GLOBAL HEALTH 2023

The Twelfth International Conference on Global Health Challenges

ISBN: 978-1-68558-112-1

September 25 - 29, 2023

Porto, Portugal

GLOBAL HEALTH 2023 Editors

Evgeniy Krastev, Sofia university St. Kliment Ohridski, Bulgaria

Hesham Ali, University of Nebraska Omaha, USA

GLOBAL HEALTH 2023

Forward

The Twelfth International Conference on Global Health Challenges (GLOBAL HEALTH 2023), held between September 25th and September 29th, 2023, continued a series of international events taking a global perspective on population health, from national to cross-country approaches, multiplatform technologies, from drug design to medicine accessibility, including everything under mobile, ubiquitous, and personalized characteristics of new age population.

Recent advances in technology and computational science influenced a large spectrum of branches in approaching population health. Despite significant progresses, many challenges exist, including health informatics, cross-country platforms interoperability, system and laws harmonization, protection of health data, practical solutions, accessibility to health services, and many others. Technological progress, personalized medicine, ambient assistance, and pervasive health, complement patient needs. A combination of classical and information-driven approaches is being developed, where diagnosis systems, data protection mechanisms, remote assistance and hospital-processes are converging.

We take here the opportunity to warmly thank all the members of the GLOBAL HEALTH 2023 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to GLOBAL HEALTH 2023. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the GLOBAL HEALTH 2023 organizing committee for their help in handling the logistics of this event.

We hope that GLOBAL HEALTH 2023 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress related to global health challenges.

GLOBAL HEALTH 2023 Chairs

GLOBAL HEALTH 2023 Steering Committee

Matthieu-P. Schapranow, Hasso-Plattner-Institut für Digital Engineering gGmbH, Germany
Lorena Parra Boronat, Instituto Madrileño de Investigación y Desarrollo Rural, Agrario y Alimentario And Universitat Politecnica de Valencia, Spain
Ah-Choo Koo, Research Centre for Interactive Media Faculty of Creative Multimedia, Multimedia University, Malaysia
Evgeniy Krastev, Sofia University St. Kliment Ohridsky, Bulgaria
Ozgu Can, Ege University, Turkey
Monique Chyba, University of Hawaiʻi at Mānoa, USA

GLOBAL HEALTH 2023 Publicity Chairs

Laura Garcia, Universitat Po, litecnica de Valencia, Spain
Lorena Parra Boronat, Universitat Politecnica de Valencia, Spain

GLOBAL HEALTH 2023
Committee

GLOBAL HEALTH 2023 Steering Committee

Matthieu-P. Schapranow, Hasso-Plattner-Institut für Digital Engineering gGmbH, Germany
Lorena Parra Boronat, Instituto Madrileño de Investigación y Desarrollo Rural, Agrario y Alimentario And Universitat Politecnica de Valencia, Spain
Ah-Choo Koo, Research Centre for Interactive Media Faculty of Creative Multimedia, Multimedia University, Malaysia
Evgeniy Krastev, Sofia University St. Kliment Ohridsky, Bulgaria
Ozgu Can, Ege University, Turkey
Monique Chyba, University of Hawaiʻi at Mānoa, USA

GLOBAL HEALTH 2023 Publicity Chairs

Laura Garcia, Universitat Po, litecnica de Valencia, Spain
Lorena Parra Boronat, Universitat Politecnica de Valencia, Spain

GLOBAL HEALTH 2023 Technical Program Committee

Wael Albayaydh, University of Oxford, UK
Mohamed Basel Almourad, Zayed University, Dubai, UAE
Kyota Aoki, Utsunomiya University, Japan
Marzieh Ayati, University of Texas Rio Grande Valley, USA
Samuel Bamgbose, Prairie View A&M University, USA
Panagiotis D. Bamidis, School of Medicine - Aristotle University of Thessaloniki, Greece
Oresti Banos, University of Granada, Spain
Valérie Barbier, Mater Research Institute, Australia
Vilmos Bilicki, University of Szeged, Hungary
Ozgu Can, Ege University, Turkey
Chih-Lin Chi, University of Minnesota, USA
Javier Civit, COBER S.L., Seville, Spain
Sandra Costanzo, University of Calabria, Italy
Steven A. Demurjian, The University of Connecticut, USA
Jitender Singh Deogun, University of Nebraska - Lincoln, USA
Ivan Evgeniev, Technical UniversityofSofia, Bulgaria
Azeema Fareed, COMSATS, Pakistan
Marta Fernandez, Official College of Nursing of Madrid, Spain
John Gialelis, University of Patras, Greece
Marie-Pierre Gleizes, IRIT, France
Alexandra González Aguña, University of Alcala, Spain
Adela Grando, Arizona State University | College of Health Solutions / Mayo Clinic, USA
David Greenhalgh, University of Strathclyde, Glasgow, UK
Vida Groznik, University of Primorska & University of Ljubljana, Slovenia
Teresa Guarda, Universidad Estatal Peninsula Santa Elena - UPSE / Universidad de las Fuerzas Armadas - ESPE / ALGORITMI Research Centre | ESPE | UPSE, Ecuador

Paulo Sousa Teixeira, Polythecnic Institute of Cávado and Ave, Portugal
Patrícia Isabel Sousa Trindade Silva Leite, Polythecnic Institute of Cávado and Ave, Portugal
Vasile Stoicu-Tivadar, University Politehnica Timișoara, Romania
Jiangwen (Javon) Sun, Old Dominion University, USA
Shorabuddin Syed, University of Texas for Health Science Center at San Antonio / University of Arkansas for Medical Sciences, USA
Les Sztandera, Thomas Jefferson University, USA
Vahid Taslimitehrani, Realtor.com, USA
Masatsugu Tsuji, Kobe International University / Osaka University, Japan
Niruwan Turnbull, Mahasarakham University, Thailand
Gary Ushaw, Newcastle University, UK
Santosh Vijaykumar, Northumbria University, UK
Massimo Villari, Università di Messina, Italy
Vivian Vimarlund, Linköping University, Sweden
Dongwen Wang, Arizona State University, USA
Mudasser F. Wyne, National University, USA
Farouk Yalaoui, University of Technology of Troyes, France
Erliang Zeng, University of Iowa, USA
Peng Zhao, Institute for Data Science & Informatics University of Missouri Columbia, USA
Qiang Zhu, University of Michigan - Dearborn, USA
Dina Ziadlou, International Society for Telemedicine and e-Health, ISfTeH, Switzerland / Colorado Technical University, USA

Copyright Information

Table of Contents

Implementation of Chatbots in Mental Healthcare: A Human Factor Perspective

Zhenqi Zhao
Department of Business Information Technology
University of Twente
Enschede, Netherlands
e-mail: z.zhao-2@student.utwente.nl

Onur Asan
School of Systems and Enterprises
Stevens Institute of Technology
Hoboken, USA
e-mail: oasan@stevens.edu

Mo Mansouri
School of Systems and Enterprises
Stevens Institute of Technology
Hoboken, USA
e-mail: mmansour@stevens.edu

***Abstract*—The global prevalence of mental health disorders, accentuated by the COVID-19 pandemic, emphasizes an urgent need for efficient solutions. This paper explores the potential of chatbots in addressing the mental health situation from a human factor perspective. Using stakeholder analysis, we identified and categorized various entities that influence or are influenced by the integration of chatbots in mental healthcare. A shaping forces analysis revealed driving factors such as increasing student mental health needs, existing therapy system limitations, and technological advancements. While chatbots serve as promising alternative solutions for mental health crisis, they come with challenges. This study offers a fresh perspective on understanding the interaction between chatbots and mental healthcare. It underscores the effects of this transition not only on directly affected stakeholders but also on participants who might be indirectly influenced within the system.**

***Keywords*—Chatbot; Mental Health; Artificial Intelligence; Human Factors.**

I. Introduction

Nearly 970 million individuals worldwide grapple with mental illnesses, with the most common symptoms of anxiety and depression [1]. The outbreak of COVID-19 in 2020 exacerbated the situation [2]. Within a year, the number of people with anxiety disorders increased by 26% and the number of people dealing with depressive disorders grow by 28% [3]. Despite the availability of effective treatments, existing mental health care solutions remain insufficient for the global population. Chatbot applications, however, offer a promising solution.

Recent years have seen a growing interest in the integration of chatbots within the healthcare domain, leading to a variety of applications [4]. In an effort to mitigate the mental health crisis and ensure more accessible care, several public projects have been launched. In Europe, for instance, the Mental Health Monitoring Through Interactive Conversations (MENHIR) project was initiated, offering 24-hour mental health support via a chatbot [5]. Similarly, Ulster University introduced a chatbot called iHelpr, designed primarily to facilitate users in self-assessing stress, anxiety, and depression levels [6]. However, these tools are not yet broadly adopted.

This paper is structured into 5 sessions. Following this introduction, Section 2 covers background on mental health challenges, chatbots in the mental health field, and advancements in related technologies. Section 3 presents a stakeholder analysis. Section 4 discusses the shaping forces influencing chatbot adoption, and Section 5 concludes with findings and recommendations for future work.

II. Background

The background section provides an overview of the current global mental health situation, the integration of chatbots in mental health, and the technological advances that power chatbots.

A. Current Mental Health Situation

Mental health conditions have become an increasing concern globally. Roughly one in eight individuals worldwide cope with mental disorders, with anxiety and depressive disorders being most common [7]. In the U.S., nearly 22.8% of adults faced mental illness in 2021 [8]. Despite the Netherlands having a slightly lower rate, the state of mental health is at its poorest in two decades, with 15% of the population admitting to psychological issues according to the Netherlands Statistical Office (CBS).

The COVID-19 pandemic only exacerbated these mental health challenges. Many studies point out a surge in depression, anxiety, and stress during the pandemic [2]. Factors such as enforced self-isolation and disruptions in daily routines might be attributed to increased loneliness, anxiety, insomnia, and even self-harm or suicidal tendencies [9].

Overall, the rising incidence and awareness of mental health issues reflect the need for timely solutions to address these challenges.

B. Chatbots in Mental Health

Conversational agents have been defined as ”software systems that mimic interactions with real people” [10] through various means such as text, spoken language, and gestures. A subset of these agents, chatbots, have gained attention in the healthcare sector due to their accessibility and efficiency. For instance, “Wysa” is a chatbot that engages users in written dialogues, recognizing their emotions and guiding them to build emotional resilience skills [11]. Another example, Emohaa, functions as a generative dialogue platform that facilitates open-ended conversations about emotional concerns, offering emotional support [12].

 ISBN: 978-1-68558-112-1

However, current healthcare chatbots serve as supplementary tools rather than replacing medical professionals [13]. The majority of these chatbots develop based on decision trees. Only a small percentage utilize more advanced machine learning techniques [14]. Relying heavily on decision trees limits user inputs to predefined phrases and words and constrains the user's initiative in the conversation [4].

C. Related Technologies

The field of Artificial Intelligence (AI) has witnessed rapid advancements in recent years, notably advancing chatbot capabilities through two elements: Machine Learning (ML) and Natural Language Processing (NLP) [15]. Machine learning is a statistical technique for interpreting data, recognizing patterns, and utilizing historical conversation to generate appropriate responses [16]. Natural language process focuses on analyzing the nature of human language and facilitating machine comprehension and interpretation of user inputs [15].

Within healthcare, natural language learning has a multi-faceted role, including interpreting user utterances, identifying significant changes, analyzing emotion, and extracting entities. Combined with ML, it can be used to predict or pinpoint behaviors in real-time, like identifying self-harm or suicide risk from interaction [15]. On the other hand, natural language learning can process unstructured patient notes and medical reports and transcribe patient discussions to provide unstructured data for facilitating relevant research [15].

However, the integration of chatbots into the mental health arena faces several challenges. One primary concern is the reliability of responses [4]. Due to the inherent design of machine learning, machine learning does not always guarantee consistent or desired outputs. In healthcare, particularly mental health, this unpredictability can pose risks. The integration of chatbots into daily clinical practice presents another challenge [16]. Gaining clearance from regulatory bodies often requires a thorough and extensive evaluation process. Lastly, the quality of chatbots is a salient issue. Many users feel that current chatbot dialogues lack depth and clarity, leading to confusion [17].

III. Stakeholder Analysis

A stakeholder analysis was conducted to better understand the diverse interests and potential impacts of involved groups. This section encompasses identifying possible stakeholders and classifying them based on the power-interest matrix.

A. Identification of Stakeholders

The current practice of chatbots in mental therapy is limited, necessitating a deeper exploration of their implications for stakeholders. Identifying the relevant stakeholders is the primary action. This study emphasizes three primary dimensions: the individual, organizational, and national levels, as illustrated in Figure 1.

At the individual level, the foremost affected stakeholders are the patients and the therapists. Subsequently, patients' family members bear indirect influences. Studies have shown that mental illness not only has the potential to cause physical illness in patients but also has the potential to jeopardize their lives. For instance, bipolar disorder has been linked to a higher cardiovascular disease rate [18]. Moreover, the study has demonstrated an association between suicidality and chronic insomnia [19]. Given the chronic and demanding nature of illness, the family caregivers often experience burnout and negative emotion, thereby imperiling their own well-being [20]. The widespread use of chatbots in the therapy field will grant more patients and their family members easier access to relevant information and treatments, helping their recovery. The health chatbots support therapists in managing their own health [21]; however, this trend may pose challenges to therapists in terms of job opportunities.

Figure 1. Identification of Stakeholders

On the organizational level, both insurance companies and universities are likely to be affected. With the growing recognition of mental health, people are leaning towards insurance options covering therapy costs. Given that chatbots typically charge less than traditional face-to-face therapists due to the lack of need for physical workspace and human resource saving, this could lead to reducing expenses for insurance companies. Simultaneously, for universities, these digitized conversations offer invaluable firsthand data for research purposes.

Lastly, from a national perspective, governments should also focus on the mental health system. According to World Health Organization, over 700,000 people were estimated to have committed suicide in 2019 [7], which is one of the leading causes of death among young people [22]. In the U.S., a significant 16.5% of teenagers aged from 6 to 17 were diagnosed with mental disorders in 2016 [23]. It underscores the necessity for governments to address both physical and mental well-being of their citizens.

 ISBN: 978-1-68558-112-1

B. Mapping of Stakeholders

Utilizing the power-interest matrix, stakeholders are categorized into four different groups based on their relative influences and interests in adapting the system. This mapping is depicted in Figure 2.

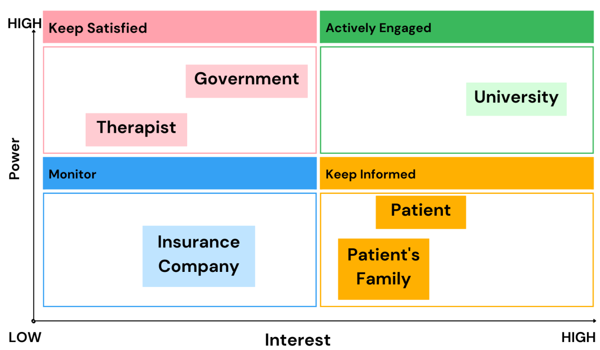


Figure 2. Power-Interest Matrix of Stakeholders

The government, a stakeholder at the national level, wields considerable influence but has a limited interest in chatbot development in the mental health field. The surge in mental health issues can be attributed to various causes, so the government has alternative solutions to address these issues. For instance, mental health challenges resulting from lockdown or quarantine are likely to subside post-pandemic. Therapists, on the other hand, have significant influence due to two primary reasons. First, developing chatbots requires extensive real data from past therapy sessions. Second, the sensitive nature of therapy demands rigorous validation before its widespread. However, many therapists favor direct interaction with patients and value immediate feedback, explaining their limited interest in chatbots.

Despite outstanding achievements in artificial intelligence and the deployment of chatbots by industries, the unclear commercial potential has limited private sector investment in this domain. In contrast, extensive research has been conducted in the academic sector, positioning universities as influential stakeholders. The wealth of data serves as an enticing factor for universities as well. Conversely, insurance companies exhibit both limited influence and interest due to a minor portion of insurance companies' claim expenses.

Patients and their families stand to benefit from the implementation of chatbots that may reduce associated costs. The limited understanding and social prejudices associated with mental disorders often deter many from seeking help. Chatbots offer a more private consultation environment, mitigating these concerns. However, due to the limited resources available to patients and their families, their influence remains modest.

IV. Shaping Forces Analysis

This shaping forces analysis examines key drives behind implementation of chatbots in the mental health field. As shown in Figure 3, three primary forces were discussed, including the demands of potential patients (students), the drawbacks of the current therapy system, and the burst of technologies.

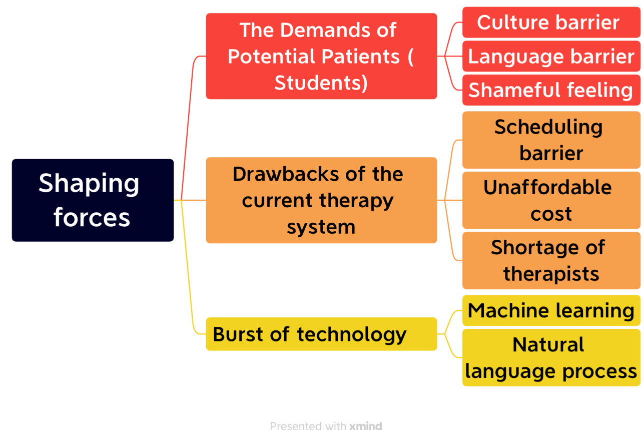


Figure 3. Shaping Forces Analysis

A. The Demands of Potential Patients (Students)

The attractiveness of diverse cultures has seen an uptick in students pursuing education abroad. However, this transition, coupled with rigorous academic demands, induces negative emotional states such as homesickness and loneliness. For instance, 45% of Chinese students at Yale exhibited depression symptoms, with 29% displaying anxiety symptoms [24]. Contributing factors include loss of familiar support networks, daily language barriers, and cultural adjustments [25]. Despite these challenges, international students often hesitate to seek help because of unfamiliar practice processes, cultural barriers, shameful feelings, and linguistic challenges.

Chatbots offer a potential solution by providing conversation in the student's native language. By leveraging data sourced from the native country, chatbots can effectively mitigate cultural barriers. Furthermore, the private platform can significantly reduce the stigma.

B. Drawbacks of the Current Therapy System

While seeking a recommendation for a cancer specialist is relatively straightforward, finding the right therapist is often more complex. Individuals' perceptions and feelings toward therapists can vary greatly, so it is necessary to make several attempts before finding a compatible fit. This journey, from searching for information to booking sessions, often becomes tedious, causing loss of motivation. Common barriers include the challenges of scheduling appointments and the overwhelming uncertainty of where to find help [25]. Another significant concern is cost. Notably, mental health treatments are not generally covered by private insurance policies, making it unaffordable for numerous families [26]. To compound these challenges, some countries, like China, face a shortage of qualified therapists and social workers [26].

 ISBN: 978-1-68558-112-1

Chatbots, with their 24/7 availability, can efficiently overcome the challenges of appointment scheduling, potentially reducing therapy costs, and bridging the professional shortage gap.

C. *Burst of Technologies*

Figure 4 shows the number of publications related to Artificial Intelligence in mental health domain from 2013 to 2022, sourced from Scopus. This data was gathered using search terms: (”Mental Health” AND (”Artificial Intelligence” OR ”Machine Learning” OR ”Natural Language Process”)). The search was confined to the past decade and restricted to English language publications.

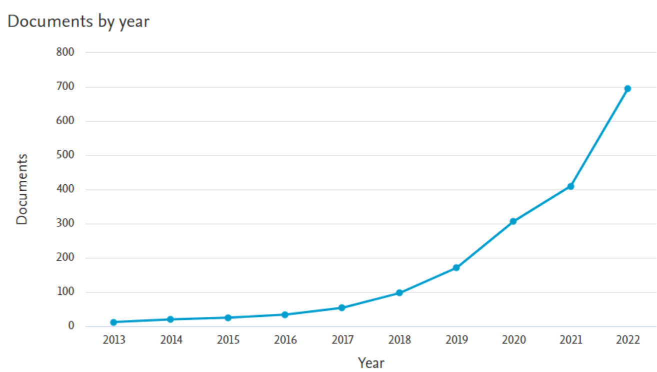


Figure 4. Publications Per Year

As mentioned before, the innovation of technology will be the gasoline for developing more advanced chatbots. A notable surge in publications is evident from 2017 onwards, offering a glimpse into the prospective growth trend in this area.

V. Conclusion and Future Work

The mental health situation is critical. As patient numbers surge, the apparent insufficiency of professionals becomes more pronounced. Despite society’s efforts to diminish the stigma associated with mental illness, many individuals still hesitate to seek help due to stigma. Additionally, the expense of therapy remains a financial burden for many. Chatbots offer a potential solution by providing 24-hour remote sessions. While earlier studies largely centered on the technological aspects of incorporating chatbots into mental healthcare or evaluated their impact solely from patient or therapist viewpoints, this paper broadens the scope. It not only considers the patients and therapists but also seeks to identify and analyze other stakeholders who may indirectly be impacted by, or have an influence on, the system.

However, potential challenges must be acknowledged. Data security stands out as a paramount concern. The related data requires meticulous storage to avoid any breaches and strict regulations to prevent any misuse. Another central issue is trust. Besides data security, patients must trust the advice offered by the chatbot and be willing to open up to chatbots. Since effective therapy hinges on trust, its absence could risk the therapeutic process being ineffective.

Future advancements should explore the potential of Large Language Models (LLMs) and Explainable AI. Large language models have demonstrated proficiency in simulating human-like conversations and achieving tasks in other domains. The integration could foster the development of AI-driven chatbots in therapeutic contexts. Explainable AI, meanwhile, can interrupt the decision-making process, minimizing inappropriate responses.

In conclusion, introducing chatbots in the mental health sector could promise unforeseen shifts, presenting both unique opportunities and challenges for the industry and patients.

