UNIVERSITY OF CALIFORNIA,
IRVINE

Holistic Health Monitoring and Personalized Intervention for Well-being Promotion

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Electrical Engineering and Computer Science

by

Milad Asgari Mehrabadi

Dissertation Committee:
Professor Nikil Dutt, Chair
Professor Amir M. Rahmani
Professor Yuqing Guo

2022

DEDICATION

To Mom, Dad and Mina.

TABLE OF CONTENTS

Page

LIST OF FIGURES **v**

LIST OF TABLES **vii**

ACKNOWLEDGMENTS **ix**

VITA **xi**

ABSTRACT OF THE DISSERTATION **xv**

1 Introduction **1**

2 Sleep Monitoring **5**
2.1 Background 6
2.2 Instrument Validation Study 8
2.2.1 Background 8
2.2.2 Methods 9
2.2.3 Results 16
2.2.4 Discussion 25
2.3 Pregnant women's sleep patterns before and during the COVID-19 pandemic 30
2.3.1 Background 30
2.3.2 Data collection 30
2.3.3 Data preprocessing 31
2.3.4 Statistical analysis 31
2.3.5 Results 32
2.4 Conclusions 32

3 Mental health **35**
3.1 The Effect of COVID-19 and Lockdown on Mental Health in California . . . 36
3.1.1 Background 36
3.1.2 Method 37
3.1.3 Results 38
3.1.4 Discussion 42

4 Cardiovascular Biomarkers Monitoring **44**
4.1 Blood Pressure Waveform Reconstruction from PPG using Cycle Generative Adversarial Networks . . . 46
4.1.1 Background . . . 46
4.1.2 Related Work . . . 48
4.1.3 Material and Methods . . . 48
4.1.4 Dataset . . . 48
4.1.5 Experimental Results . . . 55
4.2 Detection of COVID-19 Using Heart Rate and Blood Pressure: Lessons Learned from Patients with ARDS . . . 60
4.2.1 Background . . . 60
4.2.2 Methods . . . 62
4.2.3 Results . . . 65
4.2.4 Discussion . . . 68
4.3 Conclusion . . . 70

5 Physical Activity Recommendation **71**
5.1 Background . . . 71
5.2 Methods . . . 73
5.2.1 Participant recruitment . . . 73
5.2.2 Study Procedures . . . 73
5.2.3 Exercise organization . . . 74
5.2.4 Recommendation System Design . . . 75
5.2.5 Statistical Analysis . . . 83
5.3 Results . . . 84
5.3.1 Improvement in performed exercise duration trends . . . 84
5.3.2 Minutes in light- and moderate-intensity . . . 86
5.3.3 Aggregated weekly performance . . . 89
5.3.4 Recommended exercises . . . 91
5.3.5 Participants feedback . . . 93
5.4 Discussion . . . 93
5.4.1 Principal Results . . . 93
5.4.2 Limitations . . . 93
5.4.3 Comparison with Prior Work . . . 94
5.5 Conclusion . . . 95

6 Summary and Conclusions **96**

Bibliography **98**

Appendix A Mental health disorders **110**

Appendix B Abbreviations **119**

LIST OF FIGURES

Page

1.1 Top 100 low-dimension narratives extracted from Facebook posts (a) and sleep and anxiety sub-graph (b). . . . 3
1.2 Topics related to anxiety and stress. . . . 3

2.1 Sleep extraction multi-modal cross checking. . . . 14
2.2 Watch data processing pipeline. . . . 14
2.3 Bland-Altman plots for total sleep time, sleep efficiency, and wake after sleep onset gathered by the Oura ring and the actigraphy device. Subjects' actigraphy minus Oura ring discrepancies on sleep parameters (y-axis) are plotted compared with actigraphy (x-axis). Biases, upper, and lower agreement limits are marked. In addition, the satisfactory ranges are plotted as the dashed lines. SE: sleep efficiency; TST: total sleep time; WASO: wake after sleep onset. 17
2.4 Bland-Altman plots for total sleep time, sleep efficiency, and wake after sleep onset gathered by the Samsung watch and the actigraphy device. Subjects' actigraphy minus Samsung watch discrepancies on sleep parameters (y-axis) are plotted compared with actigraphy (x-axis). Biases, upper, and lower agreement limits are marked. In addition, the satisfactory ranges are plotted as the dashed lines. SE: sleep efficiency; TST: total sleep time; WASO: wake after sleep onset. . . . 20
2.5 Bland-Altman plot for sleep onset latency estimated by the Oura ring. SOL: sleep onset latency. . . . 24
2.6 Bland-Altman plot for sleep onset latency estimated by the Samsung watch. SOL: sleep onset latency. . . . 25
2.7 Trends in sleep. The daily mean with 95% confidence intervals of daily total sleep time (TST) and wake after sleep onset (WASO) (n = 22–32) during the eight-week data collection period (56 days). These figures are obtained from [105]. . . . 32
2.8 Subjective evaluations of sleep. Participants' (n = 23) subjectively assessed quality of sleep before and during the pandemic-related restrictions in a scale from 0 to 100 the higher value indicating a better quality of sleep. These figures are obtained from [105]. . . . 33

3.1 Pearson correlation of mental health disorders in 2020. . . . 40
3.2 Distribution of total mental health disorder reports per age group (3.2a) and per each disorder for 26-35 years old population (3.2b). . . . 41

4.1 Correlation between BP and PPG signal. 47
4.2 BP estimation pipeline. 50
4.3 CycleGAN overview from [148]. 53
4.4 Generator network of CycleGAN. 54
4.5 Discriminator network of CycleGAN. 54
4.6 PPG to ABP overview using CycleGAN. 55
4.7 Transformed PPG signal (a) and ABP signal (b) to 256 by 256 images for a sample user and their corresponding reconstructed images using CycleGAN (c) and (d). 56
4.8 SBP (a) and DBP (b) prediction error. 57
4.9 SBP (a) and DBP (b) Bland-Altman plots. 57
4.10 SBP (a) and DBP (b) distributions and their corresponding mean values. . . 58
4.11 SBP (a) and DBP (b) per-subject Bland-Altman plots. 59
4.12 True and reconstructed (fake) ABP signal for 2 subjects. 59
4.13 The distribution comparison of average HR between the positive and negative test results for each age group. The dashed lines represent the quartiles. . . . 66
4.14 The performance of the model in terms of accuracy (a), AUC (b) and f1 score (c) using test data with respect to the number of included days. 67
4.15 The 2-dimensional representation of test data using t-SNE considering the entire test data. 68

5.1 Exercise duration setup. 76
5.2 The overview of the system. 77
5.3 An overview of the developed applications for users' cellphones and Samsung smartwatches. 78
5.4 An example page of the ZotCare dashboard. 79
5.5 Reward function with respect to the HRs during the exercise. 83
5.6 Average weekly duration of exercise for each group. The bars represent standard error. 85
5.7 Performed vs. recommended duration of exercise per each subject. The line represents the fitted HLM model to each subject. 86
5.8 Normal Q-Q plot for normality of the residuals for exercise performed duration. 87
5.9 Trends in minutes in light and moderate-intensity exercise over time. 88
5.10 Dot plot visualization of between-subject variations compared to the mean trend. 88
5.11 Normal Q-Q plot for normality of the residuals for minutes in light/moderate-intensity exercise. 89
5.12 Aggregated number of weekly exercise duration performed by the subjects. . 90
5.13 Trends in weekly exercise duration. 91
5.14 Normal Q-Q plot for normality of the residuals for the weekly exercise duration. 92
5.15 Average cumulative rewards for subjects in Level 1 and Level 2. 92

LIST OF TABLES

Page

2.1 Participants' background information. 10
2.2 Mean, SD, 95% CI, and paired t test results for the actigraphy and the Oura ring sleep parameters in a sample of 45 healthy adults. 17
2.3 Bias and agreement limits based on Bland-Altman plots for the actigraphy and the Oura ring. 18
2.4 Mean, SD, 95% CI, and paired t test results for the actigraphy and the Samsung watch sleep parameters in a sample of 35 healthy adults. 19
2.5 Bias and agreement limits based on Bland-Altman plots for the actigraphy and the Samsung watch. 19
2.6 Mean, SD, and average mean differences (the actigraphy minus the Oura ring) for 23 women (141 samples) and 22 men (125 samples). 21
2.7 Mean, SD, and average mean differences (the actigraphy minus the Samsung watch) for 16 women (65 samples) and 19 men (69 samples). 22
2.8 Pearson correlation between the actigraphy, ring, and smartwatch with the corresponding P values for the considered sleep attributes. TST: total sleep time, SE: sleep efficiency and WASO: wake after sleep onset. 22
2.9 Mean, SD, 95% CI, and paired t test results for the actigraphy versus Oura ring estimates of sleep onset latency. 23
2.10 Mean, SD, 95% CI, and paired t test results for the actigraphy versus Samsung watch estimates of sleep onset latency. 23
2.11 Bias and agreement limits based on Bland-Altman plot of the sleep onset latency for the actigraphy and the Oura ring. 23
2.12 Bias and agreement limits based on Bland-Altman plot of sleep onset latency for the actigraphy and Samsung watch. 24

3.1 Reported values of each disorder per age group in 2019 and 2020. 39

4.1 BHS standard ranges. 51
4.2 Average performance of our proposed model. 55
4.3 Our results in regards to BHS standard ranges. 57
4.4 Comparison of our CycleGAN-based model performance with prior works. . 58
4.5 Age distribution of subject with different COVID-19 test results. 62
4.6 The architecture of the proposed neural network. 64
4.7 Point Biserial correlation of statistical features and COVID-19 test results (* shows significant correlation). 65

4.8 95% confidence intervals of HR, DBP and SBP for each COVID-19 test result group. 65

5.1 Summary of fixed effects of the HLM fitted to the exercise performed duration. 85
5.2 Summary of random effects of the HLM fitted to the exercise performed duration. 85
5.3 Summary of fixed effects of the HLM fitted to the minutes in light/moderate-intensity exercise. 87
5.4 Summary of random effects of the HLM fitted to the minutes in light/moderate-intensity exercise. 87
5.5 Summary of fixed effects of the HLM fitted to the weekly exercise duration. . 90
5.6 Summary of random effects of the HLM fitted to the weekly exercise duration. 90

ACKNOWLEDGMENTS

I would like to thank my advisors Professor Nikil Dutt and Professor Amir Rahmani, for their support and guidance on my work during the course of my Ph.D. Their continuous support has allowed me to explore the research topics freely, and their vision and motivation have inspired me to push my limit and become a better version of myself every day.

Furthermore, I would like to thank my thesis committee member Professor Yuqing Guo for agreeing to serve on my candidacy exam and defense committee, for her valuable time, dedication, contributions, and helpful rigorous feedback. I would also like to thank many other researchers and anonymous reviewers for their very helpful discussion and feedback, which have helped improve my research.

I am thankful to Professor Salma Elmalaki and Professor Hung Cao, and I would like to thank Professor Athina Markopoulou for her support and help during my Ph.D.

I would like to thank Dr. Hamed Soroush and Dr. Shantanu Rane for giving me an internship opportunity at the PARC Research center and for their invaluable advice, guidance, and patience during my internship.

Most importantly, I cannot begin to express my sincere thanks to my family. My Mom, Dad, and Mina, without their unconditional love and support of whom, I would have never been where I am today. I am forever grateful to them for providing me with the opportunity to pursue my education and always encouraging me to seek a better life. I am grateful for my wonderful sister, Dr. Mina Asgari, who was always there when I had medical-related research questions, and my grandparents, who were always the source of support.

I am truly blessed to have Sara as an endless source of joy and love through the darkest moments. She has been a source of encouragement and motivation, and I could not have finished this work without her constant support, confidence, understanding, patience, and faith in me. I would also like to thank her wonderful parents for their support during the last few past years.

I am also grateful to my friends and fellow lab-mates, who were always there for me throughout my journey.

This dissertation was supported by the Academy of Finland through the SLIM Project under grant numbers 316810 and 316811 and in part by the U.S. National Science Foundation (NSF) through the UNITE Project under grant number SCC CNS-1831918.

I would like to acknowledge IEEE for giving me permission to include Chapter 4 of this dissertation parts of which was originally published in IEEE Xplore.

Portions of Chapter 4 were previously published as “Novel Blood Pressure Waveform Reconstruction from Photoplethysmography using Cycle Generative Adversarial Networks, EMBC, 2022, Mehrabadi, M.A., Aqajari, S.A.H., Zargari, A.H.A., Dutt, N. and Rahmani, A.M.”.

Permissions to reuse the text were granted by IEEE and co-authors.

Portions of Chapter 4 were previously published as "Detection of covid-19 using heart rate and blood pressure: Lessons learned from patients with ards, EMBC, 2021, Mehrabadi, M.A., Aqajari, S.A.H., Azimi, I., Downs, C.A., Dutt, N. and Rahmani, A.M.". Permissions to reuse the text were granted by IEEE. The co-author listed in this publication directed and supervised research which forms the basis for the dissertation.

