, virtual support collaborative communities with common goals and interests flourish. These online communities make salient a context-relevant social identity that motivates behaviors that facilitate compliance to the public commitment, and hence, more effective goal pursuit [33], and play an important role in the participants' weight

loss effort [34]. In addition to these studies, another body of research has indicated that individuals who receive support from a group or community are more likely to maintain their weight loss long-term [35]. Accountability can also play a vital role in keeping motivation up in the weight loss journey. Building a support structure around oneself, whether it is through hiring a personal trainer, joining a support group, or seeking advice from friends and family. This will help in holding oneself accountable and keeping them motivated.

Although there is limited research on the effectiveness of a weight loss accountability circle, the available evidence suggests that it is a promising approach for weight management. In a study examining the effectiveness of online weight management communities, participants who received social support and accountability through an online platform were more likely to achieve their weight loss goals than those without support. Another study found that participants who received social support through a mobile phone-based weight management program were more likely to adhere to healthy behaviors and lose weight [36]. Figure 3 depicts SureMediks community support and accountability screenshots.

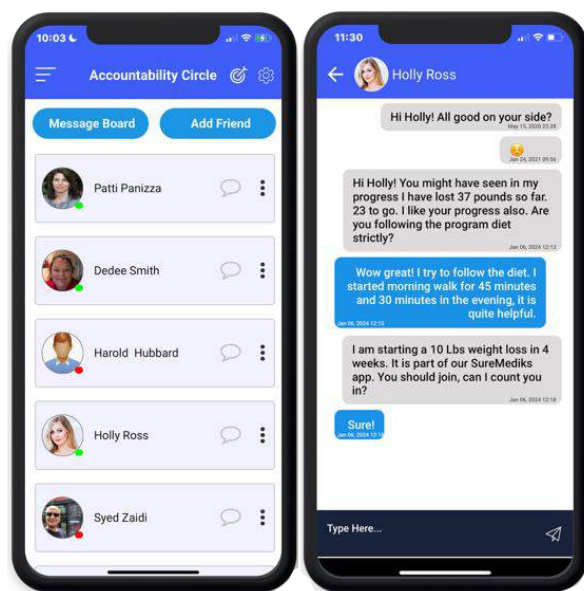


Figure 3: Example of SureMediks accountability circle. The accountability circle members can communicate, see one another progress and engage in gamification challenges.

H. Instant and interactive communication with weight loss provider

Several studies have demonstrated the benefits of timely guidance and feedback in weight loss. A meta-analysis of 22 randomized controlled trials found that interventions that included timely feedback were associated with greater weight loss than those without feedback [37]. Timely guidance and feedback help individuals to stay on track with their weight loss goals and to make adjustments as needed. It also provides a sense of accountability, which can be motivating for

individuals trying to lose weight. There are several approaches to providing timely guidance and feedback for weight loss. One approach is to use technology-based interventions, such as mobile apps and wearable devices that track physical activity and food intake [38]. Studies have shown that these interventions can be effective in facilitating weight loss and improving adherence to healthy behaviors [39]. Instant communication with patients such as text messages, video consultation, or any other forms of communication as means of behavioral prompting, support, and self-monitoring is more successful in their weight loss [40]. Such instant communication without proper feedback may not be effective [41], which further endorses our idea of providing instant feedback through the providers' guidance as a patient steps on the scale.

I. Diet and nutrition manager

The previous studies support the use of managing diet and nutrition as an effective weight loss strategy. Keeping a food journal can increase awareness of eating habits, identify problem areas, and promote accountability. Technology-based approaches to food journaling can make the process more convenient for individuals and may improve adherence. The studies suggest that keeping a food journal is associated with greater weight loss and improved dietary intake. In a randomized controlled trial, participants who kept a food journal lost twice as much weight as those who did not keep a food journal [42]. Similarly, in a study of overweight women, those who kept a food journal lost more weight and consumed fewer calories than those who did not keep a food journal [43]. Technology-based approaches to food journaling, such as mobile phone applications and online websites, have also been found to be effective in promoting weight loss [44]. Consistent tracking is a significant predictor of weight loss, resulting in an additional seven pounds of weight loss over the course of the program suggesting the intervention successfully achieved clinically and significantly long-term weight loss [45].

J. Physical activity tracker

Numerous studies have shown a positive correlation between exercise and weight loss. A meta-analysis of randomized control trials conducted [46], found that exercise combined with calorie restriction leads to more weight loss than calorie restriction or exercise alone. Additionally, the American College of Sports Medicine (ACSM) recommends that individuals aiming for weight loss should engage in at least 150-250 minutes of moderate-intensity exercise per week [47]. One effective way to monitor exercise and track weight loss is by keeping a journal. A weight loss and exercise journal can help individuals stick to their goals and track their progress. According to a study [48], participants who kept tracking their physical activity they lose weight and their motivation increases, as they stopped tracking and monitoring, they started gaining weight and their motivation

declined. The journaling can also serve as a source of motivation and a reminder of why the individual embarked on the weight loss journey in the first place.

To assess the efficacy of our proposed AI-based DTx platform, SureMediks, we developed a prototype of the platform and set it up for a field trial. The implemented features and expert system's knowledge base were derived from a large research body and field trials mentioned previously in this section. In this paper we report summary of the field trial and the results.

II. METHOD

This section describes our AI-based platform, SureMediks, field validation covering participants details, procedures and measurements.

A. Platform

The platform, SureMediks, consists of the following key elements: 1) An Internet-connected body composition scale to get patient's weight and related body metrics, 2) A mobile application through which patients receive tailored guidance, education, motivation, communicate with the HCP, interactive with accountability circle members for community support and visually can see the weight loss progress, 3) An AI agent acting as an expert system, 4) Healthcare providers' interaction as an human interface with patients, 5) A dashboard for the HCP to view patients' weight loss progress and interact with the patients, and 6) Optimizing algorithms running in the background.

B. Participants and weight loss goals

A participant sample of 1137 people of age 21 years and older from the USA, Canada, UK, and Australia were invited through emails and a weblink to participate in this field study. They were provided with key screening questions if they were determined and committed to losing weight that year, ready to be strictly focused on weight loss, ready and committed to be on a low-calorie trackable diet with daily trackable physical activity.

Finally, 391 participants took part in the trial from start to end. Of the 391 participants, 59% of the participants were female and 41% were male. Their education level, marital status, and other socioeconomic factors were not part of our selection criterion. However, their current weight, BMI, and age were among the primary concerns as we wanted to have diversity in age and weight buckets. They were diverse in geography and lifestyle; however, these aspects were common; acceptability of using technology in their weight loss effort, determination to lose weight, owning a smart device, phone, or tablet, and having access to the Internet, wired or through cellular data.

Their start (baseline) mean weight, μ_{Start} , was 124.6 Kg with a standard deviation, σ_{Start} , of 31.57 Kg, and a wide range of 65-181 Kg weight distribution as depicted in Table 1. Mean age of the participants, μ_{Age} , was 43.56 years with

a standard deviation, σ_{Age} , of 12.60 years, and the range of 21-71 years. Their BMI mean, μ_{BMI} , was 43.9 Kg/m² with SD, σ_{BMI} , of 8.5 Kg/m², $30 > \text{BMI} > 25$ was considered overweight and $\text{BMI} \geq 30$ was considered obesity as per World Health Organization (WHO) generic guidelines. The weight loss goal was 10% of the start weight however we set a stretch goal of 15% as the majority of the participants insisted on raising the bar.

C. Procedure and measures

The participants were provided with a WiFi-enabled smart body composition weighing scale. A mobile app for Apple and Android devices was made available at the Apple App Store and Google Play store which the participants could download free of cost on their respective devices.

The study coordinators and coaches collaborated with the participants through a dashboard. The coaches had their own dashboards which they could log in and manage, communicate, and monitor the participants' progress, food intake, and physical activity. Figure 4 shows the high-level architecture of our implementation. We created six groups of 391 participants with six different weight buckets. Bucket 1 with participants of 65-85kg of weight, Bucket 2 for 86-105kg weight, Bucket 3 for 106-125kg, Bucket 4 for 126-145kg, Bucket 5 for 146-165kg, and Bucket 6 for the participants with the weight of 166-181kg. These six weight buckets had 61, 78, 83, 60, 66, and 43 participants respectively, totaling 391 participants.

The patients' involvement process starts with downloading and installing the app used in this study on their smart devices (phones or tablets). After opening the app, the first step of their setup was to register the smart scale by scanning the scale ID on the back of the scale using the scanning button in the app. Alternatively, they could manually enter the scale ID. Associating the scale ID to each participant's profile is very important as all the key metrics transferred via the Internet from the scale represented the participant's data. Next, they enter their information, including age, height, preferred units (Lbs./inches, Kg/cm), and physical activity level. The desired weight (15% less than the current weight), and duration in which 15% loss to be achieved, i.e., 24 weeks, and their group or bucket number from the list available in the app. After the participants have completed their signup process on their app they are automatically added to their respective coach's dashboard. As they step on the scale the very first time their baseline (weight, BMI, fat mass, muscle mass, and basal metabolic rate) is established automatically, no manual intervention was needed. At that time, the intelligent agent sent them their weekly goals, which they could see in their app and also on the scale. These weekly goals were estimated by the Khokhar WL formula. The curve tension, in the beginning, is moderate, 0.75, and gets adjusted dynamically up and down based on their weight loss performance. Logically one can think of it as hopping from one curve to a more suitable curve dynamically. This is all adjusted by the intelligent agent. If

the participants were struggling with one curve, they were moved to the relatively easier curve adaptively. This event was noted in the dashboard as, a sub-goals reassignment. We were interested in the correlation between the percentage of weight loss and the number of sub-goal reassignments.

Each time the participants stepped on the scale they received feedback from the intelligent agent on how they were doing and what they needed to be cautious about. Along with the automatic instant education, guidance, and feedback through AI, their coaches also interacted with them through the text messages which the participants received on their app and could respond back to the coach and video calls through the dashboard. This way the participants had two communication channels with the coach; one through text messages and the other through video calls. Only the coach could initiate the video call. The AI agent had the knowledge base of the weight loss approach. Our objective was to adopt a generic approach to make this study relevant to all types of weight loss approaches. The participants were on a low-calorie diet, designed and supervised by the AI and the coaches. The recommended food items were shown to the participants in the app, they could select the food from that list. If they ate any food out of the list, food library, they could manually enter it in the app. If the participants didn't enter their food in the food journal, they were sent a reminder by the AI agent. Their consumed food got translated into macronutrients, shown to them in the app, and also sent to the dashboard for AI processing and for coaches' records. The participants were guided to a physical activity of their choice from the menu in the app. The physical activity was also tracked by AI and the coach. The participants were asked to step on the scale twice a week; once a week at the least. This was just a guideline; they could step on the scale daily if they wanted. Most participants stepped on the scale two times a week. The participants could view their weight loss progress through charts in their app which included weight loss tracking and key body composition metrics and their coaches could view the same on their dashboards. Based on the information received on the dashboard the coach could send additional informational links on diet and physical activities. The coaches focused on the participants' metabolic rates and weekly weight loss. If the desired weekly goals were not achieved by the participants, the AI agent and the coach sent them strict calorie intake guidelines and instructions.

The participants were encouraged to create their own accountability circle by picking their circle members from any six groups; they were not limited to their own groups. We wanted to record the correlation between the number of accountability members and their weight loss percentage. For each participant, the number of members in their accountability circle was noted by the system automatically. These accountability groups acted as community support groups and proved to be helpful; helping out one another by providing additional moral support. They could interact with one another through an accountability circle and share their progress to keep themselves and the others, motivated.

We set conditions with the AI agent such that if a participant is gaining weight instead of losing, we wanted to be notified by an alert by a text message so that we could intervene timely. This was a great tool to handle unwanted exceptions. The participants were very positive and took immediate measures to correct their diet approach, physical activities, and lifestyle in general.

Through our proposed platform we encouraged participants to take part in the challenges (gamification) which they could initiate within their accountability circle. The app facilitated the challenge creation and follow-up till completion. There were total of six challenges to lose 3% in each challenge. The platform noted how many challenges a participant took part in, as we wanted to see the correlation of these challenges to overall weight loss.

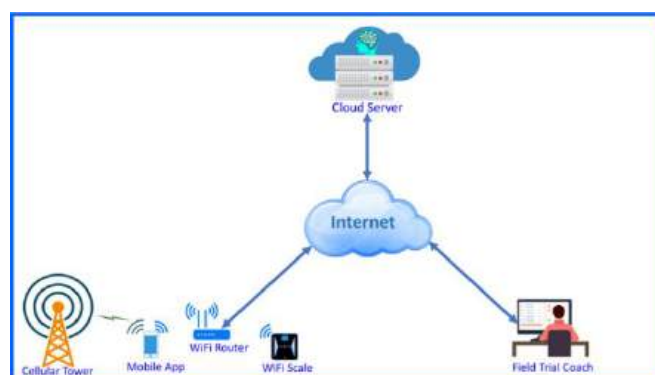


Figure 4: High-level implementation of the architecture: the participants have the scale and app, whereas the coaches have dashboards.