References

[1] Institute of Health Metrics and Evaluation, May. 2022., “Global Health Data Exchange (GHDx).” [Online]. Available: https://vizhub.healthdata.org/gbd-results/
[2] C. Moreno et al., “How mental health care should change as a consequence of the covid-19 pandemic,” *The lancet psychiatry*, vol. 7, no. 9, pp. 813–824, 2020.
[3] World Health Organization, “Mental health and covid-19: early evidence of the pandemic’s impact: scientific brief, 2 march 2022,” Tech. Rep., 2022.
[4] L. Laranjo et al., “Conversational agents in healthcare: a systematic review,” *Journal of the American Medical Informatics Association*, vol. 25, no. 9, pp. 1248–1258, 2018.
[5] Z. Callejas, K. Benghazi, M. Noguera, M. I. Torres, and R. Justo, “Menhir: Mental health monitoring through interactive conversations,” *Procesamiento del Lenguaje Natural*, vol. 63, pp. 139–142, 2019.
[6] G. Cameron et al., “Assessing the usability of a chatbot for mental health care,” in *Internet Science: INSCI 2018 International Workshops, St. Petersburg, Russia, October 24–26, 2018, Revised Selected Papers 5*. Springer, 2019, pp. 121–132.
[7] World Health Organization, “Suicide worldwide in 2019: global health estimates,” Tech. Rep., 2021.
[8] Substance Abuse and Mental Health Services Administration, “Key substance use and mental health indicators in the united states: Results from the 2020 national survey on drug use and health,” Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, Tech. Rep., 2021.
[9] A. Kumar and K. R. Nayar, “Covid-19 and its mental health consequences,” *Journal of Mental Health*, vol. 30, no. 1, pp. 1–2, 2021.
[10] C. Khatri, A. Venkatesh, B. Hedayatnia, R. Gabriel, A. Ram, and R. Prasad, “Alexa prize state of the art in conversational AI,” *AI magazine*, vol. 39, no. 3, pp. 40–55, 2018.
[11] B. Inkster, S. Sarda, and V. Subramanian, “An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study,” *JMIR mHealth and uHealth*, vol. 6, no. 11, p. e12106, 2018.
[12] S. Sabou et al., “A chatbot for mental health support: exploring the impact of emohaa on reducing mental distress in china,” *Frontiers in Digital Health*, vol. 5, p. 1133987, 2023.
[13] M. Jovanović, M. Baez, and F. Casati, “Chatbots as conversational healthcare services,” *IEEE Internet Computing*, vol. 25, no. 3, pp. 44–51, 2020.
[14] A. A. Abd-Alrazaq, M. Alajlani, A. A. Alalwan, B. M. Bewick, P. Gardner, and M. Househ, “An overview of the features of chatbots in mental health: A scoping review,” *International Journal of Medical Informatics*, vol. 132, p. 103978, 2019.
[15] K. Denecke, A. Abd-Alrazaq, and M. Househ, “Artificial intelligence for chatbots in mental health: opportunities and challenges,” *Multiple perspectives on artificial intelligence in healthcare: Opportunities and challenges*, pp. 115–128, 2021.
[16] T. Davenport and R. Kalakota, “The potential for artificial intelligence in healthcare,” *Future healthcare journal*, vol. 6, no. 2, p. 94, 2019.
[17] A. A. Abd-Alrazaq, M. Alajlani, N. Ali, K. Denecke, B. M. Bewick, and M. Househ, “Perceptions and opinions of patients about mental health chatbots: scoping review,” *Journal of medical Internet research*, vol. 23, no. 1, p. e17828, 2021.

 ISBN: 978-1-68558-112-1

[18] D. J. Kupfer, "The increasing medical burden in bipolar disorder," *Jama*, vol. 293, no. 20, pp. 2528–2530, 2005.
[19] K. C. Cukrowicz, A. Otamendi, J. V. Pinto, R. A. Bernert, B. Krakow, and T. E. Joiner Jr, "The impact of insomnia and sleep disturbances on depression and suicidality." *Dreaming*, vol. 16, no. 1, p. 1, 2006.
[20] R. K. Chadda, "Caring for the family caregivers of persons with mental illness," *Indian journal of psychiatry*, vol. 56, no. 3, p. 221, 2014.
[21] A. Palanica, P. Flaschner, A. Thommandram, M. Li, and Y. Fossat, "Physicians' perceptions of chatbots in health care: cross-sectional web-based survey," *Journal of medical Internet research*, vol. 21, no. 4, p. e12887, 2019.
[22] World Health Organization, "World mental health report: transforming mental health for all," Tech. Rep., 2022.
[23] D. G. Whitney and M. D. Peterson, "Us national and state-level prevalence of mental health disorders and disparities of mental health care use in children," *JAMA pediatrics*, vol. 173, no. 4, pp. 389–391, 2019.
[24] X. Han, X. Han, Q. Luo, S. Jacobs, and M. Jean-Baptiste, "Report of a mental health survey among chinese international students at yale university," *Journal of American college health*, vol. 61, no. 1, pp. 1–8, 2013.
[25] H. Forbes-Mewett and A.-M. Sawyer, "International students and mental health," *Journal of International Students*, vol. 6, no. 3, pp. 661–677, 2019.
[26] A. Carbonell, J.-J. Navarro-Pérez, and M.-V. Mestre, "Challenges and barriers in mental healthcare systems and their impact on the family: A systematic integrative review," *Health & social care in the community*, vol. 28, no. 5, pp. 1366–1379, 2020.

 ISBN: 978-1-68558-112-1

Advancing Healthcare Tourism in Malaysia through the Implementation of the Flagship Medical Tourism Hospital Programme

Farizal Jaafar*, Lily Jasmin Abd Halim †, Ch'ng Zi Man †, Devaraj Subramaniam ‡, and Marale Atechian §
*Acting Chief Executive Officer, Malaysia Healthcare Travel Council, Kuala Lumpur, Malaysia
Email: farizal.j@mhtc.org.my
†Corporate Strategy, Malaysia Healthcare Travel Council, Kuala Lumpur, Malaysia
Email: lily.jasmin@mhtc.org.my; chng.zm@mhtc.org.my
‡APAC Consulting Team, IQVIA Solutions Malaysia, Malaysia
Email: devaraj.subramaniam@iqvia.com
§Principal Consultant, Joint Commission International, Riyadh, Kingdom of Saudi Arabia
Email: MAtechian@jcrinc.com

***Abstract*—Access to quality healthcare and advanced treatments has led to a growing demand for medical tourism. In 2019, an estimated 15 to 20 million individuals globally travelled abroad for medical purposes and the growth of this market is expected to double in the coming years. Patients are drawn to seeking healthcare and medical treatment abroad, primarily due to several factors, including the potential for cost savings, shorter waiting times for treatments, and a wider range of quality healthcare options. Recognising this trend, numerous countries are intensifying their efforts to develop and promote their respective medical tourism industries, with the goal of becoming the preferred destination for patients seeking treatments abroad. Notably, in certain countries, government agencies are spearheading initiatives to accelerate industry growth, which holds significant potential to contribute to the countries' economic development. Asia emerges as the principal beneficiary of the industry, garnering substantial demand from both domestic and international patients. In Malaysia, the Malaysia Healthcare Travel Council and the Ministry of Health Malaysia have developed the 5-year Malaysia Healthcare Travel Industry Blueprint with the aim to provide "the best healthcare travel experience". As part of the strategic initiatives outlined in the blueprint, a first-of-its-kind Flagship Medical Tourism Hospital Programme was designed as an effort to reinforce Malaysia's position as a leader in the medical tourism market, improve healthcare quality, and contribute to economic growth. This paper aims to discuss the Flagship Medical Tourism Hospital Programme and the assessment methodology employed for Malaysian hospitals participating in this programme. Additionally, strategies for capacity building and revenue growth within the programme, and benefits towards the nation and beyond are reviewed. The programme's success in developing flagship medical tourism hospitals within Malaysia positions it as an exemplary model for other nations seeking to establish their own thriving medical tourism industry.**

***Index Terms*—medical tourism; medical travel; healthcare destination; Malaysia healthcare.**

I. INTRODUCTION

The search for quality healthcare has propelled individuals to travel across international borders for medical treatment, also known as medical tourism. Driven by the needs of an increasingly elderly population for medical procedures with high quality service and challenges such as high treatment costs and long waiting times, the tourism and medical sectors of many countries have developed strategies to seize opportunities within this market [8]. In 2019, it was estimated that approximately 15 to 20 million individuals globally travelled abroad for medical purposes, contributing to a total market size of around US$ 33 billion [7]. The growth of this market is expected to double in the coming years, driven by the increasing demand from patients. While the market experienced a 10% annual growth rate from 2013 to 2018, it is projected to accelerate to approximately 20% annually from 2021 to 2027 [1]. The global medical tourism market stands poised to present remarkable opportunities for both the healthcare industry and nations at large, thereby holding the potential to revolutionise the way healthcare services are delivered and positively impact economic growth on a global scale.

Asia emerges as the principal beneficiary of the flourishing medical tourism industry, garnering substantial demand from both domestic and international patients. By 2025, the Asia-Pacific (APAC) region is forecasted to become the largest medical tourism market globally, accounting for approximately 34% of the global medical tourism revenue [2], with the primary customer groups originating from Asia Pacific (APAC), Europe, and the Middle East [1]. The region's favourable attributes, including advanced healthcare infrastructure, cutting-edge medical technologies, and a convergence of skilled healthcare professionals, position it as a highly sought-after destination for medical tourism. This trend not only signifies Asia's ascendancy in the global healthcare landscape but also highlights its pivotal role in driving economic growth and development through the influx of medical tourists from around the world.

In Malaysia, the government plays a significant role in growing the medical tourism market. The Malaysia Healthcare Travel Council (MHTC) [3], an agency established under the

 ISBN: 978-1-68558-112-1

purview of the Ministry of Health Malaysia [6], tasked to facilitate, and promote the healthcare travel industry, under the "Malaysia Healthcare" brand. In November 2021, MHTC [3] launched the Malaysia Healthcare Travel Industry Blueprint 2021 to 2025, officiated by the Minister of Health Malaysia [3], [4], [6]. The blueprint aims to enhance the healthcare travel ecosystem, strengthen the Malaysia Healthcare brand, and expand Malaysia's healthcare offerings to more targeted markets [4]. The aspiration is to provide the "Best Malaysia Healthcare Travel Experience by 2025" to healthcare travellers and achieve revised industry revenue of MYR 2.4 billion (US$ 513.4 million) from medical bills and spillover effect of more than MYR 9.6 billion (US$ 2 billion) to other industries (including ancillaries and tourism spending) by 2025 [1], [4]. Aligning to the Industry Blueprint, 'Malaysia Healthcare' remains committed in strengthening Malaysia's position as the leading global healthcare destination with synergistic public-private partnership, from government stakeholders to private industry players [4].

As an integral part of the five-year Malaysia Healthcare Travel Industry Blueprint, the FMTH Programme plays an instrumental role in transforming the healthcare travel ecosystem. With the debut of the FMTH Programme in the country, MHTC [3] aims to reinforce Malaysia's position as a safe and trusted destination for high-quality healthcare services and seamless end-to-end patient experience [4]. Malaysia's healthcare travel industry is unique in a sense that all private healthcare providers are stringently monitored by the Ministry of Health. The FMTH Programme will serve as the future torchbearer for Malaysia Healthcare, not only locally but also internationally, significantly contributing to the country's export service, with expectations for Malaysian healthcare to be significantly on par or above renowned international hospitals, namely in Thailand, Singapore, Korea, Taiwan, and more. This paper aims to introduce the FMTH Programme, discuss the assessment methodology employed for Malaysian private hospitals participating in this programme, strategies for capacity building and revenue growth within the programme, and benefits towards the nation and beyond that would advance Malaysia's position within the global medical tourism market.

The rest of the paper is structured as follows. Section 2 discusses the mechanisms of the FMTH Programme, its assessment methodology, the key enablers for capacity building and revenue growth, as well as the broader impact on Malaysia's healthcare travel industry and the nation. Section 3 concludes the paper and lists future work directions.

II. Discussion

A. Flagship Medical Tourism Hospital Programme Mechanism

The FMTH Programme acts as a catalyst to drive and elevate Malaysia's private healthcare services, in order to accelerate the medical tourism growth [4]. MHTC [3] aims to identify flagship medical tourism hospitals in the country and accelerate their capability building and revenue growth [4]. The goal of the programme is to establish and garner international recognition for the Flagship Medical Tourism Hospitals [4]. The national programme spans across five years, from 2021 to 2025, to achieve key targets of direct healthcare investment and increase healthcare travellers' revenue, which directly drives a four-fold multiplier effect to ancillaries and the tourism industry [9].

The FMTH Programme involves the invitation of 22 leading private hospitals from Klang Valley, Malacca, Penang, and Johor Bahru, which are part of MHTC's Elite Members, for participation [4]. The Members who agreed to participate are required to go through stringent assessments before identifying the four FMTH finalists in 2022. The assessments undertook data-driven methodology covering 51 metrics across the three key programme pillars of Medical Excellence, Service Excellence, and International Branding (see Figure 1). As a result, four hospitals were selected as FMTH finalists, namely Institut Jantung Negara or National Heart Institute in Kuala Lumpur, Island Hospital in Penang, Mahkota Medical Centre in Melaka, and Subang Jaya Medical Centre in Selangor [4].

The FMTH finalists are then required to undergo the three-year Acceleration phase for capability building and revenue growth acceleration from 2023 to 2025 [4]. The implementation focus during the Acceleration phase includes programme management, training and development by the industry experts, activation of programme enablers and global brand profiling of the programme and Flagship finalists [4]. The growth and development of the Flagship finalists will be continuously assessed against the international best practices and benchmarks based on the three strategic programme pillars.

The four Flagship finalists will be required to undergo a final round of assessment and selection in the third quarter of 2025 [4]. The Flagship hospital(s) will be eventually identified in the fourth quarter of 2025 [4]. The highest recognition of the Flagship hospital(s) will then be announced by the Honourable Prime Minister of Malaysia in 2025 [4].

B. Assessment Methodology

The overall assessment and selection of the FMTH Programme is carried out by an independent and specialised Programme Management Advisor (IQVIA Solutions Malaysia) and Programme Assessment Advisor (Joint Commission International (JCI)) [4]. The assessments are conducted based on data submission and rigorous onsite assessments (including leadership interviews, patient journey interviews, facility and environment of care visits, healthcare travellers' interviews and validation of data and documents), focusing on the three strategic pillars – Medical Excellence, Service Excellence and International Branding.

1) Medical Excellence: The overall selection process for the pillar of medical excellence involved a comprehensive review of specific parameters to ensure the highest standards of patient safety and quality of care. These included patient safety requirements such as clinical documentation adequacy, falls for outpatient care, medication error rate, return visits to the

Copyright (c) IARIA, 2023. ISBN: 978-1-68558-112-1

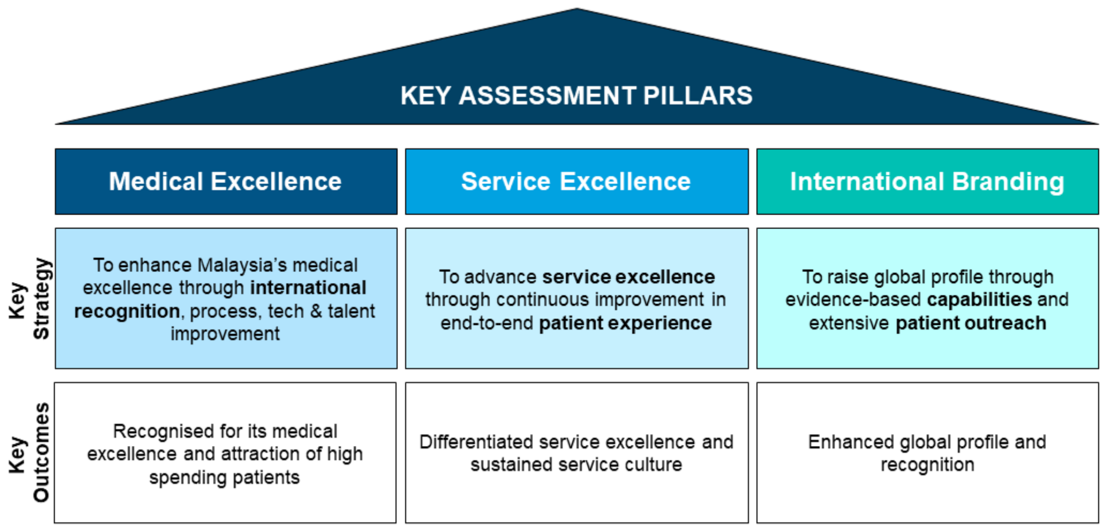


Fig. 1. Strategic pillars of the FMTH Programme.

emergency room, postponement of surgeries, unplanned return to the operating room, cardiopulmonary arrests outside critical care, healthcare-acquired infection rate, unplanned transfers to acute care for one-day surgeries, procedural sedation rate, complications in invasive procedures and 30-day postoperative mortality rate [4].

In addition, the process considered the availability of innovative medical technologies, international lab accreditation, centre of excellence international recognition, ongoing staff performance evaluation, job-related training hours per healthcare personnel and the percentage of medical staff who graduated from top global programmes [4].

Thereafter, an onsite assessment of the hospital's performance was conducted, covering the following five key dimensions [4]:

- **Strategy and Culture of Safety** is a key dimension demonstrating the commitment and support of the governing entity and the executive leadership to the strategic initiatives of the organization. Within the context of the FMTH, medical tourism is considered as a strategic goal, and it should be part of the governing entity priorities and a standing discussion item on the meetings' agenda. This dimension should also reflect the amount of investment and resources allocated to medical tourism.

  In addition, medical travellers need to receive care with the utmost safety. A culture of safety is key in providing a high-quality care and to create an environment where errors become learning points that drive improvement. This culture aims towards high reliability and pursuit of zero harm.

- **Healthcare traveller's medical journey** follows the experiences of medical travellers to evaluate the hospital's performance against international standards. It is a key approach to sequentially follow the course of care, treatment, and services received by the medical travellers from preadmission through post discharge and recovery. It evaluates the preadmission screening, assessments, the course of care, treatment and services provided to the medical travellers by and within the hospital by reviewing the medical travellers' medical record, conducting various interviews for medical travellers and the healthcare providers. The journey assesses the interrelationships between and among disciplines including the review of the pre-admission education, instruction, and the follow-up care post discharge, and addresses the communication methods including language preference and cultural considerations. Travellers' privacy and confidentiality is paramount and an expectation that the hospital should guarantee within this journey.

- **Environment of care and facility review** is a key dimension to ensure that the physical environment of the hospital is safe, secure, and does not pose a physical risk to patients, families, staff, and visitors. The hospital should have measures implemented to limit access to sensitive areas of the hospital.

  Hospital fires remain one of the top hazards for any healthcare organisation. Therefore, the expectation is to ensure the implementation of a robust fire safety programme including the early detection, containment and suppression of fire, proper maintenance of fire safety equipment and clear, unobstructed fire exits. This dimension also focuses on the safe maintenance and operation of the medical equipment and utilities.

 ISBN: 978-1-68558-112-1

The organisation should have redundancies in place in relation to the utility system to support essential services and ensure it is operated effectively and efficiently to meet the needs of the patient. The treatment preparation areas are functionally separated and respect the industry standards for safe medication preparation.

- **Internationally recognised Centres of Excellence (CoE)** involves recognition of effective and high-quality service and management tools using evidence-based, sponsored clinical practice guidelines. This dimension focuses on the commitment of the organisation to excellence whereby the collaboration of programmes throughout the organisation acts as a testimony towards the centre of excellence, and contains the structure, resources, and support it needs to be successful and to provide the desired clinical outcome.

  All patients with the same medical condition or disease receive the same high-quality care, treatment, and services as reflected in the programme's scope of services and mission. Medical research and development, including participation in clinical trials and scientific publications, play an important role in advancing medical knowledge and improving patients' medical outcomes. These series of initiatives in building the international CoE strive to stay at the forefront of medical innovation and contribute to the overall progress of healthcare at the global level.

- **Information management and technology in healthcare** can significantly improve patient safety by automating and streamlining healthcare services for the seamless transition of patient health information and offering safety mechanisms that have the potential to reduce the risk of error. For example, medication errors can be greatly reduced through the implementation of a computerised prescribing mechanism and the use of barcodes for medication administration to ensure adequate patient identification.

  Integration of new technology with the existing systems is an important criterion and aids the hospital's assessment of opportunities for the optimisation of existing processes, including those that could be enabled by new technology. When mobile devices are used for texting, emailing, or other communications of patient data and information, the hospital should implement processes to ensure quality of patient care and maintain security and confidentiality of patient information. It is an expectation that the hospital develops, maintains, and tests any programmes to account for planned and unplanned downtimes and to ensure adequate continuum of care for medical travellers.

2) Service Excellence: The hospital's performance in service excellence is measured based on five key dimensions [4]:

- **Patient centricity performance** - Patient centricity is the compass that guides hospitals towards personalised care, empowering patients with dignity and involvement, fostering trust, and ultimately improving health outcomes. By placing patients at the heart of their services, hospitals cultivate an environment of empathy, collaboration, and quality, elevating the overall healthcare experience. This dimension evaluates hospital's initiatives to constantly enhance patient satisfaction and retention. Metrics assessed include patient satisfaction score, patient retention rate, and patient complaints.

- **Operational performance** - Operational performance is the backbone of hospital service, ensuring efficient processes, timely care delivery, and optimal resource management. By prioritising operational excellence, hospitals can enhance patient safety, streamline workflows, and maximise their ability to meet the ever-evolving healthcare needs of their communities. The dimension covers process efficiency in various departments in the hospitals including patient service department, outpatient clinics, pharmacy department and radiology department/ diagnostic clinic. Metrics assessed include waiting time for phone/ email enquiries, Specialist Outpatient Clinics (SOC) waiting time, and diagnostic turn-around time.

- **Innovation and technology enablers** - Innovation and technology enablers are vital in hospital services, revolutionising healthcare delivery, enhancing diagnostic accuracy, and enabling remote care. By embracing advancements, hospitals can improve efficiency, patient outcomes, and access to care, paving the way for a brighter and more interconnected future of healthcare. This dimension evaluates the level of technology adoption in enhancing patient satisfaction. Metrics assessed include process automation/digitalisation, accessibility of patient communication/request platform, and internet connectivity.

- **Service offering enablers** - Service offering enablers play a crucial role in hospital services by meeting diverse patient needs. By continually evolving and enhancing their service offerings, hospitals can provide comprehensive, tailored care that addresses the unique requirements of every individual, fostering improved patient satisfaction and outcomes. The dimension covers the initiatives that hospitals have implemented to further enhance patient journey in and outside of hospitals. Metrics assessed include spectrum of language interpretation, travel arrangements, and on-arrival service.

- **Infrastructure enablers** - Infrastructure is the cornerstone of effective hospital services, providing the physical foundation for quality care. Robust infrastructure ensures seamless operations and creates a conducive environment for healthcare professionals to deliver timely and comprehensive treatments, ultimately enhancing patient experiences. This dimension evaluates the quality of the infrastructure as a place of living. Metrics assessed include facility hygiene, availability of international patient centres, prayer rooms, quality of staff and facilities, availability and variety of restaurants, and availability of leisure facilities (e.g., convenience stores, recreational spaces).

3) International Branding: The hospital's performance in international branding is measured based on four key dimensions [4]:

- **Global presence/ branding performance** - A global presence in medical tourism allows hospitals to attract patients from around the world. By establishing themselves as trusted destinations for international healthcare, these hospitals attract affluent patients seeking high-quality medical services abroad. This dimension evaluates the hospital's branding effort in the target medical tourist countries. Metrics assessed include international recognition and awards, number of countries with brand presence, and the conversion rate from international enquiries to facility visits.

- **Corporate Social Responsibility (CSR) and health diplomacy performance** - CSR and health diplomacy initiatives are essential as they demonstrate commitment to ethical practices, community engagement, and sustainable healthcare. By actively addressing social and environmental concerns as well as engaging in health diplomacy, these hospitals not only enhance their reputation but also contribute positively to the local communities they operate in. They promote inclusive healthcare access, foster long-term partnerships, and contribute to diplomatic efforts in addressing health challenges at both the national and international levels. The dimension covers the hospital's investment in building reputation through CSR and health diplomacy. Metrics assessed include CSR initiative numbers, spend, impact, geographical coverage of target countries, and the effectiveness of health diplomacy activities.