VITA

Milad Asgari Mehrabadi

EDUCATION

Doctor of Philosophy in Computer Science and Engineering **2022**
University of California Irvine *Irvine, CA*

Master of Science in Computer Science and Engineering **2018**
University of California Irvine *Irvine, CA*

Bachelor of Science in Computational Engineering **2015**
Sharif University of Technology *Tehran, Iran*

WORK EXPERIENCE

Data Science Intern **2021**
Capital One *San Francisco, California*

Research Intern **2019**
Xerox PARC *Palo Alto, California*

Data Science Intern **2017**
Bitglass *Campbell, California*

RESEARCH EXPERIENCE

Graduate Research Assistant **2016–2022**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **2017–2020**
University of California, Irvine *Irvine, California*

REFEREED JOURNAL PUBLICATIONS

A Descriptive Comparative Pilot Study: Association between use of a self-monitoring device and sleep and stress outcomes in pregnancy **2022**
J. Auxier, MA. Mehrabadi, AM. Rahmani, A. Axelin, Computers, Informatics, Nursing Journal (CIN)

The Causality Inference of Public Interest in Restaurants and Bars on COVID-19 Daily Cases in the US: A Google Trends Analysis **2021**
MA. Mehrabadi, N. Dutt, AM. Rahmani, JMIR Public Health and Surveillance

Pregnant women's daily patterns of well-being before and during the COVID-19 pandemic in Finland: longitudinal monitoring through smartwatch technology **2021**
H. Niela-Vilen, J. Auxier, E. Ekholm, F. Sarhaddi, MA. Mehrabadi, A. Mahmoudzadeh, I. Azimi, P. Liljeberg, AM. Rahmani, A. Axelin, PLOS ONE

A Technology-Based Pregnancy Health and Wellness Intervention (Two Happy Hearts): Case Study **2021**
T. Jimah, H. Borg, P. Kehoe, P. Pimentel, A. Turner, S. Labbaf, M. Asgari Mehrabadi, A M Rahmani, N. Dutt, Y. Guo, JMIR formative research

Sleep Quality Prediction During the Menstrual Cycle based on Daily Sleep Diary Reports **2021**
N. Sattari, M. Asgari Mehrabadi, A. Aqajari, J. Zhang, K. Simon, E. Alzueta, T. Duali, M. de Zambotti, A M. Rahmani, F. Baker, S. Mednick, SLEEP

Sleep Tracking of a Commercially Available Smart Ring and Smartwatch Against Medical-Grade Actigraphy in Everyday Settings: Instrument Validation Study **2020**
MA Mehrabadi, I. Azimi, F. Sarhaddi, A. Axelin, H. Niela-Vilén, S. Myllyntausta, S. Stenholm, N. Dutt, P. Liljeberg, AM. Rahmani, JMIR Mhealth Uhealth

Link prediction in multiplex online social networks **2017**
M. Jalili, Y. Orouskhani, M. Asgari, N. Alipourfard, and M. Perc, Royal Society Open Science

REFEREED CONFERENCE PUBLICATIONS

Novel Blood Pressure Waveform Reconstruction from Photoplethysmography using Cycle Generative Adversarial Networks **2022**

MA. Mehrabadi, SAH. Aqajari, AHA. Zargari, N. Dutt, AM. Rahmani, IEEE Engineering in Medicine & Biology Society (EMBC)

Detection of covid-19 using heart rate and blood pressure: Lessons learned from patients with ards **2021**

MA. Mehrabadi, SAH. Aqajari, I. Azimi, CA. Downs, N. Dutt, and AM. Rahmani, IEEE Engineering in Medicine & Biology Society (EMBC)

Two Happy Hearts: A Technology-Based Personalized Health Self-Management Model for Pregnant Women **2021**

T. Jimah, H. Borg, P. Kehoe, S. Labbaf, MA. Mehrabadi, AM. Rahmani, N. Dutt and Y. Guo, American Nursing Informatics Association

Towards Smart and Efficient Health Monitoring Using Edge-Enabled Situational-Awareness **2021**

S. Shahhosseini, A. Kanduri, MA. Mehrabadi, E. Kasaeyan Naeini, D.o Seo, S. Lim, AM. Rahmani, N. Dutt, IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)

Sleep Quality Prediction During the Menstrual Cycle based on Daily Sleep Diary Reports **2021**

N. Sattari, MA. Mehrabadi, A. Aqajari, J. Zhang, K. Simon, E. Alzueta, T. Duali, M. de Zambotti, AM. Rahmani, F. Baker, and S. Mednick, SLEEP meeting

pyEDA: An Open-Source Python Toolkit for Preprocessing and Feature Extraction of Electrodermal Activity **2021**

SAH. Aqajari, E. Kasaeyan Naeini, MA. Mehrabadi, S. Labbaf, N. Dutt and AM. Rahmani, International Conference on Ambient Systems, Networks and Technologies (ANT)

A Digital Health Approach to Promote Emotional Well-Being in Pregnant Women: The Two Happy Hearts Case Study **2021**

T. Jimah, H. Borg, MA. Mehrabadi, S. Labbaf and Y. Guo, Association of Women's Health, Obstetric and Neonatal Nurses

SCIBORG: Secure Configurations for the IoT Based on Optimization and Reasoning on Graphs (Best Paper Award) **2020**

H. Soroush, M. Albanese, M. A. Mehrabadi, I. Iganibo, M. Mosko, J. Gao, D. Fritz, S. Rane and E. Bier, IEEE Conference on Networking and Security (CNS)

Dynamic iFogSim: A Framework for Full-Stack Simulation of Dynamic Resource Management in IoT Systems **2020**

D. Seo, S. Shahhosseini, MA. Mehrabadi, B. Donyanavard, S. Lim, N. Dutt, and A. Rahmani, IEEE International Conference on Omni-layer Intelligent Systems (COINS)

SOFTWARE

Python, C/C++, R, Java, TensorFlow, Docker, CUDA, Git, MongoDB

ABSTRACT OF THE DISSERTATION

Holistic Health Monitoring and Personalized Intervention for Well-being Promotion

By

Milad Asgari Mehrabadi

Doctor of Philosophy in Electrical Engineering and Computer Science

University of California, Irvine, 2022

Professor Nikil Dutt, Chair

Well-being is a crucial factor in human lives and society insofar as it is an indicator of satisfaction. Within the pillars of well-being, we favor sleep, physical activity, and mental health because these can represent body health. Furthermore, the entire world has been affected widely by a global virus pandemic which could significantly impact societies with vulnerable factors of well-being. Hence, we have investigated the effect of COVID-19 as one of the representatives of threats to social well-being. Parties interested in illness prevention and health promotion may find it helpful to measure, monitor, and promote well-being. Through the advancement of the Internet of the Things (IoT), it is now possible to monitor health outcomes and biomarkers in everyday free-living conditions without needing to proceed to labs or clinical settings. Taking the above into consideration, we can organize the main contribution of this dissertation into two components.

First, we examine the trends and patterns of sleep and mental health disorders at a population level. To do so, we evaluate the sleep parameters of the Oura ring and the Samsung Gear Sport watch compared with a medically approved actigraphy device in a midterm everyday setting, where users engage in their daily routines. We used home-based sleep monitoring to examine the sleep characteristics of 45 healthy people (23 women and 22 men) for 7 days. Then we investigate the sleep trends of 38 pregnant women during the COVID-19 lock-

down in Finland. The subjects used the Samsung Gear Sport smartwatch, and their sleep data was recorded. Subjective sleep reports were obtained using a smartphone app designed specifically for this study. Later, we analyze different mental health disorder reports before and during the pandemic and discuss the most vulnerable population. The benefit of such investigations is that capturing real-time information and public attitudes would facilitate policymakers to monitor public health and social wellness.

In the second part, we focus on individual-level analyses. We use Machine Learning and Deep Learning techniques to monitor, reconstruct, evaluate, and forecast various tasks utilizing individuals' data and biomarkers. We begin by reconstructing the blood pressure signal. Continuous blood pressure (BP) monitoring can help individuals manage their chronic diseases such as hypertension, requiring non-invasive measurement methods in free-living conditions. Recent approaches fuse Photoplethysmograph (PPG) and electrocardiographic (ECG) signals using different machine and deep learning approaches to estimate BP non-invasively; however, they fail to reconstruct the complete signal, leading to less accurate models. We propose a cycle generative adversarial network (CycleGAN) based approach to extract a BP signal known as ambulatory blood pressure (ABP) from a clean PPG signal. Our approach uses a cycle generative adversarial network that extends the GAN architecture for domain translation and outperforms state-of-the-art methods by up to 2x in BP estimation.

Next, we focus on patients diagnosed with acute respiratory distress syndrome (ARDS) who are in more life-threatening circumstances when it comes to COVID-19, resulting in severe respiratory system failure. We investigate the behavior of COVID19 on ARDS patients by utilizing simple vital signs. We analyze the long-term daily logs of blood pressure (BP), and heart rate (HR) associated with 150 ARDS patients admitted to five University of California academic health centers to distinguish subjects with COVID-19 positive and negative test results. In addition to the statistical analysis, we develop a deep neural network model to extract features from the longitudinal data. Our deep learning model achieved 0.81 area under

the curve (AUC) to classify the vital signs of ARDS patients infected with COVID19 versus other ARDS diagnosed patients. Finally, we designed and developed a study to recommend personalized exercises to non-pregnant subjects to increase their physical activity level. We developed smartphone and smartwatch applications to collect, monitor, and suggest exercises using a contextual multi-arm bandit framework. This study includes constructing and developing a personalized model that predicts or recommends different actions depending on individual user biofeedback.

Chapter 1

Introduction

Well-being is a crucial factor in human lives and society as it indicates that people believe their lives are going satisfactorily. Satisfactory living conditions (e.g., job) are essential to happiness and monitoring these situations is critical for public policy [11]. Although there is no unique definition for well-being, the Centers for disease control and prevention (CDC) defines some minimal characteristics, including the presence of positive emotions and moods (e.g., happiness), the absence of negative emotions (e.g., anxiety), emotional well-being (mental health), physical well-being (body health), and social well-being [11].

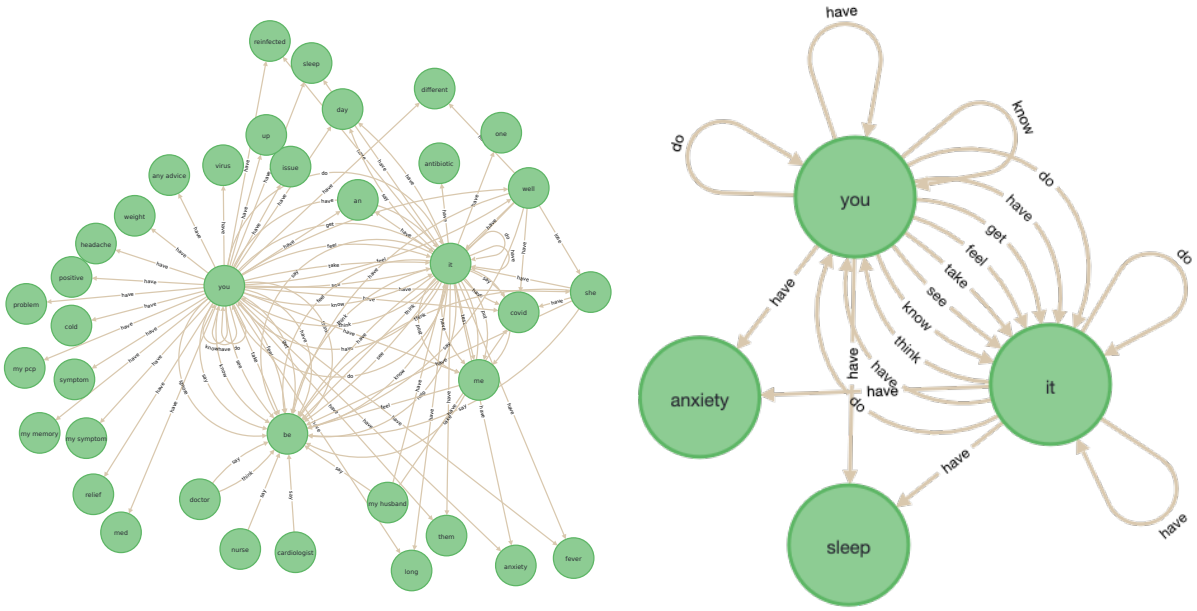
In this dissertation, we focus on physical health, mental health, and social health as these are explicitly mentioned in the World Health Organization (WHO) Constitution [45]: "Health is a state of complete physical, mental, and social well-being and not merely the absence of disease and infirmity."

The well-being domains, as mentioned earlier, are strongly interconnected, and it is challenging to separate them. However, we favor mental health, sleep, and physical activity, as the main pillars of well-being, since these can be the indicators of body health. In addition, the entire world has been affected significantly by a global virus pandemic. The first case of

this virus, known as SARS-CoV-2, was reported in China in December 2019, and the first case outside China was discovered in January 2020 [2]. In February, the World Health Organization called this virus COVID-19 [57]. In this dissertation, we have also investigated the effect of COVID-19 as one of the representatives of social well-being, which was introduced above as one of the main pillars.

In addition, our preliminary analysis on SurvivorCorps data [9] supported that COVID-19 was impactful on these aforementioned pillars. SurvivorCorps is a COVID-19-related organization dedicated to supporting COVID-19 survivors and researches in the filed. It consists of a Facebook group containing over 170,000 users as of October 1st, 2021. People use this group to share their experiences and feelings about this virus. These data captures experiences which cannot be detected using Electronic Health Records (EHR). We used `Relatio` [23] to extract narratives for Facebook posts. Hence, we only focused on only entities and their corresponding relation. Fig. 1.1a illustrates the extracted narratives. As Fig. 1.1b shows, anxiety and sleep related topics has been widely discussed in this dataset. In addition, we performed a topic modeling using DocScan algorithm [128]. We found some clusters regarding anxiety and stress which were supporting the impact of COVID-19 (Fig. 1.2).

Monitoring and quantifying such pillars would be the next challenge. Many parties interested in illness prevention and health promotion may find it helpful to measure, monitor, and promote well-being. Empowering people to gain more control over and improve their health is known as health promotion [11]. Thanks to the Internet of the Things (IoT) advances, it is now possible to monitor health outcomes and biomarkers in everyday free-living conditions without needing to proceed to labs or clinical settings [92, 21, 94, 67, 20]. Therefore, there is enormous growth in the number of consumer-grade medical devices for physical and psychological health, emotional awareness, and sleep quality [108, 121]. Of these devices, only 5% have been formally validated, and around 10% are well-supported, which means they are



(a) (b)

Figure 1.1: Top 100 low-dimension narratives extracted from Facebook posts (a) and sleep and anxiety sub-graph (b).

(a) (b)

Figure 1.2: Topics related to anxiety and stress.

being used in research [108].

Consequently, the accuracy and validation of these non-invasive, consumer-grade devices play an essential role. In this dissertation, we first evaluated the accuracy of a commercial smart ring and smartwatch in sleep [92], then we used these devices for the rest of the studies. These devices have been validated in other aspects by our colleagues in our group

for physical activity [106] and heart rate and heart rate variability [34].