Apart from the community support through the accountability circle, support from the coaches and the AI agent, the participants received a daily motivational quote from the system which would show up on their smart device screen at the scheduled time; different times for the different countries, as we wanted to deliver these quotes in the morning. The quotes were selected by the AI agent based on their progress or the challenges they were facing. This feature of our program was very well received by all participants.

On the completion of 26 weeks of the trial, their weights were noted in the system. The statistics of their weight loss progress are shown in Table 1. In this study, participants received feedback from an intelligent agent (ES) based on their current and historical data, each time they stepped on the scale along with education and guidance through the ES. The flow of ES is depicted in Figure 9. Two sample feedbacks are shown in Figure 5. Coaches also interacted with participants through text messages and video calls. The participants followed a low-calorie diet recommended by AI-based feedback mechanisms and the coaches, with food items shown in the app. Physical activity was chosen from a menu and tracked by AI and coaches. Participants were encouraged to step on the scale at least twice a week and could track their progress through charts in the app. The coaches focused on

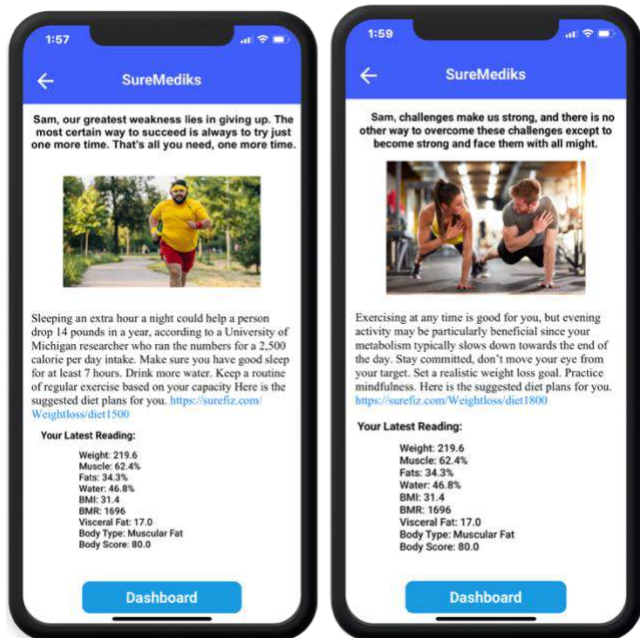


Figure 5: Samples of guidance from the ES, first part is motivational and second part is feedback and guidance.

metabolic rates and weekly weight loss, providing additional guidelines if goals were not achieved. Participants formed accountability circles for support and motivation, and alerts were set up to notify if weight gain occurred. Participants were proactive in making corrections to their diet, physical activities, and lifestyle based on feedback and guidance from the ES.

SureMediks, encouraged participants to engage in challenges within their accountability circle, facilitated by the app. There were six challenges to lose 3% weight each, and the platform tracked the number of challenges participants took part in. In addition to community support, participants received daily motivational quotes selected by the AI agent based on their progress or challenges. After 26 weeks, participants' weekly weights were noted and their weight loss progress statistics were analyzed using MS Excel data analysis tools.

D. Results

The detailed weight loss statistics of each of the six buckets is as follows: For Weight Bucket1, 65-85 Kg, the mean weight loss, μ_{wl1} , was 10.1 kg, standard deviation, $\sigma_{wl1} = 3.4$ kg, mean weight loss percentage of 13.3, with a 95% confidence interval (CI) of 12.18% -14.38%, and BMI loss (drop) of 4.3 points. For Weight Bucket2, 86-105 Kg, the mean weight loss, μ_{wl2} , was 13.6 Kg, standard deviation, $\sigma_{wl2} = 4.4$ Kg, mean weight loss percentage of 14.2, with a 95% confidence interval (CI), 13.20% -15.19 %, and BMI loss (drop) of 5.2 points. For Weight Bucket3, 106-125 Kg, the mean weight loss, μ_{wl3} , was 15.9 Kg, standard deviation, $\sigma_{wl3} = 5.2$ Kg, mean weight loss percentage of 14.0, with a 95% confidence interval (CI), 13.03% - 14.96%, and BMI

loss (drop) of 5.9 points. For Weight Bucket4, 126-145 Kg, the mean weight loss, μ_{wl4} , was 19.1 Kg, standard deviation, $\sigma_{wl4} = 5.9$ Kg, mean weight loss percentage of 14.5, with a 95% confidence interval (CI), 13.41% - 15.58%, and BMI loss (drop) of 6.8 points. For Weight Bucket5, 146-165 Kg, the mean weight loss, μ_{wl5} , was 19.4 Kg, standard deviation, $\sigma_{wl5} = 6.8$ K, mean weight loss percentage of 12.53, with a 95% confidence interval (CI), 11.45% - 13.60%, and BMI loss (drop) of 6.7 points. For Weight Bucket6, 166-181 Kg, the mean weight loss, μ_{wl6} , was 25.5 Kg, standard deviation, $\sigma_{wl6} = 7.3$ Kg, mean weight loss percentage of 14.8, with a 95% confidence interval (CI), 13.54% - 16.07% , and BMI loss (drop) of 8.6 points.

Overall, for all 391 participants, 65-181kg, the mean weight loss, μ_{wl} , 17.27 Kg, with standard deviation, $\sigma_{wl6} = 7.0$ Kg, mean weight loss percentage of 13.89, with a 95% confidence interval (CI), 13.45% – 14.35%, and BMI loss (drop) of 8.6 points. The p-value was significant, $p < 0.0001$, for all results, confidence interval (CI), 13.54% - 16.07% ,

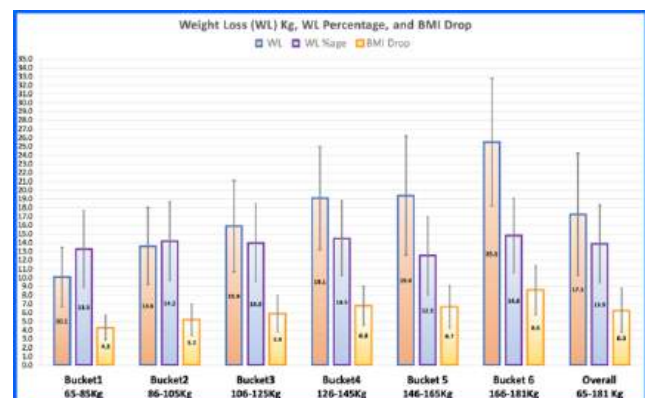


Figure 6: Higher BMI drop with larger weight buckets. the weight loss percentage is similar across all the buckets.

and BMI loss (drop) of 8.6 points. Table 1 shows this data in tabular form. Figure 6 depicts the key results: mean and standard deviation of weight loss, weight loss percentage, and BMI loss (drop).

Figure 7 shows the weekly plotted mean weight loss progress in kilo grams of all the buckets combined (391 participants). In this plot, the Amber curve depicts the weekly weight loss progress for the period of the trial and the Blue line shows the weekly predicted mean weight of the participant per the Khokhar Weight Loss formula (Equation 1). The predicted weight loss curve could serve as the trend curve as well. Figure 8 depicts a typical weight loss progress during the field trial. Both figures demonstrate the predictability and effectiveness of the Khokhar WL Formula as well. The weight loss progress in Amber color line adheres closely with the Blue line, validating the Khokhar WL formula.

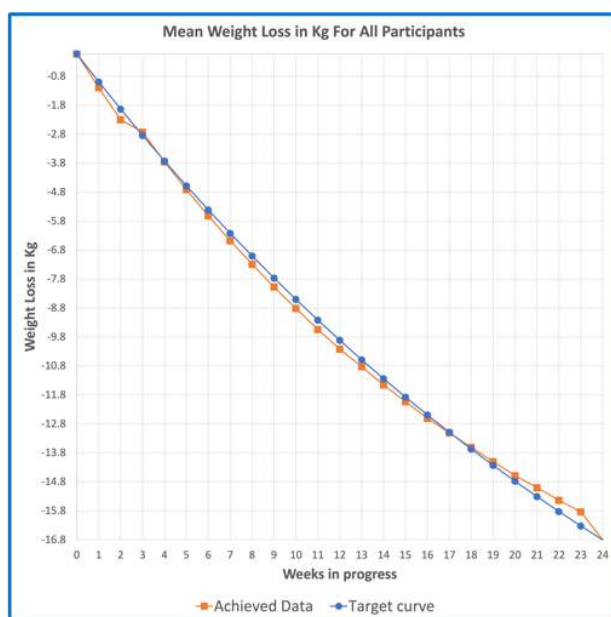


Figure 7: Weekly weight loss progress of all 391 participants. average weekly weight loss was 0.71 Kg.

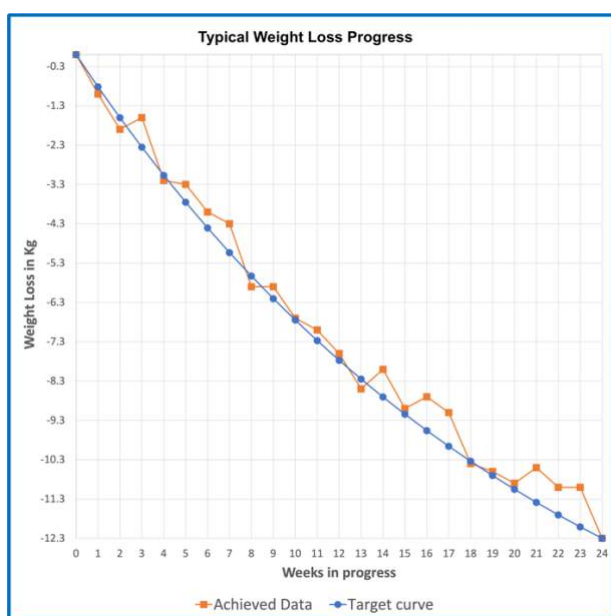


Figure 8: A typical weight loss progress of individual participants.

III. DISCUSSION

This study of digital therapeutics for obesity management in general, and weight loss in particular, suggests that digital platforms are efficient for a weight loss program and can surpass the target of the benchmark of 5%–10% of initial body weight [49-50]. Our goal was to achieve a 10% weight loss of the baseline weight, however, our participants insisted on a higher percentage of weight loss, hence we set a stretch weight loss goal of 15% of the initial weight. Intrinsically the study also manifests that determination plays a vital role in a weight loss effort [51-52]. More significantly, the current

study has shown that a bit higher weight loss goal may be beneficial. This is a significant result because there is debate over the wisdom of setting higher weight loss goals [53-54]. However, we agree that the weight loss goals must be reasonable and coherent with the underlying weight loss approach and methodology. There were several driving factors in this study's weight loss outcome: creating short-term weekly goals and dynamically adjusting them, AI guidance, extensive communication and guidance from the coaches, motivation, accountability, and community support, intelligent food journaling, gamification, and physical activity tracking.

Long-term weight loss goals can often seem daunting and overwhelming to individuals seeking to adopt healthier lifestyles. Dividing a large goal into weekly short-term goals made it less overwhelming and increased the participants' sense of control and confidence in the process. Defining the smaller weekly goals resulted in a quick outcome, in a week, precisely. The weekly outcomes helped adaptively adjust diet, physical activity, motivational messaging, and coaches' interaction, which in turn drove a steady weight loss during the study. Additionally, setting smaller goals aided in habit formation. Achieving short-term goals reinforced positive behavior, making it more likely for that behavior to become a habit. Over time, small changes in behavior accumulated and lead to a healthier lifestyle overall [55]. The digital therapeutic platform monitored the short-term goals and accomplishments of the participants. If a participant missed the short-term goals (lagged behind) twice in a row, the short-term goals for that participant were reassigned to a relatively easier set of goals. Table 2 shows a strong negative correlation (-0.679) of weight loss and the number of times sub-goals (short-term goals) were reassigned. It suggests that those participants who missed fewer sub-goals achieved more weight loss.

Every time participants stepped on the scale the AI agent provided them with instant feedback and guidance. Along with other factors, AI's timely guidance played an important role to achieve an overall of 13.9% weight loss. Previous studies [56] support our claim that the implementation of timely guidance and feedback in weight loss interventions leads to better outcomes. Along with the AI guidance and feedback the coaches' timely and asynchronous guidance further added in achieving the weight loss goal. This body of research further suggests that the use of smartphone applications for weight loss providing frequent guidance (weekly) and feedback lead to greater weight loss than participants without the applications. This is consistent with our study as well. The extensive communication initiated by the coaches kept the diet and physical activity of the participants consistent with their goals and addressed any issues and difficulties they were facing. The key source of success for surpassing the weight loss goal of 10% of initial body weight and approaching 15% weight loss, the stretch goal, is effective communication via AI, text messaging, and video consultation. Such extensive required communication

between the healthcare provider and the patient is not in practice, primarily because of the lack of infrastructure or framework needed for this practice.