- **Marketing & promotion enablers** - Marketing and promotion enablers help raise awareness, attract international patients, and showcase the hospital's unique offerings. Effective marketing strategies and promotion efforts position the hospital as a preferred destination for medical tourism, increasing visibility, credibility, and ultimately driving patient volume, revenue, and sustainable growth. This dimension evaluates the effectiveness of the hospital's overall medical tourism marketing strategy. Metrics assessed include international marketing budget, number of online and offline channels utilised, and quality of marketing messaging among others.

- **Sustainable Development Goals (SDGs) and Environmental Sustainable Governance (ESG) performance** - The incorporation of SDGs and ESG principles into hospitals can have significant impacts in promoting global sustainability, enhancing hospital's brand reputation, strengthening community engagement and contributing to the long term wellbeing of patients, staff as well as the environment. The evaluation dimensions include the alignment of hospital's strategies, policies, and actions to both the SDGs and ESG goals as well as the integration into hospital's practices. Metrics assessed include the effectiveness of implementation and impact measurements.

C. Accelerating Capacity Building and Revenue Growth

Throughout the Acceleration phase from 2023 to 2025, programme management plays a key role in strengthening and elevating the bar of excellence in delivering exceptional end-to-end services to patients of the finalists, further reinforcing Malaysia's position as a safe and trusted destination for healthcare. The components of programme management include the charting and activation of the 2023 to 2025 Flagship Acceleration Plan including Key Performance Index (KPI) from each finalist as well as the programme mentoring and monitoring by the appointed advisors [4].

Furthermore, the Acceleration phase will be guided by the global industry experts via a series of training and development programmes [4]. This includes the C-suites coaching on healthcare innovation and transformation as well as the customised training modules for the hospital personnel. This initiative directly supports the achievement of KPI targets set in the finalists' Acceleration plan.

FMTH also plays a catalytic role in expediting international recognition of the Flagship finalists. The active commitment and participation of the finalists in delivering medical and service excellence will eventually lead towards cultivating international branding. Extensive initiatives on global brand profiling for FMTH have commenced through evidence-based capabilities and patient outreach, some of which have strong presence within the government ministries [4].

In support of the finalists' Acceleration plan, the finalists are granted with several enablers, as listed below. The successful activation of all the enablers is the result of the collaborative effort from the various government ministries and agencies [4].

- **Investment Tax Allowance (ITA)** - Hospital investment is supported through the provision of additional ITA on qualifying capital expenditure including healthcare

 ISBN: 978-1-68558-112-1

technology.

- **Fast Track Facilitation** - Finalists' applications are subject to expedited approvals to support development milestones as per the acceleration journey, with a designated committee that oversees the approval process related to renewals, facility improvement, onboarding niche specialists, among many others.

- **Healthcare Technology Sandbox** - Hospitals are granted flexibility of testing concepts that support medical and service excellence or improve patient experience in a sandbox. The tested concepts may be related to use of digital health & healthcare innovation.

- **Programme mentors and advisors** - Finalists gain access to programme advisors and industry experts that guide them on growth and development, programme monitoring and mentoring, KPI tracking, and more.

- **National endorsement and recognition** - The four Flagship finalists were announced by the Minister of Health, Malaysia in March 2023. The Flagship Hospital(s) will eventually be recognised and awarded by the Honourable Prime Minister of Malaysia in 2025.

D. Advancing Healthcare Industry and the Nation

In addition to the participating hospitals, the FMTH Programme brings long-term benefits to the development of Malaysia's healthcare industry and the nation. The national programme provides the country's hospitals and medical professionals with global exposure and opportunities to establish standing and reputation as the top global medical tourism destination. The programme also enables the nation's healthcare industry to challenge the position of popular medical tourism destinations (i.e., Thailand and Singapore) in the APAC region and to future-proof Malaysia and the region.

The economic impact of the FMTH Programme is immense. In terms of direct economic impact on the health tourism industry, the programme is expected to contribute 30% - 35% of the total healthcare traveller's revenue of MYR 2.4 billion (US$ 513.3 million) in 2025 [1]. This is almost double the growth in comparison to the achievement in 2022 [1]. In terms of indirect economic impact, approximately MYR 9.6 billion (US$ 2 billion) in revenue from economic spill over (ancillary and tourism spend) is forecasted in 2025, representing a MYR 4.4 billion (US$ 941.2 million) increase from 2022 [1]. In addition, domestic and foreign direct investments to boost medical tourism in Malaysia by 2025 are estimated to be MYR 1.0 billion (US$ 213.9 million) and MYR 500-700 million (US$ 106.9 - 149.7 million), respectively [5]. Hospital construction, operational expansion and healthcare digitalisation also contribute to the creation of more jobs, employment opportunities for locals, tax generation and fiscal impacts.

III. CONCLUSION AND FUTURE WORK

The FMTH Programme is an innovative and game-changing initiative implemented by MHTC [3] to raise Malaysia's healthcare profile as home to globally recognised healthcare icons. Through strategic public-private partnerships, strong leadership culture and unwavering commitment, this groundbreaking programme successfully elevates the finalists into world-class healthcare facilities.

The programme impact extends beyond the individual finalists, stimulating various sectors of the economy and contributing to overall industry development. By attracting health tourists and their companions, Malaysia experiences increased revenue in tourism, hospitality, transportation, ancillary services and more within the healthcare travel ecosystem. This eventually creates employment opportunities and fosters the growth of local businesses, establishing a virtuous cycle of economic prosperity.

In summary, the FMTH Programme is revolutionising the Malaysia Healthcare landscape, showcasing Malaysia as a world-renowned and credible global healthcare brand. The programme's success in developing flagship medical tourism hospitals within the country positions it as an exemplary model for other nations seeking to establish their own thriving medical tourism industry. The programme's accomplishments serve as a source of inspiration and guidance for these countries, providing them with valuable insights and best practices to emulate.

REFERENCES

[1] Malaysia Healthcare Travel Council Informatics, 2023. MHTC Healthcare Travel Informatics Analysis.

[2] IQVIA Solutions Malaysia, 2023. IQVIA Market Prognosis and Proprietary Forecast.

[3] Malaysia Healthcare Travel Council, 2023. Malaysia Healthcare Travel Council. Malaysia Healthcare Travel Council (MHTC). URL https://www.mhtc.org.my/discover-mhtc/ (accessed 9.11.23).

[4] Malaysia Healthcare Travel Council, 2021a. Malaysia Healthcare Travel Industry Blueprint 2021-2025.

[5] Malaysia Healthcare Travel Council, 2021b. MHTC Flagship Programme Development Feasibility Outcome.

[6] Ministry of Health Malaysia, 2020. Official Portal of Ministry of Health Malaysia. URL https://www.moh.gov.my/index.php/pages/view/137?mid=14 (accessed 9.11.23).

[7] Z. H. Perry, L. Apel-Sarid, L. Salzer, A. Toker, Medical Tourism in a Socialized Health Care System—The Need for Strict Systemic Supervision. RHS 4, 68, 2019, https://doi.org/10.22158/rhs.v4n2p68

[8] Polaris Market Research, 2020. Medical Tourism Market Size, Share - Industry Trends Report 2020-2026. Polaris. URL https://www.polarismarketresearch.com/index.php/industry-analysis/medical-tourism-market (accessed 9.11.23).

[9] World Travel & Tourism Council, 2019. Travel & Tourism Economic Impact 2019. World Travel & Tourism Council, London, United Kingdom.

Considerations for Applying MediaPipe to Gait Analysis

Comparison with Commercial Software

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

***Abstract*— MediaPipe, which enables skeletal analysis using videos of walking subjects without the use of markers, can be easily introduced into rehabilitation sites. Because the video used for analysis is captured from a smartphone or video camera, the viewpoint is obtained from a single camera. Therefore, the skeletal coordinates cannot be recognized during analysis and the obtained coordinates are relative values. In this study, we used data obtained from MediaPipe to calculate stride length, walking speed, knee height change, and ankle angle and compared them with commercially available software. During the measurements, a pseudo-motor restriction was applied 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.**

Keywords-MediaPipe; skeletal analysis; smartphone; 3D analysis.

I. INTRODUCTION

Measures are urgently required to prepare for a rapidly aging population. Falling is a significant problem among the elderly, as it causes them to be bedridden and places a heavy burden on their caregivers [1]-[8]. Therefore, motion analyses have been conducted using insoles [9][10] and mat-like pressure sensors arranged two-dimensionally [11], wearable devices [12], and images [13]. For gait analysis, measurements using multiple cameras with attached markers have been used in rehabilitation facilities, as typified by the Vicon system [14][15]. A camera called Kinect [16]-[19] has also been used to analyze the movement of a camera linked to game software. However, problems remain, such as the need for an expensive dedicated system, space for recognizing the markers, and an operator who is 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, affordable and easy-to-use software is available. The introduction of a system requires continuous cost. For this reason, it is currently in a state where it cannot be sufficiently spread.

Software that can perform skeleton authentication includes OpenPose [20][21] developed by Carnegie Mellon University and MediaPipe [22]-[24] released by Google. Both use deep learning and have a high certification system.

In our previous work [25], we presented a basic application of MediaPipe in the field of rehabilitation. Furthermore, for the use of a walking assist device, we reported that the effect continued even approximately 5 min after the walking assist device was removed. In this paper, we report the results of additional research on the accuracy and application range of walking parameters obtained using MediaPipe. If the analysis results from front filming can be utilized, data captured in hallways can also be used. In this study, we performed the analysis using front filming. MediaPipe, which can use Python, has the potential to be used by healthcare and welfare professionals who are not analysis experts. The ability to analyze videos from the front view using MediaPipe can also enable filming in rehabilitation rooms and hallways of hospitals and facilities; thus, healthcare and welfare professionals can use it themselves.

Section II describes the experimental methodology, including the software used and the commercially available equipment and software. Section III shows the results using MediaPipe and commercial software. Section IV discusses the results obtained with the two types of software. Section V presents the conclusions.

This study was approved by the Ethics Committee on Research with Humans as Subjects of the Teikyo University of Science.

 ISBN: 978-1-68558-112-1

II. EXPERIMENTS

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 using the tool ORPHE ANALYTICS [26] to confirm the accuracy of the calculation results obtained from the 3D coordinate data obtained using 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 device that can be fixed to the shoelace. The data obtained from these sensors could be analyzed to display various analysis results. A photograph of the ORPHE CORE® attached to the shoelaces of a shoe is shown in Figure 1.

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

Owing to the limitations of the laboratory, we could not use 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 using ORPHE ANALYTICS from the front, we used an iPhone with ORPHE TRAC 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 using the free software AG Desktop Recorder [27]. This screen was loaded into MediaPipe, which was 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 used to calculate the change in the difference between the knee and ankle. The angle of the ankle was calculated using vectors connecting the ankle and knee and the ankle and toe.

Figure 2 shows examples of measurements using MediaPipe. The image on the left shows the measurement without motion restriction, and that on the right shows that with motion restriction.

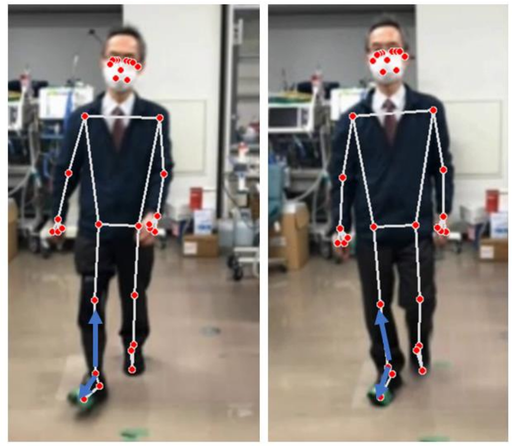


Figure 2. Examples of measurement results.

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

III. EXPERIMENTAL RESULTS

A. *Measurement of Strides*

The right-foot ankle trajectory measured with the MediaPipe is shown in Figure 3. Only one round trip was used 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 used in this study 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 walking to the change in direction was obtained.

The z-axis values for walking when approaching the camera are shown in Figure 4.

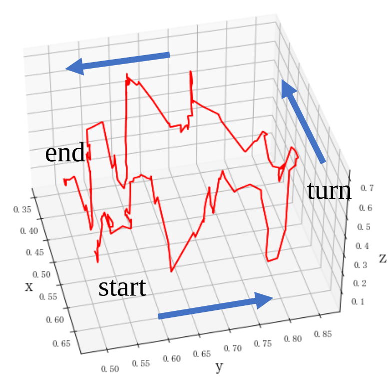


Figure 3. Right-foot ankle trajectory measured with MediaPipe.

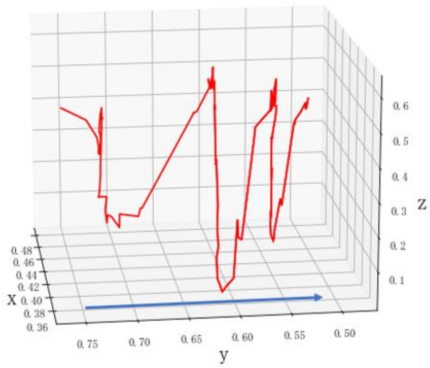


Figure 4. Z-axis values for walking when approaching the camera.

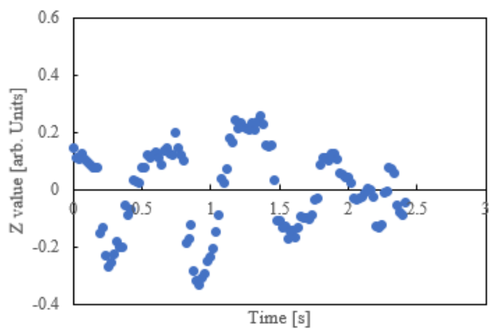


Figure 5. Result of modifying the effect of walking direction.

Numerical data were displayed in Excel, and the inclination due to the walking direction was obtained; the corrected results are shown in Figure 5. Using this diagram, we considered the point corresponding to the landing to be the minimum value based on the change in amplitude. As the figure shows, 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 landing point of the foot 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. In the right leg with restricted motion, the stride length was larger owing to the hip motion.

The left and right stride lengths obtained from ORPHE were 0.75 and 1.0 m, respectively, which were 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.

B. *Results of knee height measurements*

Figure 6 shows the results of MediaPipe for the changes in the right and left knee height during walking. 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 shown from the beginning of walking to turning, considering the effect of rotation.

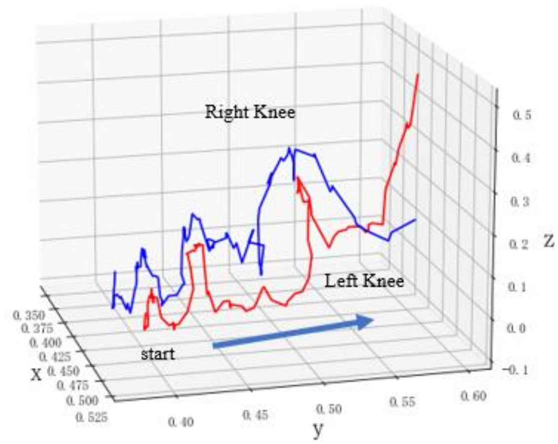


Figure 6. Height of the knee position evaluated using MediaPipe.

The results of the ORPHE ANALYTICS measurement of knee height are shown in Figure 7. 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 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 direction of gait during the measurement. When comparing knee heights, we used not only moving images but also changes in the x-axis direction, which is characteristic of a change in direction, and deleted data from points in the range that appeared to indicate a change in the turn direction.

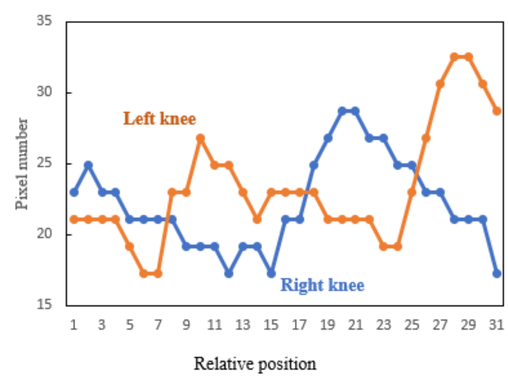


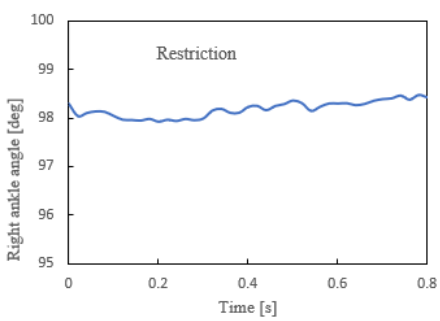
Figure 7. Knee height measurements with ORPHE ANALYTICS.

C. Measurement Results of the ankle angle

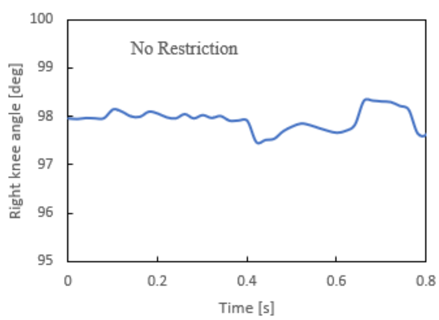
The results of the evaluation of the right ankle angle with and without pseudo-motor restriction are shown in Figure 8. (a) shows results with motion restriction added and (b) without. For the right foot with pseudo-motor restriction, almost no change in the ankle angle was observed during walking, whereas for the right foot without motor restriction, the amplitude of the angle widened in the last part of the gait, although it was very slight.

For comparison, Figure 9(a) shows the left ankle angle without pseudo-motor restriction, and (b) shows the left ankle angle change without restriction. The values were large owing to the shooting angle. No characteristic waveform changes were observed in the left foot. This may reflect the difference in flexion and dorsiflexion of the participant's left and right feet.

The ORPHE ANALYTICS data were not directly displayed as an ankle angle, but the Euler angle obtained from the accelerometer was considered to correspond to it. The angle changed abruptly at regular intervals, which was considered to correspond to the kicking of the foot. The plastic fixture was used to hold the shoelaces in place; therefore, the changes may have been large and different, but we considered that more absorption changes could be measured with the plastic fixture than with MediaPipe.

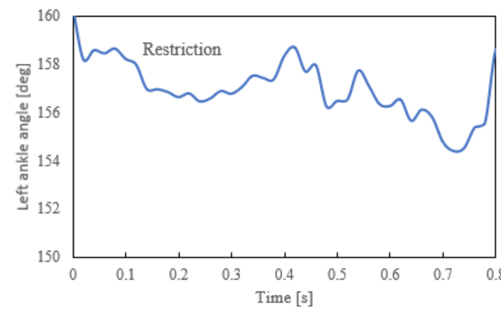


(a)

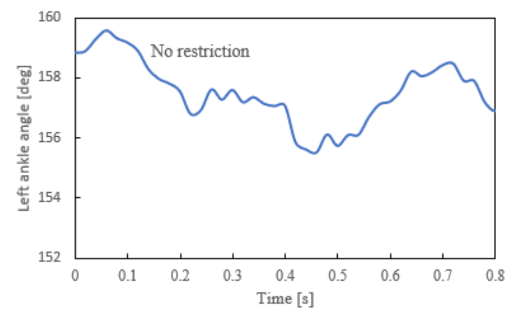


(b)

Figure 8. Angles of the right ankle without restriction.



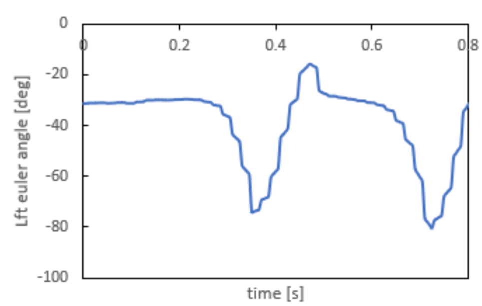
(a)



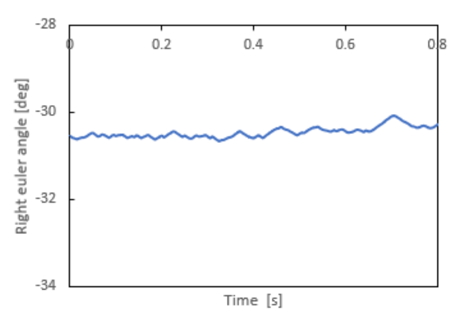
(b)

Figure 9. Angles of the left ankle without restriction.

The ORPHE ANALYTICS data did not directly display this as an ankle angle, but the Euler angle obtained from the accelerometer was considered to correspond to it. The results of the motion restriction are shown in Figure 10. (a) and (b) for the left and right ankles, respectively.



(a)



(b)

Figure 10. Euler angle of the right ankle.

The angle changed abruptly at regular intervals, which was considered to correspond to the kicking of the foot. The angle was already approximately 30° because a plastic fixture was used to hold the shoelaces in place. Although a large change in the measurement angle may have been measured owing to the fixing method, a more absorbing change was considered to be measured compared with MediaPipe.

IV. DISCUSSION

In this experiment, the main reason for the difficulty in 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 not possible, and 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, but 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 have not provided sufficient accuracy. However, owing to the limitation of the number of pixels, it was more difficult to handle than MediaPipe.

Because only one participant was used in this measurement, the data were limited to a specific individual. It would be important 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.

V. CONCLUSIONS

The values obtained through 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.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Numbers JP20K11924and JP23K11207.

REFERENCES

[1] Y. Uchida, T. Funayama, and Y. Kogure, “Investigation of the Application of MediaPipe to Gait Analysis,” in *GLOBAL HEALTH 2022*, pp. 1-6, IARIA, 2022. ISBN: 978-1-61208-995-9.

[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 Assitive Technology for the Life and Safty Critical Applocations: 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, pp.3239, 2018.

[7] V. Bucinskas et al., “Wearable Feet Pressure Sensor for Human Gait and Falling Diagnosis,” Sensors, pp. 5240, 2021.

[8] P. M. Riek, A. N. Best and R. Wu, “Validation of Inertial Sensors to Evaluate Gait Stability,” Sensors, vol. 23, pp. 1547, 2023.

[9] T. Funayama, Y. Uchida and Y. Kogure, “Assessment of Walking Condition Using Pressure Sensors in the Floor Mat,” in *GLOBAL HEALTH 2022*, IARIA, pp. 7-12 , November 2022, ISBN: 978-1-61208-995-9.

[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] 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,” vol. 259, pp. 29-36, 2022.

[12] 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, pp. 234, 2020.

[13] S. Majumder et al., “Smart Homes for Elderly Healthcare-Recent Advances and Research Challenges,” Sensors, vol. 17, pp. 2496, 2017.

[14] 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.

[15] 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.

[16] P. Plantard, E. Auvinet, A. S. Le Pierres and F. Multon,”Pose Estimation with a Kinect for Ergonomic Stuidies: Evaluation of the Accuracy Using a Virtual Mannequin,” Sensors, pp. 1785-1803, 2015.

[17] 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 & Postture, vol. 68, pp. 193-200, 2019.

[18] 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 Spaial Margin of Stability ad Spationtemporal Vaiables,” Sensors, vol. 21, pp. 2104, 2021.

[19] 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.

[20] M. Ota, H. Tateuchi, T. Hahiguti and N. Ichihasi, “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.