The main contribution of this dissertation can be organized into two components. The first two chapters examine the validity and accuracy of consumer-grade sleep trackers as well as trends and patterns of sleep and mental health disorders at a population level. The benefit of such investigations is that capturing reliable real-time information and public attitudes would facilitate policymakers to monitor public health and social wellness. The second part of this dissertation which encloses the last two chapters, focuses on individual level analyses. We employ Machine Learning and Deep learning approaches to monitor, reconstruct, analyze and predict different tasks using signals and biomarkers of individual subjects. These two chapters concentrate on designing and developing a personalized model which can predict or recommend different tasks based on the biofeedback of individual users.

The rest of this dissertation is organized as follows. In Chapter 2, we focus on the validity of the Oura smart ring [7] and Samsung smartwatch [6] in measuring sleep features and quality [92]. Then we investigate the effect of COVID-19 on sleep quality measurements of pregnant women which is collected by these devices [105]. In Chapter 3, we analyze the trend of mental health disorder reports before and during the COVID-19 pandemic to investigate the social well-being and endangered population. In Chapter 4, we focus on advanced Machine Learning and Deep Learning methods for bio-signal synthesis and processing. First, we propose a model to synthesize the entire ambulatory blood pressure (ABP) signal from a photoplethysmograph (PPG) signal using a cycle generative adversarial network (CycleGAN) framework [91]. Then we proposed a deep learning method leveraging convolutional neural networks (CNN) and long short-term memory (LSTM) to investigate the behavior of COVID-19 in patients with acute respiratory distress syndrome (ARDS) [94]. Finally, in Chapter 5, we present an exercise recommendation system based on reinforcement learning that uses biomarkers and the user's context to recommend a personalized walking exercise that enhances the user's aerobic capacity.

Chapter 2

Sleep Monitoring

Assessment of sleep quality is essential to address poor sleep quality and understand changes. Owing to the advances in the Internet of Things and wearable technologies, sleep monitoring under free-living conditions has become feasible and practicable. Smart rings and smartwatches can be employed to perform mid- or long-term home-based sleep monitoring. However, the validity of such wearables should be investigated in terms of sleep parameters. Sleep validation studies are mostly limited to short-term laboratory tests; there is a need for a study to assess the sleep attributes of wearables in everyday settings, where users engage in their daily routines.

The first part of this chapter aims to evaluate the sleep parameters of the Oura ring along with the Samsung Gear Sport watch in comparison with a medically approved Actigraphy device in a midterm everyday setting, where users engage in their daily routines.

In the second part of this chapter, we were interested in monitoring sleep behavior of pregnant women after evaluating the effectiveness of Internet of Things–based devices and smartwatches technology for monitoring sleep. The purpose of this study was to look at daily characteristics of well-being in pregnant women before and during the national stay-at-home

restrictions imposed in Finland due to the COVID-19 pandemic.

2.1 Background

Sleep is a multifaceted and dynamic phenomenon that indicates individuals' overall health and well-being and is affected by a variety of factors such as behavioral habits, stress, and disorders [69, 58]. Sleep disturbances are common across different population groups (e.g., older people and pregnant women) and negatively impact body functions, including the cardiovascular and immune system and hormonal release [28, 135]. Such sleep problems need to be investigated thoroughly to reduce the associated health risks and complications. Monitoring sleep quality is a vital step in this regard when the individuals' sleep parameters are tracked [109].

Sleep quality assessment methods have been conventionally performed in clinical settings by monitoring users' biological signals and body movements. Polysomnography (PSG), the gold standard method used for sleep analysis, is enabled by the continuous monitoring of different cardiorespiratory and neurophysiological indicators [27]. Owing to PSG's complex and multichannel data collection, this method is limited to short-term hospital or laboratory-based monitoring. Actigraphy is another well-established method enabled by a 3D accelerometer that captures the movements of a limb to monitor sleep [18]. This method has been shown to be accurate enough compared with PSG in a healthy subject population [115, 88, 117, 119], although the results might be inaccurate when the subjects are individuals with sleep disorders [18, 119, 136]. In addition, other studies conducted with large populations have shown an agreement between actigraphy and PSG in total sleep time (TST), wake after sleep onset (WASO), and sleep efficiency (SE) parameters [119, 89]. On the other hand, some studies have considered the validity of actigraphy's sleep onset latency (SOL) compared with PSG [119, 124] and showed that actigraphy consistently underestimated SOL in comparison with

PSG. This method is more convenient than PSG because it allows users to wear actigraphy in everyday settings (i.e., days to weeks), although conventional medical-grade actigraphy devices are still infeasible for long-term studies (i.e., months to years) because of their size, design, and battery life issues.

Advancements in consumer wearable technology provide opportunities to extend sleep monitoring to mid- or long-term home-based health care applications using low-power, miniaturized, and fashionable wearables [74, 110, 25]. Wearable electronics and the Internet of Things–based systems are growing dramatically and are expected to revolutionize health care delivery and outcomes [56, 97]. In particular, smart rings will most likely become popular in sleep studies. Longer battery life, elegant design, and sophisticated embedded sensors in such rings have enabled them to be used not only in clinical trials (instead of medical-grade actigraphy) but also in different population-based studies [72, 78]. Such devices offer continuous data collection of body movements and vital signs in everyday settings. The data can be utilized to continuously monitor the sleep disturbances of individuals for an extended period [24].

Sleep monitoring using consumer wearables such as wrist-worn activity trackers, smartwatches, and smart rings necessitate valid sleep data collection and data analysis to provide accurate sleep parameters. Various studies have investigated wrist bands in terms of sleep monitoring accuracy across different population groups. For example, the validation of sleep data of 7 different commercial activity trackers was assessed by conducting data collection for 2 days on healthy adults [48]. In other studies, the sleep estimation of Fitbit devices [42, 102, 59], Jawbone [41, 43, 38], and physical activity monitors [127] has been investigated against actigraphy, PSG, or both in overnight tests on healthy adolescents and individuals with obstructive sleep apnea. These studies focused on the sleep quality assessment of wearables by tracking a set of non-staging sleep parameters, including TST, SOL, WASO, and SE [95, 101, 71, 37]. Regarding smart ring validation, there is one study that has validated

the Oura smart ring against PSG in an overnight laboratory setup [44]; however, there is no previous research in the literature validating a smart ring against actigraphy in the mid- or long-term. Furthermore, these earlier validation studies are limited to laboratory settings and/or overnight (i.e., single night) data collection. The effect of home-based health monitoring, where the users might be involved in different conditions and environments, is ignored in these validation studies. Therefore, the results obtained could be inaccurate for long-term and remote monitoring.

2.2 Instrument Validation Study

2.2.1 Background

In this study, we aim to assess the validity of sleep data acquired by a smart ring, Oura, in comparison with a medically approved actigraphy device. We utilize the Oura ring as a compact and relatively small device with a user-friendly design. In addition, we assessed the Samsung Gear Sport smartwatch against actigraphy to compare the accuracy of Oura ring in the detection of different sleep attributes. In general, because watches and rings are worn in different parts of the subject's hand, they respond differently to signal logging disturbances, such as motion artifacts. The devices were tested in a 7-day monitoring study, approved by the ethical committee, where the sleep data of 45 healthy individuals were monitored. Participants were asked to use the devices 24 hours for 7 days and carry out their daily routines as usual. We compared TST, SOL, WASO, and SE obtained from the Oura ring, Samsung watch, and ActiGraph. The parameters obtained by the 2 consumer-grade wearables (i.e., the ring and the watch) were evaluated with the sleep parameters extracted from actigraphy using paired t-tests, Bland-Altman [30] plots, and Pearson correlation. The parameters were investigated considering the gender of the participants as a dependent variable. Finally, we

conclude the section with a discussion of our obtained results and the validity of sleep data of the wearables in everyday settings.

2.2.2 Methods

Participants and Recruitment Recruitment was performed in southern Finland from July to August 2019. In earlier validation studies of commercial devices, the sample sizes varied between 20 and 40. Therefore, we aimed at a target sample of 40 people. The recruitment started with convenience sampling by personally contacting a few students and staff members of the University of Turku. Afterward, snowball sampling was used until the target sample size was reached; 6 additional participants were enrolled because of expected missing data. We aimed for variation among participants by age, weight, physical activity, education, and lifestyle as related to sleep and stress levels.

A sample of healthy individuals between 18 and 55 years of age was enrolled. Potential participants were excluded if they had (1) a diagnosed cardiovascular disease, (2) restrictions regarding physical activity, (3) symptoms of an illness at the time of recruitment (i.e., flu symptoms including sore throat, runny nose, cough, and fever), or (4) any restrictions on using the devices at work. In a face-to-face meeting with the interested individuals, researchers described the purpose of the study and the wearable devices. They were asked to wear the Gear Sport smartwatch, Oura ring, and ActiGraph wristband for 1 week in their normal daily life. Each participant provided written informed consent. Altogether, 46 participants, including 23 women and 23 men, participated in the study. A participant (male) was excluded from the analysis because he did not wear the actigraphy device. Therefore, the final sample size was 45 (23 women, 22 men). Table 2.1 shows the participants' background information. The table includes 42 participants, as the background information of the 3 participants is missing.

Table 2.1: Participants' background information.

Characteristics		Value
Age (years), mean (SD)		
	Women	31.5 (6.6)
	Men	33 (6)
BMI, mean (SD)		
	Women	24.4 (5.6)
	Men	25.5 (2.9)
Expected sleep (daily hours), mean (SD)		
	Women	7.35 (1.0)
	Men	7.17 (1.05)
Physical activity, n (%)		
	Almost daily	12 (27)
	Once a week	9 (20)
	> Once a week	21 (47)
Working status, n (%)		
	Working	32 (71)
	Unemployed	1 (2)
	Student	8 (18)
	Other	1 (1)

Ethics The study was conducted according to the ethical principles based on the Declaration of Helsinki and the Finnish Medical Research Act (#488/1999). The study protocol received a favorable statement from the ethics committee (University of Turku, Ethics committee for Human Sciences, Statement #44/2019). The participants were informed about the study, both orally and in writing, before obtaining their consent. Participation was voluntary, and all participants had the right to withdraw from the study at any time and without giving any reason. To compensate for the time used for the study, each participant received a €20 ($23) gift card to the grocery store at the end of the monitoring period when returning the devices.

Data Collection Our data collection for 1 week included 4 approaches for monitoring participants' sleep. We utilized 3 devices (i.e., 2 wearable and 1 actigraphy device) to continuously capture sleep data and a self-report form by which subjective measures were collected. Samsung Gear and ActiGraph were worn in the wrist, and the Oura ring was worn

in one of the fingers of the nondominant hand; thus, all 3 devices were on the same hand. The participants completed a short background questionnaire at the meetings. They were also asked to report their sleep times, such as bedtime, waking up time, and naps, during the 7-day study period via a structured self-report (i.e., daily log) form. They were also asked to report other events during the study, such as device removal from the wrist or if specific events occurred (e.g., visiting a hospital because of a disease). The self-report data were used to interpret the actigraphy data and mitigate possible errors; such a correction was necessary for this study because the actigraphy was selected as the baseline sleep monitoring method. In addition to the verbal instructions, participants were given a written guideline for using the devices.

The Oura ring [7] was the first wearable device investigated in this study. The Oura ring is a commercial sleep tracker device that uses acceleration and gyroscope data, photoplethysmogram (PPG) signal, and body temperature to estimate sleep parameters, heart rate variability, respiratory rate, and intensity of physical activity. The ring is lightweight (4-6 g) and easy to use. It also has an acceptable battery life, that is, the battery lasts up to 1 week in regular use. The ring is connected to the Android or iOS Oura mobile app via Bluetooth. The data are automatically sent to the mobile app and transferred to the cloud server. The data can be accessed from the mobile app or the cloud server. In this study, we extracted the sleep data of participants from the Oura cloud.

In addition to the Oura ring, we used the Samsung Gear Sport watch [6], which is an open-source smartwatch that enables remote health monitoring. The watch includes a PPG sensor and an inertial measurement unit through which PPG signal, acceleration, and gyroscope data can be collected continuously. The data are processed to extract various variables, including heart rate, sleep duration, and step counts. The Gear Sport watch runs open-source Tizen operating system, enabling customized data collection. In this study, we programmed the watch to collect sleep parameters, PPG data, and hand movement data during the

monitoring. The PPG and hand movement data were utilized to validate the sleep events (detailed in the Data Analysis section). Moreover, we also developed an app for the watch to send the collected data to our server via Wi-Fi.

For actigraphy, we used the wGT3X-BT device by ActiGraph. The wGT3X is a noncommercial triaxial accelerometer that measures the wrist's acceleration in 3 orthogonal axes at 80 Hz. This device is waterproof, and its battery life is approximately 3 weeks. The device does not provide any feedback to the participants about their activity or sleep. The acceleration data collected by the device were utilized to obtain the estimates of sleep parameters.

Data Analysis

Data analysis included the sleep parameter extraction from the collected data and the statistical analysis leveraged to evaluate the ring and watch.

Actigraphy Raw data from the actigraphy device were downloaded to a computer and converted into 60-second epochs using the ActiLife software (version 6.13) [12] provided by the manufacturer (ActiGraph). We used the Cole-Kripke algorithm [36] to define each epoch as sleep or wake. This algorithm was selected because it has been validated in the adult population using wrist-worn accelerometers. The ActiGraph algorithm that is available in the ActiLife software was then used to detect the sleep periods and estimate sleep attributes. Using the Troiano wear time validation algorithm [137], the auto sleep period detection algorithm detects nonwear bouts, ignores nonwear periods greater than a day, and nonwear periods that have almost all zeros (5 or more epochs of nonzeros). The nonwear periods that remain are defined as sleep time. Sleep data were systematically checked, cleaned, and sleep periods that did not represent true sleep times were deleted. These deletions included sleep periods with nonwear time during evenings or mornings that the algorithm had incorrectly scored as sleep, daytime sleep periods, and sleep periods outside the actual measurement

week.