The motivation was another driving force that directed the participants' behavior. The motivation to lose weight was derived from various internal and external factors such as achieving better health (this was the case for the majority of the participants), self-esteem, and better physical appearance. Above all, all the participants had been trying to lose weight for a while and had spent a significant amount of money but they did not have a framework to stay motivated and guided. As the participants started seeing weight loss success short-term (weekly), it reinforced their motivation and made a positive impact on their cognitive and emotional well-being [57]. Since the participants had high levels of intrinsic motivation consistently, they got engaged in the required physical activities and initiated healthy social modeling behavior, complied with the diet plans, and exhibited greater levels of peer connectedness using the accountability circle feature in the smartphone app. [58]. Motivation is not a one-and-done incentive; one must stay motivated to reach their desired goals. Regular reinforcement through daily motivational messages, implied motivation in each AI feedback message and guidance, frequent cognitive-behavioral interventions from the coaches, and regular weigh-ins contributed to sustaining motivation leading to their weight loss goal.

Accountability and community support played a significant role for the participants in achieving their weight loss goals with a correlation coefficient of 0.524 ($p < 0.05$) between weight loss percentages and the number of accountability members in the participants' accountability circle. Accountability means being responsible or answerable for something you do. In the context of this study, this meant that the participants held themselves responsible for their diet, exercise, and staying focused in achieving their weight loss goal. It also meant being accountable to the members of their accountability circle. It kept track of their actions and helped them stay on track. Accountability is one of the most effective ways to achieve desired outcomes [59]. Such support and interventions motivate people to make changes in their diet and physical activity, resulting in a significant role in the process of weight loss. When the participants felt supported, pushing themselves to make lifestyle changes was easier. The impact of accountability circle feedback during a weight-loss competition between men and women played an important role. The study showed that regular behavior feedback led to more significant degrees of weight loss overall. The accountability circle feedback led to a sense of increased stakeholder focus and a notable shift in participants' internal focus resulting in long-term weight loss changes [60]. The participants had agreed during the recruiting and reaffirmed that they would enter the food they consumed in the food journal provided in the mobile app. They adhered to the commitment. Journaling their food assisted them in becoming more mindful of their food habits, reducing their

impulsive eating or extreme dieting behaviors, and leading to improvements in healthier food choices. Food intake journaling facilitated weight loss by increasing dietary awareness and reducing caloric intake while increasing physical activity [61]. This approach provided the participants a constant reminder, keeping foodless habitual rather than a source of enjoyment. Journaling has broader use beyond weight loss. Food journals not only positively influenced motives to adhere to healthy food choices, but also supported in identifying dietary triggers [62] and an improved relationship with food. Along with the direct benefits of food journaling to the participants, the reporting of their food consumption to the coaches and the platform further enriched the guidance and feedback delivered to the participants. Similarly, tracking and reporting physical activity played a motivational role in the participants' weight loss [63]. Quantifiable activity tracking data, such as calorie burn or weekly changes in preferred routines resulting from tracking metrics, encouraged the participant's regular self-monitoring behavior adoption [64].

In summary, for weight and obesity management, a comprehensive approach with the optimal use of technology is effective and should be incorporated in weight loss practices and clinics. Inducing motivation and determination into the patients along with frequent and effective communication will keep them in the right direction regardless of the rate at which they lose weight. Technology is the best vehicle to deliver effective and sustained results.

IV. CONCLUSION AND FUTURE WORK

Consistent weight loss needs a multidisciplinary approach. Determination, motivation, effective communication, diet, physical activity, accountability, and tailored guidance and education are vital elements. Digital therapeutics for obesity have the potential to significantly improve patient adherence and treatment outcomes and can deliver a framework where these key elements asynchronously and coherently work for the best patient-HCP engagement and optimal patient outcome. It is a promising way to address the global pandemic of obesity and warrants significant investment for further development. AI plays a vital role in delivering tailored guidance and education to the patient and catalyze the effectiveness of DTx. With a properly designed and operated digital therapeutics platform surpassing the benchmark of 10% weight loss in 24 weeks is feasible with an effective diet and physical plan along with the vital elements of a multidisciplinary approach, which a DTx platform can deliver effectively using ES.

Our future work is focused on studying how SureMediks can effectively complement medical weight loss with Glucagon-Like Peptide (GLP-1) and similar weight loss medication and post metabolic surgery weight loss.

ACKNOWLEDGMENT

We would like to thank our participants and their commitment to their dedicated time. Their compliance and dedications to our field trial were commendable. We would also like to thank Mr. Kannan Palaniswami for his efficient technical support and Ms. Helga Andersen for her moral support for completing the field trial and compilation of this paper.

REFERENCES

- [1] S. Khokhar, J. Holden, C. H. Toomer, L. Whitby, "Efficacy of an AI-Based Weight Loss Digital Therapeutics Platform: A Multidisciplinary Perspective" The First International Conference on AI-Health (AIHealth 2024) IARIA, Mar. 2024, ISBN: 978-1-68558-136-7.
- [2] World Health Organization. Obesity. (2021). <https://www.who.int/news-room/facts-in-pictures/detail/6-facts-on-obesity> [Accessed May 5, 2024].
- [3] Milken Institute. Economic impact of excess weight now exceeds \$1.7 trillion. Science Daily. (2018). <https://www.sciencedaily.com/releases/2018/10/181030163458.htm> [Accessed May 5, 2024].
- [4] Overweight and Obesity Statistics. National Institute of Diabetes and Digestive and Kidney Diseases. (2020). <https://www.niddk.nih.gov/health-information/health-statistics/overweight-obesity> [Accessed May 5, 2024].
- [5] C. Boutari, C.S. Mantzoros, "A 2022 update on the epidemiology of obesity and a call to action: as its twin COVID-19 pandemic appears to be receding, the obesity and dysmetabolism pandemic continues to rage on," *Metabolism*, vol. 133, pp. 155-217, 2022.
- [6] G.A. Bray, et al., "The science of obesity management: An endocrine society scientific statement," *Endocr Rev.*, vol. 39 no. 2, pp. 79-132, 2018.
- [7] Gambineri, A., Laudisio, D., Marocco, C., Radellini S., Colao, A., Savastano, S. (2019) Obesity Programs of nutrition, Education, Research and Assessment (OPERA) group. Female infertility: which role for obesity? *Int J Obes Suppl.* 9(1):65-72.
- [8] N. Stein, K.A. Brooks, "Fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults," *JMIR Diabetes*, vol. 2, no. 2, pp. e28, 2017
- [9] E. M. Forman, et al., "Using artificial intelligence to optimize delivery of weight loss treatment: Protocol for an efficacy and cost-effectiveness trial." *Contemp Clin Trials*. Vol. 124, pp. 107-129, 2022.
- [10] C.A. Pellegrini, A.F. Pfammatter, D.E. Conroy, B. Spring, "Smartphone applications to support weight loss: current perspectives," *Adv Health Care Technol.* vol. 1, pp. 13-22, 2015.
- [11] R.A. Krukowski, J. Harvey-Berino, T. Ashikaga, C.S. Thomas, N. Micco, "Internet-based weight control: the relationship between web features and weight loss," *Telemed J E Health.*, vol. 14, no. 8, pp. 775-782, 2008.
- [12] L. Hebden et al., "Mobile health intervention for weight management among young adults: A pilot randomized controlled trial," *J. Hum. Nutr. Diet.*, vol. 27, no. 4, pp. 322-332, 2014.
- [13] J.S. Hong, C. Wasden, D.H. Han, "Introduction of digital therapeutics," *Comp. Methods Programs Biomed.*, vol. 209, pp. 106-319, 2021.
- [14] N. Hinchliffe, M.S. Capehorn, M. Bewick, J. Feenie, "The potential role of digital health in obesity care," *Adv. Ther.* vol. 39, no. 10, pp. 4397-4412, 2022.
- [15] L.J. Appel et al., "Comparative effectiveness of weight-loss interventions in clinical practice," *N Engl J Med.* vol. 365, no. 21, pp. 1959-1968, 2011.
- [16] J. Antoun, H. Itani, N. Alarab, A. Elshamawy, "The effectiveness of combining nonmobile interventions with the use of smartphone apps with various features for weight loss: systematic review and meta-analysis," *JMIR Mhealth Uhealth.* vol. 10, no. 4, p.e35479, 2022.
- [17] R.W. Jeffery, R.R. Wing, R.R., Mayer, "Are smaller weight losses or more achievable weight loss goals better in the long term for obese patients?" *J. Consult. Clin. Psychol.*, vol. 66, no. 4, pp. 641-645, 1998.
- [18] K. Uetake, N. Yang, "Success breeds success: Weight loss dynamics in the presence of short-term and long-term goals," Canadian Centre for Health Economics. 2017.
- [19] E.S. Wise, J. Felton, M.D. Kligman, "Exponential decay modeling can define parameters of weight loss trajectory after laparoscopic Roux-en-Y gastric bypass," *Am J Surg.* vol. 216, no. 1, pp.120-123, 2017.
- [20] S. Khokhar, "Methods and systems for interactive weight management," U.S. Patent No 11,353,358. Washington, DC: U.S. Patent and Trademark Office. 2022
- [21] R.A. Krukowski, J. Harvey-Berino, T. Ashikaga, C.S. Thomas, N. Micco, "Internet-based weight control: the relationship between web features and weight loss," *Telemed J E Health.*, vol. 14, no. 8, pp. 775-782, 2008.
- [22] E.M. Forman, et al., "Can the artificial intelligence technique of reinforcement learning use continuously-monitored digital data to optimize treatment for weight loss?," *J Behav Med.* vol. 42, no. 2, pp. 276-290, 2019.
- [23] T.S. Sayed, "Application of expert systems or decision-making systems in the field of education," *Information technology in industry*, vol. 27, no. 3, pp. 1176-1183, 2021.
- [24] C.L. Chi, W. Nick Street, J.G. Robinson, M.A. Crawford, "Individualized patient-centered lifestyle recommendations: an expert system for communicating patient specific cardiovascular risk information and prioritizing lifestyle options.," *Journal of biomedical informatics*, vol. 45, no. 6, pp. 1164-1174, 2012.
- [25] P. Valsalan, N. U. Hasan, U. Farooq, M. Zghaibeh, I. Baig, "IoT Based Expert System for Diabetes Diagnosis and Insulin Dosage Calculation. Healthcare (Basel, Switzerland)," vol. 11, no.1, pp. 12-21, 2022.
- [26] P. Jirků, M. Andel, P. Hájek, P. Expertní, "Systém v diagnostice diabetu [An expert system for the diagnosis of diabetes]," *Casopis lekaru ceskych*, vol. 125, no. 2, pp. 49-52, 1986.
- [27] L. Hu, et al., "Challenges of conducting a remote behavioral weight loss study: Lessons learned and a practical guide," *Contemp Clin Trials*. 108:106522, 2021.