[21] Y. Saiki et al., “Reliability and validity of OpenPose for measuring hip knee ankle angle in patients with knee osteoarthritis,” Scientific Reports, vol. 13, pp. 3297, 2023.

[22] V. Bazarevsky et al., “BlazePose: On-device Real-time Body Pose tracking,” arXiv:2006.1204v1 [cs.CV] 2020.

[23] 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, pp. 119-124, 2022, ISSN:2582-7839.

[24] J.-L. Vhung, L.-Y. Ong and M-C. Leow, “Comparative Analysis of Skelton-Based Human Pose Estimation,” Future Internet, vol.14, pp. 380, 2022.

[25] Y. Uchida, T. Funayama, and Y. Kogure, “Possibility of Gait Analysis with MediaPipe and Its Application in Evaluating the Effects of Gait-assist Devices”, International J. of Advances in Life science, vol.15, no 1&2, pp. 45-55, 2023.

[26] 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, pp. 331, 2023.

[27] AG Desktop Recorder: https://streamgaga.com/how-to-use-ag-desktop-recorder, retrieved: August, 2022.

Exploring the Potential of a Wrist-Worn Optical Sensor for Measuring Daily Life Activities

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

Yasutaka Uchida
Dept. of Life Science
Teikyo University of Science
Tokyo, Japan
e-mail: uchida@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

***Abstract*— In recent years, wearable sensor devices that can be worn in daily life have rapidly gained popularity. The ability to monitor daily life, through a prolonged assessment, enables the detection of subtle changes in health status. Evaluating individuals in their usual daily settings provides an assessment that cannot be obtained within a hospital environment. However, there are not many research data on how activities specifically and objectively affect health. Therefore, in this study, we utilized an optical sensor device worn on the wrist to measure daily activities during individuals' daily routines. We studied heart rate and Fourier analysis and examined the relationship between condition and activity. As a result, the following aspects were deemed necessary to consider during measurement: 1) Analyze the data, taking into account not only the unstable period until the device stabilizes but also the time when subjects are operating the device. Consider the excluded data time and set the measurement time accordingly. 2) In addition to frequencies of 1Hz, 0.3Hz, 0.1Hz, also include analysis of the low-frequency ranges, such as 0.01Hz. 3) In case of significant variations in optical sensor data caused by arm movements, during abrupt and rapid changes (such as when the measured values from the accelerometer sensor exceed a certain threshold), data is excluded from the analysis of the light sensor. 4) Divide and analyze the data based on units, such as 10 seconds, 1 minute, 10 minutes, and investigate the changes in a time series. 5) Compare changes during activities and rest periods, and within the subject's activities. Examine the variations in the time required for cool-down and recovery. 6) Consider external factors, such as the influence of natural light, fluorescent lights, and vibrations from cars or trains. This study suggests that tailored measurement and analysis for various activities and environments are crucial in order to utilize optical sensor for health promotion and rehabilitation in daily life activities.**

Keywords- Optical Sensor; Activities of Daily Living; Spectrum Analysis; Self-Therapy.

I. Introduction

To enhance the quality of life of individuals with health-related issues, it is crucial to examine the reciprocal relationship between health conditions and activities of daily living and to make necessary adjustments in their daily lives. Circadian rhythms also affect health, and activities such as sleeping, eating, and outdoor activities affect circadian rhythms [1]–[3]. Effective self-management of activities and maintenance of health conditions are essential for self-therapy and improving the overall quality of life. However, grasping the relationship between health condition and daily activities is not a straightforward task. Life activities are not simply judged as inherently good or bad for one’s health. They are influenced by various factors, such as the individual's health condition, intensity of activities, habituation, personal adaptation, and environment, and depend largely on the balance of these factors. There are situations similar to those of workaholics, in which individuals may become engrossed in their activities and find it difficult to pay attention to their health. Compared to objective conditions, such as fever and coughing, it is more difficult to accurately assess subjective conditions, such as pain and fatigue. The ability to objectively perceive fluctuations in health conditions during their daily lives and comprehend the activities that affect their mind and physical well-being holds significant potential for enhancing health management among individuals facing health challenges [4]–[8]. Moreover, it is valuable for rehabilitation in daily living [9]–[12]. Recently, several monitoring and support devices and systems using wearable sensors have been researched, developed, and commercialized for older adults and individuals with health issues [13]–[16]. Wearable devices that employ sensors, such as acceleration, temperature, and pressure sensors are becoming increasingly popular [17].

We have studied methods for assessing body changes during activities. In addition, the relationship between daily

 ISBN: 978-1-68558-112-1

activities and heart rate has been examined using an optical sensor device [18]–[21]. Therefore, in this study, we conducted an exploratory investigation using a wrist-worn wearable device equipped with optical sensors in the context of individuals' daily routines. Activities and health conditions were recorded using a cloud-based service, whereas simultaneous measurements were conducted using a wearable device. We studied the heart rate and Fourier analysis and examined the relationship between the condition and activity. We suggest certain aspects of the measurement methodology for use during daily activities.

This study was approved by the Ethics Committee on Research with Humans as Subjects of the Teikyo University of Science. Section II describes the experimental method, Section III describes the results, Section IV presents the discussion, and Section V presents the conclusions and future work.

II. EXPERIMENTAL METHOD

A. Devices

The device used for the measurement was a Maxim Integrated MAXREFDES103, which was worn on the wrist as a wristwatch. The measurements can be obtained using a PC and an Android device. Three LEDs (green, red, and infrared) were used as optical lights. The green LED uses two diodes and outputs the green and green2 data. It incorporates an Arm Cortex-M4F embedded processor and supports Bluetooth connectivity for data transfer. The Sampling frequency was set to 25 Hz. This device outputs four types of optical light data (green, green2, red, and infrared) in CSV format, three-axis accelerometer data (x, y, and z), heart rate, transcutaneous arterial oxygen saturation (SpO_2), and timestamp information. The heart rate was calculated and provided as a value using MAXREFDES103. In this study, we used the green light sensor data, heart rate, and timestamps obtained from the CSV output.

Figure 1. Device used (MAXREFDES103).

B. Measurement

The measurements were conducted during the participants' daily life activities, as determined by his judgment, using a MAXREFDES103 device. The corresponding situations were simultaneously recorded. The subject was a single man in his 60s, and the data obtained from this individual were used for the analysis. Information on health conditions, stress levels, activity details, activity duration, and activity location from Google Form records were used in the analysis. Health conditions and stress levels were recorded on a scale of 1 to 10, with 1 representing the best state and 10 representing the worst perceived state. A free-text field was also included.

C. Analysis Method

The analysis excluded the initial two minutes of data obtained from the device, comprising 3000 data points, and the final 1000 data points. This exclusion was made to account for the time required for the data to stabilize and the time spent by the subject to operate the device. Thus, a fixed duration of time at the start and end of the measurement period, which included the time spent by the subject operating the clock, was excluded from the analysis. The Fourier analysis was performed using the fft function from the NumPy library in Python, specifically numpy.fft.fft.

III. RESULTS

A. Optical data used in the analysis

We examined the data outputted in four CSV files based on three types of light: green, infrared, and red. The green LED utilizes two different diodes and outputs data for green and green 2. Although the intensity of the stronger light differed from one implementation to another, green light was used because it was often relatively strong in spectral intensity. Figure 2 shows a graph of the output data obtained using optical light. Infrared (hereafter referred to as IR) refers to the near-infrared light. Figure 3 shows an example graph of the green output data, limiting the number of data points to 500.

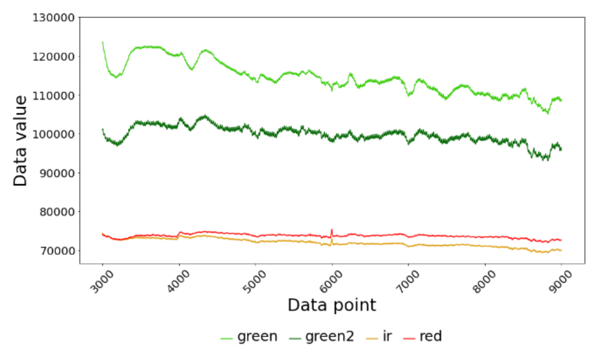


Figure 2. An example of output data by optical light type.

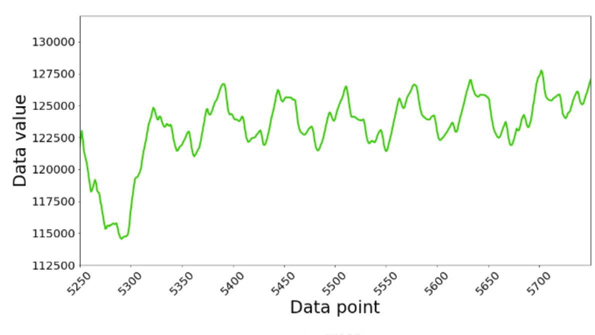


Figure 3. Green output data only 500.

There were periods of rapid changes in the output data within short durations, as well as periods with minimal variations.

B. Data Interval Used for Analysis

Measurements during daily activities are often performed over long periods. This results in a large amount of data. The obtained long-duration data were divided, and the frequencies that were deemed to be influenced by heart rate were identified through Fourier analysis and compared with the heart rates calculated from the device. The frequency with the highest spectral intensity in the range of 0.9 Hz to 2.5 Hz by Fourier analysis was considered to be the frequency that was influenced by the heart rate. The division was performed based on the number of data points after the start of the measurement, specifically at intervals of 250, 500, 1000, 2000, 4000, 6000, 8000, 10000, 15000, 20000 and 30000 data points. Each interval of 250 data points corresponded to approximately 10 seconds. An example of this is shown in Figure 4.

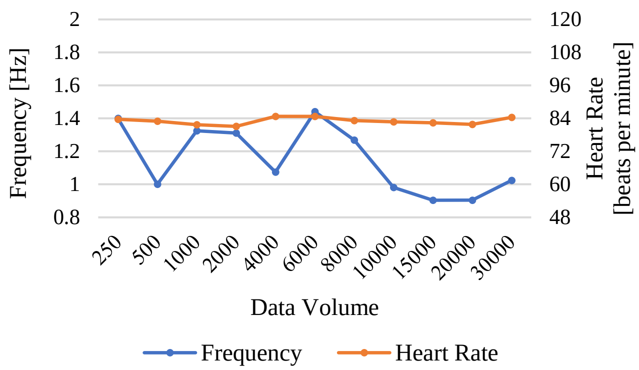


Figure 4. 0.9-2.5 Hz Frequency at maximum spectral and heart rate.

As the number of data points increased, the difference between the heart rates calculated from the device and the actual heart rates increased. With 250 data points, the heart rates closely matched, and this was often the case for up to 1000 data points. Although there were instances where the heart rates matched, even with a larger number of data points, inconsistencies became more frequent when the number of data points exceeded 2000. This may be attributed to factors, other than the heart rate mixing in at 0.9-2.5 Hz, such as increased measurement time, an increase in the number of different heart rates (i.e., more variation in heart rate values) and the impact of Fourier analysis, among other considerations.

C. Characteristic Frequency Bands

The number of data points was set to 1000, 2000, 4000, 6000, 8000, 10000, 15000, and 20000 or more, each of which was Fourier-analyzed to examine the spectral intensity and frequency band characteristics. Peaks were observed near 1.5 Hz, 0.3 Hz, 0.1 Hz, and occasionally below 0.1 Hz with increments of 0.01 Hz, with an occasional peak around 0.01 Hz below 0.1 Hz. Figure 5 and 6 show sample diagrams of the relationship between the spectral intensity and frequency based on Fourier analysis for 1000 and 15,000 data points, respectively. Figure 7 shows an example of the low-frequency region down to 0.20 Hz for 15000 data points.

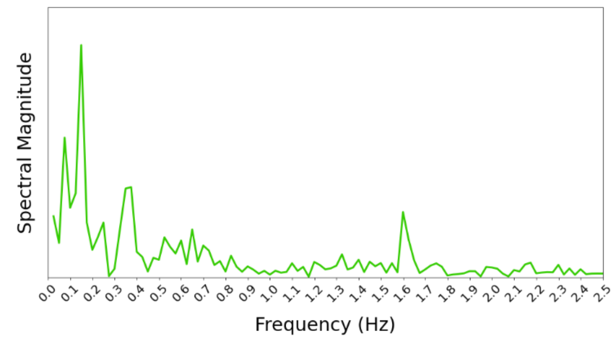


Figure 5. An example with 1000 Data Points.

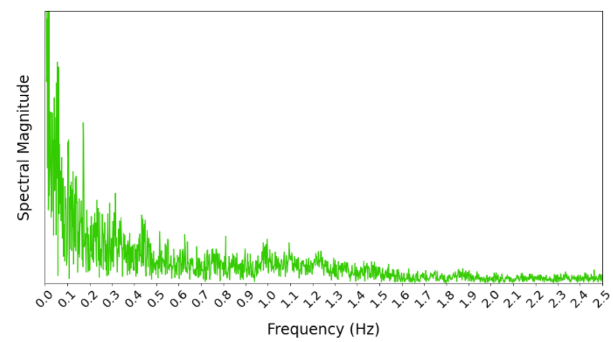


Figure 6. An example with 15000 Data Points.

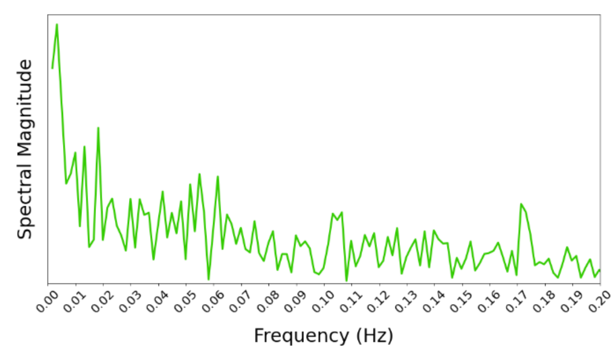


Figure 7. An example with the Low-Frequency Section with 15000 Data Points.

D. Frequency Characteristics due to Health Condition and Stress

The recorded measurements on a 10-point scale for health condition and stress levels were compared in terms of heart rate between states 8 and 9, indicating poor health and high stress, and between states 4 and 5, representing normal conditions. Because there were no recorded measurements for the states rated 3 or below, relatively good states 4 and 5 were used for comparison. Figure 8 shows the relationships between perceived health conditions, stress, and heart rate.

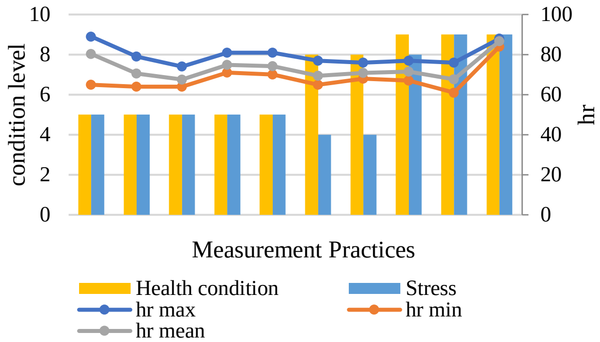


Figure 8. Heart rate related to health condition and stress.

There have been times when both self-perceived health condition and stress level were nine, and during those times, we noticed that the maximum, minimum, and mean heart rates were all high. However, a distinct correlation between the heart rate and these factors could not be established. Even when the patient's health was relatively good, the maximum heart rate remained high.

We examined self-perceived health conditions, stress, and the results of the Fourier analysis (see Figure 9). We extracted the frequencies at the peak spectral intensity within the ranges of 0.01-0.05 Hz, 0.05-0.15 Hz, 0.2-0.4 Hz, and 0.9-2.5 Hz. The numbers on the X-axis of the graph represent the stress levels on the right side of the graph, ranging from 1 to 10, and the health conditions on the left side of the graph, ranging from 1 to 10. No obvious features were observed over the entire frequency range.

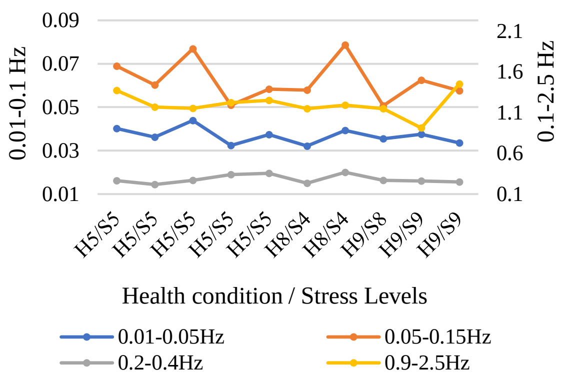


Figure 9. Fourier analysis data related to health condition and stress.

We compared the spectrum intensity and frequency plots from the Fourier analysis. Figures 10 and 11 represent poor and good health, respectively. Figures 10 and 11 show approximately 5000 data points.

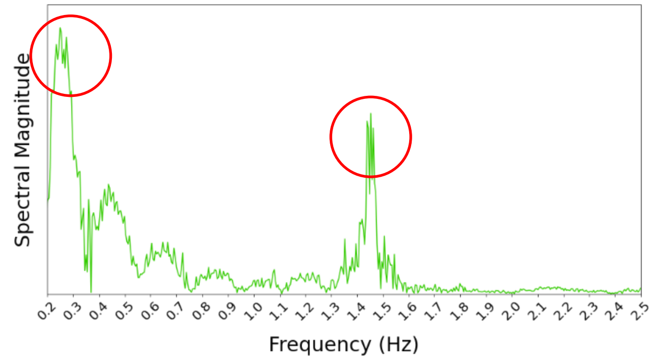


Figure 10. Fourier analysis results for poor health.

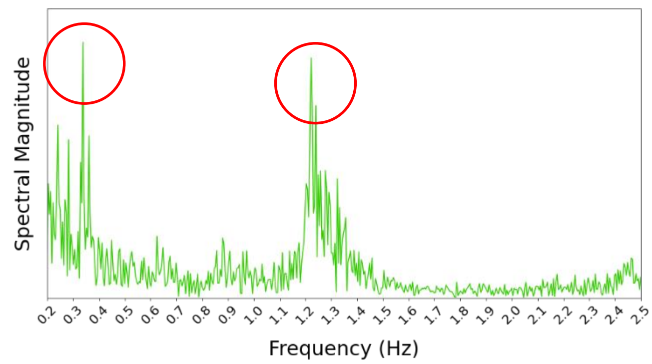


Figure 11. Fourier analysis results for good health.

Although there were no common characteristics across all instances, during periods of poor health, the peaks in the frequency range associated with heart rate and respiration appeared jagged and fluctuating, rather than well defined. However, during times of good health, there were also instances where sharp peaks were observed during periods.

E. Frequency Characteristics due to Bathing

The effects of bathing on each frequency band were also examined. The Fourier results before and after bathing, analyzed at 6000 number of data points, are shown in Figures 12 and 13.

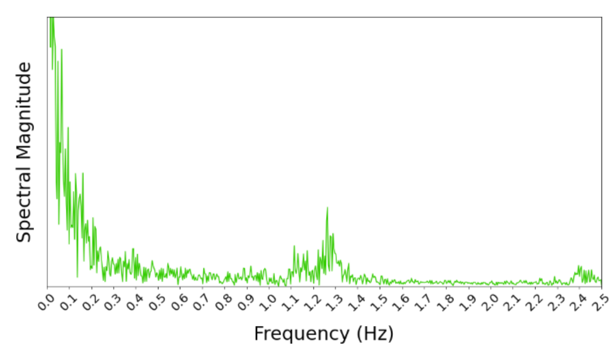


Figure 12. Fourier analysis before bathing.

 ISBN: 978-1-68558-112-1

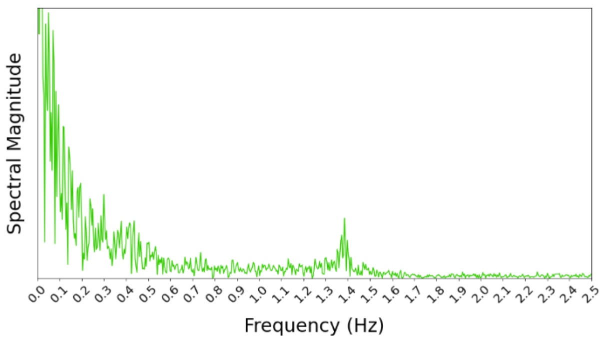


Figure 13. Fourier analysis after bathing.

After extracting the frequencies at maximum spectrum in the frequency bands 0.01-0.05 Hz, 0.05-0.15 Hz, 0.2-0.4 Hz, and 0.9-2.5 Hz, it was observed that the frequencies after bathing were higher in all frequency bands. This is illustrated in Figure 14. Owing to the potential differences in the Fourier analysis results between long durations with a large number of data points and short durations with a small number of data points, we extracted and examined the frequencies at the peaks of the maximum spectral intensity when analyzing the data with 250 and 1000 data points. The results are summarized in Table 1.

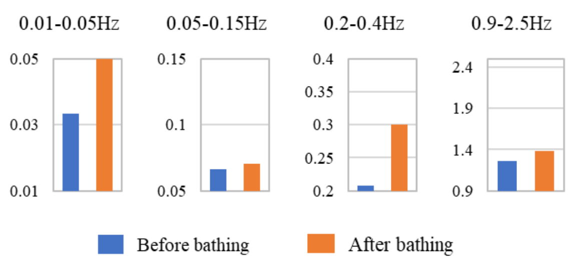


Figure 14. Frequency of the strongest of the spectrum in each frequency band before and after bathing.

TABLE I. FREQUENCY DIFFERENCES BASED ON DATA POINTS.

	250 data	1000 data	6000 data
0.2-0.4Hz Before Bathing	0.20	0.37	0.21
0.2-0.4Hz After Bathing	0.20	0.30	0.30
0.9-2.5Hz Before Bathing	1.20	1.27	1.27
0.9-2.5Hz After Bathing	1.30	1.02	1.39

In the case of 250 data points, for the frequency range of 0.2-0.4 Hz, the frequencies were 0.20 Hz before bathing and 0.20 Hz after bathing, and for the frequency range of 0.9-2.5 Hz, the frequencies were 1.20 Hz before bathing and 1.30 Hz after bathing. In the case of 1000 data points, for the frequency range of 0.2-0.4 Hz, the frequencies were 0.37 Hz before bathing and 0.30 Hz after bathing, and for the frequency range of 0.9-2.5 Hz, the frequencies were 1.27 Hz before bathing and 1.02 Hz after bathing. The frequency varied depending on the data segmentation method used.

F. Frequency Characteristics due to driving a car

The raw data obtained during car driving are shown in Figure 15, and the Fourier analysis is shown in Figure 16. During car driving, the wrist wearing the device moves frequently because of steering wheel manipulation.

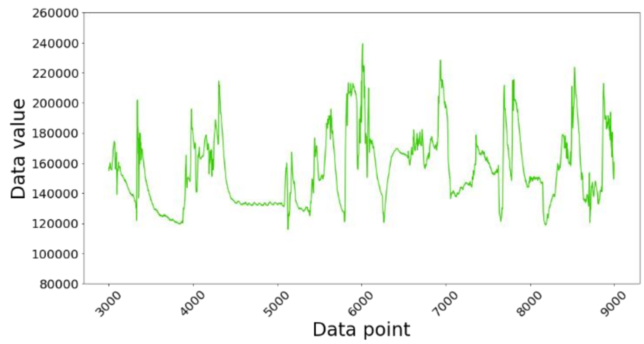


Figure 15. The raw data during car driving.