Wearables We used the application programming interface provided by the Oura ring and the Samsung watch to extract different semistructured data for our analyses. The Oura ring provides JavaScript Object Notation files, including the sleep parameters per night. The 3 main types of sleep parameters provided by the ring are (1) parameters related to different levels of sleep and nonstaging sleep, including the start and end of sleep, the number of awakenings, total awakening time, and sleep onset, (2) scores to measure the quality of sleep in different stages, and (3) average heart rate for every 5 min during sleep. In this study, we only investigated the nonstaging sleep parameters because of the limitation of the baseline actigraphy method.

In contrast, the Gear Sport watch provides a data record when the user's status changes; for example, the status changes from wake to sleep. We used these records to extract sleep events per night and validated the sleep events using the heart rate and hand movement data collected by the watch. Validation was performed to prevent the misdetection of sleep events owing to not wearing the device. For example, the watch was not used (no movement) for 1 hour, but a sleep event was detected by mistake. In this regard, we recorded a window of 30-second PPG signal when a sleep event started and ended. The sleep event was considered valid if valid heart rate values were detected from the PPG signals. In addition, we considered the hand movement magnitude for validation if the PPG signal was invalid because of practical issues. Finally, we cross-checked the sleep events with the step count data (reported by the watch) and corrected or discarded the sleep events if there was no match between the data. Fig. 2.1 shows an example of manual cross check we did to detect different sleep parameters using sleep events, steps and heart rate data.

It should be noted that the watch could not detect a few sleep events because of technical and practical issues during the monitoring. For example, the sleep event was missed because the watch's turn-off button was pressed accidentally during the night. This issue mostly

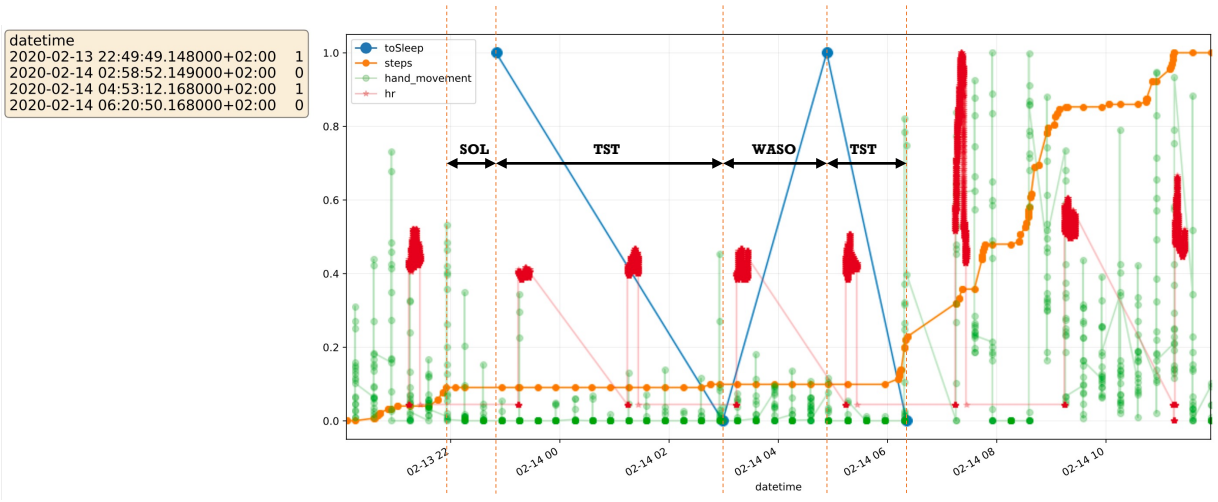


Figure 2.1: Sleep extraction multi-modal cross checking.

occurred during the monitoring, as the watch and actigraphy were worn on the same hand close to each other. As the watch could not record the sleep events, we removed 21 nights of data out of 181 (11.6%) of the watch for the sake of an unbiased comparison between the actigraphy and watch.

Using the actual valid sleep events, we calculated WASO, TST (in minutes), and SE (%) per night. Since the watch does not provide SOL explicitly, we calculate such a feature based on the difference between the start of the actual sleep and the last time the subject had steps. A summary of the processing pipeline is illustrated in Fig. 2.2.

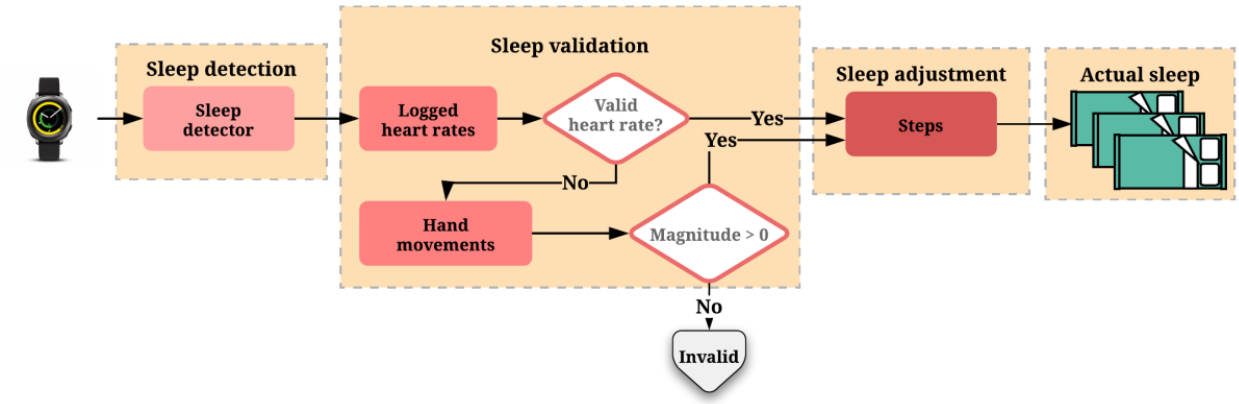


Figure 2.2: Watch data processing pipeline.

Statistical Analysis

We report the mean, SD, and 95% CI of the sleep parameters collected by the Oura ring, Samsung watch, and ActiGraph. The difference between the ring (or the watch) and the ActiGraph was also computed using two-tailed paired t tests to test the null hypothesis. In our context, the null hypothesis is that the true mean difference between the two measurements is 0 [142]. Due to the interest in observing the paired differences between values reported by ring (or the watch) and ActiGraph (baseline), the paired t test was utilized. In addition, we used the Bland-Altman plot to illustrate and estimate the agreement between the devices. These methods provided mean differences (bias) and SD of the differences between the ring (or the watch) and the actigraphy, lower and upper agreement limits, and 95% CI of the mean differences.

The sign of mean differences indicates underestimation or overestimation of the ring (or the watch) compared with the actigraphy: a negative bias shows an overestimation, whereas a positive bias indicates an underestimation.

The satisfactory difference between the ring (or the watch) and the actigraphy data was selected as $\leq$ 30 min for TST and WASO and $<$ 5% for SE, similar to other studies in the literature [41, 44, 96]. We investigated the ratio of the samples within these satisfactory ranges. Moreover, we also investigated gender as a dependent variable in the validity of sleep parameters using t tests, considering the mean differences between the ring (or the watch) and the actigraphy.

Finally, to analyze the linear relationship between actigraphy and the ring (or the watch) corresponding sleep measurements, we performed Pearson correlation tests on pairwise sleep attributes of the actigraphy and the ring (or the watch).

2.2.3 Results

A total of 45 subjects (23 women and 22 men) participated in this study. The subjects were 33.1 years old, on average, with an SD of 6.4 years. In total, we recorded 284 valid available days by actigraphy; however, after matching the corresponding available days of the ring (or the watch), we had fewer valid days for the analysis. As discussed in the Methods section, in this study, we exploited 4 different sleep attributes. Although the results regarding SOL are not conclusive (because SOL of the actigraphy is unreliable [124]), for the sake of comparison, we report such results in addition to the other sleep parameters in this section.

Comparisons Between Ring and Equivalent Actigraphy Sleep Measures

To validate the Oura ring against actigraphy, we matched the available dates of the ring with the corresponding dates of actigraphy. In total, for all the participants, sleep data of 266 days (i.e., 5.91, SD 1.32 days per subject) were included in the analysis.

The mean, SD, and 95% CI of the extracted sleep parameters are presented in Table 2.2. The table also shows the paired t test values of these parameters with their corresponding P values. Bland-Altman plots were used to show the agreements between the 2 measures. Fig. 2.3 depicts the agreement between the ring and actigraphy for the TST, WASO, and SE. The bias and lower and upper agreement limits for these parameters are also summarized in Table 2.3.

As shown in Table 2.2, the ring significantly overestimated the actigraphy (t_{265}=-6.26; $P <$ .001) in the estimation of TST. On the basis of 2.3, this overestimation in TST is, on average, 15.27 (SD 39.68) min (95% CI -20.07 to -10.47). Of 266 total samples, 14 fell outside the agreement range (lower limit -93.04 min, upper limit 62.50 min). The mean difference of TST between the actigraphy and ring fell within the satisfactory range, and 65.03% of the

Table 2.2: Mean, SD, 95% CI, and paired t test results for the actigraphy and the Oura ring sleep parameters in a sample of 45 healthy adults.

Parameter		Mean (SD)	95% CI	t-value (df)	P value
Total sleep time (min)					
	t-test	N/A	N/A	-6.26 (265)	< .001
	Actigraphy	419.04 (78.31)	409.59-428.5	N/A	N/A
	Oura ring	434.31 (72.14)	425.6-443.02	N/A	N/A
Sleep efficiency (%)					
	t-test	N/A	N/A	3.69 (265)	< .001
	Actigraphy	90.47 (5.1)	89.86-91.09	N/A	N/A
	Oura ring	89.13 (6.28)	88.38-89.89	N/A	N/A
Wake after sleep onset (min)					
	t-test	N/A	N/A	10.03 (265)	< .001
	Actigraphy	43.57 (27.28)	40.28-46.86	N/A	N/A
	Oura ring	26.17 (24.98)	23.15-29.18	N/A	N/A

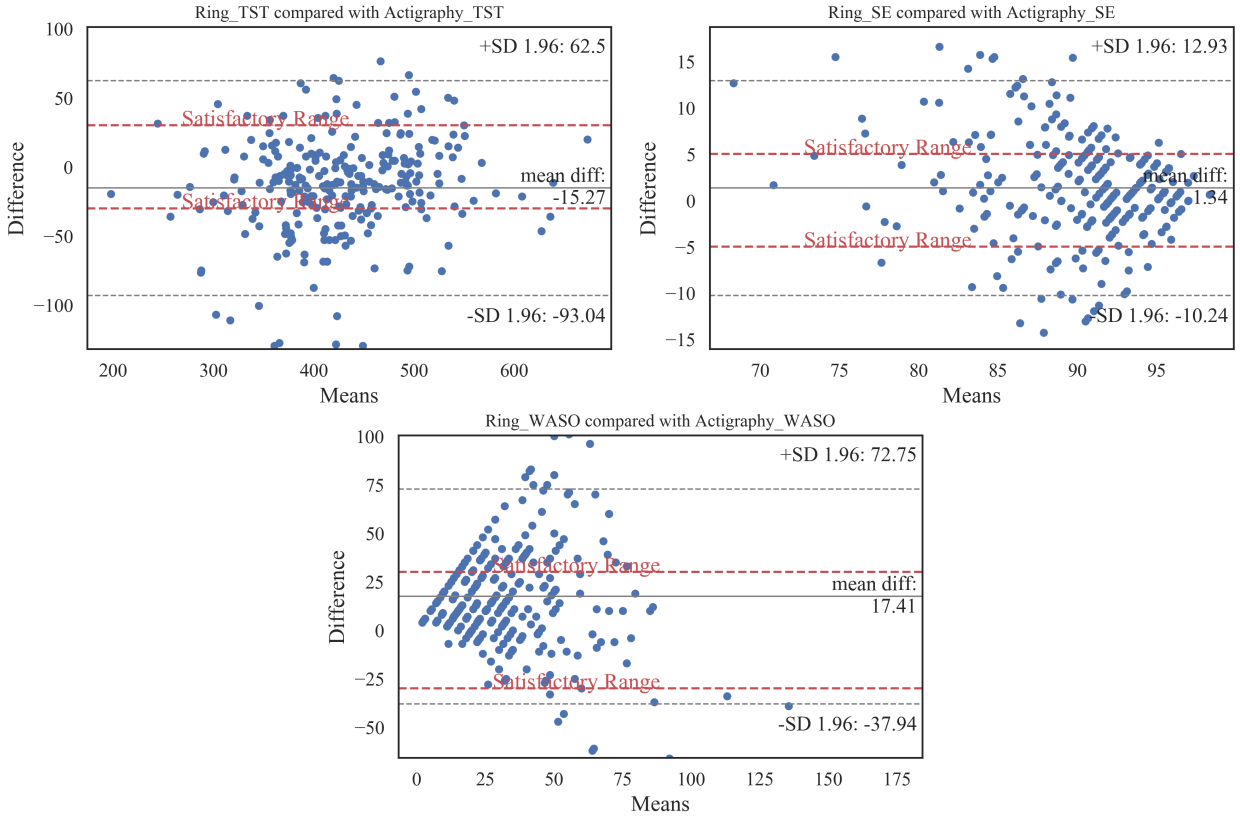


Figure 2.3: Bland-Altman plots for total sleep time, sleep efficiency, and wake after sleep onset gathered by the Oura ring and the actigraphy device. Subjects' actigraphy minus Oura ring discrepancies on sleep parameters (y-axis) are plotted compared with actigraphy (x-axis). Biases, upper, and lower agreement limits are marked. In addition, the satisfactory ranges are plotted as the dashed lines. SE: sleep efficiency; TST: total sleep time; WASO: wake after sleep onset.

Table 2.3: Bias and agreement limits based on Bland-Altman plots for the actigraphy and the Oura ring.

Parameter	Mean difference (SD)	Lower and upper agreement limits
Total sleep time (min)	-15.27 (39.68)	-93.04, 62.5
Sleep efficiency (%)	1.34 (5.91)	-10.24, 12.93
Wake after sleep onset (min)	17.41 (28.24)	-37.94, 72.75

data samples followed the satisfactory range condition.