- [28] L.J. Appel, et al., "Comparative effectiveness of weight-loss interventions in clinical practice," *N Engl J Med.* vol. 365, no. 21, pp. 1959-1968, 2011.
- [29] G.C. Williams, V.M. Grow, Z.R. Freedman, R.M. Ryan, E.L. Deci, "Motivational predictors of weight loss and weight-loss maintenance," *J Pers Soc Psychol.* vol. 70, no. 1, pp.115-126, 1996.
- [30] S. Soini, P. Mustajoki, J.G. Eriksson, "Long-term weight maintenance after successful weight loss: Motivational factors, support, difficulties, and success factors," *Am J Health Behav.* vol. 42, no. 1, pp. 77-84, 2018.
- [31] D.S. West, et al., "A motivation-focused weight loss maintenance program is an effective alternative to a skill-based approach," *Int. J. Obes. (Lond).*, vol. 35 no. 2, pp. 259-269, 2011.
- [32] R.R. Wing, R.W. Jeffery, "Benefits of recruiting participants with friends and increasing social support for weight loss and maintenance," *J Consult Clin Psychol.* vol. 67, no. 1, pp. 132-138, 1999.
- [33] T.W. Bradford, S.A. Grier, G.R. Henderson, "Weight loss through virtual support communities: A role for identity-based motivation in public commitment," *Journal of Interactive Marketing.* vol. 40, no. 1, pp. 9-23, 2017.
- [34] K.O. Hwang, et al., "Social support in an Internet weight loss community," *Int J Med Inform.* vol. 79, no. 1, pp. 5-13, 2010.
- [35] D.F. Tate, R.R. Wing, R.A. Winett, "Using Internet technology to deliver a behavioral weight loss program," *JAMA.* vol 285, no. 9, pp. 1172-1177, 2001.
- [36] M. Islam, T. Poly, B. Walther, Y. Li, "Use of mobile phone app interventions to promote weight loss: Meta-Analysis," *JMIR Mhealth Uhealth.* vol. 8, no.7, p. e17039, 2020.
- [37] L.E. Burke, J. Wang, M.A. Sevvick, "Self-monitoring in weight loss: a systematic review of the literature" *J Am Diet Assoc.* vol. 111, no.1, pp. 92-102, 2010.
- [38] J.G. Thomas, D.S. Bond, "Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: Implications for JITAI optimization," *Health Psychol.* vol. 34S, pp. 1261-1267, 2015.
- [39] E.M. Winpenny, et al., "Changes in physical activity, diet, and body weight across the education and employment transitions of early adulthood: A systematic review and meta-analysis," vol. 21, no. 4, p. e12962, 2020.
- [40] K. Patrick, et al., "A text message-based intervention for weight loss: randomized controlled trial," *J Med Internet Res.* vol. 13, no. 11, p. e1, 2009.
- [41] G. Turner-McGrievy, D. Tate, "Tweets, apps, and pods: results of the 6-month mobile pounds off digitally (mobile pod) randomized weight-loss intervention among adults," *J. Med. Internet Res.*, vol. 13, no. 4, pp. 1-14, 2011.
- [42] J.F. Hollis et al., "Weight loss maintenance trial research group. weight loss during the intensive intervention phase of the weight-loss maintenance trial," *Am. J. Prev. Med.*, vol. 35, no. 2, pp. 118-126, 2008.
- [43] A. McTiernan, "Exercise effect on weight and body fat in men and women," *Obesity (Silver Spring)*, vol. 15, no. 6, pp. 1496-512, 2007.
- [44] M.C. Carter, V.J. Burley, C. Nykjaer, J.E. Cade, "Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial," *J Med Internet Res.* vol. 15, no. 4, p. e32, 2013.
- [45] J.S. Ingels, R. Misra, J. Stewart, B. Lucke-Wold, S. Shawley-Brzoska, "The effect of adherence to dietary tracking on weight loss: using hlm to model weight loss over time," *J Diabetes Res.* 6951495, 2017.
- [46] D.L. Swift, N.M. Johannsen, C.J. Lavie, C.P. Earnest, T.S. Church, "The role of exercise and physical activity in weight loss and maintenance," *Prog Cardiovasc Dis.* vol. 56, no. 4, pp. 441-447, 2014.
- [47] J.E. Donnelly, et al., "Appropriate physical activity intervention strategies for weight loss and prevention of weight regain for adults," *Med Sci Sports Exerc.* vol. 41, no. 2, pp. 459-471, 2009.
- [48] K. Frie, J. Hartmann-Boyce, S. Jebb, S., J. Oke, P. Aveyard, "Patterns in weight and physical activity tracking data preceding a stop in weight monitoring: observational analysis," *J. Med. Internet Res.*, vol. 22, no. 3, pp. 157-190, 2020.
- [49] Bethesda, "Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults --The evidence report," *National Institutes of Health. Obes Res.* 6, suppl. 2:51S-209S, 1998.
- [50] R.B. Kumar, L.J. Aronne, "Efficacy comparison of medications approved for chronic weight management," *Obesity (Silver Spring).* vol. 23 suppl. 1:S4-7, 2015.
- [51] K. Elfhag, S. Rössner, "Who succeeds in maintaining weight loss? A conceptual review of factors associated with weight loss maintenance and weight regain," *Obes Rev.* vol. 6, no. 1, pp. 67-85, 2005.
- [52] P.J. Teixeira, M.N. Silva, J. Mata, A.L. Palmeira, D. Markland, "Motivation, self-determination, and long-term weight control," *Int J Behav Nutr Phys Act.* vol. 9, no. 22, 2012.
- [53] G.E. Ames, et al., "Changing weight-loss expectations: a randomized pilot study," *Eat Behav.* vol. 6 no. 3, pp. 259-269, 2005.
- [54] J.A. Linde, R.W. Jeffery, E.A. Finch, D.M. Ng, A.J. Rothman, "Are unrealistic weight loss goals associated with outcomes for overweight women?," *Obes Res.* vol. 12, no. 3, pp. 569-576, 2004.
- [55] R.R. Wing, D.F. Tate, A.A. Gorin, H.A. Raynor, J.L. Fava, "A self-regulation program for maintenance of weight loss," *N Engl J Med.* vol. 355, no. 15, pp. 1563-1571, 2006.
- [56] M.R. VanDellen, E.S. Kuhl, M.M. Filbin, W.S. Tse, "Providing frequent group feedback improves success in weight loss competition between men and women," *Health Commun.* vol. 34, no. 8, pp. 895-902, 2019.
- [57] R.M. Ryan, E.L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being," *Am Psychol.* vol. 55, no. 1, pp. 68-78, 2000.
- [58] K. Patrick, et al., "Outcomes of a 12-month web-based intervention for overweight and obese men," *Ann Behav Med.* vol. 42, no. 3, pp. 391-401, 2011.
- [59] K. Chhabria, K.M. Ross, S.J. Sacco, T.M. Leahey, "The assessment of supportive accountability in adults seeking obesity treatment: psychometric validation study," *J Med Internet Res.* vol. 22, no. 7, p. e17967, 2020.

- [60] T.M. Leahey, R. Kumar, B.M. Weinberg, R.R. Wing, "Teammates and social influence affect weight loss outcomes in a team-based weight loss competition," *Obesity* (Silver Spring). vol. 20, no. 7, pp. 1413-1418, 2012.
- [61] M.J. Hutchesson, M.E. Rollo, R. Callister, C.E. Collins, "Self-monitoring of dietary intake by young women: online food records completed on computer or smartphone are as accurate as paper-based food records but more acceptable," *J Acad Nutr Diet*. vol. 115, no. 1, pp. 87-94, 2015.
- [62] M. Bastien, P. Poirier, I. Lemieux, J.P. Després, "Overview of epidemiology and contribution of obesity to cardiovascular disease," *Prog Cardiovasc Dis*. vol. 56, no. 4, pp. 369-381, 2014.
- [63] T.A. Wadden, J.S. Tronieri, M.L. Butryn, "Lifestyle modification approaches for the treatment of obesity in adults," *Am Psychol*. vol. 75, no. 2, pp. 235-251, 2020.
- [64] G.M. Turner-McGrievy, et al., "Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mHealth weight loss program," *J Am Med Inform Assoc*. vol. 20, no. 3, pp. 513-518, 2013.

Table 1: Weight loss statistics of 391 participants in their weight buckets and the overall weight loss. Weight units are in Kg

| | Weight Bucket 1 65-85 Kg | Weight Bucket 2 86-105 Kg | Weight Bucket 3 106-125 Kg | Weight Bucket 4 126-145 Kg | Weight Bucket 5 146-165 Kg | Weight Bucket 6 166-181 Kg | Overall |
|---|-----------------------------|---------------------------------|----------------------------------|----------------------------------|----------------------------------|-------------------------------------|---|
| No. of participants | 61 | 78 | 83 | 60 | 66 | 43 | 391 |
| Start weight mean, μ_{Start} | 76.0 | 95.8 | 113.5 | 135.2 | 154.8 | 172.1 | 124.6 |
| Start weight SD, σ_{Start} | 6.1 | 3.3 | 6.1 | 5.8 | 5.2 | 4.5 | 31.57 |
| End weight range | 52-79 | 69-94 | 85-119 | 101-131 | 118-151 | 132-159 | 52-159 |
| End weight mean, μ_{End} | 65.9 | 82.2 | 97.6 | 115.6 | 135.4 | 146.6 | 107.21 |
| End weight SD, σ_{End} | 6.2 | 6.6 | 6.8 | 7.2 | 8.6 | 8.1 | 24.7 |
| Weight loss range | 4-18 | 4-21 | 5-26 | 5-29 | 7-32 | 7-37 | 4-37 |
| Weight loss mean, $\mu_{Wt,loss}$ | 10.1 | 13.6 | 15.9 | 19.1 | 19.4 | 25.5 | 17.27 |
| Weight loss SD, $\sigma_{Wt,loss}$ | 3.4 | 4.4 | 5.2 | 5.9 | 6.8 | 7.3 | 7.0 |
| Weight loss percentage | 13.29 | 14.20 | 14.00 | 14.50 | 12.53 | 14.81 | 13.89 |
| %age weight loss 95% CI & p-value | 12.18 - 14.38 p < 0.0001 | 13.20 - 15.19 p < 0.0001 | 13.03 – 14.96 p < 0.0001 | 13.41 – 15.58 p < 0.0001 | 11.45 – 13.60 p < 0.0001 | 13.54- 16.07 p < 0.0001 | 13.45- 14.33 p < 0.0001 |
| Start BMI mean | 31.9 | 36.7 | 42.2 | 48.6 | 52.6 | 57.4 | 43.9 |
| Start BMI SD | 3.0 | 4.1 | 4.7 | 6.0 | 6.5 | 6.2 | 8.5 |
| End BMI mean | 27.6 | 31.5 | 36.3 | 41.8 | 45.9 | 48.8 | 37.6 |
| End BMI SD | 2.7 | 3.8 | 4.1 | 5.8 | 5.7 | 5.6 | 8.5 |
| BMI loss | 4.3 | 5.2 | 5.9 | 6.8 | 6.7 | 8.6 | 6.3 |

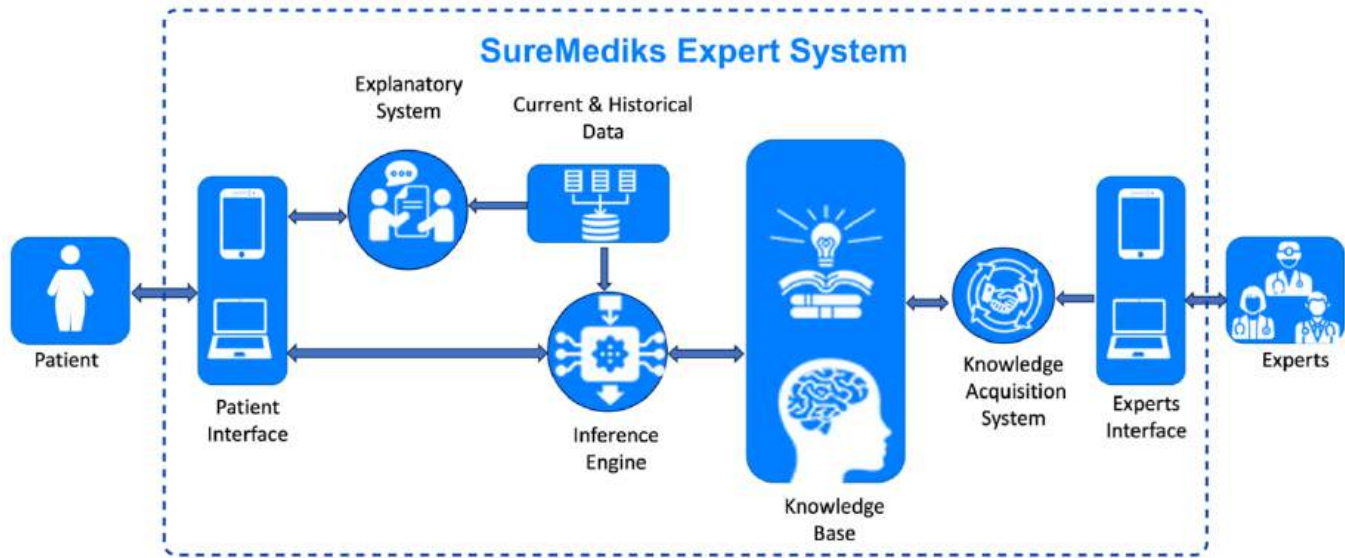


Figure 9: Distributed Architecture and the operation flow of our SureMediks Expert System.

| Table 2: Bivariate correlation between study’s additional parameters along with p-values | | | | | | |
|--|--------------------|------------------------|------------------------|-------------------------------|--------------------------|--------|
| | Weight loss %age | Age | Sub-goals reassignment | Accountability circle members | Challenges participation | Gender |
| Weight loss %age | 1 | | | | | |
| Age | -0.103 p < 0.05 | 1 | | | | |
| Sub-goals reassignment | -0.679 p < 0.05 | 0.029 p = 0.56 | 1 | | | |
| Accountability circle members | 0.523 p < 0.05 | - 0.050 p = 0.32 | -0.556 p < 0.05 | 1 | | |
| Challenges participation | 0.639 p < 0.05 | - 0.125 p < 0.05 | -0.380 p < 0.05 | 0.348 p < 0.05 | 1 | |
| Gender | 0.101 p < 0.05 | 0.039 p = 0.43 | 0.0284 p = 0.57 | -0.023 p = 0.64 | 0.060 p = 0.24 | 1 |

Considerations of Application Examples of Gait Analysis Using MediaPipe

Yasutaka Uchida

Dept. of Life Science
Teikyo University of Science
Tokyo, Japan
e-mail: uchida@ntu.ac.jp

Tomoko Funayama

Dept. of Occupational Therapy
Teikyo University of Science
Yamanashi, Japan
e-mail: funayama@ntu.ac.jp

Eiichi Ohkubo

Dept. of Life Science
Teikyo University of Science
Tokyo, Japan
e-mail: ohkubo@ntu.ac.jp

Yoshiaki Kogure

Professor Emeritus
Teikyo University of Science
Tokyo, Japan
e-mail: kogure@ntu.ac.jp

Yasunori Fujimori

Dept. of Rehabilitation
Seirei Yokohama Hospital
yasu6622@outlook.jp

Ryota Kimura

Dept. of Rehabilitation
Seirei Yokohama Hospital
ryota151@infoseek.jp

Kohei Kimura

Dept. of Rehabilitation
Seirei Yokohama Hospital
13rp13@g.seirei.ac.jp

Abstract— MediaPipe, which enables skeletal analysis utilizing videos of walking subjects without markers, can be easily introduced into rehabilitation sites. The viewpoint is obtained from a single camera because the video utilized for analysis is captured from a smartphone or video camera. Therefore, the skeletal coordinates cannot be recognized during analysis, and the obtained coordinates are relative values. Here, we employed data obtained from MediaPipe to calculate stride length, walking speed, knee height change, and ankle angle and compared them with commercially available software. A pseudo-motor restriction was applied during the measurements by wearing a supporter on the right knee. We found that the presence of motion restriction and various parameters during gait can be obtained by combining the confirmation of gait trajectory with 3D analysis and clarifying the measurement range. In addition, patients undergoing rehabilitation were asked to wear walking aids, and the walking patterns before, during, and after wearing them were analyzed with MediaPipe to confirm the effectiveness of the walking-assisted aid attachment.