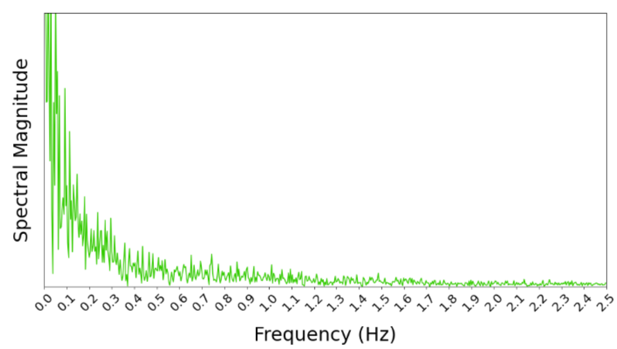


Figure 16. Fourier analysis data during car driving.

Figure 16 shows repeated significant fluctuations. The effect of each frequency band was also examined. No clear peaks were found in all frequency ranges, including 0.01-0.05 Hz, 0.05-0.15 Hz, 0.2-0.4 Hz, and 0.9-2.5 Hz.

IV. DISCUSSION

In this study, it was observed that as the number of data points increased, there was a tendency for a larger difference between the calculated heart rate from the device and the frequency at the maximum spectral intensity. Several factors are considered to contribute to this phenomenon. First, as the measurement duration increases, the device may move, resulting in unstable data acquisition. Second, factors other than the pulse rate can interfere with the frequency range of 0.9-2.5 Hz. Additionally, an increased variety of pulse rates and a higher occurrence of different pulse rate values can occur, leading to variations in pulse rate measurements. Finally, the characteristics of the Fourier analysis, such as the influence of rapid changes within short time intervals on the entire spectrum, can also play a role.

The spectral intensity and frequency band characteristics were examined using a Fourier analysis with different numbers of data points (1000, 2000, 4000, 6000, 8000, 10000, 15000, and 20000 or more). When using a dataset of 1000 data points, discernible peaks were observed at approximately 1.5 Hz, 0.3 Hz, and 0.1 Hz. One thousand data points were equivalent to 40 s. When using a larger number of data points, such as 15000, it was occasionally observed that peaks near 0.01 Hz were present. It is considered that frequencies around 1.5 Hz are influenced by the pulse rate, whereas frequencies around 0.3 Hz are influenced by respiration. The peak near 0.1 Hz could potentially be associated with Myer wave-related sinus arrhythmia (MWSA) derived from blood pressure. The frequency range of 0.15-0.45 Hz represented high-frequency (HF) components influenced by parasympathetic nervous control, the range of 0.04-0.15 Hz represented low-frequency (LF) components influenced by sympathetic nervous control, and frequencies below 0.1 Hz might be related to myogenic and neurohumoral factors [22]–[24]. However, it should be noted that measurements taken during daily activities are influenced by various factors and do not necessarily accurately represent specific biological information. Hayano cautioned that applying the association between short-term heart rate variability measured under strictly controlled conditions and autonomic function to long-term heart rate variability recorded during free activity often leads to erroneous interpretations [25].

With the widespread adoption of wearable devices, measuring daily activities and physiological signals has become easier. Therefore, in addition to correlating physiological information, it is important to consider activities, perceived health conditions, or stress levels during daily life to enhance rehabilitation and lifestyle interventions. In post-exercise rehabilitation, cool-down is important, and in cardiovascular rehabilitation, it is typically set 5 to 10 min [26]. The cool-down duration may increase with higher levels of fatigue. Cool-down is the process of returning a fatigued body and mind to their original state and promoting recovery. Rest is crucial in daily life, and there are instances in which the body unconsciously rests, even during activity. When measuring the extent of activity and the necessary cool-down, a segmentation method based on 10-minute intervals may be potentially. Ten minutes is 15000 data points, and the fact that a peak around 0.01 Hz was sometimes seen may be one guide to the division method of 15000 data according to 10-min intervals.

In terms of health condition and stress, sharp peaks that were not observed during periods of poor health were observed during periods of good health. It is possible that during poor health conditions, the heart rate is not stable and the heart rate variation increases, whereas during good health conditions, there are times when the heart rate is stable at a certain level. In this study, no clear relationship was found between the Fourier analysis results and health conditions. Nonetheless, physical conditions are related to biological information. It is well known that there is a correlation between stress and biological information, as observed in white-coat hypertension. However, the manifestation of this relationship varies between individuals. As there was no change in biological information, does not mean there was no change in health conditions. The awareness of health conditions also varies among individuals. Factors such as individual differences in manifestation, the relationship between subjective awareness of health conditions and stress, a low correlation between biological information and manifestation, challenges in measurement methods, and devices not picking up information may contribute to these observations.

In the investigation before and after bathing, with 6000 data points, the frequency at the peak spectrum was higher in all frequency ranges of 0.01-0.05 Hz, 0.05-0.15 Hz, 0.2-0.4 Hz, and 0.9-2.5 Hz compared to before bathing. However, with 1000 data points, it was lower in the frequency ranges of 0.2-0.4 Hz and 0.9-2.5 Hz. The frequency ranges of 0.01-0.05 Hz and 0.05-0.15 Hz were excluded from the analysis due to the limited number of data points. It can be inferred that the frequency is higher after bathing; however, depending on the extracted data to be analyzed, such as when movement is involved, the results may not match.

No distinctive results were observed while driving. This may be due to external factors other than the individual, such as the repeated movement of the upper limbs while wearing the device during steering wheel operation and the vibration of the vehicle. Considering that driving always occurs outdoors, natural light may also have an impact. However, the fact that a device is affected also implies that it affects a person. Considering the external factors that can influence a person, it is important to conduct further investigation. Additionally, by dividing and analyzing the data in short intervals, such as 10 s, instead of long durations, it may be possible to obtain distinctive data when the upper limbs wearing the device are in a stationary state.

V. CONCLUSION AND FUTURE WORK

Measurements during activity cannot be directly extrapolated from results obtained during rest. However, long-term measurements are beneficial for rehabilitation in daily life. It is important to consider what kind of activity, perceived health condition, and stress the individual is undergoing, and how they respond to them. In doing so, the following points will be important to keep in mind. In addition to frequencies of 1 Hz, 0.3 Hz, and 0.1 Hz, the analysis should include low-frequency ranges, such as 0.01 Hz. For long-term measurements, data points should be excluded from the analysis if there are significant device movements or rapid changes This implies that there are changes that exceeding a certain threshold within a specific time frame. If accelerometer data is also collected, any time data in which the accelerometer readings surpass a certain threshold will be excluded from the analysis. The data will be divided and analyzed at different time intervals: 10 s, 1 min, and 10 min. It is important to compare changes due to activity and rest as well as changes during the subject's activity, examine the changes in the time required for recovery, including cool-down, and consider the influence of factors such as natural light, fluorescent lighting, and external stimuli, such as vibrations from cars or trains should be considered. These are our suggestions. Human behavior is diverse and the impact of activities on the body is influenced by individual preferences, personalities, and characteristics. Therefore, it was not possible to categorize them mechanically. However, with the advent of big data utilization, it is possible to make objective judgments from the flexible data of activities. We intend to continue investigating the relationship between activities and health conditions to contribute to our understanding of lifestyle pathology and health promotion.

ACKNOWLEDGMENT

This study was supported by JSPS KAKENHI, Grant Number JP23K11207.

REFERENCES

[1] Y. Leng, T. Blackwell, and P. M. Cawthon, "Association of circadian abnormalities in older adults with an increased risk of developing Parkinson disease, " JAMA Neurol, vol. 77, No. 10, pp. 1270-1278, 2020.

[2] W. H. Walker, J. C. Walton, A. C. DeVries, and R. J. Nelson, "Circadian rhythm disruption and mental health", Translational Psychiatry, vol. 10, No. 28, 2020. https://doi.org/10.1038/s41398-020-0694-0

[3] K. Wulff, S. Gatti, J. G. Wettstein, and R. G. Foster, "Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease", Nature Reviews Neuroscience, vol. 11, pp. 589-599, 2010.

[4] T. Funayama, Y. Kogure, R. Kimura, N. Homma, and Y. Uchida, "Application of the human-machine interface technology to occupational therapy", Proceedings of 16th International Congress of the World Federation of Occupational Therapists", PCI-19-21, Yokohama, Japan, June, 2014.

[5] W. B. Arfi, I. B. Nasr, G. Kondrateva, and L. Hikkerova, "The Role of Trust in Intention to Use the IoT in eHealth: Application of the Modified UTAUT in a Consumer Context," Technological Forecasting and Social Change, vol. 167, 120688, pp. 1-33, 2021.

[6] E. Mbunge, B. Muchemwa, S. Jiyane, and J. Batani, "Sensors and healthcare 5.0: transformative shift in virtual care through emerging digital health technologies," Global Health Journal, vol. 5, Issue 4, pp. 167-177, 2021.

[7] J. O. O. Aguirre, J. R. Campos, G. A. Hernández, I. M. Cano, L. R. Mazahua, and J. L. S. Cervantes, "Remote healthcare for elderly people using wearables: A Review," Biosensors, vol. 12, Issue 73, pp. 2022. doi : 10.3390/bios12020073

[8] F. J. S. Thilo, S. Hahn, R. J. G. Halfens, and J. M. G. A. Schols, "Usability of a wearable fall detection prototype from the perspective of older people-A real field testing approach, " Journal of Clinical Nursing, vol. 28, Issue 1-2, pp. 310-320, 2018. doi: 10.1111/jocn.14599

[9] B. Marques, J. McIntosh, A. Valera, and A. Gaddam, "Innovative and assistive ehealth technologies for smart therapeutic and rehabilitation outdoor spaces for the elderly demographic," Multimodal Technol. Interact., vol. 4, Issue 76, 2020. doi: 10.3390/mti4040076.

[10] K. Kamiya et al., "Gait speed has comparable prognostic capability to six-minute walk distance in older patients with cardiovascular disease, " European Journal of Preventive Cardiology, vol. 25, No. 2, pp. 212–219, 2018. doi: 10.1177/2047487317735715,

[11] A. Jalal, M. Batool, and K. Kim, "Stochastic Recognition of Physical Activity and Healthcare Using Tri-Axial Inertial Wearable Sensors," Applied Sciences, vol. 10, No. 20, 7122, 2020. https://doi.org/10.3390/app10207122

[12] H. Feng et al., "Association between accelerometer-measured amplitude of rest–activity rhythm and future health risk: a prospective cohort study of the UK Biobank, " The Lancet Healthy Longevity, vol. 4, Issue 5, e200-e210, 2023.

[13] N. Davoody and M. Hägglund, "Care professionals' perceived usefulness of ehealth for post-discharge stroke patients," Exploring Complexity in Health, pp. 589-593, 2016. doi: 10.3233/978-1-61499-678-1-589.

[14] S. Díaz, J. B. Stephenson, and M. A. Labrador, "Use of wearable sensor technology in gait, balance, and range of motion analysis," Applied Sciences, vol. 10, Issue 1, 234, 2020. doi: 10.3390/app10010234

[15] F. Muheidat, and L. Tawalbeh, "In-home floor based sensor system-smart carpet- to facilitate healthy aging in place (AIP)," IEEE Access, vol. 8, pp. 178627-178638, 2020. doi: 10.1109/ACCESS.2020.3027535

[16] A. Torku, A. P. C. Chan, E. H. K. Yung, J. Seo, and M. F. A. Afari, "Wearable sensing and mining of the informativeness of older adults' physiological, behavioral, and cognitive responses to detect demanding environmental conditions," Environment and Behavior, vol. 54, Issue 6, pp. 1005-1057, 2022. doi: 10.1177/00139165221114894

[17] K. Meng, X. Xiao, W. Wei, G. Chen, A. Nashalian, S. Shen, X. Xiao, and J. Chen, "Wearable Pressure Sensors for Pulse Wave Monitoring," Advanced Materials, vol. 34, Issue 21, 2109357, pp. 1-23, 2022. https://doi.org/10.1002/adma.202109357

[18] T. Funayama, Y. Kogure, K. Hori, and Y. Uchida, "A pilot study on the relationship between daily life and biological information," Human Interface Society in Japanese, vol. 22, No. 2, pp. 31-34, 2020, ISSN 2188-6652.

[19] T. Funayama, Y. Uchida, and Y. Kogure, "Assessment of Walking Condition Using Pressure Sensors in the Floor Mat," Global Health 2022, pp. 7-12, 2022.

[20] 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,Issue 5, pp. 33-43, 2023.

[21] 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.

[22] H. M. Stauss, "Identification of blood pressure control mechanisms by power spectral analysis," Clinical and Experimental Pharmacology and Physiology, vol. 34, No. 4, pp. 362-368, 2007, https://doi.org/10.1111/j.1440-1681.2007.04588.x.

[23] R. Luke, M. J. Shader, and D. McAlpine, "Characterization of Mayer-wave oscillations in functional near-infrared spectroscopy using a physiologically informed model of the neural power spectra," Neurophotonics, vol. 8, Issue 4, pp. 041001-1 - 041001-9,2021. https://doi: 10.1117/1.NPh.8.4.041001

[24] J. Krohova et al., "Multiscale Information Decomposition Dissects Control Mechanisms of Heart Rate Variability at Rest," Entropy, vol. 21, Issue 5, 526, 2019.

[25] J. Hayano, "Long–term Heart Rate Variability : Benefits and Pitfalls," Japanese journal of biofeedback research, vol. 2, No. 46, pp. 83-90, 2019, ISSN 03861856.

[26] M. Ambrosetti et al., "Secondary prevention through comprehensive cardiovascular rehabilitation: From knowledge to implementation. 2020 update," European Journal of Preventive Cardiology, vol. 8, No. 5, pp. 460-495, 2021.

Common Data Model for Interoperability of Observational Health Data: Bulgarian Diabetes Register Pharmacology Case Study

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

Petko Kovachev
Faculty of Math. and Informatics
Sofia University St. Kliment Ohridski
Sofia, Bulgaria
e-mail: az@petko.info

Simeon Abanos
Faculty of Math. and Informatics
Sofia University St. Kliment Ohridski
Sofia, Bulgaria
e-mail:simeonabanos@gmail.com

Dimitar Tcharaktchiev
Department of Medical Informatics
Medical University
Sofia, Bulgaria
e-mail: dimitardt@gmail.com

***Abstract*—This paper demonstrates the potential of a standard common data model to facilitate access to observational data and extract knowledge. The common data model enables platform interoperability for computational health technologies allowing assessment of the burden caused by the pharmacology costs on the healthcare system. It helps understanding the trends and effects in using different classes of drugs for diabetes treatment by exploring clinical data from the Bulgarian diabetes register. Unlike most regularly published reports on diabetes prevalence, the research results are obtained from a population-based study rather than applying aggregated statistical estimates. The Bulgarian Diabetes Register is a public common data model implementation allowing to overcome platform interoperability problems. It contains the latest and complete dataset of outpatient records of 501,065 distinct patients with diabetes in Bulgaria in 2018. The pharmacology case study reports new results for better assessment of the cost burden created by prescribing drugs for diabetes. Two major groups of drugs are considered- drugs for treatment of diabetes and related comorbidities. Novel drug diabetes therapies are just evolving in 2018, while the Metformin prescriptions prevail significantly. The costs are evaluated both at patient-centric level 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-observational data; platform interoperability; Common Data Model; diabetes register; pharmacology cost analysis.

I. INTRODUCTION

Modern healthcare more than ever depends on computer information technologies for data processing and exchange of medical information. Huge amounts of data are generated by people and digital devices that participate in the execution of almost all business processes in the healthcare system. For example, national diabetes registers maintain data describing the health status of diabetics. Such clinical data are collected routinely during health care procedures under real-world conditions. This kind of clinical data is collectively referred to as observational health data. Electronic Health Records (EHR) are used to accumulate 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 [1]. Unlike the EHR, an Electronic Medical Record (EMR), such as the Outpatient Record (OpR) provide a narrower view of the patient's medical history than the EHR because it is maintained by a single Healthcare Provider (HP) [2]. Similarly to 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 enable sharing of knowledge, skills and experience through communication between the actors in the healthcare system, provide a basis for research and education, satisfy organizational and legal requirements [3]. Nowadays, all of these opportunities for utilizing EHRs cannot be fully exploited. The reason is the lack of platform and data 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 [4]. Data exchange schemas and standards for reference models have been introduced for sharing EHR data across clinicians, patients and communities [5] [6] [7] . 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 [8]. Foundational

 ISBN: 978-1-68558-112-1

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 [9] [10]. 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 such data without human intervention [11] [12]. 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) [13]. 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 [14] [15]. 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 diabetics (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 in order to receive aggregated results after 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 [16] [17] [18] 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 [19]. 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 [10] [20]. 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 [16] introduces its own standard organization and representation of EHR data for a distributed network of nine population- based Clinical Research Networks of data contributors (14 billion diagnoses, 2.6 billion medication orders and 9.8 billion laboratory results) [21]. 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, enrollment, pharmacy, demographics, lab, death and vital signs (more than 365 million unique patient identifiers, 16 billion pharmacy dispensings, 15 billion unique medical encounters, 45 million laboratory test results) [17] [22]. This CDM is extensible to any data source because data is represented as detailed as possible. Thus, the Sentinel CDM is flexible about demands for running data queries in any type of analysis. Queries are processed in a distributed pattern as follows. 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. 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

 ISBN: 978-1-68558-112-1

the EHDEN project of the EU (118 EHDEN data partners, more than 1,12 billion unique patient identifiers) [23]. 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 cost, 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 has allowed us to map to OMOP CDM health data from 6,887,876 outpatient records issued in Bulgaria to patients 501,065 with diabetes during their encounters to GPs or HPs in 2018 [15]. Meta data of the thus obtained OMOP CDM of the Bulgarian Diabetes Register are published in the EHDEN Portal (Figure 1).

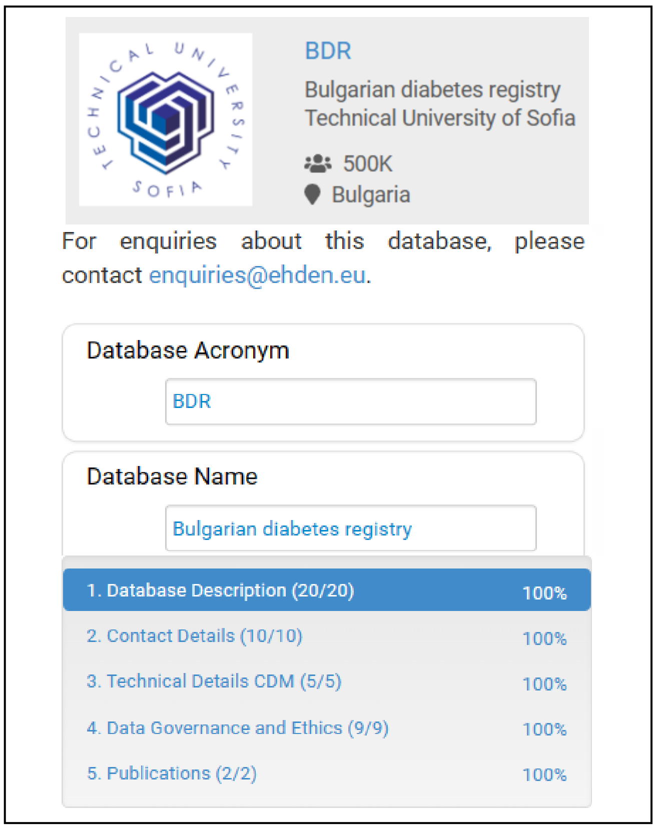


Figure 1. Link to the OMOP CDM of BDR inside the EHDEN Portal.

The distribution of diabetics (Type 1 and Type 2) relative to the population of the corresponding administrative region is displayed in Figure 2. This figure shows that most of the people living in the northern part of the country and especially, in the north-west part, suffer from 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.

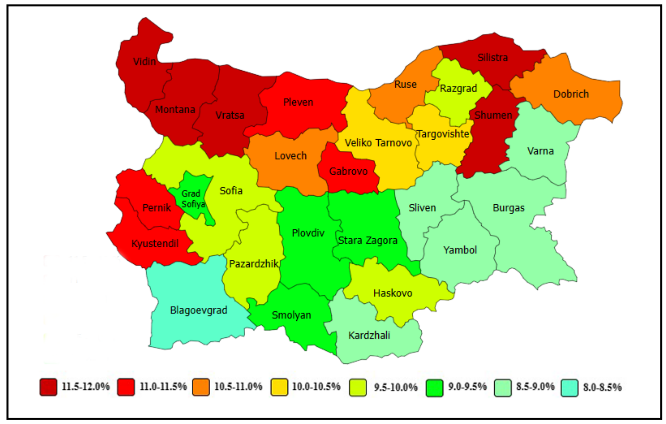


Figure 2. Distribution of patients with diabetes in Bulgaria in 2018.

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

These records contain administrative data and encoded 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 drug codes for medications that are reimbursed
- ✓ Encodings for examinations and procedures
- ✓ Codes describing specialized health care
- ✓ Codes describing hospitalization need
- ✓ Codes for planned consultations,
- ✓ Laboratory tests and medical imaging

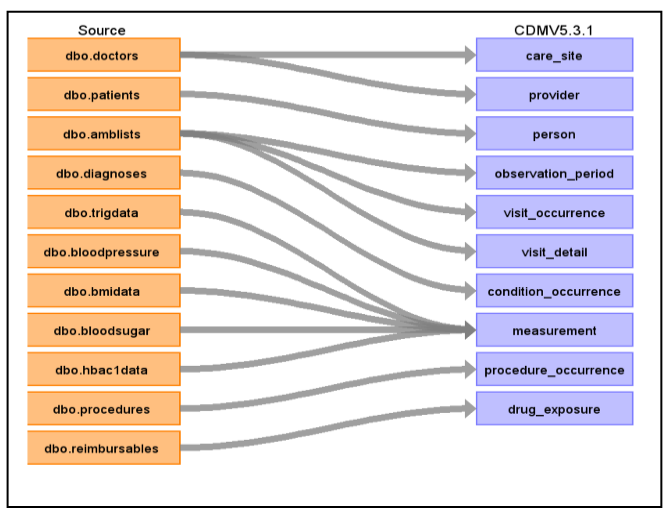


Figure 3. Mapping of outpatient records to OMOP CDM.

 ISBN: 978-1-68558-112-1

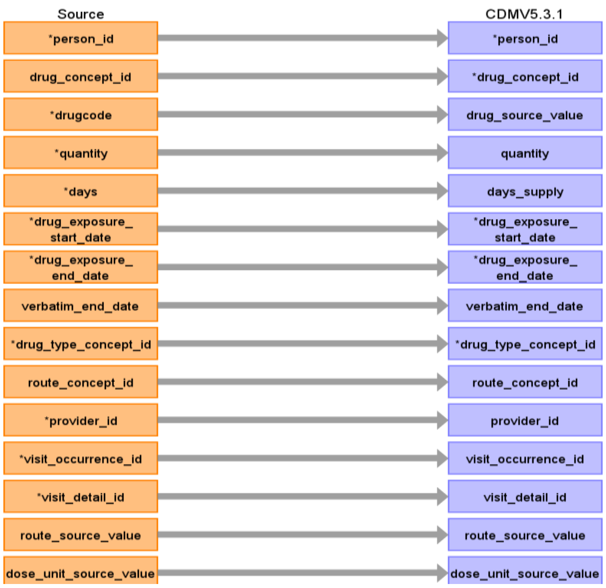


Figure 4. Mapping to table *drug_exposure* of the OMOP CDM.