On the other hand, in terms of WASO, the Oura ring significantly underestimated (t_{265}=10.03; $P < .001$) the actigraphy by, on average, 17.41 min (95% CI 13.99 to 20.82). Out of 266 samples, 17 fell outside the agreement limits (lower limit -37.94 min, upper limit 72.75 min). In terms of the satisfactory range, the mean difference fell within the range and covered 69.92% of the total samples.

In addition, the Oura ring underestimated SE compared with the actigraphy by 1.34% on average (95% CI 0.63 to 2.06). This underestimation was significant, as shown in Table 2.2 (t_{265}=3.69; $P < .001$). The mean difference in SE between the Oura ring and the actigraphy fell within the satisfactory range (< 5%), along with 65.78% of samples (including 44 out of 45 subjects). Moreover, 18 samples fell outside the agreement limits (lower limit -10.24%, upper limit 12.93%).

Comparisons Between Watch and Equivalent Actigraphy Sleep Measures

Similar to the ring validation, we considered the available dates for the actigraphy with corresponding data collected by the Samsung watch. As mentioned in the Wearables section, we removed the technically invalid watch data that occurred because of practical issues during the monitoring. Therefore, there were fewer sleep data from the watch than the other devices. After the matching procedure and invalid data removal, the number of subjects for the watch validation was 35 (19 men and 16 women), with 134 data samples (3.82, SD 1.50 days per subject). Table 2.4 summarizes the mean, SD, and 95% CI of the Samsung watch

and the actigraphy with the corresponding available dates for different sleep parameters.

Table 2.4: Mean, SD, 95% CI, and paired t test results for the actigraphy and the Samsung watch sleep parameters in a sample of 35 healthy adults.

Parameter		Mean (SD)	95% CI	t-value (df)	P value
Total sleep time (min)					
	t-test	N/A	N/A	-3.54 (133)	< .001
	Actigraphy	409.29 (81.43)	395.38-423.21	N/A	N/A
	Samsung watch	431.81 (82.21)	417.76-445.85	N/A	N/A
Sleep efficiency (%)					
	t-test	N/A	N/A	-6.49 (133)	< .001
	Actigraphy	90.40 (5.05)	89.54-91.26	N/A	N/A
	Samsung watch	94.84 (7.03)	93.64-96.04	N/A	N/A
Wake after sleep onset (min)					
	t-test	N/A	N/A	10.26 (133)	< .001
	Actigraphy	42.23 (23.43)	38.23-46.24	N/A	N/A
	Samsung watch	10.96 (30.46)	5.76-16.17	N/A	N/A

In addition, we performed paired t tests for the sleep parameters of the 2 devices. The results are shown in Table 2.4. As shown in this table, the t test values for all considered sleep parameters were statistically significant ($P < .001$). The positive and negative sign of the t value denotes the underestimation and overestimation of actigraphy by the watch, respectively. Bland-Altman plots showing TST, WASO, and SE agreements between the actigraphy and the watch are also illustrated in Fig. 2.4. Moreover, bias and lower and upper agreement limits of sleep parameter outcomes by the actigraphy and the watch are summarized in Table 2.5.

Table 2.5: Bias and agreement limits based on Bland-Altman plots for the actigraphy and the Samsung watch.

Parameter	Mean difference (SD)	Lower and upper agreement limits
Total sleep time (min)	-22.51 (73.24)	-166.07, 121.04
Sleep efficiency (%)	-4.44 (7.88)	-19.89, 11.01
Wake after sleep onset (min)	31.27 (35.15)	-37.62, 100.15

As shown in Fig. 2.4, the watch overestimated the actigraphy in TST, on average, by 22.51 min (95% CI -35.08 to -9.95). Among the 134 samples, 9 were beyond the agreement limits (lower limit -166.07 min, upper limit 121.04 min). The mean difference of the actigraphy's and the watch's TST was within the satisfactory range; however, less than 50% (52/134, 38.8%) of the samples were within this satisfactory range.

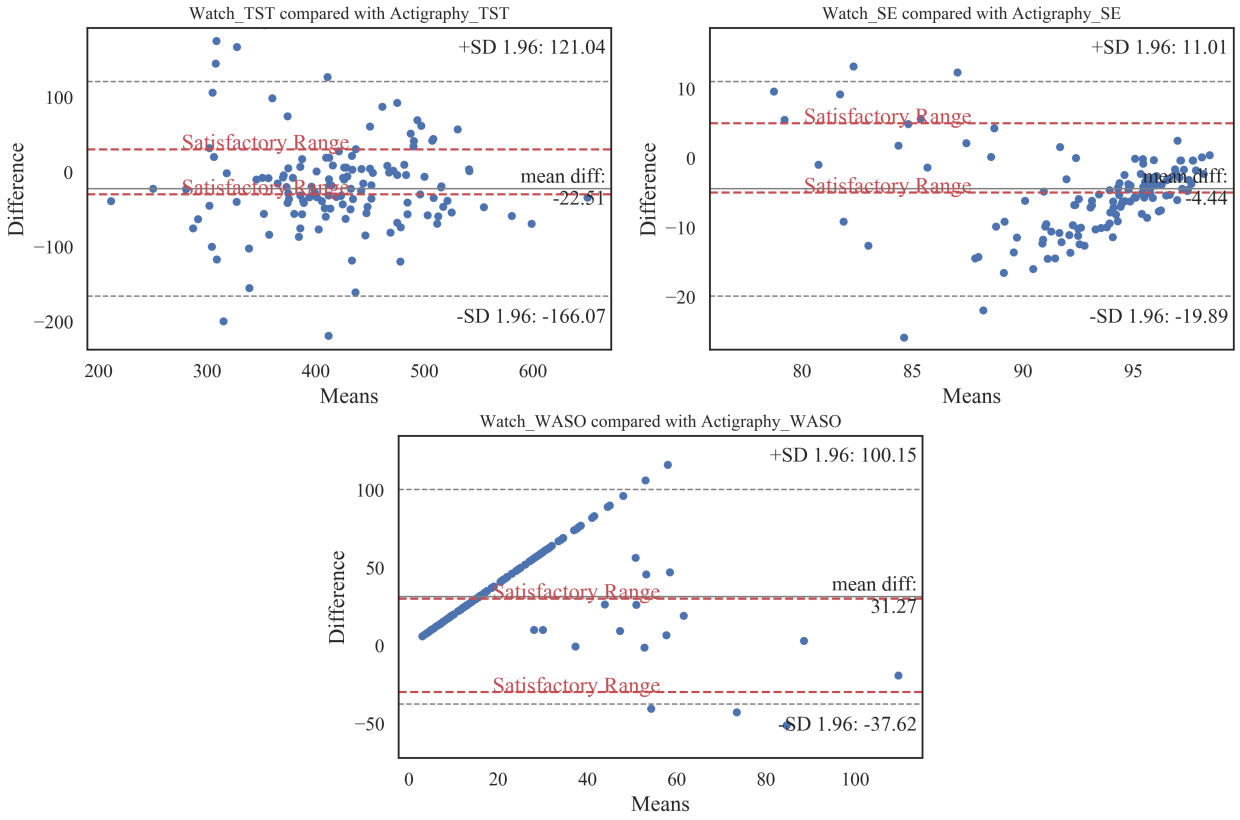


Figure 2.4: Bland-Altman plots for total sleep time, sleep efficiency, and wake after sleep onset gathered by the Samsung watch and the actigraphy device. Subjects' actigraphy minus Samsung watch discrepancies on sleep parameters (y-axis) are plotted compared with actigraphy (x-axis). Biases, upper, and lower agreement limits are marked. In addition, the satisfactory ranges are plotted as the dashed lines. SE: sleep efficiency; TST: total sleep time; WASO: wake after sleep onset.

In addition to TST, the Samsung watch overestimated SE by 4.44% (95% CI -5.79 to -3.09) compared with the actigraphy; 8 samples fell outside the agreement limits (lower limit -19.89%, upper limit 11.01%), with 42.53% of the samples within the satisfactory range.

On the other hand, the watch underestimated WASO by 31.27 min on average (95% CI 25.24 to 37.3). Only 9 samples were outside of the agreement limits (lower limit -37.62 min, upper limit 100.15 min), and 45.52% of the samples were within the satisfactory range.

Gender-Dependent Changes in the Mean Differences Between the Actigraphy and the Ring (or the Watch)

We also considered the gender of the participants to determine if the mean difference in sleep parameters differed between female and male groups. Table 2.6 shows the mean and SD of each sleep attribute of the actigraphy and the ring and the difference between these devices for male and female groups, separately.

The average of the mean difference between the TST of the actigraphy and the Oura ring did not differ between the male and female groups (t_{530}=0.99; $P = .32$). However, the mean differences of the other sleep parameters (i.e., SE and WASO) were significant between female and male participants ($P < .001$ and $P = 0.004$).

Table 2.6: Mean, SD, and average mean differences (the actigraphy minus the Oura ring) for 23 women (141 samples) and 22 men (125 samples).

Parameter		**Mean (SD)**			*t*-**value** (*df*)	*P* **value**
		Actigraphy	**Oura ring**	**Differences**		
Total sleep time (min)						
	t-test	N/A	N/A	N/A	0.99 (530)	.32
	Women	429.67 (70.25)	442.66 (64.67)	-12.98 (37.94)	N/A	N/A
	Men	407.05 (85.21)	424.89 (78.94)	-17.84 (41.39)	N/A	N/A
Sleep efficiency (%)						
	t-test	N/A	N/A	N/A	-4.33 (530)	< .001
	Women	90.64 (4.93)	90.73 (5.16)	-0.09 (5.86)	N/A	N/A
	Men	90.29 (5.31)	87.33 (6.9)	2.96 (5.55)	N/A	N/A
Wake after sleep onset (min)						
	t-test	N/A	N/A	N/A	2.86 (530)	.004
	Women	44.9 (30.08)	22.87 (20.7)	22.03 (29.19)	N/A	N/A
	Men	42.07 (23.75)	29.88 (28.69)	12.19 (26.16)	N/A	N/A

Similarly, we compared the mean differences of the sleep parameters between the actigraphy and the watch for the male and female groups. Table 2.7 summarizes such differences for each sleep parameter. As shown in Table 2.7, there was a significant difference between the mean differences of the male and female groups for TST ($P < .001$), SE ($P = .01$), and WASO ($P = .01$).

Table 2.7: Mean, SD, and average mean differences (the actigraphy minus the Samsung watch) for 16 women (65 samples) and 19 men (69 samples).

Parameter		Mean (SD)			t-value (df)	P value
		Actigraphy	Samsung watch	Differences		
Total sleep time (min)						
	t-test	N/A	N/A	N/A	3.48 (266)	$< .001$
	Women	427.08 (73.76)	427.67 (74.76)	-0.59 (65.67)	N/A	N/A
	Men	392.54 (85.22)	435.7 (89.04)	-43.16 (74.01)	N/A	N/A
Sleep efficiency (%)						
	t-test	N/A	N/A	N/A	2.39 (266)	.01
	Women	90.82 (4.88)	93.6 (7.92)	-2.78 (8.04)	N/A	N/A
	Men	90.0 (5.2)	96.01 (5.9)	-6.0 (7.4)	N/A	N/A
Wake after sleep onset (min)						
	t-test	N/A	N/A	N/A	-2.40 (266)	.01
	Women	42.49 (24.33)	18.64 (39.75)	23.85 (42.54)	N/A	N/A
	Men	41.99 (22.73)	3.73 (14.76)	38.26 (24.36)	N/A	N/A

Correlations

We also investigated the possible linear relationship between the actigraphy and the ring (or the watch) data, using the Pearson correlation test. The correlation value (r) ranges from -1 to 1, where ∓ 1 implies an exact linear relationship. The correlation values and their P values are shown in Table 2.8. We considered $P < .01$ to be significant for the correlation values between the sleep parameters of the actigraphy and the ring (or the watch).

Table 2.8: Pearson correlation between the actigraphy, ring, and smartwatch with the corresponding P values for the considered sleep attributes. TST: total sleep time, SE: sleep efficiency and WASO: wake after sleep onset.

Devices	**Pearson correlation with the actigraphy,** r					
	TST	P**-value**	**SE**	P**-value**	**WASO**	P**-value**
Oura ring	0.86	$< .001$	0.47	$< .001$	0.41	$< .001$
Samsung watch	0.59	$< .001$	0.17	.04	0.16	.06

As shown in Table 2.8, comparing TST of actigraphy with TST of the ring and TST of the watch, we found a significantly high correlation between the actigraphy and the ring ($r = 0.86$; $P < .001$). In contrast, the correlation between the actigraphy and the watch was $r = 0.59$ ($P < .001$).

With regard to SE, there was a correlation between actigraphy and the ring ($r = 0.47$; $P < .001$). In addition, the correlation between the actigraphy and the watch was acceptable

($r = 0.17$; $P = .04$), but not as high as that of the ring.

For the WASO validation, there was a significant correlation between the actigraphy and the ring ($r = 0.41$; $P < .001$). However, our analysis showed a nonsignificant correlation between WASO of the actigraphy and WASO of the watch ($r = 0.16$; $P = .06$).

SOL Comparison Across Devices

As previously mentioned, SOL results were not conclusive since SOL of actigraphy is unreliable. We report SOL separately in the following: mean, SD, 95% CI, and paired t test results of the SOL for comparison between the actigraphy and the Oura ring (or Samsung watch) are presented in Tables 2.9 and 2.10. Bland-Altman plots showing the SOL agreements between the actigraphy and the ring (or the watch) are illustrated in Fig. 2.5 and Fig. 2.6. Details of these plots are summarized in Tables 2.11 and 2.12.

Table 2.9: Mean, SD, 95% CI, and paired t test results for the actigraphy versus Oura ring estimates of sleep onset latency.