Keywords- MediaPipe; skeletal analysis; smartphone; 3D analysis; walking-assisted aid attachment; spectrometry.

I. INTRODUCTION

Measures are urgently required to prepare for a rapidly aging population. Falling is a significant problem among older people, as it causes them to be bedridden and places a heavy burden on their caregivers [1]-[8]. Therefore, motion analyses have been conducted utilizing insoles [9]-[11] and mat-like pressure sensors arranged two-dimensionally [12][13], wearable devices [14][15], and images [16][17]. For gait analysis, measurements utilizing multiple cameras with attached markers have been utilized in rehabilitation facilities, as represented by the Vicon system [18][19]. A camera called Kinect [20][22] has also been utilized to analyze the movement of a camera [22][23] linked to game software. However, problems remain, such as the need for an expensive dedicated system, space for recognizing the markers, and an operator familiar with the dedicated software. The rapid spread of smartphones has facilitated the capturing of pictures anytime and anywhere, and the threshold for capturing pictures has decreased. Moreover, excellent paid and easy-to-utilize software is available. However, introducing a system requires continuous expenditure, which is currently in a state where it cannot be sufficiently spread.

Software that can perform skeleton authentication includes OpenPose [24][25], developed by Carnegie Mellon University, and MediaPipe [26]-[30], released by Google. Both apply deep learning and have an advanced certification system.

At Global Health 2022, we presented a basic application of MediaPipe in rehabilitation. Furthermore, for the utilization of a walking-assist device, we reported that the effect continued approximately 5 min after the device was removed. This paper reports additional research results on the accuracy and application range of walking parameters obtained utilizing MediaPipe. If the analysis results from front filming can be utilized, data captured in hallways can also be utilized. Here, we performed the analysis utilizing front filming. MediaPipe, which can utilize Python, can be used by healthcare and welfare professionals who are not analysis experts. The ability to analyze videos from the front view via MediaPipe can also enable filming in rehabilitation rooms and hallways of hospitals and facilities; thus, healthcare and welfare professionals can utilize it themselves. This paper reports additional analysis results regarding the accuracy and applicability of parameters obtained from MediaPipe during walking and the results of walking analysis conducted on a participant who had atherothrombotic cerebral infarction.

Section II describes the experimental methodology, including the software utilized and the commercially available equipment and software. In Section III, we compare the results of measurements conducted on healthy individuals utilizing MediaPipe and commercially available software with the analysis results obtained utilizing MediaPipe during rehabilitation sessions for participants with atherothrombotic cerebral infarction. Section IV discusses the results obtained from the two types of software and the measurement results of individuals during rehabilitation, comparing the assessments from therapists with the results obtained from MediaPipe. Section V presents the conclusions.

II. EXPERIMENTS

We developed a method for analyzing MediaPipe data using a male in his 60s, represented by subject A, and validated it with a male in his 70s, represented by subject B, in rehabilitation training.

A. Establishment of Analysis Method

The participant was a male in his 60s. During the measurement, his right knee was fixed with a supporter to pseudo-restrict his movement, and a comparison was made utilizing the tool ORPHE ANALYTICS ® [31] to confirm the accuracy of the calculation results obtained from the 3D coordinate data obtained via MediaPipe. This software enabled us to attach ORPHE CORE®, which utilizes acceleration and angular rate meters, to the instep of a shoe using a special attachment that can be fixed to the shoelace. The data obtained from these sensors could be analyzed to

display various results. A photograph of the ORPHE CORE® attached to the shoelaces is presented in Figure 1.



Figure 1. ORPHE CORE® attached to shoelaces using an attachment.

Owing to the limitations of the laboratory, we could not utilize the timed up-and-go method, in which the participant stands from a seated position in a chair, walks around a cone 3 m away, and sits down again while being observed and photographed from the lateral direction. Therefore, to enable analysis utilizing ORPHE ANALYTICS from the front, we utilized an iPhone with ORPHE TRACK installed to receive acceleration signals from ORPHE CORE via Bluetooth; simultaneously, the data of the walking state were uploaded to the cloud service.

A video of the walking condition displayed on the ORPHE ANALYTICS screen was recorded at 720p utilizing the free AG Desktop Recorder software. This screen was loaded into MediaPipe and operated using Jupyter Notebook in Python to obtain 3D data corresponding to 33 locations on the Land Marker. From these data, we extracted data for the left and right hips, knees, ankles, and toes. Based on these data, Python displayed the trajectories of the knees and other parts of the body in 3D. In addition, Microsoft Excel was employed to calculate the change in the difference between the knee and ankle. The ankle angle was calculated with vectors connecting the ankle and knee and the ankle and toe.

Figure 2 presents examples of measurements utilizing MediaPipe. The image on the left depicts measurement without motion restriction, and the image on the right depicts that with motion restriction as indicated by the yellow circle.

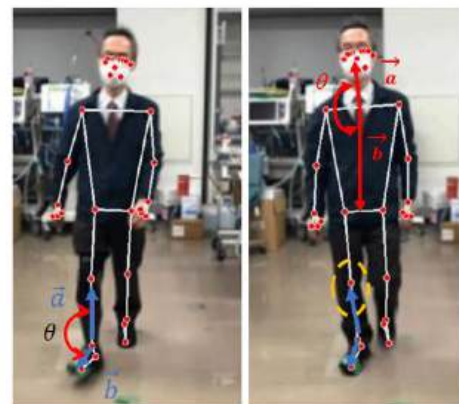


Figure 2. Examples of measurement results.

The supporter restricting movement is worn on the right knee, although it is difficult to see from the photo.

B. Gait Training Measurement

Gait was measured in subjects who are hospitalized and receiving training in physical therapy (PT) and occupational therapy (OT). The participant who underwent measurements at the hospital is a single individual with the following specific characteristics: a male in his 70s, right-handed. His diagnosis was atherothrombotic cerebral infarction (right putamen, corona radiata, subcortical parietal lobe); his disabilities were left hemiplegia, dysarthria, attention disorder, and left hemispatial neglect; his medical history included cerebral contusion, hypertension, and a left femoral neck fracture. Before the stroke, he was independent in his daily activities and occasionally went shopping at a local supermarket. Rehabilitation comprising PT, OT, and speech and hearing therapy commenced the day following the stroke. PT consisted of gait training five times a week for 40 min each. Sixteen days after the stroke, a gait evaluation utilizing a robot was conducted. Owing to the history of cerebral contusion, it is possible that the decline in motor function is not limited to the left side of the body but may also be present on the right side.

Figure 3 presents a photograph of the Orthobot® equipped for walking-assisted attachment.

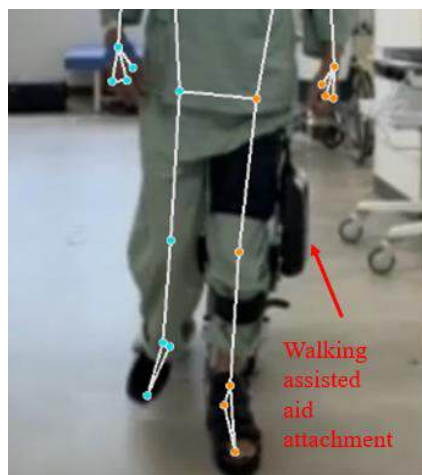


Figure 3. Photograph of walking with walking-assisted aid attachment.

In this paper, we refer to the subject of the analysis method as subject A and the subject of the gait training as subject B.

This study was approved by the Ethics Committee on Research with Humans as Subjects of the Teikyo University of Science. This experiment was conducted in accordance with the Declaration of Helsinki.

III. EXPERIMENTAL RESULTS

A. Analysis Method

A-1. Measurement of Strides

The right-foot ankle trajectory of subject A measured with the MediaPipe is illustrated in Figure 4. Only one round trip was utilized in the analysis. This is because plotting the trajectory of a round-trip walking state would cause the trajectories to cross each other, making them difficult to read. Because the camera is fixed, the coordinate data are x and y values corresponding to the 2D screen, except for the z-axis coordinates in the depth direction, which are relative, a characteristic of MediaPipe. Therefore, the data for one round trip were adopted here because performing a simple analysis is difficult. The amplitude increased until the change of direction occurred, indicating that the z-axis value did not change significantly during the change in direction. The area from the start of the walk to the change in direction was obtained.

The z-axis values for walking when approaching the camera are presented in Figure 5.

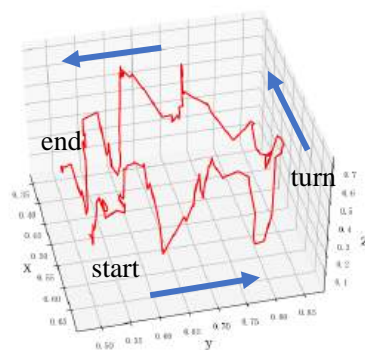


Figure 4. Right-foot ankle trajectory measured with MediaPipe.

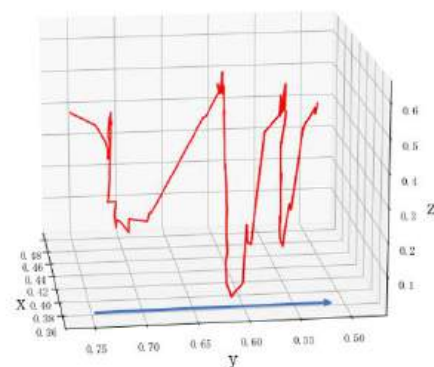


Figure 5. Z-axis values for walking when approaching camera.

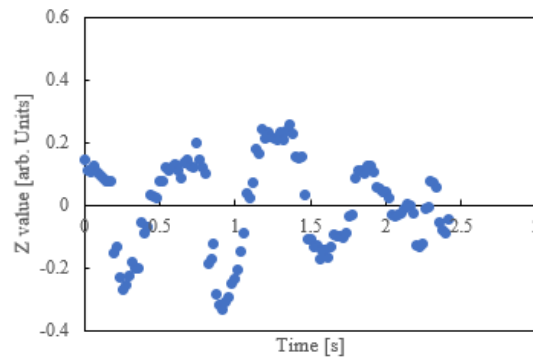


Figure 6. Result of modifying effect of walking direction.

Numerical data were displayed in Excel, and the inclination due to the walking direction was obtained; the corrected results are presented in Figure 6. Utilizing this diagram, we considered the point corresponding to the landing to be the minimum value based on the change in amplitude. As the figure demonstrates, the amplitude increased as the participant approached the camera, and the center of the amplitude also increased. Therefore, the center of the amplitude was approximated as increasing with a linear function, and the difference from the coordinate data was considered. The minimum value was set as the foot's landing point when the amplitude varied periodically, although a certain variation was observed. The actual measurement was obtained from the screen position, and the stride length was determined as the distance between the landing points. Walking speed was calculated from the respective times.

In the MediaPipe, the stride length was 0.80–0.90 m, and the velocity obtained was 0.8 m/s. The stride lengths of the left and right legs were 0.70 and 0.80 m, respectively. The stride length was larger in the right leg, with restricted motion owing to the hip motion.

The left and right stride lengths obtained from ORPHE were 0.75 and 1.0 m, respectively, larger than the values obtained from MediaPipe. In both cases, the value for the right leg was larger. The walking speeds on the left and right sides were 0.78 and 0.76 m/s, respectively, which were almost the same.

A-2. Calculation of Knee Height

Figure 7 presents the results of the changes in the right and left knee heights using MediaPipe during the gait of subject A. Red indicates the right knee with limitation of motion by the supporter, and blue indicates the left knee without limitation of motion. Here, the results are also presented from the beginning of walking to turning, considering the effect of rotation.

The results of the ORPHE ANALYTICS measurement of knee height are presented in Figure 8. The upper-left corner of the screen is the origin, and the maximum y-axis corresponding to the vertical direction is represented by 352 pixels. Therefore, the height of the right knee, which is a small

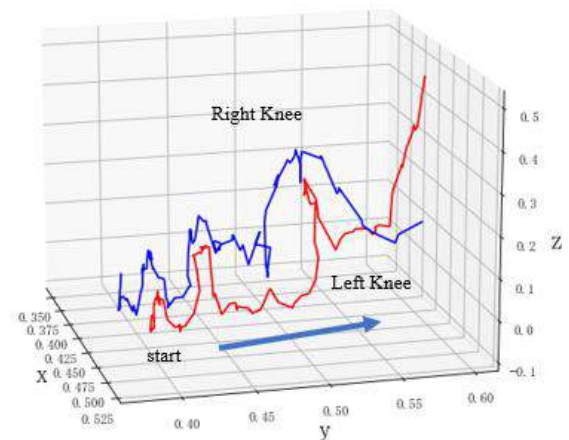


Figure 7. Height of knee position evaluated with MediaPipe.

value in the figure, had a higher value. The horizontal axis represents the number of measurement points for data analysis and not the time axis. The y-axis value for the x-axis, which corresponds to the direction of motion, changed significantly when the participant changed the gait direction during the measurement. When comparing knee heights, we utilized moving images and changes in the x-axis direction, which is characteristic of a change in direction. We deleted data from points in the range that appeared to indicate a change in the turn direction.