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

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

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, Glimepiride, Gliclazide, Tolbutamide,Chlorpropamide, Tolazamide
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
T9	Statin-Dependent therapies	Simvastatin, Lovastatin, Ravastatin , Fluvastatin, Atorvastatin, Cerivastatin, Rosuvastatin, Ppitavastatin

The existing literature distinguishes several distinct classes among the drugs for diabetes treatment [19] [24]. These classes are presented in Table 1 where the custom *Code* introduced for shortness and 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 to be the most modern and promising [19]. This is another reason to find out what is the share of sales of these drugs. Similar interest represents the distribution of sales of drugs prescribed for treatment of diabetes comorbidities. For convenience in referencing these drugs we introduce the drug encodings displayed in Table 2 for the most frequently encountered comorbidities among patients with diabetes. By means of these codes, 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, C03, C07, C08, C09, C10
A1	Antithrombotic agents	B01
N	Nervous system disorders	N01-N07
G	Urological disorders	G04
S	Ophthalmolotical disorders	S01
L	Endocrine disorders	L02
M	Ttreatment of bone diseases	M05
R	Asthma drug categories	R03

In addition to table *drug_exposure* the analytical code in this study makes use of tables *person*, *condition_occurence*, *observation_period*, *visit_occurence* of the CDM. The results of executing this code 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 (4.43% male and 5.35% female). Unlike other public data, 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 based on 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 viewpoints. 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 drugs prescribed to diabetics in Bulgaria in 2018 is 160,766,702 euros where 96,171,943 euros is the amount for prescribed drugs from Table 1. It makes about 321 euros per diabetic patient, where 129 euros and 192 euros are spent on the average for drugs for treatment of diabetes comorbidities

(Table 2) and the diabetes itself (Table 1). Accordingly, 59.82% of the TC are for drugs prescribed for diabetes treatment (Table 1), where 61.51% is the share of the insulin class of drugs.

In the beginning, we have explored 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 5 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 kind of drugs are usually prescribed for initial treatment of Type 2 diabetes and besides, the number patients with this diabetes type prevail significantly over the patients with Type 1 diabetes. This explains the peak value in the prescriptions for Metformin drugs.

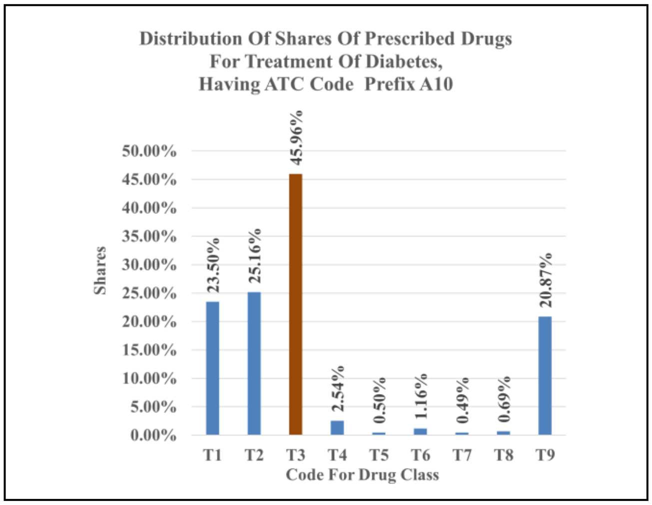


Figure 5. Shares of prescriptions for diabetes treatment.

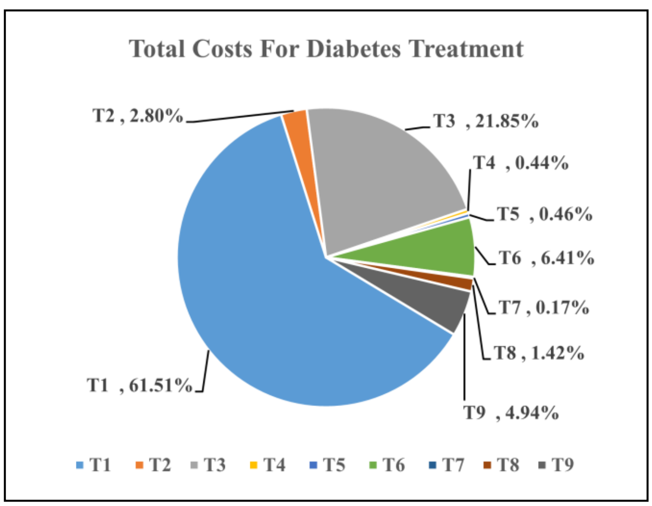


Figure 6. Total costs of drugs for diabetes treatment.

In terms of costs the shares of the drugs in Table 1 change as it is displayed in Figure 6. 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 5. 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.

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 are the costs for treatment of the most frequently 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.

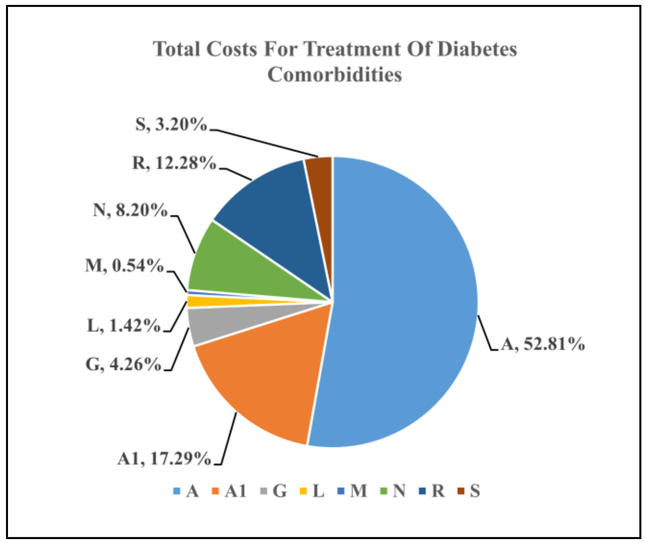


Figure 7. Total costs of drugs for treatment of diabetes comorbidities.

Figure 7 confirms that drugs for cardiovascular disorders and drugs with antithrombotic agents (code A and A1 in Table 2) have the greatest weight (70.10%) in the TC for treatment of comorbidities. The drugs for treatment of asthma of (code R in Table 2) are at the second place (12.28%) 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 8.2% 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.

In conclusion, note that Table 1 and Table 2 entries do not exhaust all the drugs classes prescribed to diabetics. For example, drugs that are given to diabetics but not mentioned in these two tables are drugs for treatment of rare diseases or disorders caused by immune deficiency. Most of these drugs are rather expensive and represent a huge burden in the overall amount reimbursed to patients for treatment of diabetes (135% of the TC of drugs from Table 1 and Table 2). In case we add these extra costs to the TC then we get average 675 euros per diabetic patient expenses for prescribed drugs.

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.

The pharmacology case study considered here is just one example of the potential for exploring the health data. 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 [24] 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 this publication is 58.90% against the above quoted percentage 61.51%. Another match is established in the reported share of cardiovascular drug costs with respect to all drug costs 33.5% against 34.2 % found in our study. There is, however, a great difference in the average cost per diabetic patient, 1066 euros against 675 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. Since the finance reports of NHIF are public [25], 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 (96,171,943 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 quantitates 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. 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 to overcome 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 the 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. 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 80-10-8/11.04.2023 of the Scientific Research Fund of Sofia university St. Kliment Ohridski.

REFERENCES

[1] K. Dipak, "Electronic Health Record Standards," Yearbook of Medical Informatics, vol. 45, no. 01, pp. 136-144, 2006.

[2] L. Spooner and K. Pesaturo, "The Medical Record," in Fundamental Skills for Patient Care in Pharmacy Practice, Jones & Bartlett Learning, 2014.

[3] B. Blobel, "Interoperable EHR Systems–Challenges, Standards and Solutions," European Journal for Biomedical Informatics, vol. 14, no. 2, pp. 10-19, 2018.

[4] HIMSS, Dictionary of Healthcare Information Technology Terms, Acronyms and Organizations, 5th ed., CRC Press, 2019, p. 75.

[5] HL7, "HL7 FHIR Release 5," HL7, 26 March 2023. [Online]. Available: http://hl7.org/fhir/. [Accessed 11 June 2023].

[6] 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 June 2023].

[7] 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 June 2023].

[8] HIMSS, "Interoperability in Healthcare," Healthcare Information and Management Systems Society, 2023. [Online]. Available: https://www.himss.org/resources/interoperability-healthcare.

[9] 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(1), no. 4, pp. 1-7, 2019.

[10] S. Kent et al., "Common Problems, Common Data Model Solutions: Evidence Generation for Health Technology Assessment," Pharmacoeconomics, vol. 39, no. 3, pp. 275-285, 2021.

[11] SNOMED International, "SNOMED CT," 2023. [Online]. Available: https://www.snomed.org/. [Accessed 14 June 2023].

[12] 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].

[13] R. Claire, J. Elvidge and D. Dawoud, "EHDEN - D2.7 - Workshop for HTA Community Report," Zenodo (https://doi.org/10.5281/zenodo.7550919), 2023.

[14] EHDEN, "Bulgarian Database Register(BDR)," EHDEN Portal, 2023. [Online]. Available: https://portal.ehden.eu/. [Accessed 5 June 2023].

[15] 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 June 2023].

[16] 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 June 2023].

[17] 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 June 2023].

[18] Observational Health Data Sciences and Informatics, "OMOP Common Data Model," ohdsi.org, 2022. [Online]. Available: https://ohdsi.github.io/CommonDataModel/. [Accessed 10 June 2023].

[19] A. D. Dahlén et al., "Trends in Antidiabetic Drug Discovery: FDA Approved Drugs, New Drugs in Clinical Trials and Global Sales.," Frontiers in Pharmacology vol. 12:80754, pp. 1-16, 2022.

[20] 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, vol. 255, Studies in Health Technology and Informatics, IOS Press, 2018, pp. 60 - 64.

[21] C. B. Forrest et al, "PCORnet® 2020: Current State, Accomplishments, and Future Directions," Journal of Clinical Epidemiology, vol. 129, p. 60–67, 2021.

[22] 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 June 2023].

[23] Observational Health Data Sciences and Informatics, The Book of OHDSI, https://ohdsi.github.io/TheBookOfOhdsi/, 2021.

[24] L. Scalone 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.

[25] 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 June 2023]

Exploring Episodic Future Thinking (EFT) for Behavior Change: NLP and Few-Shot In-Context Learning for Health Promotion

Sareh Ahmadi
Department of Computer Science
Virginia Tech
Blacksburg, VA, 24061, USA
email:saraahmadi@vt.edu

Edward A. Fox
Department of Computer Science
Virginia Tech
Blacksburg, VA, 24061, USA
email:fox@vt.edu, ORCiD 0000-0003-1447-6870

***Abstract*—Maladaptive health behaviors are closely linked to lifestyle-related diseases, such as obesity and type 2 diabetes. One significant factor contributing to maladaptive behavior is delay discounting, the tendency to prioritize immediate rewards over delayed ones. Episodic Future Thinking (EFT) is an intervention to reduce delay discounting and promote behavior change. EFT involves mentally simulating future events in a vivid manner, influencing decision-making and emotional well-being. Studies show EFT's effectiveness in reducing delay discounting and its potential for improving various health behaviors, including exercise and medication adherence. However, EFT's mechanisms of action and the conditions that impact its efficacy are unknown. This paper describes a study of EFT 'cue texts' to determine what makes them effective. It explains a new and efficient method to classify such texts with a few data, which can be used for further analysis to identify what characteristics of the texts lead to positive health outcomes. Classification framework is built using the FLAN-T5 large language model, with good results from zero-shot, and better results from few-shot in-context learning. This approach may be extended to address other behavioral health, wellness informatics, and technology-related approaches to global health challenges.**

Keywords*— *Episodic Future Thinking (EFT); delay discounting; maladaptive health behavior; Natural Language Processing (NLP); zero-shot learning; few-shot in-context learning.

I. INTRODUCTION

Maladaptive health behavior refers to actions or habits that are detrimental to a person's physical or mental well-being. Smoking and excessive alcohol consumption are examples of these behaviors. Maladaptive health behaviors are closely linked to lifestyle-related diseases, e.g., obesity, type 2 diabetes (T2D), cardiovascular diseases, certain types of cancer, and respiratory conditions. Promoting healthy lifestyle choices – regular exercise, a balanced diet, stress management, and avoidance of harmful substances – can significantly reduce the risk of developing these diseases. Interventions focusing on behavior change, support, and education can help individuals adopt healthier habits.

One contributing factor to maladaptive health behavior and lifestyle-related diseases is Delay Discounting (DD), which refers to the tendency to devalue delayed rewards in favor of immediate gratification [1]. Many unhealthy behaviors, and diseases related to lifestyle, are connected to DD [2]. Excessive discounting of delayed rewards is observed not only in substance-dependent individuals but also in individuals with behavioral disorders, such as pathological gambling, overeating, obesity, and Attention Deficit Hyperactivity Disorder (ADHD). Interventions could help with treating addiction and disorders linked to excessive discounting.

An increasing number of health behaviors and populations have been targeted by Episodic Future Thinking (EFT) [3] as an intervention for behavior change, aiming to decrease DD. EFT is a cognitive process that involves mentally simulating or envisioning future events in a detailed and vivid manner. It allows individuals to project themselves into the future and imagine specific situations, actions, and outcomes. EFT has garnered significant attention in recent years due to its potential role in influencing behavior, decision-making, and emotional well-being [3].

When investigating the impact of EFT on both DD and health behavior, participants commonly produce written or spoken depictions of personally significant future events. These descriptions are subsequently utilized as prompts/cues to facilitate EFT during decision-making tasks conducted in a laboratory setting or in real-world environments [4].

One study [5] focused on the association between DD and glycemic regulation, medication adherence, and eating and exercise behaviors in adults with prediabetes. It suggests that DD is a significant predictor of glycemic control and health behaviors in adults with prediabetes. Modifying DD can improve glycemic control and prevent the progression from prediabetes to T2D; interventions such as EFT may be beneficial. Another study [6] examined the effects of EFT on medication adherence in individuals with T2D, and potential mechanisms underlying these effects, such as improvements in prospective memory and DD. EFT had a positive impact on medication adherence among participants with T2D. Further research [7] focused on the long-term effects of EFT training on DD in individuals with prediabetes, as well as its impact on weight, HbA1c levels, and physical activity. Results indicate that EFT training can lead to sustained changes in DD and that

 ISBN: 978-1-68558-112-1

a combination of EFT and a low carbohydrate, low glycemic index diet can be effective for weight loss and glycemic control in individuals with prediabetes.

It has been shown in several studies that EFT is an effective intervention that can be scaled up to reduce DD and promote healthier behaviors. This study aims to enhance the effectiveness of EFT in preventing and treating T2D by gaining a better understanding of how EFT works and the factors that influence its efficacy. Previous and ongoing studies have shown significant variations in the content characteristics of EFT cues. The structure of the cues also varies, such as the extent to which they form a coherent narrative or describe achievement of health and personal goals, such as weight loss or financial planning. Additionally, cues differ in terms of imagery, vividness of events, emotional tone, and level of detail provided. We aim to build Natural Language Processing (NLP) classifiers to predict EFT content characteristics. In addition, our research aims to reduce the cost of annotating participant data for classification, and leverage more efficient and adaptable Large Language Models (LLMs). By utilizing techniques that rely on pre-trained LLMs and their ability to generalize to new tasks, we hope to pave the way for more accessible and scalable methods in NLP research. We are particularly interested in exploring the application of few-shot and zero-shot in-context learning techniques within the emerging field of instruction-tuned models.

In Section II, we discuss the application of NLP in the health domain, as well as the methodologies employed in constructing NLP classifiers. Section III delves into the zero-shot and few-shot classification of EFT data. The results are presented in Section IV, and we draw our conclusions in Section V.

II. Related Work

In recent years, NLP has been shown to help with global health challenges. In this domain, by leveraging corpora and learning approaches, NLP has demonstrated strong performance in various tasks, e.g., text mining [8], classification [9], sentiment mining [10], and information extraction [11]. In particular, NLP techniques may offer multiple perspectives in mental health research and in mental health clinical practice [12]. For instance, a study [13] explored the feasibility of automatically extracting schemas from thought records. A method for identifying the use of evidence-based psychotherapy for post-traumatic stress disorder was developed by applying NLP methods to clinical notes [14]. NLP also can address the knowledge gap in utilizing lifestyle modification data, including diet, exercise, and tobacco cessation, from Electronic Health Records (EHRs) for research purposes [15].

Furthermore, as the research in diabetes care is growing, and a considerable portion of real-world data exists in narrative form, NLP technology presents a viable solution for effectively analyzing narrative electronic data [16]. Given the success of NLP approaches, several studies have been dedicated to diabetes care and diseases [17] [18]. A high-performance NLP system [19] was developed for automatically detecting hypoglycemic events from EHR notes of diabetes patients. It can be utilized for EHR-based hypoglycemia surveillance and population studies to improve patient care and enhance research in diabetes management. A thorough thematic analysis was performed [20] to identify 12 themes of vulnerability related to the health and well-being of T2D patients by leveraging language models with high test accuracy. To understand the information needs of diabetics a classification schema for diabetes-related questions was developed by analyzing questions collected from a health website [21]. An investigation of the relationship between lifestyle counseling in primary care settings and clinical outcomes in patients with diabetes applied NLP to electronic notes [22] to retrieve and classify lifestyle modification assessments and advice.

Transformers such as BERT [23] offer a promising approach for building text classifiers, but one significant challenge lies in the amount and quality of data they require. Annotating data for training classifiers can be costly, time-consuming, and requires domain-specific knowledge. Manual annotation involves experts meticulously labeling large volumes of data, which is a resource-intensive and time-consuming task. Likewise, since many text datasets are imbalanced – with few instances of the minority category relative to those in the majority category – special care and techniques are required, as shown in our prior studies [24] [25].

Initially, fine-tuning was a dominant approach where these pre-trained language models were adapted to specific downstream tasks by further training on task-specific labeled data. While fine-tuning yields impressive results, it heavily relies on large amounts of task-specific labeled data, which could be limiting when labeled data is scarce or non-existent for certain tasks or domains. Prompt tuning [26] emerged as a response to address the limitations of fine-tuning. Instead of relying solely on task-specific labeled data, prompt tuning leverages the pre-trained language model's ability to generate text by providing input prompts. By constructing appropriate prompts, including task-specific information or instructions, language models can be fine-tuned on new tasks without the need for extensive labeled data. Zero-shot learning takes the concept of prompt tuning further by allowing language models to generalize their understanding to unseen tasks or categories. With zero-shot learning, a pre-trained language model is capable of performing tasks for which it has never been explicitly trained [27]. By leveraging auxiliary information or prompts, such as textual descriptions or instructions, zero-shot learning enables language models to classify or generate text for new categories or tasks. Few-shot in-context learning [28] [29] builds upon the zero-shot learning paradigm and focuses on adapting language models to new tasks or categories with only a small number of labeled examples (also referred to as demonstrations). In few-shot in-context learning, the language model leverages a few labeled examples to quickly learn task-specific patterns or characteristics. The training examples are concatenated and provided as a single input to the model, which suits the k-shot learning scenario. GPT3 [28] showed emergent few-shot learning by simply pre-pending examples of the task as the input to the model. During testing, the model is assessed on a new target task with k-training examples. This approach significantly reduces both computation costs and the data requirements for adaptation to new tasks, making it particularly useful when labeled data is scarce or expensive to acquire.

In the context of few-shot in-contex learning, the term “k-shot” refers to the number of labeled examples available for each task. For example, if a model is trained on a 5-shot for a classification task, it means each task is provided with only five labeled examples or demonstrations. The model then uses this limited data to make predictions when presented with new tasks during testing. The value of k can vary depending on the specific few-shot learning scenario and the available data. In this context, k=0 represents zero-shot learning, k=1 corresponds to one-shot learning, and k>1 indicates few-shot learning. It is expected that the model’s performance improves with a larger k because it can learn from more examples. Additionally, the inclusion of a prompt provides additional context, which enhances the model’s accuracy, particularly when k is small.

The capacity of LLMs to adjust to specific tasks using few-shot demonstrations (in-context learning) has been observed [30]. As the size of LLMs increases, emergent capabilities have become more apparent [31] [32]. LLMs have demonstrated the capacity to generalize to unfamiliar tasks through instruction-based learning. Instruction tuning is a novel approach in NLP that utilizes natural language instructions to enable zero-shot and few-shot performance of language models on previously unseen tasks. Instruction-tuned LLMs are fine-tuned with inputs and outputs that are instructions, using techniques such as Reinforcement Learning from Human Feedback (RLFH) [33], or instruction-tuned based on supervised fine-tuning which involves the process of refining a pre-trained language model using public benchmarks and datasets which have instruction template formats. To enhance the fine-tuning process, additional instructions are introduced, either manually created or automatically generated, to augment the training data [34] [35]. These approaches can improve the LLMs’ ability to follow instructions and safely adapt to new tasks.

III. METHOD

Given the superiority of the instruction-tuned language models [30], as the first step, we need to choose a pre-trained instruction-tuned model. The FLAN-T5 11B model (11 billion parameters, FLAN-T5 XXL) [36] outperforms the PaLM 62B model (62 billion parameters) [36], a novel transformer language model trained using the Pathways ML system [32] which is a recently developed machine learning system that allows for highly effective training. Moreover, FLAN-T5 excelled on difficult tasks in the BIG-Bench dataset [36]. Given the exceptional performance of the FLAN-T5 model and the public release of it, this model was selected. It is the instruction-tuned version of the T5 encoder-decoder model [37] that has undergone fine-tuning across a variety of tasks to follow instructions. It is able to perform zero-shot NLP tasks, as well as few-shot in-context learning tasks.

In our dataset, the participants were asked to write short texts (cues) about the events for different time frames, ranging from one month to ten years. Participants generated detailed and vivid descriptions of these events. The data used in this study originated from 18 different research studies conducted by medical research teams at two universities (Virginia Tech and the University of Buffalo). These studies specifically investigated the impact of EFT on diabetes and other relevant health-related outcomes. In total, the dataset comprises approximately 11,000 cues, each a few sentences in length. An example of selected data from one participant is as follows:

- In about 1 month, I am playing golf with my friends. We are having a great time and enjoying the company and competition. We laugh and have a great time.
- In about 3 months, I am picking my daughter up from college. I am excited she is done with school and we go to lunch at our favorite sushi restaurant and enjoy each other’s company.
- In about 6 months, I am fishing in the bay with friends. We are on a charter boat and excited to catch some nice fish. We bet on who will catch the biggest fish.

The goal is to build a classification framework to predict the topic of the cues, as well as a level/value for three categories related to imagery, featuring variation in: event vividness, episodicity, and emotional valence. Table I shows the definitions for the 14 categories.