Parameter		**Mean (SD)**	**95% CI**	*t*-**value** (*df*)	*P* **value**
Total sleep time (min)					
	t-test	N/A	N/A	-13.01 (265)	< .001
	Actigraphy	0.91 (1.37)	0.75-1.08	N/A	N/A
	Oura ring	12.84 (14.92)	11.04-14.65	N/A	N/A

Table 2.10: Mean, SD, 95% CI, and paired t test results for the actigraphy versus Samsung watch estimates of sleep onset latency.

Parameter		**Mean (SD)**	**95% CI**	*t*-**value** (*df*)	*P* **value**
Total sleep time (min)					
	t-test	N/A	N/A	-10.08 (133)	< .001
	Actigraphy	0.99 (1.38)	0.75-1.22	N/A	N/A
	Samsung watch	13.79 (14.86)	11.25-16.33	N/A	N/A

Table 2.11: Bias and agreement limits based on Bland-Altman plot of the sleep onset latency for the actigraphy and the Oura ring.

Parameter	**Mean difference (SD)**	**Lower and upper agreement limits**
Sleep onset latency (min)	-11.93 (14.92)	-41.18, 17.32

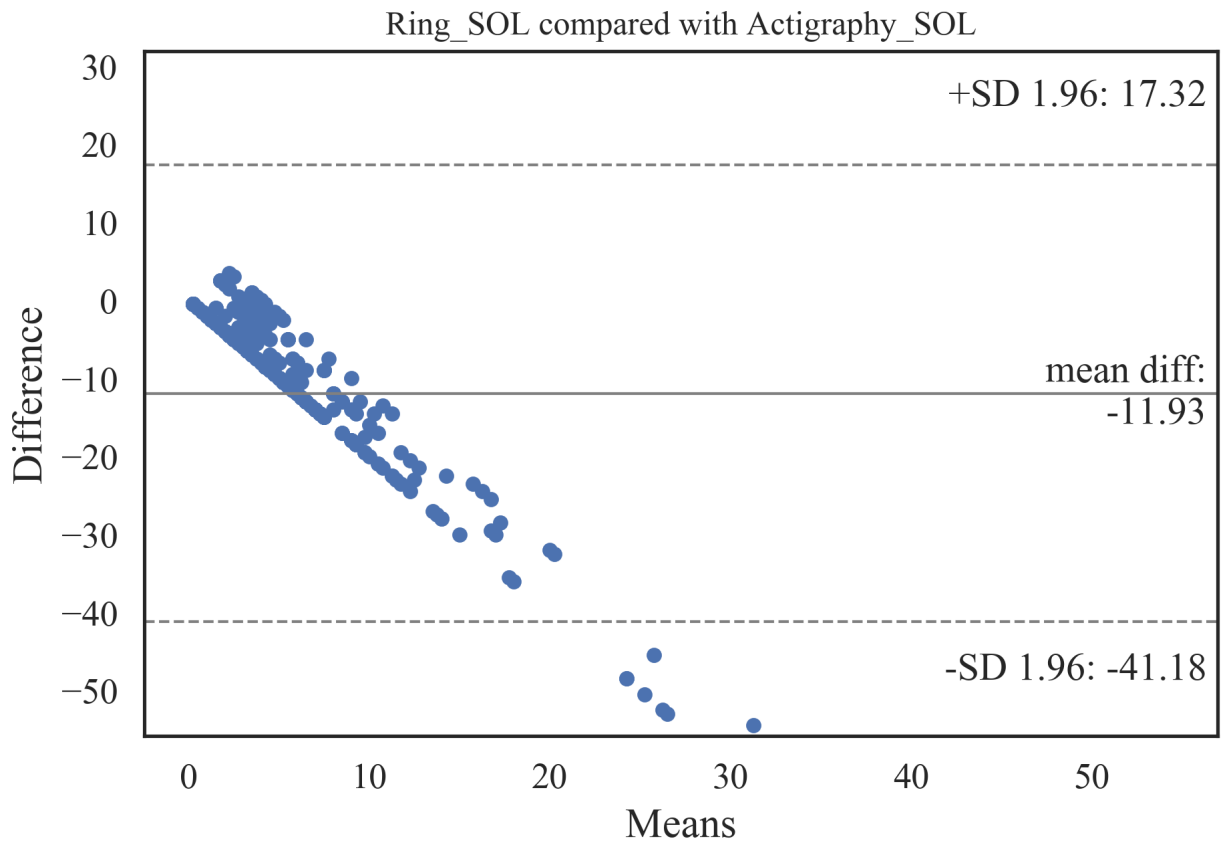


Figure 2.5: Bland-Altman plot for sleep onset latency estimated by the Oura ring. SOL: sleep onset latency.

Table 2.12: Bias and agreement limits based on Bland-Altman plot of sleep onset latency for the actigraphy and Samsung watch.

Parameter	Mean difference (SD)	Lower and upper agreement limits
Sleep onset latency (min)	-12.81 (14.65)	-41.52, 15.91

The Oura ring overestimated the SOL, on average, by 11.93 min (95% C: -13.74 to -10.13) compared with the actigraphy. Out of 266 samples, 14 fell outside the agreement limits (lower limit -41.18 min, upper limit 17.32 min). Table 2.9 shows that the overestimation of the SOL by the ring was significant (t_{265}=-13.01; $P < .001$). Similarly, the watch overestimated the SOL, on average, by 12.81 min (95% CI -15.32 to -10.29). Most of the samples (all except 2) were within the agreement limits (lower limit -41.52 min, upper limit 15.91 min).

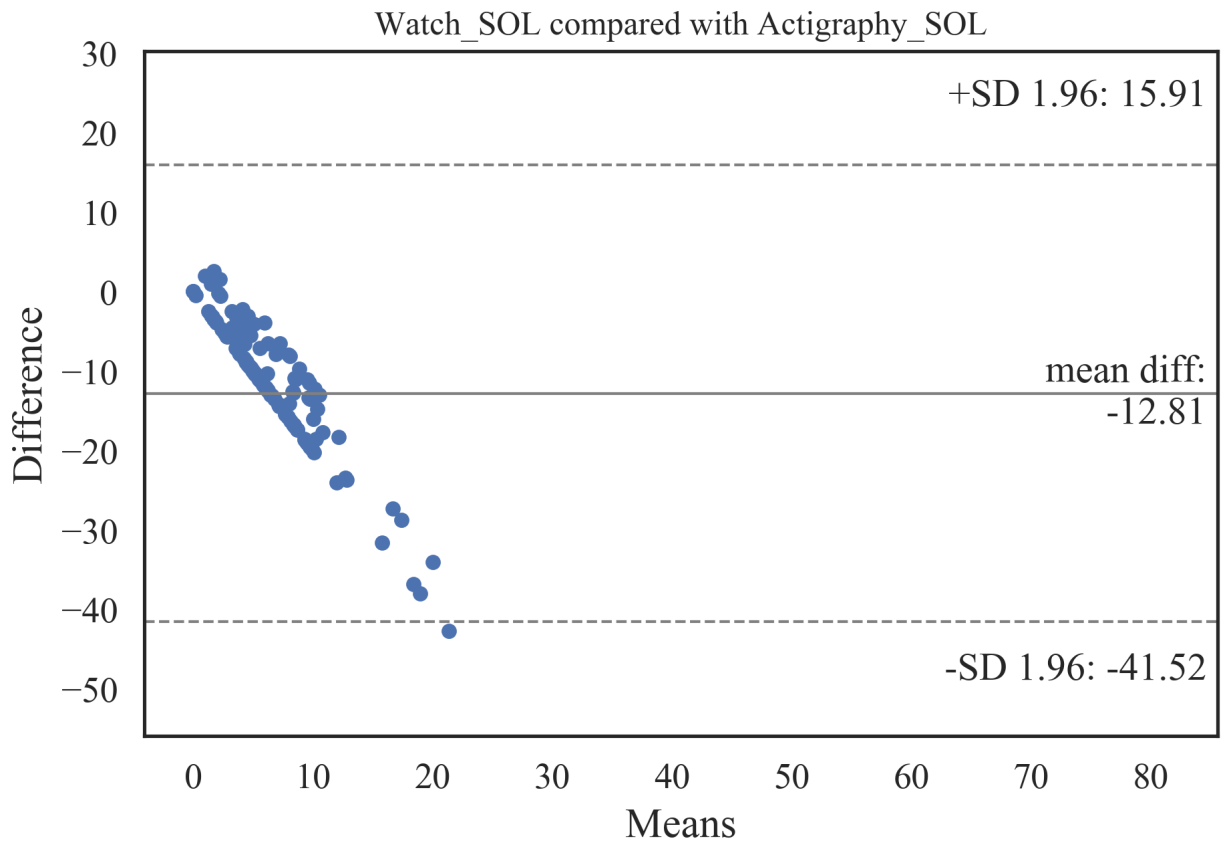


Figure 2.6: Bland-Altman plot for sleep onset latency estimated by the Samsung watch. SOL: sleep onset latency.

2.2.4 Discussion

Principal Findings

To the best of our knowledge, this is the first sleep validation study of the Oura ring and the Samsung watch performed under free-living conditions in comparison with an actigraphy method. The free-living condition allows participants to engage in their daily routines as usual during the monitoring. If commercial devices are used in trials under such free-living conditions, subjective evaluations and self-reports are insufficient to measure the validity of these devices [52, 79, 77]. It is important to test these devices against research devices in order to investigate their error margins and to standardize their software versions, minimizing

controllable measurement differences. In contrast to related work, this study investigated wearables in a 1-week home-based monitoring, providing a higher confidence level on the validity of sleep parameters reported by these wearables. We discuss the results obtained and compare them with the related sleep validation studies, most of which are limited to the laboratory settings and compared with PSG.

Our findings showed that the mean differences of TST, WASO, and SE between the actigraphy device and the Oura ring were within the satisfactory range (i.e., ≤ 30 min for TST and WASO and $< 5\%$ for SE). Within the 266 valid total nights of sleep, only 14 TST, 17 WASO, and 18 SE fell outside the agreement limits. Our results also indicated significant correlations between the TST, WASO, and SE of the ring and the actigraphy. These findings are in accordance with a previous validation study of the Oura ring carried out in a single laboratory overnight study [44].

On the other hand, we found significant differences between the means of TST, WASO, and SE of the ring and the actigraphy. In our study, the Oura ring overestimated the TST (15.27 min) and underestimated the WASO (17.41 min) and SE (1.34%). Although the differences were within the satisfactory range, our results showed more overestimation and underestimation of the Oura ring than the lab-based sleep validation study [44]. This might be explained by the difference between the studies' samples and setups. Our study included more sleep data (i.e., 225 more nights) and was performed in the house. Therefore, our results should be more accurate and have higher confidence levels in real-world applications. Unfortunately, these inaccuracies in sleep measurements in commercial devices might decrease their feasibility for clinical trials [55].

In accordance, the results showed biases in the sleep parameters provided by the Oura ring. However, the mean differences were within the satisfactory range, and only a few samples were outside the agreement limits. Therefore, the Oura ring can be acceptable for monitoring nonstaging sleep parameters under free-living conditions.

Moreover, our results indicated that the mean differences of the TST, WASO, and SE between the Samsung watch and the actigraphy were higher than the Oura ring's mean difference. The TST and SE mean differences of the watch were higher but still within the satisfactory range. However, the WASO mean difference (i.e., 31.27 min) was negligibly higher than the range. Within the 134 valid total nights of sleep detection by the watch, 9 TST, 9 WASO, and 8 SE fell outside the agreement limit. Similarly, the correlation of the watch and actigraphy was lower than the ring, as the Pearson r values of the 3 parameters were closer to 0. Consequently, the sleep parameters of the watch had acceptable mean differences and indicated significant correlations with the actigraphy, but the Oura ring outperforms the Samsung watch in terms of the nonstaging sleep parameters.

Comparison With Prior Work

In previous studies, wrist activity trackers such as Fitbit Charge HR and Jawbone UP were compared with the PSG in lab tests on healthy adults [42, 41, 127]. The devices showed good agreement with the PSG in terms of TST, WASO, and SE. This is in accordance with our results for both the Oura ring and the Samsung watch. However, the overestimations or underestimations in our findings were higher than those in previous studies. The biases are particularly significant for the Samsung watch. For example, de Zambotti et al. [42] indicated that the Fitbit Charge HR overestimates TST by 8 min and SE by 1.8% and underestimates WASO by 5.6 min. These low biases might be because of their limited setups and data collection, that is, an overnight laboratory sleep test on 32 healthy individuals.

There are a few studies performed under free-living conditions to evaluate activity trackers such as the Misfit Shine, Jawbone UP, and different models of Fitbit on healthy adults [48, 82]. Our results regarding the Oura ring highlighted the high correlations obtained by these studies. For instance, Liang et al. [82] indicated that there were high Pearson correlations between Fitbit Charge 2 and their baseline (a single-channel electroencephalogram-based

device) in terms of TST ($r = 0.94$), WASO ($r = 0.25$), and SE ($r = 0.50$). Ferguson et al. [48] considered the TST correlations between 4 activity tracker devices and a research-grade accelerometer/multi-sensor device (BodyMedia SenseWear). The authors showed that the correlations were higher than 0.82 for the devices. On the other hand, our smartwatch results showed moderate correlations for TST, WASO, and SE.

Furthermore, we considered gender as a dependent variable to evaluate whether there was a mean difference in sleep parameter changes between male and female groups. Considering the Oura ring, our results showed a nonsignificant difference between female and male groups in TST, which is similar to the findings of de Zambotti et al. [41]. Moreover, Carter et al. [35] evaluated the objective estimation of sleep parameters compared with subjective assessments. In comparison with this study, we obtained similar results in terms of objective TST. However, the watch in our study showed a significant difference in TST. Besides, both the ring and the watch indicated significant differences between female and male groups in WASO and SE ($P < .05$), which disagrees with de Zambotti et al. [41] but confirms the findings of Carter et al. [35].

Limitations

We considered using an actigraphy device as the baseline method, which is one of the limitations of this study. Our analysis was limited to TST, WASO, and SE parameters. Although we collected the SOL of the Oura ring and the Samsung watch, we could not evaluate the values, as the SOL measure of the actigraphy is unreliable [124]. The actigraphy methods are insufficient for evaluation of sleep stages (e.g., deep sleep). Therefore, future work should investigate the sleep stages provided by the ring and watch, considering a feasible PSG or electroencephalogram-based method designed for home-based monitoring.