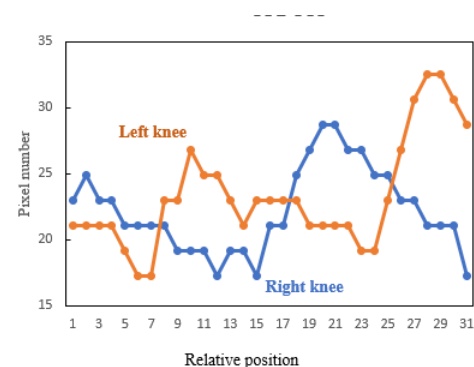


Figure 8. Knee height measurements with ORPHE ANALYTICS.

B. Feature Extraction of Gait Training Effect

B-1. Measurement Results of Blurring of Nasal Trajectory

We examined the instability of subject B undergoing rehabilitation while walking by analyzing the trajectory of the nose region. The assistant from behind is supporting the upper arm to prevent falls. Measurements were taken before, during, and after walking-assisted aid attachment. The image was captured from the front because of facility constraints. While shooting from the arrow direction is preferred, we considered the extent of assessment possible from the front image this time. The participant is wearing a mask, but MediaPipe recognizes the nose.

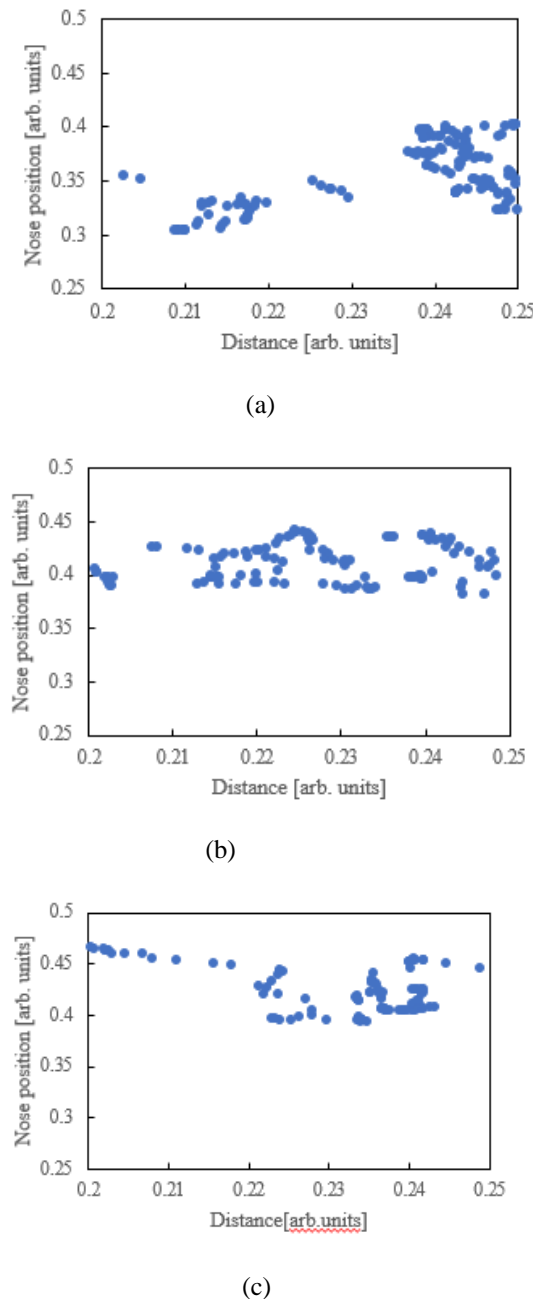


Figure 9. Blurring of nasal trajectory

Figure 9 illustrates the trajectory of the nose coordinates with respect to the walking direction. Because the walking direction is given as a relative value in MediaPipe, the same walking section was plotted for comparison. In the figure, (a) is before utilizing walking-assisted aid attachment, (b) is while utilizing walking-assisted aid attachment, and (c) is after utilizing walking-assisted aid attachment.

B-2. Measurement Results of Shoulder Tilt While Walking

Figure 10 (a) represents the subject B's results before utilizing the walking-assisted aid, (b) while utilizing the walking-assisted aid, and (c) after utilizing the walking-assisted aid. The changes in shoulder inclination during walking were determined based on the left shoulder.

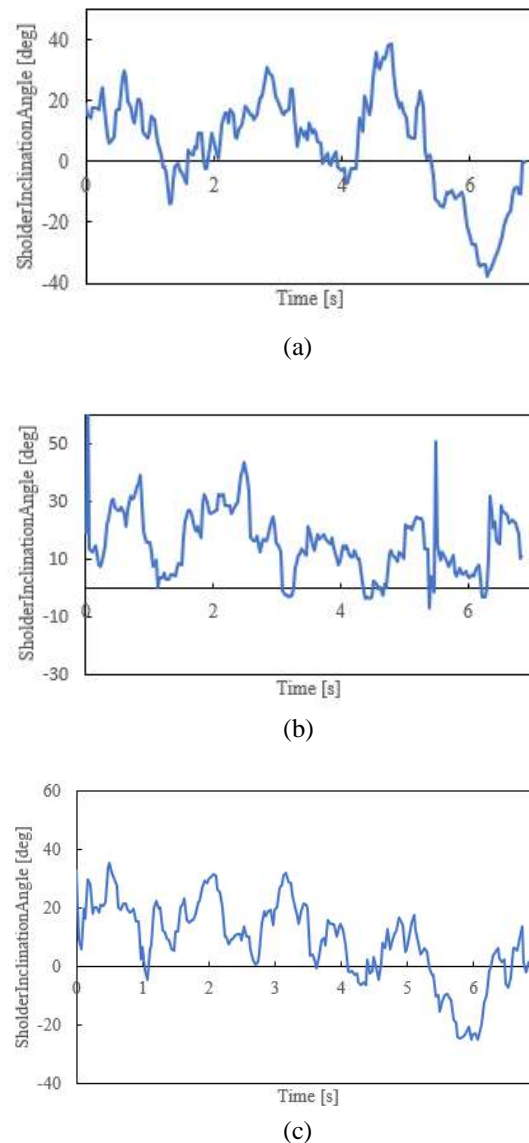


Figure 10. Shoulder tilt while walking.

B-3. Measurement Results of Hip Tilt While Walking

Figure 11 (a) represents the subject B's results before using the walking-assisted aid attachment, (b) while using the walking-assisted aid attachment, and (c) after using the walking-assisted aid attachment. The changes in hip inclination during walking were determined based on the left hip.

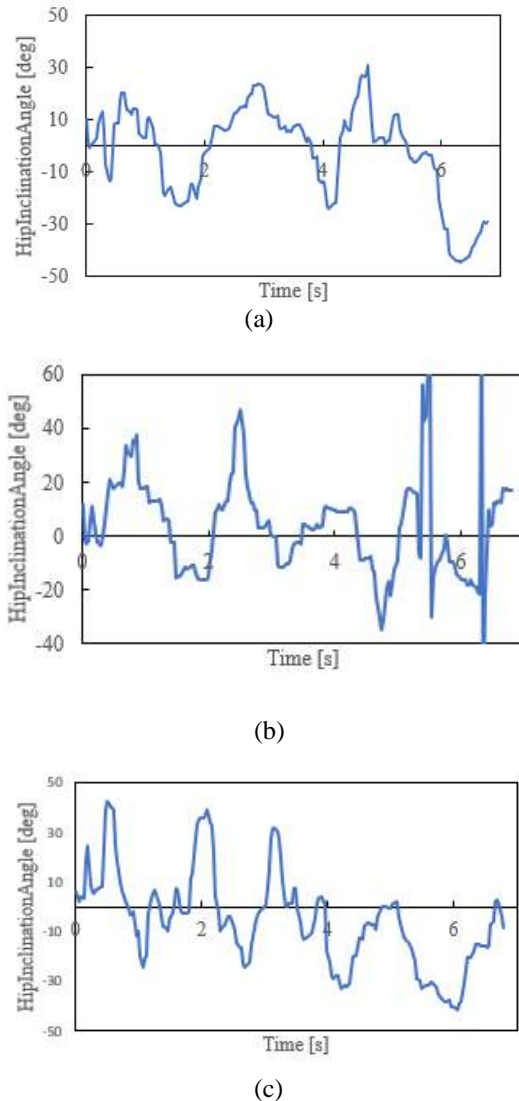


Figure 11. Hip tilt while walking.

The inclinations of the shoulder and hips demonstrated almost the same tendency.

B-4. Measurement Results of Neck Tilt While Walking

Walking affects not only the lower limbs but the whole body. This section focused on the head and neck parts of the body. An inverted pendulum model was utilized for walking. Based on the coordinate information obtained from MediaPipe, we calculated how much the subject B's neck tilts from the central axis of the torso when walking. Figure 2 depicts the line connecting the coordinates of the midpoint of the left and right shoulders with the coordinates of the midpoint of the left and right hips, and the line connecting the coordinates of the midpoint of the left and right hips with the left and right shoulders and the tip of the nose. The angle formed was defined as the neck angle, calculated with the formula for interior angles of vectors. The results are presented in Figure 12.

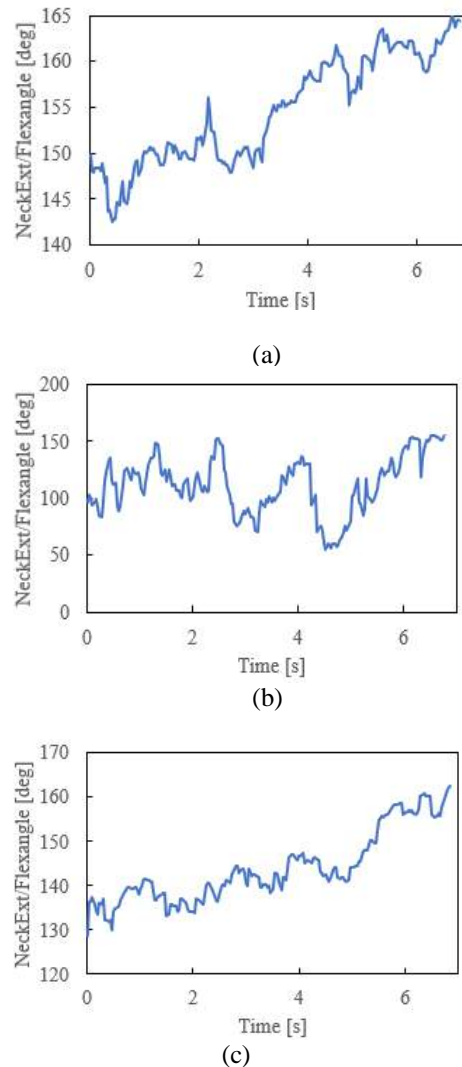
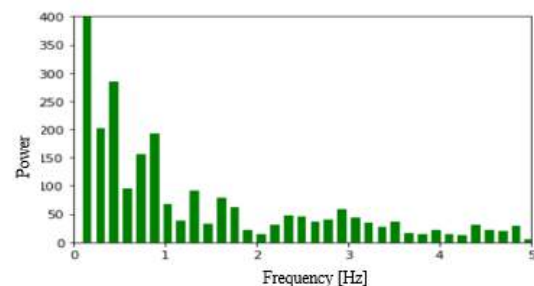


Figure 12. Neck angle while walking.

Because the vibration of the head tilt is considered to be different depending on each condition, the results of obtaining the discrete Fourier transform (DFT) spectrum are presented in Figure 13. The spectrum above 2 Hz becomes stronger after utilizing the walking aid attachment.