A subset of the data is used for manual labeling in Amazon Mechanical Turk. Each text was labeled by three different annotators to ensure the quality of the labeling process. The annotators are given the definition, and an example, for each category (Table I). Overall there are 400 labeled texts. For binary categories, we randomly sample 10 labeled texts belonging to a category (positive examples) and 10 labeled texts not belonging to that category (negative examples), i.e., 20-shots, and the 380 remaining are used for testing the few-shot and zero-shot settings. For 10-shot, 5 of the positive and 5 of the negative sampled labeled texts are used. The test data is the same (i.e., for zero-shot, 10-shot, and 20-shot). For the three-class classification, we randomly sample 10 examples from each class (30 labeled data for few-shot learning) and the 370 remaining are used for testing. For the 15-shot case, we use half of the labeled examples.

For an instruction-tuned model, the instructions and context provided to a language model are encompassed within prompts. Therefore it requires prompt engineering such that the input to the model contains well-crafted prompts, ensuring meaningful guidance, rather than blindly inputting everything without context. Prompt engineering involves the process of designing and refining prompts to effectively utilize language models for various applications [38]. Typically, the components that constitute a prompt are as follows:

- Instruction: Instruct the model on the desired actions,

 ISBN: 978-1-68558-112-1

TABLE I
CATEGORY DEFINITION

Class	Definition
Vivid: not	The text contains no details about the event. It is difficult to imagine the event. No context has been given regarding the event.
Vivid: moderately	The text contains only a few details or mostly non-specific details. The reader is left to fill in gaps, making it somewhat hard to imagine the event. More details could have been provided describing the event. Some context has been given regarding the event
Vivid: highly	The text contains sufficient and specific details so that the event described is readily and easily imaginable. A considerable amount of context has been given regarding the event.
Episodic: not	The writer primarily describes general knowledge of events or occurrences. The event is described as if the writer is not present or personally experiencing the event.
Episodic: moderately	The writer describes both personal experiences, events, and actions in addition to general facts or ideas. The writer is somewhat in the moment but also adds in a few facts or ideas.
Episodic: highly	The writer primarily describes personal experiences, events, and actions, NOT general facts or ideas. The writer is describing events as if they are currently experiencing them "in the moment". The writer provides details about their own emotions and/or what they hear, see, or feel.
Emotion: negative	Primarily contains references to negative emotions or behaviors, including sadness, crying, or anger.
Emotion: neutral	Contains references to both positive and negative emotions or behaviors or contains weak or ambiguous references to positive or negative emotions and behaviors.
Emotion: positive	Contains references to both positive and negative emotions or behaviors or contains weak or ambiguous references to positive or negative emotions and behaviors.
Health	Contains an obvious, specific reference to physical or mental health. Examples include but are not limited to improved or worse physical state or mental health, and intentional changes in behaviors to improve health and health outcomes.
Recreation	Contains obvious or specific references to engaging in an activity for leisure or fun while not working at one's job. Examples include but are not limited to sports or physical activities like running or hiking, art, movies and television, or hobbies like gardening.
Better-me	Contains obvious or specific references to "a better me", including personal development, self-improvement, making positive changes in life, achievements, hard work, or determination. May contain references to the idea that things are looking up or getting better.
Celebration	Contains an obvious, specific reference to a celebration or a celebratory event.
Food	Contains obvious or specific references to food, eating, cooking, or a meal. Eating or food is a major and essential component of the text.
Alone	Contains an obvious, specific reference to events and activities which shows being done alone.
Family	Contains obvious or specific references to family (immediate or extended). Family is a major and essential component of the text.
Partner	Contains an obvious, specific reference or mention of a romantic partner.
Friends	Contains obvious or specific references to a friend or friends (non-family members). Friends are a major and essential component of the text.
Pet	Contains obvious or specific references to a pet, not any animal.

guide its utilization of external information (if available), and outline the construction of the output.

- Context: Serves as supplementary knowledge for the model, providing additional information. They can be manually included within the prompt, obtained through a vector database using retrieval augmentation, or acquired through alternative methods.
- Input Data: Refers to the input provided by a human user (i.e., the user input or query)
- Output indicator: Denotes the starting point of the to-be-generated text

Although not all prompts incorporate these elements, a well-crafted prompt frequently incorporates two or more of them. To adapt the model to our dataset, the instruction for the model is set as the category definition followed by some demonstrations (examples) for few-shot in-context learning. For classification, a demonstration involves an input x and its corresponding ground-truth label y. The model is provided with a sequence of such demonstrations, followed by a test input. The objective is for the language model to predict the label of this final data point. The demonstrations are sampled randomly for the model. Below is an example of an input instruction template for the 3-shot setting. Fig. 1 also depicts the framework for building a classifier. For the zero-shot setting, the input prompt contains only the instructions, query, and output indicator. The block below is an example.

”””Classify the given text into three categories: not episodic, moderately episodic, and highly episodic based on the definition for each category.

- Highly episodic definition: The writer primarily describes personal experiences, events, and actions, NOT general facts or ideas. The writer is describing events as if they are currently experiencing them at the moment. The writer provides details about their own emotions and/or what they hear, see, or feel.
- Moderately episodic definition: The writer describes both personal experiences, events, and actions in addition to general facts or ideas. The writer is somewhat in the moment but also adds in a few facts or ideas.
- Not episodic definition: The writer primarily describes general knowledge of events or occurrences. The event is described as if the writer is not present or personally experiencing the event.

In about 10 years, I am blowing out the candles on my birthday cake and feel pleased that my family has gathered because they love me so much. I am smiling because my adult children are making fun of how many candles are on my cake this year. I hear one of them jokingly say it's time to call the fire department and everyone laughs with love in their voices.
Output: Highly episodic

##

Input: In 6 months, I am visiting my mother-in-law. She is down visiting from Indiana. She is staying with us for a week. We have waited for her to come down for a long time because we have not seen her in years. We are ecstatic she is here and is visiting. The kids are excited that they finally get to see her. We are going shopping and do other fun things while she is here. We are looking forward to all the excitement that happens while she is visiting.
Output: Moderately episodic

##

In about 5 years, my car is paid off. Sweet! Who doesn't love any more car payments?
Output: Not episodic

##

Input: In about 4 years, I am on vacation at the beach relaxing in the sun under a big umbrella. I am with my husband and my daughter and we are finally taking a family vacation. My husband is playing with my daughter out in the ocean, jumping waves, and helping her on a boogie board. I am back in my chair on the sand, my sunglasses are on and I am enjoying the quiet and the warmth. The sand is warm between my toes and the weather is perfect. I am feeling at peace.
Output: ”””

For few-shot learning, given that the maximum sequence length for the model is 2048 tokens, we can provide up to 30-shots i.e., we randomly sample 10 examples per class for the three-class classification. To be consistent, for the binary classes, we also randomly sampled 10 positive and 10 negative examples for each category. Therefore, we can provide the model with 20-shots for each binary category

 ISBN: 978-1-68558-112-1

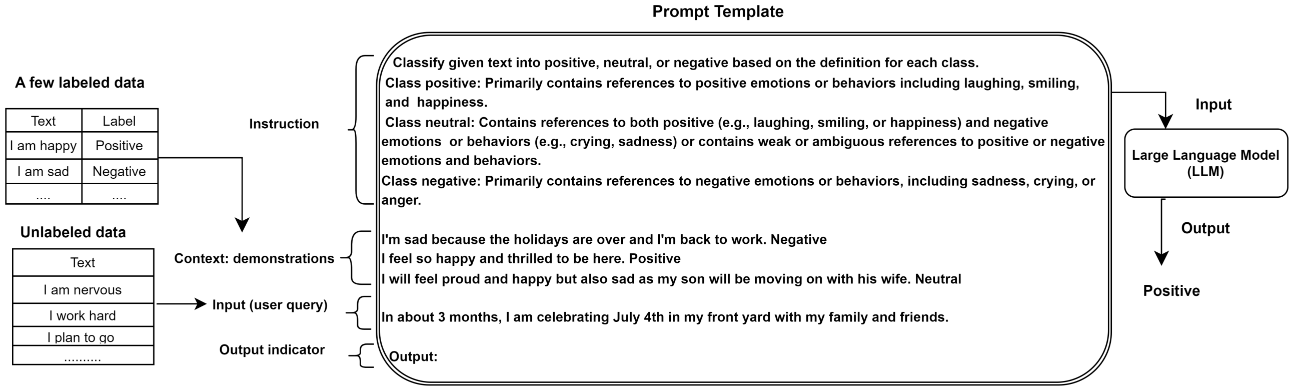


Fig. 1. Classification framework for few-shot in-context learning

(topical categories) and 30 shots for the 3-class categories (episodic, vivid, and emotion).

IV. Results

We present the results for model performance for zero-shot and few-shot in-context learning. To evaluate the classification, we measure the accuracy and macro F1-score. Each experiment is run for 3 trials and the average result is reported in Tables II and III. Table II shows the classifier performance for the three class categories while Table III shows the performance for topical (binary) categories. Performance with few-shot examples is better than the zero-shot learning approach, for all the classes, allowing the model to better comprehend and distinguish for those classes. The improvement shown in Table II through few-shot learning is impressive, i.e., for the three classes of episodicity, vividness, and emotion. Increasing the number of demonstrations from 5 to 10 per class also helped the model to improve performance. As a result, the few-shot learning paradigm demonstrates superior performance, effectively capturing underlying patterns in the data, surpassing the limitations of zero-shot learning, which relies solely on generalizing to entirely unseen classes without any labeled training samples. Our model demonstrates strong performance in the binary category for zero-shot setting, showcasing its capability to handle classification tasks effectively. However, its true capability becomes evident when faced with few-shot examples. Even with limited training data, the model exhibits superior adaptability and displays enhanced performance. These findings highlight the importance of learning from examples to enhance the classification capabilities of the model and showcase its potential for real-world applications with limited labeled data. In this study, we employed a single model for classification and observed notable advantages with few-shot learning. Overall, prompting one model provides a more efficient approach that can lead to enhanced performance and easier management of machine learning tasks. It is particularly advantageous when dealing with resource constraints, like those arising from the expense of manual labeling, especially when datasets are imbalanced and have very few positive/minority examples.

TABLE II
Performance of Flan-T5 for 3 Class Categories

category	zero-shot		15-shot		30-shot	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
episodic	60%	44%	89%	78.3%	**91.3%**	**83%**
vivid	74%	45%	81%	67.66%	**86%**	**84%**
emotion	80%	55%	81%	59.6%	**82%**	**72%**

TABLE III
Performance of Flan-T5 for binary categories

category	zero-shot		10-shot		20-shot	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
health	94%	93%	94%	93%	**94.3%**	**94.6%**
better-me	79%	77%	80.3	77	**81%**	**78%**
recreation	83%	80%	83%	81%	**85%**	**83%**
family	79%	79%	83.6%	83.6%	**85.3%**	**84%**
friend	85%	79%	89.3%	82.33%	**89.6%**	**83%**
future	96%	95%	99.6%	99.6%	**100%**	**100%**
food	55%	50%	95.6%	94.6%	**96%**	**96%**
pet	95%	84%	96.6%	87%	**98%**	**89.3%**
alone	91%	83%	91 %	83.3%	**92%**	**84%**
celebration	71%	69%	82%	79%	**82.3%**	**79.6%**
partner	84%	81%	94%	92%	**95%**	**94%**

V. Conclusion

This research examines content characteristics of EFT data generated by people who suffer from conditions such as diabetes. Little is known about how these content characteristics influence the effectiveness of EFT in promoting behavior change. We proposed to utilize a pre-trained instruction-tuned model and apply zero-shot and few-shot in-context learning for classification to predict the content and characteristics of the generated EFTs. This can then be used for in-depth analysis to pinpoint which text features contribute to positive health results. The proposed method serves as a powerful tool that addresses the barriers posed by traditional fine-tuning methods, which typically demand a large amount of labeled data. Unlike fine-tuning, few-shot in-context learning significantly reduces the data requirements, making it more accessible and applicable in scenarios where labeled data is scarce or costly to obtain. By utilizing a single pre-trained model for each classification task and adapting it to new tasks with only a few examples, this approach avoids the need for maintaining separate models for every specific classification task. This efficiency not only saves computational resources but also opens up opportunities for practical implementations across various domains and industries. In essence, few-shot in-context learning represents a new and effective NLP technique that bridges the gap between data scarcity and high-performance artificial intelligence, offering a promising pathway for further advances in health-related domains. Future work includes using other instruction-tuned LLMs that can handle longer input sequence lengths, as well as more generalization capabilities for different NLP tasks.

Acknowledgement

We would like to express our gratitude for the support provided by NIH NIDDK 3R01DK129567-02S1.

References

[1] G. J. Madden and W. K. Bickel, *Impulsivity: The behavioral and neurological science of discounting.* Amer. Psychological Assoc., 2010.

[2] W. K. Bickel, D. P. Jarmolowicz, E. T. Mueller, M. N. Koffarnus, and K. M. Gatchalian, "Excessive discounting of delayed reinforcers as a trans-disease process contributing to addiction and other disease-related vulnerabilities: emerging evidence," *Pharmacology & therapeutics*, vol. 134, no. 3, pp. 287–297, 2012.

[3] C. M. Atance and D. K. O'Neill, "Episodic future thinking," *Trends in cognitive sciences*, vol. 5, no. 12, pp. 533–539, 2001.

[4] J. M. Brown and J. S. Stein, "Putting prospection into practice: Methodological considerations in the use of episodic future thinking to reduce delay discounting and maladaptive health behaviors," *Frontiers in Public Health*, vol. 10, p. 1020171, 2022.

[5] L. H. Epstein *et al.*, "Delay discounting, glycemic regulation and health behaviors in adults with prediabetes," *Behavioral Medicine*, vol. 47, no. 3, pp. 194–204, 2021.

[6] L. H. Epstein, T. Jimenez-Knight, A. M. Honan, R. A. Paluch, and W. K. Bickel, "Imagine to remember: an episodic future thinking intervention to improve medication adherence in patients with type 2 diabetes," *Patient preference and adherence*, pp. 95–104, 2022.

[7] L. H. Epstein *et al.*, "Effects of 6-month episodic future thinking training on delay discounting, weight loss and HbA1c changes in individuals with prediabetes," *Journal of Behavioral Medicine*, vol. 45, no. 2, pp. 227–239, 2022.

[8] A. Abbe, C. Grouin, P. Zweigenbaum, and B. Falissard, "Text mining applications in psychiatry: a systematic literature review," *Int'l j. of methods in psychiatric research*, vol. 25, no. 2, pp. 86–100, 2016.

[9] W.-H. Weng, K. B. Wagholikar, A. T. McCray, P. Szolovits, and H. C. Chueh, "Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach," *BMC medical informatics and decision making*, vol. 17, no. 1, pp. 1–13, 2017.

[10] K. Denecke and D. Reichenpfader, "Sentiment analysis of clinical narratives: A scoping review," *J, Biomedical Informatics*, vol. 140, p. 104336, 2023.

[11] K. Kreimeyer *et al.*, "Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review," *Journal of biomedical informatics*, vol. 73, pp. 14–29, 2017.

[12] A. Le Glaz *et al.*, "Machine learning and natural language processing in mental health: systematic review," *J. of Medical Internet Research*, vol. 23, no. 5, p. e15708, 2021.

[13] F. Burger, M. A. Neerincx, and W.-P. Brinkman, "Natural language processing for cognitive therapy: extracting schemas from thought records," *PloS one*, vol. 16, no. 10, p. e0257832, 2021.

[14] S. Maguen *et al.*, "Measuring use of evidence based psychotherapy for posttraumatic stress disorder in a large national healthcare system," *Administration and Policy in evidence Health and Mental Health Services Research*, vol. 45, pp. 519–529, 2018.

[15] K. Shoenbill, Y. Song, L. Gress, H. Johnson, M. Smith, and E. A. Mendonca, "Natural language processing of lifestyle modification documentation," *Health Informatics J.*, vol. 26, no. 1, pp. 388–405, 2020.

[16] A. Turchin and L. F. Florez Builes, "Using natural language processing to measure and improve quality of diabetes care: a systematic review," *J. of Diabetes Science and Technology*, vol. 15, no. 3, pp. 553–560, 2021.

[17] D. H. Smith *et al.*, "Lower visual acuity predicts worse utility values among patients with type 2 diabetes," *Quality of life research*, vol. 17, pp. 1277–1284, 2008.

[18] Y. Zheng *et al.*, "Identifying patients with hypoglycemia using natural language processing: systematic literature review," *JMIR diabetes*, vol. 7, no. 2, p. e34681, 2022.

[19] Y. Jin *et al.*, "Automatic detection of hypoglycemic events from the electronic health record notes of diabetes patients: empirical study," *JMIR medical informatics*, vol. 7, no. 4, p. e14340, 2019.

[20] S. Wang, F. Song, Q. Qiao, Y. Liu, J. Chen, and J. Ma, "A comparative study of natural language processing algorithms based on cities changing diabetes vulnerability data," in *Healthcare*, vol. 10, no. 6, 2022, p. 1119.

[21] X. Zhou, Y. Ni, G. Xie, W. Zhu, C. Chen, T. Wang, and Z. Pan, "Analysis of the health information needs of diabetics in China," in *MEDINFO 2019: Health and Wellbeing e-Networks for All*, 2019, pp. 487–491.

[22] H. Zhang, S. I. Goldberg, N. Hosomura, M. Shubina, D. C. Simonson, M. A. Testa, and A. Turchin, "Lifestyle counseling and long-term clinical outcomes in patients with diabetes," *Diabetes care*, vol. 42, no. 9, pp. 1833–1836, 2019.

[23] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.

[24] S. Ahmadi, A. Shah, and E. Fox, "Retrieval-based text selection for addressing class-imbalanced data in classification," *arXiv preprint arXiv:2307.14899*, 2023.

[25] A. A. Shah, "Leveraging Transformer Models and Elasticsearch to Help Prevent and Manage Diabetes through EFT Cues," MS thesis defended 5/4/2023, Virginia Tech CS, http://hdl.handle.net/10919/115452.

[26] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–35, 2023.

[27] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[28] T. Brown *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[29] S. Min, M. Lewis, L. Zettlemoyer, and H. Hajishirzi, "MetaICL: Learning to Learn In Context," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2022, pp. 2791–2809.

[30] W. X. Zhao *et al.*, "A survey of large language models," *arXiv preprint arXiv:2303.18223*, 2023.

[31] J. Wei *et al.*, "Emergent abilities of large language models," *arXiv preprint arXiv:2206.07682*, 2022.

[32] A. Chowdhery *et al.*, "Palm: Scaling language modeling with pathways," *arXiv preprint arXiv:2204.02311*, 2022.

[33] L. Ouyang *et al.*, "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27 730–27 744, 2022.

[34] V. Sanh *et al.*, "Multitask prompted training enables zero-shot task generalization," in *ICLR 2022-Tenth International Conference on Learning Representations*, 2022.

[35] J. Wei, M. Bosma, V. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le, "Finetuned language models are zero-shot learners," in *International Conference on Learning Representations*, 2022.

[36] H. W. Chung *et al.*, "Scaling Instruction-Finetuned Language Models," *arXiv preprint arXiv:2210.11416*, 2022.

[37] C. Raffel *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *The Journal of Machine Learning Research*, vol. 21, no. 1, pp. 5485–5551, 2020.

[38] M. T. James Phoenix, *Prompt Engineering for Generative AI.* O'Reilly Media, Inc., 2024.

Towards Personalized Wellness: Detection and Analysis of Homogeneous Communities in Multi-Modal Biomedical Networks

Rama Krishna Thelagathoti and Hesham H. Ali
College of Information Science and Technology
University of Nebraska Omaha
Omaha, NE 68182, USA
email: rthelagathoti, hali@unomaha.edu

***Abstract*—In recent years, personalized wellness has attracted significant attention by many research groups. However, distinguishing between various groups associated with different health conditions, which is critical to achieve personalized healthcare, remains a complex challenge. The driving force behind this study is to harness the potential of biomedical networks to reveal homogenous groups with a high degree of similarity. Our primary goal is to uncover hidden relationships that may not be evident using traditional data analysis by discerning interaction patterns and similarities between nodes in the network. To construct a biomedical graph, we integrated mobility data and medical data, mainly Heart Rate Variability (HRV) data, for each subject, representing them as nodes. The edges between nodes reflect the similarity between subjects in both mobility patterns and HRV responses. Leveraging advanced network analysis techniques, we applied an appropriate network features to identify strongly connected homogenous groups within the graph. These groups represent subsets of subjects displaying highly similar mobility characteristics and HRV patterns. Our analysis identified five highly similar groups within the biomedical network based on the subjects mobility and HRV data. Remarkably, mobility features played a more pronounced role in the formation of these groups compared to HRV data, suggesting strong ties and similarities in mobility characteristics. These findings significantly contribute to our understanding of the interplay between movement behaviors, offering promising avenues for personalized interventions and enhanced clinical decision-making in the pursuit of advanced healthcare and improved well-being.**

***Index Terms*—Homogeneous groups; Network Analysis; Mobility Analysis; Personalized interventions.**

I. Introduction

Understanding homogeneous groups within biological networks is essential for advancing personalized therapies, designing targeted interventions, and developing individualized medicine [1] [2]. These homogeneous groups represent subsets of interconnected individuals or elements within a network who share similar characteristics, behaviors, or functional relationships [3] [4]. In other words, they are said to be cohesive in nature within the biomedical network. By analyzing the structure and dynamics of these groups, we can gain valuable insights into the complex systems of biological networks, uncover hidden patterns, and identify key factors driving their formation and functioning [5]. Cohesive groups play a pivotal role in personalized medicine and healthcare [6]. They offer a means to stratify patients into subgroups based on shared characteristics, enabling the development of tailored treatments that address the specific needs and responses of individuals within each group. Moreover, homogeneous groups can provide insights into disease subtypes, treatment response prediction, and the identification of novel therapeutic targets [6].

This study incorporates data from two modalities: mobility data collected from physical activity and medical data collected from HRV. These modalities are utilized to construct the biomedical network. Mobility data, including step counts, distance traveled, and activity patterns, offer insights into individuals' daily routines, physical activity levels, and overall mobility behaviors [7] [8]. HRV data, on the other hand, captures the variability in the time intervals between successive heartbeats, reflecting autonomic nervous system activity, physiological stress, and overall cardiovascular health [9]. The integration of mobility and HRV data allows for a comprehensive assessment of individuals' behavioral and physiological characteristics within homogeneous groups. By considering both modalities, we can gain a deeper understanding of the interplay between physical activity, physiological responses, and health outcomes. This holistic approach facilitates the identification of homogeneous groups that share not only mobility patterns but also physiological similarities, enhancing our ability to personalize interventions and treatments [10].