Another limitation of this study is that only healthy participants were included in the anal-

ysis. However, other studies have shown that the accuracy of the wearables might differ for different population groups [38, 37]. This issue may limit the generalizability of the findings. This study's future directions are to perform a home-based sleep validation study to assess the accuracy of wearables for population groups of different ages (e.g., adolescents and older people) and sleep disorders (e.g., obstructive sleep apnea). Besides, bed-based and ballistocardiograph-based sensors [118] can be used to mitigate user errors during data collection.

2.3 Pregnant women's sleep patterns before and during the COVID-19 pandemic

2.3.1 Background

In another study [105], we examined the sleep patterns of pregnant mothers in Finland during the COVID-19 pandemic. It is possible to continuously monitor and track an individual's sleep parameter data during pregnancy thanks to modern verified technologies [105]. It is crucial to figure out how pregnant women sleep on a daily basis since these patterns could be signs of how well they are doing during their pregnancy [75]. In light of the disruption to normal daily life patterns caused by the COVID-19 pandemic, health care organizations and clinical practicians have embraced the use of wearable devices and mobile applications to collect personal health parameters. The goal of this study was to examine everyday patterns of well-being (such as sleep) in pregnant women before and during Finland's COVID-19 pandemic-related stay-at-home restrictions.

2.3.2 Data collection

This study included two cohorts of pregnant women. Pregnant women with histories of preterm births (gestational weeks 22–36) or late miscarriages (gestational weeks 12–21) were recruited in the first wave between January and December 2019, and pregnant women with histories of full-term births (gestational weeks 37–42) and no pregnancy losses were recruited in the second wave between October 2019 and March 2020. This study's final sample size was 38 pregnant women [105]. Each participant received a Samsung Gear Sport smartwatch, which has demonstrated acceptable validity in everyday settings when it comes to sleep [92] (Section 2.2).

2.3.3 Data preprocessing

TST and WASO were calculated as the sleep parameters for each night using the pipeline introduced in Section 2.2. TST is the total amount of time the subject slept during the night. WASO is the amount of time spent awake after falling asleep but before waking up. We visualized various sources of data, including the steps and hand movements reported by the watch, to validate sleep intervals for such sleep events. We were able to retrieve actual sleep intervals by manually inspecting the data using such visualizations.

2.3.4 Statistical analysis

To examine trends in between-person and within-person changes in the dependent variable of interest, hierarchical linear mixed models were used. We used the notation defined by Raudenbush-Bryk, as well as Bolger-Laurenceau's recommendations [31, 112]. We examined daily values for four weeks before and four weeks during the COVID-19 pandemic (56 days), with 38 subjects in each interval. Time (day) was the single within-subject independent variable in the model, and the pandemic group (before or during the pandemic) and education level were the between-subject independent binary variables (university level or lower level education). Furthermore, to detect potential differences between the groups, the study group (high-risk or low-risk pregnancy) was used as a between-subject independent variable in the models.

To account for the outliers in the sleep measurements, we transformed the values to z-score and kept measurements within 3 (3σ 99.73% of the data). Furthermore, to test subjective data before and after the pandemic we used the paired t-test. A zero mean difference between the variables of interest was the null hypothesis. We used the complementary cumulative distribution function (CCDF) plot to show the difference between before and during the pandemic.

2.3.5 Results

The TST of pregnant women was not associated with pandemic-related restrictions (P = .266). TST, on the other hand, decreased as the pregnancy progressed (P = .021). The duration of WASO was not significantly associated with pandemic-related restrictions (P = .065).

Compared to the weeks preceding the virus outbreak, participants awoke 15 minutes later on average during the pandemic-related restrictions (P = .007). They fell asleep about 10 minutes later; however, the difference was not statistically significant (P = .0504). The participants' sleep quality did not change based on their weekly evaluations (Figs. 2.7 and 2.8).

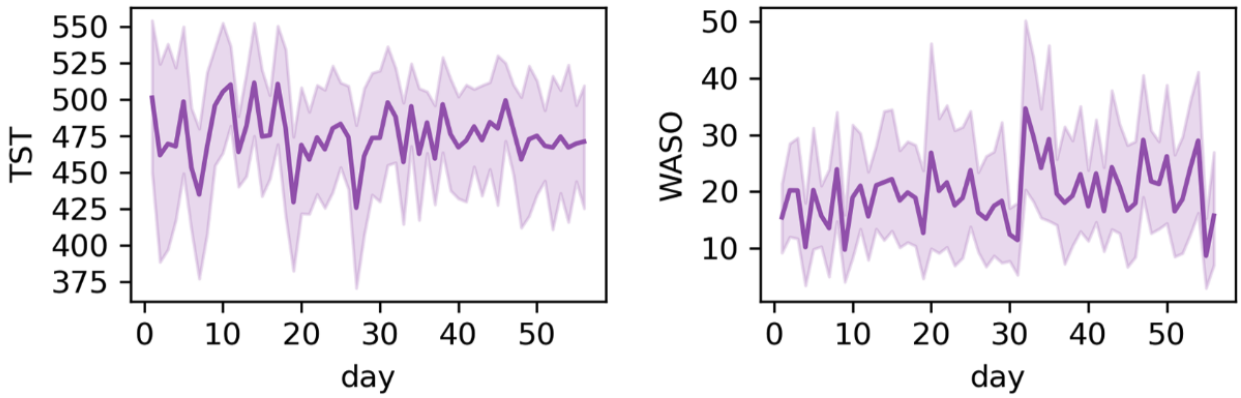


Figure 2.7: Trends in sleep. The daily mean with 95% confidence intervals of daily total sleep time (TST) and wake after sleep onset (WASO) (n = 22–32) during the eight-week data collection period (56 days). These figures are obtained from [105].

2.4 Conclusions

Sleep monitoring in free-living conditions becomes feasible and practicable using commercial devices such as smart rings and smartwatches. Notwithstanding the advances and feasibility of these wearables, their validity in terms of sleep parameters was not thoroughly investi-

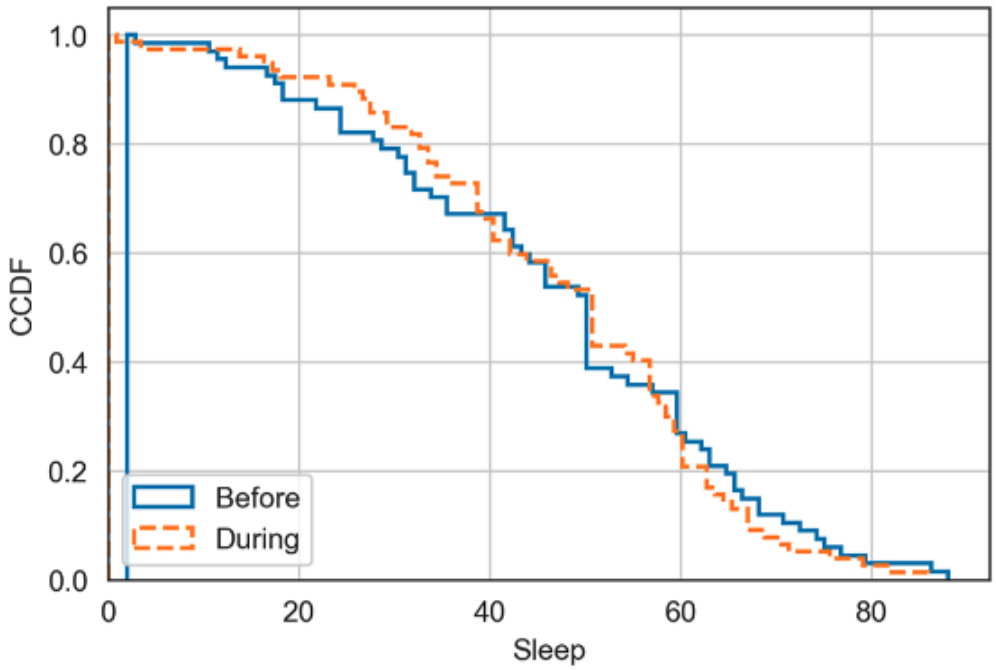


Figure 2.8: Subjective evaluations of sleep. Participants' (n = 23) subjectively assessed quality of sleep before and during the pandemic-related restrictions in a scale from 0 to 100 the higher value indicating a better quality of sleep. These figures are obtained from [105].

gated, especially for mid- to long-term studies in everyday settings. This study assessed the Oura ring and the Samsung Gear Sport watch by examining their TST, WASO, and SE under free-living conditions. The wearable devices were tested in home-based monitoring, where the sleep parameters of 45 healthy participants were tracked for 7 days. The assessment was performed in comparison with an actigraphy device, leveraging the paired t tests, Bland-Altman plots, and Pearson correlations. Sleep parameters were investigated considering the gender of the participants as a dependent variable. Our results showed that despite the statistically significant differences in the sleep parameters (i.e., TST, WASO, and SE) of both the Oura ring and the Samsung watch compared with the actigraphy device, the mean differences were within the satisfactory ranges. The sleep parameters also indicated significant correlations with actigraphy. Besides, we showed that there was no significant difference in the validation of TST between male and female groups in the Oura ring; however, both the Oura ring and the Samsung watch indicated significant differences between

the female and male groups in the estimation of WASO and SE.

Similarly, in a population sample of healthy adults, both the Oura ring and the Samsung watch had acceptable mean differences and indicated significant correlations with the actigraphy. However, the biases of the ring were considerably lower than the biases of the watch. Further validation is required to assess the validity of the sleep stages provided by the ring and the watch under free-living conditions. Moreover, future work should include the assessment of the devices for other population groups, such as individuals with sleep disorders.

Furthermore, by means of using these validated devices, we found that the COVID-19 pandemic did not appear to have a significant impact on the daily routines of Finnish pregnant women. TST decreased as the pregnancy progressed, but the pandemic-related restrictions were not associated with sleep. However, the participants' daily rhythms changed as they began to sleep later and wake up later. Working from home made it their sleep pattern more flexible since the women in the study were highly educated [100, 105].

Chapter 3

Mental health

The entire world is currently being significantly affected by a global virus pandemic. The first case of this virus, SARS-CoV-2, was reported in China in December 2019, and the first case outside China was discovered in January 2020 [93, 1]. In February, the World Health Organization named the disease caused by this virus COVID-19 [57].

Worldwide, as of July 19, 2020, there had been approximately 14,400,000 confirmed cases of COVID-19, with 604,000 deaths [4]. The United States of America, with 3,830,000 confirmed cases and 143,000 deaths, was the most affected country in the world. In some states, such as California, the numbers are still increasing, while in some other states, such as New York, the peak has passed and the average number of daily new cases is decreasing [93].

Due to the rapid spread of SARS-CoV-2, finding effective reasons for its spread can play a significant role in prevention policies. Using data mining and time series analysis methods, it is possible to investigate the impact of different phenomena on time series data [93]. For example, in economics, different studies have modeled the temporal relationships of two or more time series (e.g., the relationship between oil and gold prices) using these methods [93, 3]. Wang et al. [140] used the same causality inference methods to determine whether

a relationship exists between the main air pollutants and the mortality rate of respiratory diseases.

It is well-known that the widespread pandemic has had psychological impact on our entire society. Various population groups such as child/adolescents and elderly may experience different mental health disorder trends and it is essential to focus on more vulnerable populations (e.g., pregnant women) [130].

In this chapter, we investigate the effect of Covid-19 on mental health disorders.

3.1 The Effect of COVID-19 and Lockdown on Mental Health in California

3.1.1 Background

There have been more than 9 million cases, equating to over 90,000 deaths, of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in California since December of 2019 when the virus was detected. In an effort to prevent viral transmission, California was the first state to strictly implement a statewide lockdown and protective public health measures. It is well-documented that the the pandemic itself, coupled with social isolation, economic concerns, grief, fear and the loss of loved ones, have contributed to an increase in mental health problems. However, it is unclear if the pandemic and the public health measures used to mitigate its spread affected the rate of new diagnoses of mental health disorders such as anxiety, bipolar disorder, post traumatic stress disorder, depression, self-harm, suicide, psychosis among others. While the public health measures were designed to protect the publicly from physical illness of SARS-Co-V-2, it likely had unintended consequences for mental health. It is well-documented the environment can be protective or harmful to

one's mental health. While mental health problems are prevalent across the lifespan, there is well-known that stage of develop and environment interaction and influence the chance for psychiatric disease. In other words, the public health measures likely did not affect age groups uniformly.

The purpose of this study was to examine the influence of COVID-19 pandemic on mental health disorders across the lifespan. This retrospective review subjected data from the University of California COVID Research DataSet (UC-CORDS) collected before and after the pandemic to determine the frequency of mental health issues across the lifespan.

3.1.2 Method

Sample. Data were extracted from the UC-CORDS data set at two time periods: prior to (March-October of 2019) and during the pandemic (March-October 2020). Data were examined to between the two time periods to determine any difference in new diagnoses of mental health problems across the lifespan. This data set contains comprehensive, structured information on patients admitted to the hospitals at the University of California Health's five academic health centers (i.e., UC Davis Health, UC San Diego Health, UC Irvine Health, UCLA Health, and UCSF Health). This data set provides a wide range of information, including COVID-19 test results, and weekly values of new mental health disorders diagnosed by UC-Health hospitals. It does not necessarily contain unique subjects since one subject that developed anxiety and bipolar is recorded twice – one for anxiety and one for bipolar. Weekly values of new diagnoses of mental health disorders were extracted from UC-CORDS and further classified based on age (less than 19 (< 19), between 19 and 25 (19-25), 26-35, 36-45, 46-55, 56-65, 66-75, 76-85 and above 85 (> 85) years old).