(a)

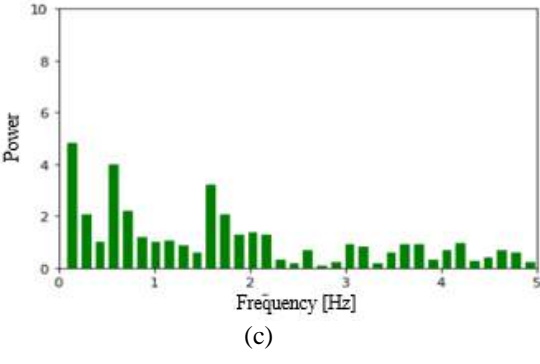
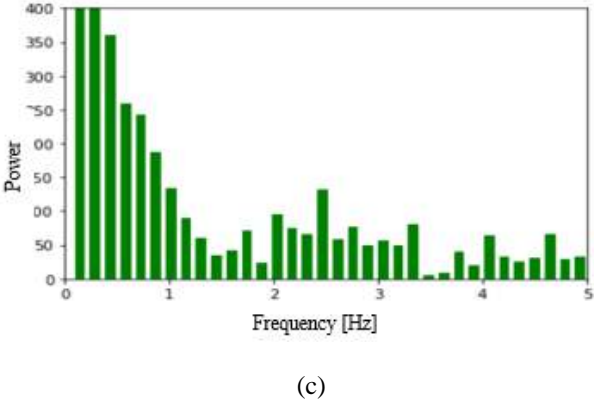
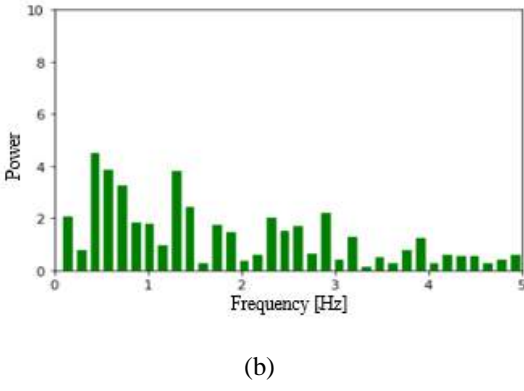
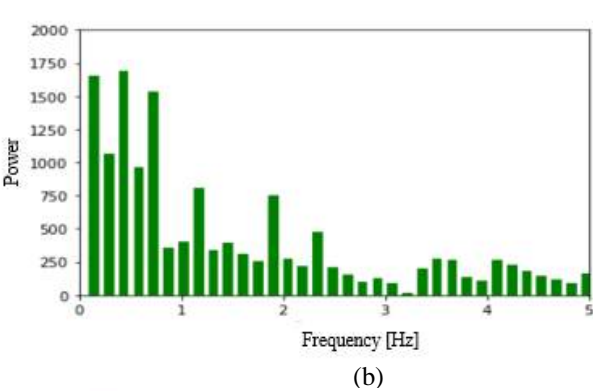


Figure 13. Neck DFT spectra.

Figure 14. Foot DFT spectra.

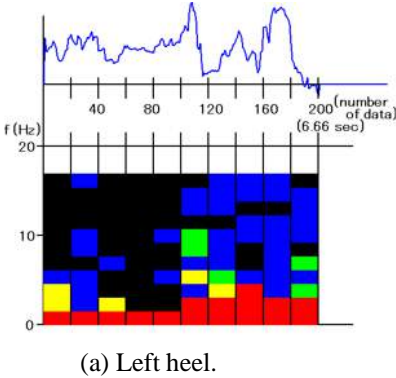
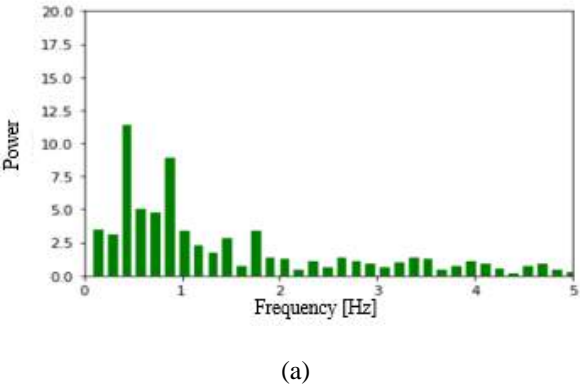
B-5. Measurement Results of Foot DTF Spectrum While Walking

DFT spectrum analysis of left foot was performed because it was observed that the left and right legs moved periodically during the subject B walking when wearing a walking aid attachment. The results are presented in Figure 14, where (a) represents the left foot, (b) is while utilizing the walking aid, and (c) is after.

The spectrum above 1.6 Hz becomes stronger after utilizing the walking aid attachment.

B-6. Determination of Rhythmic Heel Change by Time Series

Figure 15 illustrates the results of calculating a spectrogram [32] to clarify the frequency spectrum of left and right heel wobbling before and after wearing a walking aid attachment. (a) and (b) are the left and right heels before wearing them, and (c) and (d) are the left and right heels after removing the attached assisted device.



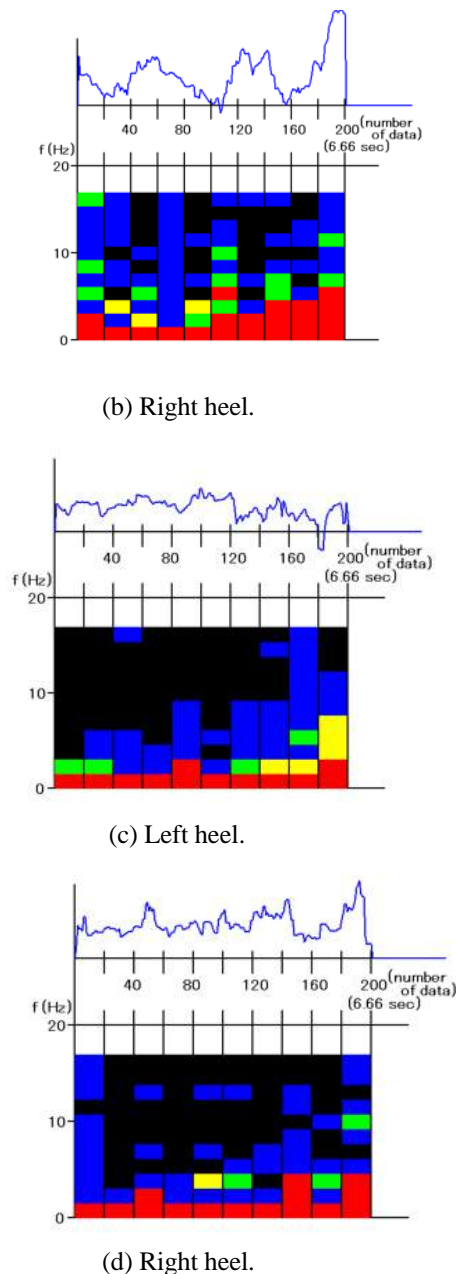


Figure 15. Spectrogram of heel position vibration

B-7. Stride and Gait Speed measurement results

The minimum points of the heel position in Figure 16 (a) to 16 (c) are considered as the ground contact point. Furthermore, a ruler was placed separately in the walking area to visually determine the heel position, and the relationship between walking step and heel position is illustrated in Figure 17. In this figure, (a), (b), and (c) respectively represent before utilizing the walking aid, during utilization, and after.

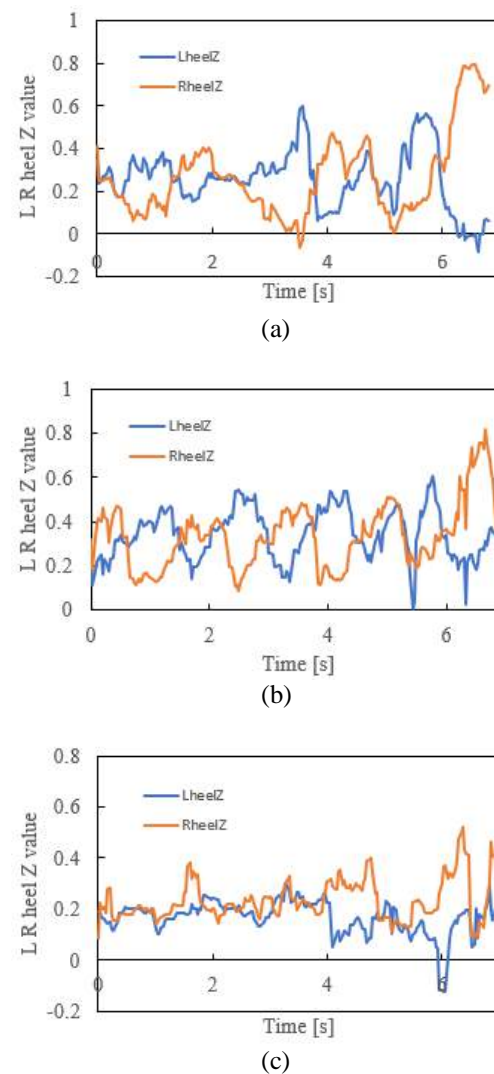
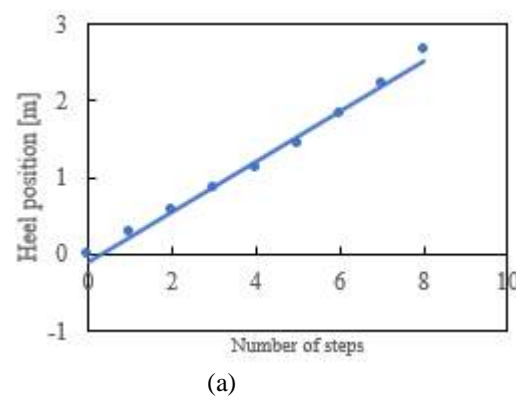


Figure 16. Positions of heels.



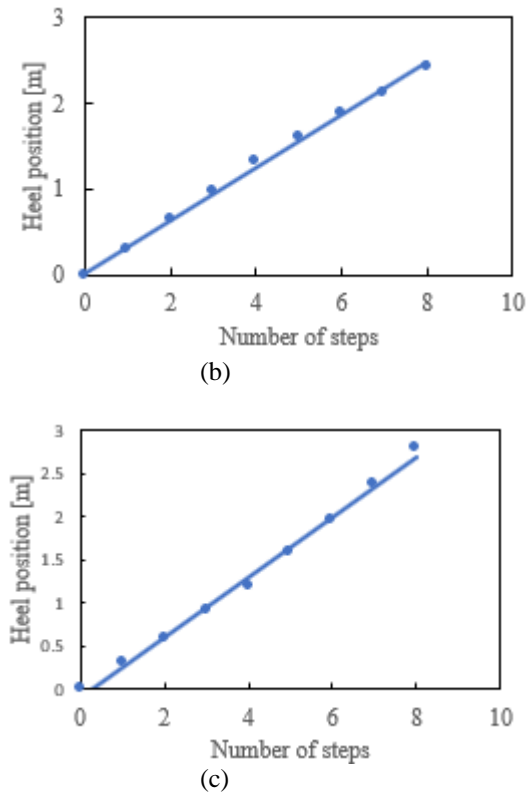


Figure 17. Heel position as function of steps.

Comparing the three figures, it is observed that the stride length improved before, during, and after utilizing the walking aid attachment.

IV. DISCUSSION

A. Differences in Measurement Results Based on Coordinate Axes Derived from Measurements in Subjects without Gait Disorders

In this experiment, the main reason for the difficulty in the analysis was that the measurement had to be performed under conditions where frequent changes in direction occurred owing to the limitations of the measurement location. Because the left and right foot coordinate values were different owing to the camera angle, simple comparison and analysis were impossible; a combination of 3D plots is considered necessary for motion analysis of the knee and ankle. In contrast, ORPHE ANALYTICS®, a commercially available software, provided data with correction. However, although it provided sufficient characteristic data of gait in terms of coordinate values, it was more difficult to handle than MediaPipe owing to the limited number of pixels; therefore, it may not have provided sufficient accuracy.

The data were limited to a specific individual because there was only one participant in this measurement. It would be necessary to increase the number of participants in the future. In addition, an accurate evaluation can be conducted

by changing the fixation position of the ORPHE CORE® to the inside of the shoe for measurement and comparison.

B. Consistency with Gait Assessment by Physical and Occupational Therapists and MediaPipe Results

B-1. Blurring of Nasal Trajectory

To evaluate the shaking while walking, we judged the state of the nose, the center of the left and right shoulders, etc., from images, and the PT and OT pointed out that the movement of the nose was the most noticeable. Table I presents the standard deviation of nasal part blur for each walking condition.

Table 1. Standard deviation of nasal part blur for each case

| Before assisted aid attachment [Ⓐ] | During assisted aid attachment [Ⓐ] | After assisted aid attachment [Ⓐ] |
|---|---|--|
| 2.90E-02 [Ⓐ] | 4.30E-04 [Ⓐ] | 1.40E-04 [Ⓐ] |

As presented in the table, the value was lowest when utilizing walking aids and remained low even after utilizing walking aids. This suggests that the effect of the walking aid continues for some time, even in persons with hemiplegia. The effect continued even after utilization, which was the same as in the case of healthy participants reported last year.

B-2. Shoulder Tilt While Walking

The PT and OT observation resulted in an assessment indicating that the right shoulder remained elevated regardless of the utilization of walking aids. The inclination derived from the coordinates of MediaPipe revealed an instantaneous state where the right shoulder dropped compared to the left shoulder within approximately 5 s from the start of walking. The photographs of these points are presented in Figure 18. However, beyond 5 s, data depicting the right shoulder dropping and abrupt peaks were observed. The instantaneous peaks were recognized as corresponding to the right shoulder dropping in relation to the landmark corresponding to the shoulder in MediaPipe.