In this study, our primary objectives are three-fold. Firstly, we aim to construct a network representation that integrates mobility and HRV data. Secondly, we seek to apply an appropriate network threshold that defines the strength of connections within the graph. By adjusting the threshold, we can identify strongly connected homogeneous groups that exhibit distinct patterns and interrelationships. Lastly, we aim to analyze the mobility patterns and HRV characteristics within these groups. The rest of the document is organized as follows. Section II presents the detailed methodology while Section III contains the obtained results. Discussion of the

 ISBN: 978-1-68558-112-1

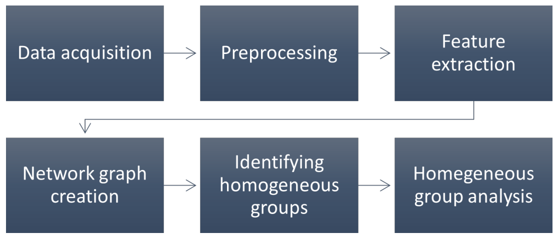


Fig. 1. The overview of the methodology.

results is presented in Section IV and Section V lists some of the limitations and future work of the study.

II. Method

A. Overview

Our study follows a five-step approach, as shown in Figure 1, to identify homogeneous groups in a biomedical network based on mobility and HRV data. We acquired multimodal data for the subjects, performed data preprocessing to clean and prepare the data, and extracted relevant features. Using these features, we constructed the network, representing subjects as nodes with edges reflecting their similarities. By applying a network threshold, we identified homogeneous groups, revealing strong ties based on mobility characteristics. The analysis emphasized the influential role of mobility features in the formation of homogeneous groups, providing valuable insights for personalized healthcare interventions.

B. Data Acquisition

The dataset utilized in this study is the HYPERAKTIV dataset, which is publicly available and can be accessed from a public repository [11]. The dataset comprises mobility data collected using wearable sensors, comprehensive diagnostic assessments, and demographic characteristics. It consists of two primary categories: individuals with Attention-Deficit/Hyperactivity Disorder (ADHD) and individuals without ADHD. But most of them were also diagnosed with other psychiatric disorders such as ADD (Attention Deficit Disorder), unipolar depression, bipolar depression, and anxiety. However, the assessment of these psychiatric disorders was made using self-reported feedback. Furthermore, the dataset includes a total of 85 patients, with 45 diagnosed with ADHD and 40 were diagnosed with other disorders in addition to ADHD. The mobility data in the dataset is captured through a wrist-worn actigraph device (Actiwatch), with a sampling frequency of 32 Hz. The motor activity data includes measurements of movement intensity in the x, y, and z-axes. In addition to the activity data, the dataset also contains HRV data. HRV data is ECG-based and recorded using a small chest-worn battery-driven device Actiheart (Cambridge Neurotechnology Ltd, England), allowing free movement and long recordings.

C. Preprocessing and Feature Extraction

The collected mobility data from the Actiwatch device was processed and organized into individual files for each participant. Each file contained raw activity measurements recorded over a span of 6 days for the ADHD group and 7 days for the non-ADHD group. However, for consistency, we considered 6 days of mobility data for all participants. The combined mobility data files were carefully examined for missing values or anomalies, and appropriate techniques were employed to handle them effectively. To gain insights into the participants' motor activity patterns, two types of features were extracted: hour-wise features and day-wise features. These features provide valuable information about the characteristics and variations in motor activity over different time intervals. The hour-wise features reveal intra-daily variations, capturing details such as activity levels during specific hours, peak activity hours, and temporal variability. On the other hand, the day-wise features capture longer-term patterns, trends, and summary statistics of activity levels over the 6-day duration.

In addition to features extracted from mobility data, we have also extracted time domain features from HRV data such as mean and median. Table 3 summarizes the list of features extracted, including mobility and HRV features. This comprehensive set of features offers a detailed representation of the participants' activity patterns at various temporal scales. They serve as a foundation for analyzing mobility patterns, exploring relationships with demographic parameters, and investigating associations with diagnostic assessment outcomes. The extracted hour-wise and day-wise features provide a robust and comprehensive representation of the participants' activity patterns, enabling in-depth analyses of their mobility behaviors.

D. Network Generation

In our study, the generation of the correlation network involves a series of steps to establish relationships and interconnections between the participants based on their mobility data. To achieve this, we employ a population analysis approach, without relying on supervised label information, to identify groups of persons with similar motor activity profiles. First, we denote all 55 participants as nodes in the correlation graph, representing each subject within the network. The objective is to connect two participants with an edge if they possess a similar motor activity profile. The similarity between subjects is quantified using the 48 hour-wise features extracted during the feature extraction phase. Specifically, we compute the pairwise Pearson correlation coefficient (ρ) between each pair of participants, which measures the linear dependence between them [12] [13].

To construct the correlation graph, we utilize the computed ρ values to create a correlation matrix that captures the degree of similarity between all pairs of users. The ρ value ranges from -1 to +1, where -1 indicates a negative correlation and +1 represents a strong positive correlation [14] [15]. Subsequently, we introduce a predefined threshold, denoted as 'k,' to identify strongly correlated pairs, thus forming a significance matrix. The significance matrix takes a value of 1 if the (ρ value between a pair of users is greater than or equal to the threshold 'k'; otherwise, it takes a value of 0. The significance matrix serves as the adjacency matrix for the

 ISBN: 978-1-68558-112-1

TABLE I: FEATURE LIST

Feature set	Feature	Count	Description
Hour-wise Features	h-m-0 to h-m-23	24	Mean (average) of activity measured by every hour (0 – 23 hours)
	h-sd-0 to h-sd-23	24	Standard deviation (SD) of activity measured by every hour (0 – 23 hours)
Day-wise Features	d-m-1 to d-m-6	6	Mean (average) of activity measured by a day (1 – 6 days)
	d-sd-1 to d-sd-6	6	Standard deviation (SD) of activity measured by a day (1 – 6 days)
HRV features	hrv1 to hrv5	5	Mean, Median, SD, Min, and Max of HRV
	Total	65	

correlation graph. In this graph, two participants (Pi, Pj) are connected by an edge if their correlation coefficient exceeds or equals 'k.' This process creates a network of strongly interconnected nodes, representing the relationships between participants based on their motor activity.

To uncover the hidden clusters or communities within the correlation graph, we apply the Louvain Clustering technique, a popular unsupervised clustering algorithm suitable for extracting clusters in biological networks [16]. The Louvain algorithm employs a random walk property to categorize nodes into different communities. These clusters represent groups of persons with similar motor activity profiles. The clustering process ensures that each community demonstrates homogeneity, with participants within the same cluster showing similar motor activity characteristics. Additionally, the clusters exhibit separation, meaning that persons in different communities exhibit distinct mobility profiles [17]. In summary, the generation of the correlation network involves computing the pair-wise Pearson correlation coefficients between subjects' motor activity data, setting a threshold 'k' to identify significant correlations, and constructing the correlation graph based on the adjacency matrix. The application of the Louvain algorithm reveals the hidden clusters representing groups of individuals with similar motor activity profiles. The resulting network provides valuable insights into the relationships and homogeneous patterns among subjects based on their mobility data.

E. Detection and Analysis of Homogeneous Groups

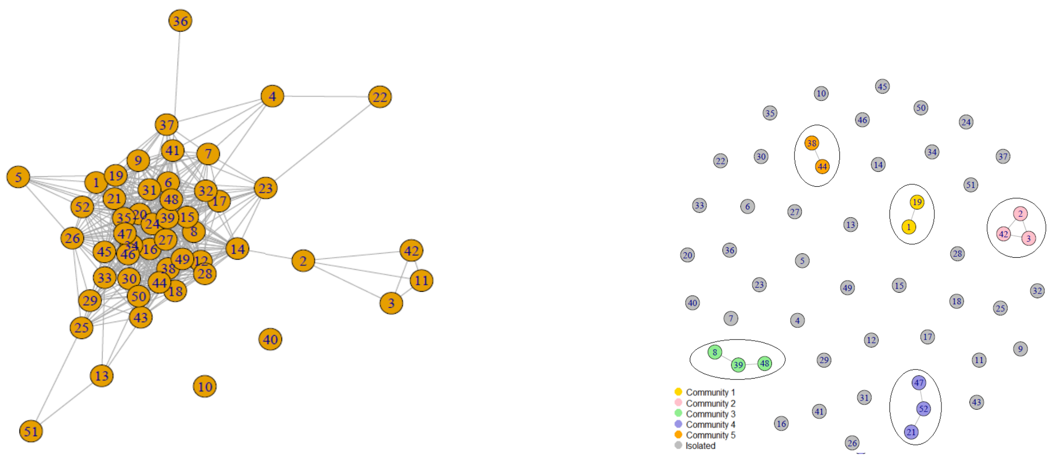
Following the generation of the correlation network, the next step is to identify homogeneous groups within the network and analyze their mobility and HRV data. To achieve this, we perform a two-step process: first, determining a correlation threshold that results in the complete disintegration of the network, and second, tweaking the network threshold to create sub-networks with cohesive connections. We initially explore the correlation threshold at which the network becomes entirely disconnected. This threshold serves as an upper limit for the subsequent network threshold adjustments. By carefully selecting a value below this disintegration threshold, we create a sub-network in which participants retain cohesive connections, thus ensuring the identification of meaningful homogeneous groups. Next, we employ the Louvain clustering algorithm on the sub-network to detect homogeneous groups of interconnected nodes. The Louvain algorithm is an efficient and widely used community detection method that optimizes the modularity of the network, effectively partitioning nodes into non-overlapping communities based on their interconnections.

Once the homogeneous groups are identified, we proceed to analyze the mobility and HRV data for each group separately. This analysis involves examining the mobility characteristics, such as average steps taken, distance traveled, and activity patterns, as well as HRV metrics, including time domain, frequency domain, and non-linear features, extracted during the feature extraction phase. Comparing the mobility and HRV data within each homogeneous group allows us to gain insights into their shared characteristics and behavior patterns. By understanding how these groups differ in their motor activity and physiological responses, we can uncover unique features that may distinguish them from other groups and potentially indicate underlying health conditions or specific behavioral traits. Furthermore, we investigate whether there are significant correlations or associations between mobility and HRV data within each homogeneous group. Such correlations could offer valuable information on how motor activity influences autonomic nervous system activity and vice versa, contributing to a deeper understanding of the interplay between physiological and behavioral aspects.

III. RESULTS

After preprocessing the initial dataset of 85 subjects, 33 subjects were excluded due to insufficient clinical and mobility data, leaving a final dataset of 52 subjects for analysis. To construct a meaningful network, we selected a correlation threshold of 0.5, resulting in a network with a reasonable level of connectivity among the subjects. However, when the threshold was increased to 0.82, the network disintegrated completely. At the initial threshold of 0.5, all subjects were connected in the network, forming the first network as shown in Figure 2(a). Subsequently, we chose a threshold of 0.8, just before the network disintegrated, to create a second network as shown in Figure 2(b). The hypothesis was that the groups identified at this resolution would represent strongly connected groups compared to other nodes. These groups demonstrated resilience even at the maximum threshold resolution, indicating significantly stronger connectivity compared to other nodes. In contrast, other nodes appeared to be either loosely or moderately connected. The groups that proved to be strongly connected at the 0.8 network resolution were identified as homogeneous groups. Their robust interconnections suggested shared behavior and characteristics within these groups, making them distinguishable from other nodes. These homogeneous groups are of particular interest as they potentially

 ISBN: 978-1-68558-112-1



(a). Correlation Network

(b). Homogeneous Groups

Fig. 2. Identification of homogenous groups

represent distinct subpopulations with unique mobility patterns and HRV responses. Analyzing the mobility patterns and HRV data within each homogeneous group revealed rich and vital information. These insights offer valuable opportunities for designing personalized interventions and tailoring healthcare strategies based on the distinct characteristics exhibited by each homogeneous group. Understanding the interplay between mobility and HRV data within these groups can lead to targeted and effective approaches to improving health outcomes and well-being for individuals within each group.

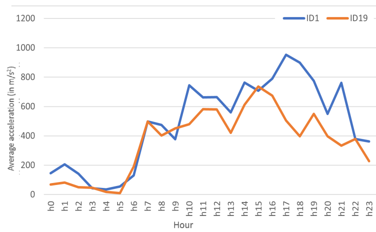
The results of our analysis shed light on the mobility patterns observed within the five homogeneous communities or clusters, represented in Figure 3(a) to Figure 3(e), at an hourly granularity level. Additionally, the HRV patterns are illustrated in Figure 3(f). These homogeneous communities are formed based on various factors, including hourly mobility patterns, sleep and wakeup patterns, average mobility acceleration (in m/s2), and overall day-wise mobility acceleration (in m/s2). Analyzing the hour-wise mobility patterns within each homogeneous community reveals distinct sleep, wakeup, and daily routine characteristics. Each community exhibits unique patterns in their sleep schedules. For instance, subjects in community 1 maintain a consistent sleep routine between midnight and 5 am, while community 3 displays a consistent sleep pattern from midnight to 6 am. On the other hand, community 5 follows a sleep routine from 2 am to 7 am, and community 4's sleep routine is observed between 1 am and 6 am. These variations in sleep patterns contribute to the formation of distinct homogeneous groups within the network.

Furthermore, the wakeup and day start routines also exhibit noticeable differences among the communities. As individuals wake up, there is a marked increase in their mobility levels. In community 1, there is a significant rise in mobility between 5 am and 7 am, while community 4 shows an increase between 7 am and 9 am. Community 3 exhibits an upsurge in mobility between 6 am and 8 am, and community 5 demonstrates a slightly delayed wakeup schedule between 9 am and 12 pm. These distinct patterns in wakeup times further differentiate the homogeneous communities, highlighting the importance of considering individual sleep and morning routine behaviors. During daytime hours, most of the communities engage in relatively active mobility activities, with little variation in their average mobility acceleration, except for community 2. Community 1 stands out with the highest average acceleration compared to the other groups, indicating a more intense level of mobility. On the other hand, communities 3 and 5 exhibit slightly lower average mobility acceleration, suggesting a comparatively less intense but still active level of mobility. Interestingly, community 2 demonstrates the lowest mobility acceleration among the groups, accompanied by abnormal mobility patterns. This finding indicates the presence of unique and potentially abnormal mobility characteristics within community 2, which warrants further attention and validation.

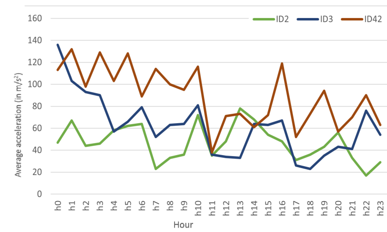
In terms of HRV patterns, the total average heart rate shows no significant difference among the subjects, but there is noticeable variation in the maximum heart rate measured over the total data collection period. Specifically, subjects ID47 and ID52 demonstrate significantly lower HRV compared to other individuals within the homogeneous groups. However, it is observed that mobility patterns exert a stronger influence in the formation of communities compared to HRV patterns.

In conclusion, our analysis of the mobility patterns within the five homogeneous communities reveals distinctive sleep, wakeup, and daily routine characteristics among the subjects. The variations in these patterns contribute to the formation of homogeneous groups, highlighting the relevance of individual sleep behaviors and daytime mobility activities. Additionally, the identification of abnormal mobility patterns within community 2 underscores the importance of considering such anomalous behaviors for further investigation and validation. These findings provide valuable insights into the mobility behaviors and daily routines within each community, facilitating the design of personalized interventions and healthcare strategies tailored to the specific characteristics exhibited by

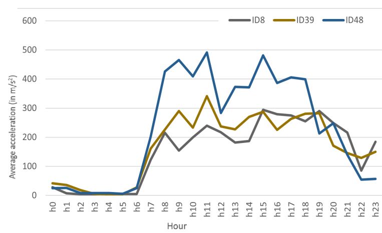
 ISBN: 978-1-68558-112-1



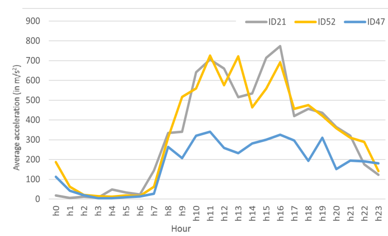
(a). Community 1



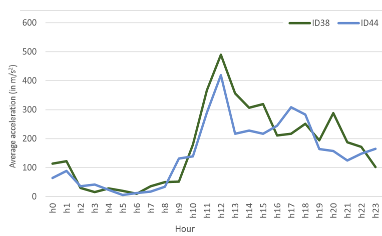
(b). Community 2



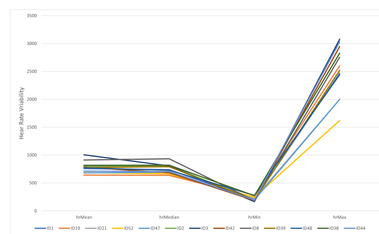
(c). Community 3



(d). Community 4



(e). Community 5



(f). HRV patterns of all communities

Fig. 3. Mobility and HRV data Analysis

these homogeneous groups.

IV. Discussion

Identifying homogeneous groups within biological networks is a critical task with significant implications for developing personalized therapies and individualized medicines. In many cases, universal medications and group therapies do not provide optimal results for all individuals. Therefore, the development of personalized and targeted interventions is essential. The fundamental objective of identifying homogeneous groups in biological networks is to enable the customization of treatments and interventions based on the unique characteristics of each group. Each homogeneous group identified in our analysis exhibits two important characteristics: homogeneity within the group and distinctiveness from other homogeneous groups. This means that individuals within a homogeneous group share similar patterns and behaviors, while also differing significantly from individuals in other groups. Understanding these distinctions allows for the design of interventions that specifically cater to the needs and characteristics of each group.

A unique aspect of our methodology is that it does not rely on the inclusion of class labels to detect homogeneous groups. This data-driven approach eliminates the need for manual human annotation of class labels, saving considerable time and effort during the data collection process. By solely relying on the data itself, our methodology efficiently identifies homogeneous groups based on the inherent patterns and relationships within the data. This data-driven approach enhances the objectivity and accuracy of the analysis. Additionally, our methodology has the potential to identify subject groups with abnormal or peculiar mobility and HRV patterns. An example of such an abnormal group is community 2, where individuals exhibit minimal and abnormal mobility patterns throughout the day. The identification of such abnormal groups is crucial as it can highlight individuals who may require specific attention, interventions, or further investigation due to their unique characteristics. By recognizing these abnormalities, healthcare professionals can tailor interventions to address the specific needs of these individuals, potentially improving their health outcomes.

V. Limitations and Future Work

There are several limitations to be acknowledged in our study. First, the sample size of our dataset was relatively small, consisting of only 52 subjects. This limited sample

 ISBN: 978-1-68558-112-1

size may impact the generalizability of our findings and limit the scope of interpretations. A larger and more diverse dataset would provide a more comprehensive understanding of homogeneous groups and their mobility patterns within the biomedical network. Another limitation arises from the selection of the correlation threshold. The choice of threshold can introduce bias into the analysis, as it determines the strength of connections within the network. Different thresholds may yield different results, and selecting the optimal threshold is subjective. Future studies should explore robust methods to determine an appropriate threshold or consider alternative approaches that are less dependent on this parameter.

To overcome the limitations mentioned above, future work should focus on expanding the dataset to include a larger number of subjects. A more extensive and diverse sample would enhance the representativeness and generalizability of the findings. Additionally, incorporating longitudinal data would provide insights into the temporal dynamics of homogeneous groups and their mobility patterns over time. Furthermore, it would be valuable to integrate clinical parameters and additional biomarkers into the analysis. Including clinical data, such as medical history, comorbidities, and medication usage, can provide a more comprehensive understanding of the factors influencing the formation of homogeneous groups. The incorporation of other biomarkers, such as genetic information or biochemical markers, could offer deeper insights into the underlying physiological and molecular mechanisms related to mobility patterns. Moreover, future research could explore advanced network analysis techniques to gain a more nuanced understanding of the relationships within the biomedical network. For example, community detection algorithms beyond the Louvain clustering algorithm could be employed to uncover finer-grained subgroups within homogeneous groups. Additionally, considering the dynamic nature of the data, time-evolving network analysis methods could be employed to capture changes and transitions in homogeneous group formations over time.

VI. Conclusion

In this study, we developed a graph model to take advantage of mobility data and medical data to identify groups with potential similar properties. Our approach focused on the identification of homogeneous groups within a biomedical network using multi-modal data. By constructing a correlation network, we identified homogeneous groups that exhibited strong interconnections based on their mobility patterns. The analysis revealed distinct patterns and characteristics within these homogeneous groups, providing valuable insights for personalized therapies and individualized interventions. Furthermore, the analysis of hour-wise mobility patterns highlighted unique sleep, wake-up, and daily routine characteristics within each group. Additionally, the examination of average mobility acceleration revealed variations among the homogeneous groups, with some demonstrating higher or lower levels of mobility. The data-driven approach used in our methodology eliminated the need for manual class label annotations, enhancing the efficiency and objectivity of the analysis. Moreover, the identification of subjects that didn't belong to any of highly similar groups emphasized the potential of our methodology to detect individuals with different mobility characteristics, warranting further investigation and personalized attention. For future directions, we plan to expand our analysis by incorporating additional clinical parameters and other medical information for each individual in the study.

References

[1] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[2] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics reports*, vol. 659, pp. 1–44, 2016.

[3] J. Yang, J. McAuley, and J. Leskovec, "Detecting cohesive and 2-mode communities indirected and undirected networks," in *Proceedings of the 7th ACM international conference on Web search and data mining*, 2014, pp. 323–332.

[4] R. Rotabi, K. Kamath, J. Kleinberg, and A. Sharma, "Detecting strong ties using network motifs," in *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017, pp. 983–992.

[5] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[6] I. D. Yalom and K. Rand, "Compatibility and cohesiveness in therapy groups," *Archives of General Psychiatry*, vol. 15, no. 3, pp. 267–275, 1966.

[7] E. A. Finkelstein et al., "Effectiveness of activity trackers with and without incentives to increase physical activity (trippa): a randomised controlled trial," *The lancet Diabetes & endocrinology*, vol. 4, no. 12, pp. 983–995, 2016.

[8] Y. Yuan, *Characterizing human mobility from mobile phone usage*. University of California, Santa Barbara, 2013.

[9] B. M. Saykrs, "Analysis of heart rate variability," *Ergonomics*, vol. 16, no. 1, pp. 17–32, 1973.

[10] A. Sallaberry, Y.-c. Fu, H.-C. Ho, and K.-L. Ma, "Contact trees: Network visualization beyond nodes and edges," *PloS one*, vol. 11, no. 1, p. e0146368, 2016.

[11] S. A. Hicks et al., "Hyperaktiv: An activity dataset from patients with attention-deficit/hyperactivity disorder (adhd)," in *Proceedings of the 12th ACM Multimedia Systems Conference*, 2021, pp. 314–319.

[12] I. Cohen et al., "Pearson correlation coefficient," *Noise reduction in speech processing*, pp. 1–4, 2009.

[13] J. Benesty, J. Chen, and Y. Huang, "On the importance of the pearson correlation coefficient in noise reduction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 757–765, 2008.

[14] R. Thelagathoti and H. Ali, "A population analysis approach using mobility data and correlation networks for depression episodes detection," *Ann Depress Anxiety*, vol. 9, no. 1, p. 1112, 2022.

[15] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: appropriate use and interpretation," *Anesthesia & analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018.

[16] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Generalized louvain method for community detection in large networks," in *2011 11th international conference on intelligent systems design and applications*. IEEE, 2011, pp. 88–93.

[17] R. K. Thelagathoti and H. H. Ali, "The comparison of various correlation network models in studying mobility data for the analysis of depression episodes." in *BIOSIGNALS*, 2022, pp. 200–207.