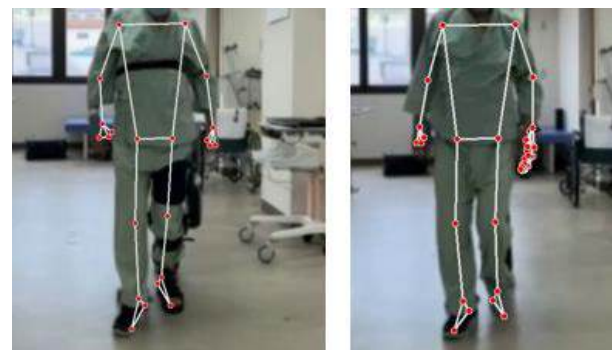


Figure 18. Photographs of right shoulder dropping.

It was decided to calculate and compare the area ratios to examine the discrepancies with the assessment. Because the data from MediaPipe were obtained at regular intervals, calculating the area by integrating the shoulder inclination angles over time was believed to enable a more accurate comparison. The area was determined by integrating the tilt angle up to 6.8 s, which is the same time for all measurements. Based on before walking-assisted aid attachment, the ratio was 1.0:4.0:3.3. The left side is decreasing because the area becomes a positive value in all cases when integrated. Based on the opinions from PT and OT, we reconsidered that the data before and during the utilization of the walking aid on the screen were calculated to be small. The results are greatly affected by the fact that the integrated area is small because the swing is large in the negative direction at a distance close to the camera. When we examined the video being analyzed in MediaPipe in detail, we found that after 5 s, it became clear that there was a possibility that the points of the marks were not sufficiently recognized. Based on this result, the area integrated over the 5 s before the sudden change was the smallest when utilizing a walking aid attachment.

B-3. Hip Tilt While Walking

Observation of the PT and OT revealed that the left hip joint was located lower than the right hip because the center of gravity shifted outward during the stance phase when the right lower limb was bearing the weight. Therefore, the position of the right hip joint is higher than that of the left hip joint. After walking-assisted aid attachment treatment, the body's weight-bearing ability when the hip joint extends during the late left stance has improved. Hence, it is now possible to bear weight during the early right stance, reducing lateral bending of the left trunk. The assessment was that the hip joint position was relatively higher on the right than on the left. The observation is correlated with the evaluation from Figure 11, which assessed the inclination of the hip based on the MediaPipe data, noting that the top amplitude is relatively small. During the utilization of walking-assisted aid, support is achieved during the left hip extension, equivalent to the stance phase late in the gait cycle. Hence, a significant amplitude is observed in the hip joint movement. This effect persists even in the lowermost part of Figure 11 after utilizing walking-assisted aid, indicating that the impact continues over time.

Figure 19 is a photograph comparing the degree of heel rise. The image illustrates the state of walking when the walking aid device is removed.

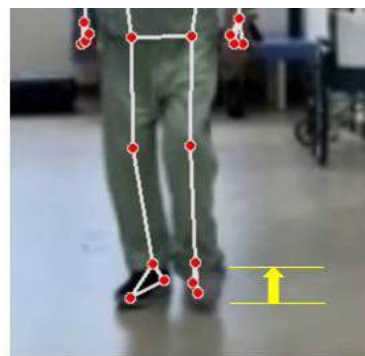


Figure 19. Photograph comparing heel height.

To determine whether the therapist's idea of improved weight bearing capacity in the late left stance phase could be judged by the heel rise, that is, the change in the final time of kicking the floor, we calculated the heel height with MediaPipe. However, it was impossible to determine that the heel position was higher after utilizing the robot than before. No evidence was found to support the therapist's consideration.

C. Neck Tilt While Walking

In measurements of a healthy individual last year, the utilization of walking aids resulted in a strong low-frequency signal spectrum below 0.5 Hz regarding the vertical oscillation of the neck during walking. However, in the case of the participant undergoing rehabilitation, spectra below 0.5 Hz in the low-frequency region were also observed; however, relatively, signal spectra of 2–3 Hz were evident. It is believed that the utilization of walking-assisted aid attachment resulted in periodicity during walking.

D. Spectrogram

The participant walked both before and after attaching the walking aid device, and spectrograms were analyzed after removing the device. Before utilizing the aid, relatively high-frequency components were observed, possibly due to instability. However, during walking without the aid, there was a reduction in relatively high-frequency components, while low-frequency components, indicative of stable walking, became predominant. Although spectrogram calculations were performed only once in this instance, making it difficult to generalize the findings, it is believed that the effectiveness of spectrograms in gait analysis has been demonstrated.

E. Gait Speed

The walking speed calculated from the slope of the step count and heel position presented in Figure 20, along with the time count from the video, indicated improvements before, during, and after the utilization of the walking aid, with speeds of 0.30 m/s, 0.36 m/s, and 0.45 m/s, respectively. This suggests that the reduction in foot instability during walking, possibly due to the utilization of the walking aid, could be one contributing factor.

V. CONCLUSIONS

The values obtained via calculation from MediaPipe, which can display skeletal certification, were compared with those of commercially available gait measurement systems to investigate the differences. The study revealed that the effects of different angles of video recording during gait should be considered in programming and in determining the results obtained with MediaPipe. However, MediaPipe can be an effective tool for determining walking conditions when the cost of implementing the system and the data required are limited.

Analysis results were obtained from assessments from PT and OT, and skeletal conditions and skeletal coordinates were obtained from media pipes regarding the walking status of participants with atherothrombotic cerebral infarction before, during, and after wearing walking aid devices. Although some results differed from the assessment because the videos were from all sides, we could obtain data that supported the PT and OT assessments in large part. By accumulating data in the future, we will be able to rehabilitate the media pipe. The application of this technology is expected to expand.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Numbers JP20K11924 and JP23K11207. We would like to thank the patients and the staff of Seirei Yokohama Hospital their cooperation in this study.

REFERENCES

- [1] Y. Uchida, T. Funayama, E. Ohkubo, and Y. Kogure, "Considerations for Applying MediaPipe to Gait Analysis," *GLOBAL HEALTH* 2023, pp. 12-17, IARIA.
- [2] L. G-Villanueva, S. Cagnoni, and L. Ascari, "Design of a Wearable Sensing System for Human Motion Monitoring in Physical Rehabilitation," *Sensors*, vol. 13, pp. 7735-7755, 2013.
- [3] Y.-L. Zheng et al., "Unobtrusive Sensing and Wearable Devices for Health Informatics," *IEEE Trans. Biomedical Engineering*, vol. 61, pp. 1538-1554, 2014.
- [4] M. M. Alam and E. B. Hamida, "Surveying Wearable Human Assistive Technology for the Life and Safety Critical Applications: Standards, Challenges and Opportunities," *Sensors*, pp. 9153-9209, 2014.
- [5] M. J. Deen, "Information and Communications Technologies for Elderly Ubiquitous Healthcare in a Smart Home," *Personal and Ubiquitous Computing*, pp. 573-599, 2015.
- [6] S. Hong and K. S. Park, "Unobtrusive Photoplethymographic Monitoring Under the Foot Sole while in a Standing Posture," *Sensors*, 3239, 2018.
- [7] V. Bucinskas et al., "Wearable Feet Pressure Sensor for Human Gait and Falling Diagnosis," *Sensors*, 5240, 2021.
- [8] P. M. Riek, A. N. Best, and R. Wu, "Validation of Inertial Sensors to Evaluate Gait Stability," *Sensors*, vol. 23, 1547, 2023.
- [9] T. Funayama, Y. Uchida, and Y. Kogure, "Assessment of Walking Condition Using Pressure Sensors in the Floor Mat," *The Eleventh, International Conference on Global Health Challenges*, Valencia, Spain, November 15, 2022, *global_health_2022_2_30_70023*.
- [10] T. Funayama, Y. Uchida, and Y. Kogure, "Detection of motion restriction with smart insoles," *Sensors & Transducers Journal*, Vol. 259, Issue 5, pp. 61-68, 2022.
- [11] S. Kim et al., "Assessing physical abilities of sarcopenia patients using gait analysis and smart insole for development of digital biomarker," *Scientific Reports*, 13:10602, 2023.
- [12] Y. Uchida, T. Funayama, K. Hori, M. Yuge, N. Shinozuka, and Y. Kogure, "Possibility of Detecting Changes in Health Conditions using an Improved 2D Array Sensor System," *sensors & Transducers*, Vol. 259, pp. 29-36, 2022.
- [13] T. Funayama, Y. Uchida, and Y. Kogure, "Step Measurement Using a Household Floor Mat and Shoe Sensors," *International Journal on Advances in Life Sciences*, vol. 15, No. 1&2, pp. 33-43, 2023.
- [14] S. Diaz, J. B. Stephenson, and M. A. Labrador, "Use of Wearable Sensor technology in Gait, Balance, and Range of Motion Analysis," *Applied Sciences*, vol. 10, 234, 2020.
- [15] T. Funayama, Y. Uchida, E. Ohkubo, and Y. Kogure, "Exploring the Potential of a Wrist-Worn Optical Sensor for Measuring Daily Life Activities," *GLOBAL HEALTH* 2023, pp.18-24, IARIA.
- [16] S. Majumder et al., "Smart Homes for Elderly Healthcare-Recent Advances and Research Challenges," *Sensors*, vol. 17, 2496, 2017.
- [17] K. Sato, Y. Nagashima, T. Mano, A. Iwata, and T. Toda, "Quantifying normal and parkinsonian gait features from home movies: Practical application of a deep learning-based 2D pose estimator," *PLOS ONE*, <https://doi.org/10.1371/journal.pone.0223549>, 2019.
- [18] M. Windolf, N. Gotzen, and M. Morlock, "Systematic Accuracy and Precision analysis of Video motion Capturing Systems-Exemplified on The Vicon-460 system," *Journal of Biomechanics*, vol. 41, pp. 2776-2780.
- [19] T. B. Rodrigues, D. P. Salgado1, C. O. Cathain, N. O'Connor, and N. Murray, "Human Gait Assessment Using a 3D Marker-less Multimodal Motion Capture System," *Multimedia Tools and Applications*, vol. 79, pp. 2629-2651, 2020.
- [20] P. Plantard, E. Auvinet, A. S. Le Pierres, and F. Multon, "Pose Estimation with a Kinect for Ergonomic Studies: Evaluation of the Accuracy Using a Virtual Mannequin," *Sensors*, pp. 1785-1803, 2015.
- [21] R. A. Clark, B. F. Mentiplay, E. Hough, and Y. H. Pus, "Three-Dimensional Cameras and Skeleton Pose Tracking for Physical Function Assessment: A Review of Use, Validity, Current Developments and Kinect Alternatives," *Gait & Posture*, vol. 68, pp. 193-200, 2019.
- [22] Y. Ma, K. Mithratatne, N. Wilson, Y. Zhang, and X. Wang, "Kinect v2-Based Gait Analysis for Children with Cerebral Palsy: Validity and Reliability of Spatial Margin of Stability and Spatiotemporal Variables," *Sensors*, vol. 21, 2104, 2021.

- [23] D. Imoto, S. Hirano, M. Mukaino, E. Saitoh, and Y. Otaka, "A Novel Gait Analysis System for Detecting Abnormal Hemiparetic Gait Patterns during Robo-assisted Gait Training : A Criterion Validity Study among Healthy Adults," *Frontiers in Neurorobotics*, 16:1047376, 2022
- [24] M. Ota, H. Tateuchi, T. Hahiguti, and N. Ichihashi, "Verification of validity of gait analysis systems during treadmill walking and running using human pose tracking algorithm," *Gait and Posture*, vol. 85, pp. 290-297, 2021.
- [25] Y. Saiki et al., "Reliability and validity of OpenPose for measuring hip knee ankle angle in patients with knee osteoarthritis," *Scientific Reports*, vol. 13, 3297, 2023.
- [26] V. Bazarevsky et al., "BlazePose: On-device Real-time Body Pose tracking," arXiv:2006.1204v1 [cs.CV] 2020.
- [27] G. Kaur, G. Jaju, D. Agawal, K. Lyer, and C. M. Prashanth, "Implementation of Geriatric Agility Detection Using MediaPipe Pose," *International Journal of Recent Advances in Multidisciplinary Topics*, vol. 3, 119, 2022, ISSN:2582-7839.
- [28] Y. Uchida, T. Funayama, and Y. Kogure, "Investigation of the Application of MediaPipe to Gait Analysis," *GLOBAL HEALTH 2022*, pp. 1-6, IARIA, 2022. ISBN: 978-1-61208-995-9.
- [29] Y. Uchida, T. Funayama, and Y. Kogure, "Possibility of Gait Analysis with MediaPipe and Its Application in Evaluating the Effects of Gait-assist Devices," pp. 44-54, IARIA, *International Journal on Advances in Life Sciences*, vol. 15, no. 1 & 2, 2023.
- [30] J.-L. Nhung, L.-Y. Ong, and M-C. Leow, "Comparative Analysis of Skelton-Based Human Pose Estimation," *Future Internet*, vol.14, 380, 2022.
- [31] Y. Uno et al., "Validity of Spatio-Temporal Gait Parameters in Healthy Young Adults Using a Motion-Sensor-Based Gait Analysis System (ORPHE ANALYTICS) during Walking and Running," *Sensors*, vol. 23, 331, 2023.
- [32] L. Xie et al., "TRLS: A Time Series Representation Learning Framework via Spectrogram for Medical Signal Processing," arXiv:2401.05431v1[eeess.SP] 2024.