The UC-CORDS data set contains different diagnoses coded using SNOMED vocabulary [8]. For analysis purposes, we have categorized various mental health disorders into 12 bins

(Table A.1). *Bipolar*, *post-traumatic stress disorder (PTSD)*, *depression*, *mood*, *psychosis*, *self-harm*, *suicide*, *personality*, and *anxiety* are straightforward. *Child/adolescent behavioral* contains disorders related to adolescents, such as hyperactivity disorder. *Substance-induced mental health* consist of the disorders developed due to the use of substances of abuse or medications (e.g., delusional disorders); however, the disorders in usage are categorized as *Substance use*.

Analysis. Statistical analyses were performed using R (version 4.0.3). Paired t-tests were performed to test differences in the occurrence of each mental health disorder for each age group. A two way ANOVA was performed to assess for between group differences. A P value of < 0.05 was considered significant.

3.1.3 Results

The occurrence of new diagnoses of mental health disorders for 2019 and 2020 based on age are reported in Table 3.1. Changes in the occurrence of mental health diagnoses were observed for those $<$ 19, 19-25, 26-35, 56-65, and 76-85 years old. The 26-35 age group had the largest increase in new diagnoses, See Fig. 3.2a.

In addition, the two-way ANOVA test for 2019 and 2020 showed the significant effect of different age groups over time ($P < .001$ and $P = .004$ respectively), justifying different age groups with different intercepts and slopes in both 2019 and 2020.

Due to the nature of the UCCORDS data set, each subject could be diagnosed with more than one type of mental health disorder. We performed a Pearson correlation to compare the occurrence of weekly report with respect to the disorder category for 2020. Fig. 3.1 shows the heatmap of such correlations.

Since the most significant changes happened for age groups 26-35, the rest of the analysis

Table 3.1: Reported values of each disorder per age group in 2019 and 2020.

Age	Year	Anxiety	Bipolar	Child/adolescent behavioral	Depression	Mood	PTSD	Personality	Psychosis	Self-harm	Substance use	Substance-induced	Suicide	Total	P-val
<19	2019	352	10	308	169	44	10	1	11	10	6	5	22	948	.002
	2020	458	27	272	252	62	38	3	9	31	4	10	46	1212	
19-25	2019	847	94	178	595	125	53	14	92	20	10	31	58	2117	$< .001$
	2020	1093	139	170	694	132	65	21	140	35	17	65	74	2645	
26-35	2019	2035	178	278	1283	224	104	20	225	26	23	141	82	4619	$< .001$
	2020	2585	294	287	1533	316	167	26	349	42	55	233	138	6025	
36-45	2019	2433	286	168	1603	268	127	23	255	22	31	193	103	5512	.32
	2020	2526	285	173	1481	330	167	23	292	43	44	245	103	5712	
46-55	2019	2339	320	116	1771	278	159	16	277	16	34	234	101	5661	.40
	2020	2273	256	112	1547	323	161	18	338	23	53	252	101	5457	
56-65	2019	2786	374	115	2443	385	167	15	355	29	45	251	102	7067	.006
	2020	2618	301	104	1967	402	118	21	389	20	53	263	114	6370	
66-75	2019	2363	225	71	2509	301	114	6	198	10	17	105	56	5975	.17
	2020	2307	228	59	2024	397	89	9	260	21	11	88	96	5589	
76-85	2019	1058	88	29	1206	169	35	1	76	4	3	12	11	2692	.036
	2020	911	36	33	1028	177	25	1	118	3	4	11	23	2370	
>85	2019	311	10	12	337	57	9	2	26	0	0	1	6	771	.18
	2020	262	11	8	271	53	1	1	57	0	0	3	8	675	

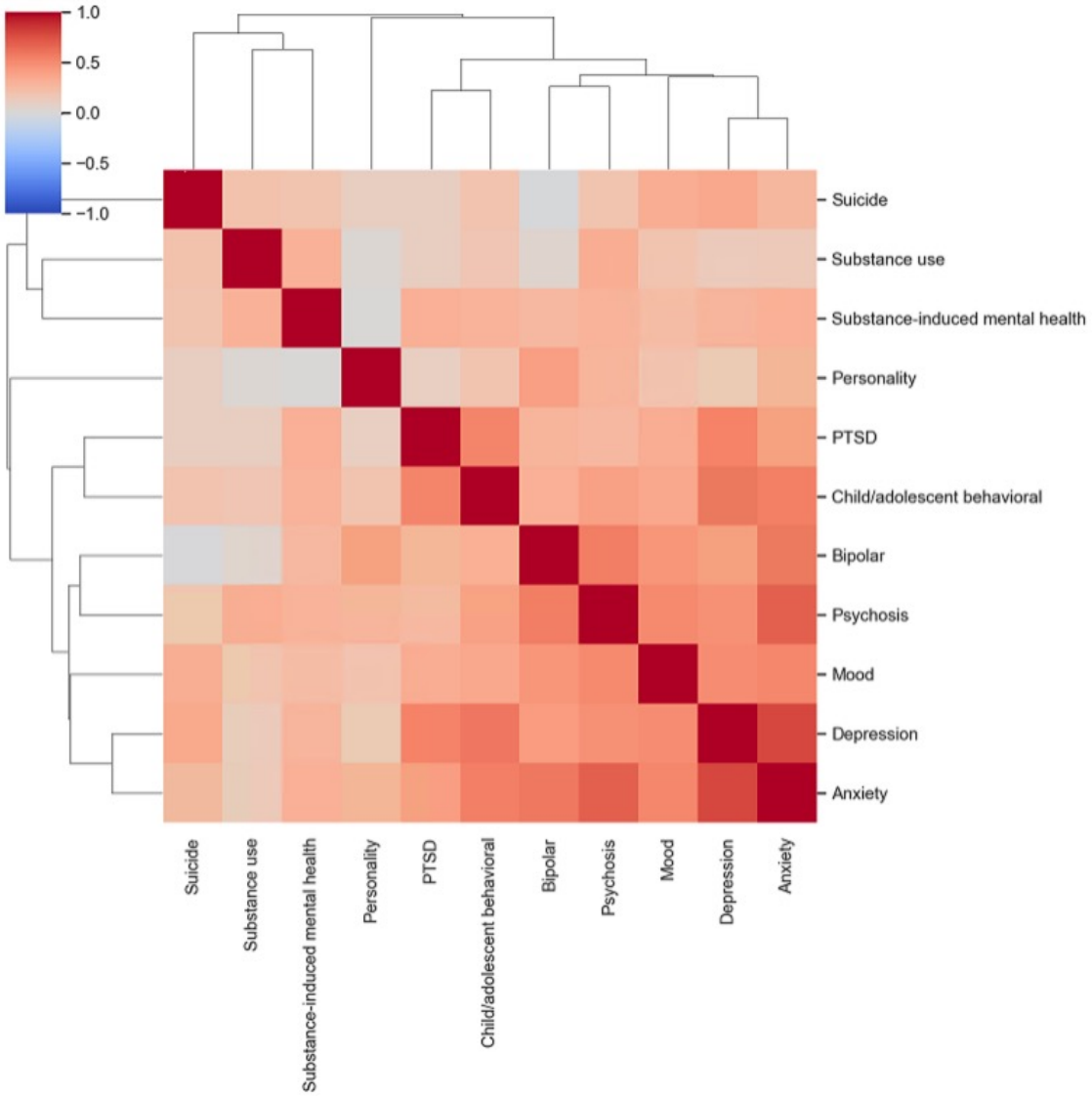
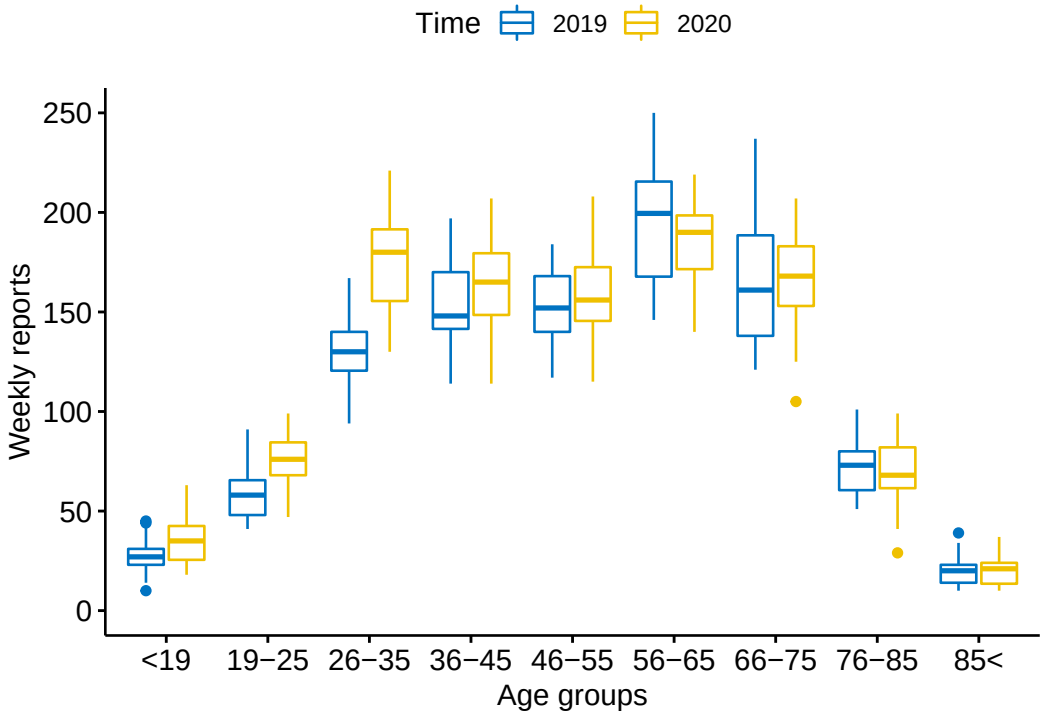


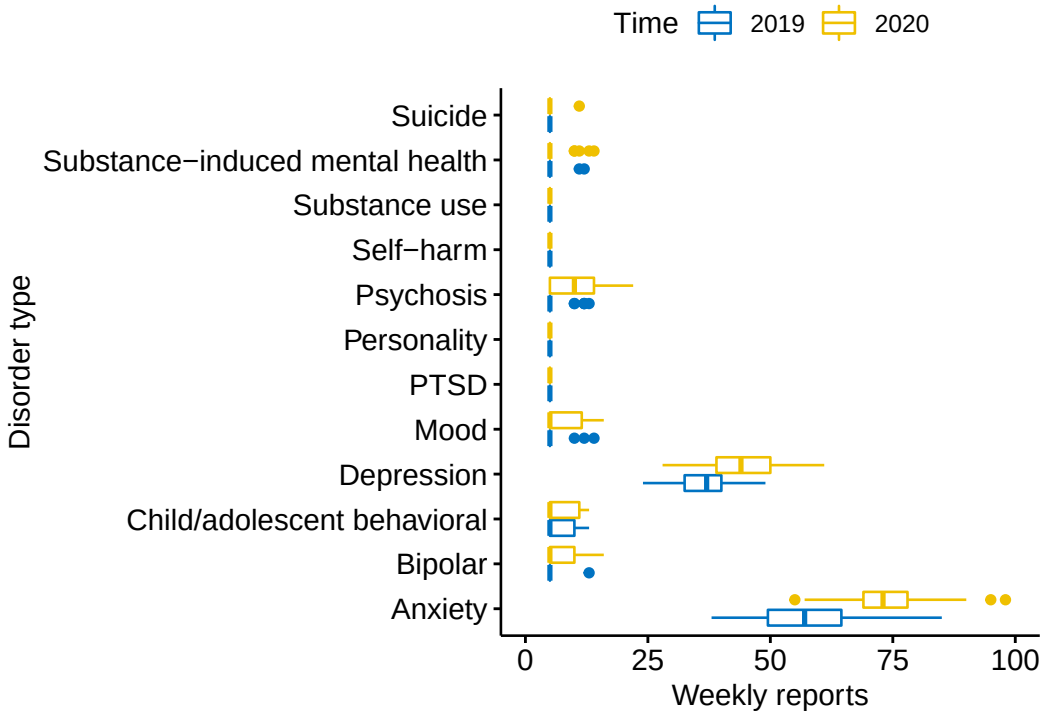
Figure 3.1: Pearson correlation of mental health disorders in 2020.

focuses on this group. Similarly, a paired t-test has been performed for each disorder group. Anxiety ($P < .001$), bipolar ($P = .02$), depression ($P < .001$), mood ($P = .03$), psychosis ($P = .002$) had significant changes in 2020 compared to 2019 (Fig.3.2b).

Likewise, to see if all diagnoses behave the same as the time changes (longitudinally), we have used a two-way ANOVA test for 2019 and 2020 separately. For 2019 and 2020, all the P-values are significant, meaning different diagnoses will have different intercepts and slopes for 2019 and 2020 ($P < .001$).



(a)



(b)

Figure 3.2: Distribution of total mental health disorder reports per age group (3.2a) and per each disorder for 26-35 years old population (3.2b).

3.1.4 Discussion

Herein we report changes in clinician diagnosed mental health disorders across age groups. Our findings are consistent with others who report an increase in mental health issues during the COVID-19 pandemic, especially anxiety and depression [113, 138, 39]. Specifically, we report an increase in anxiety, depression, bipolar disorder, mood disorder, and psychosis among individuals aged 26-35 during the lockdown phase of the COVID-19 pandemic in California.

The strength of our observations includes clinician diagnosis that is recorded in the electronic health record and as such did not rely on self-report measures of symptoms. We also observed a reduction in many mental health disorders, especially among adults older than 55. The reason for this is unclear, but may be due to not seeking health care for fear of becoming infected with SARS-CoV-2, as it was widely publicized that advancing age and the presence of cardiovascular commodities increased rick of serious illness. California was the first to initiate a lockdown and as such would be the first to experience any sequelae from those safety measures–in short, those living in California would not have the benefit of learning from another state about the impact of a lockdown.

Those ages 26-35 demonstrated the greatest increase in seeking care for a mental health disorder during the lockdown phase of the pandemic. There may be several explanations for this, including caring for and financial supporting a younger family, as well as stress and worry about loss of older family members such as parents and grandparents. Financial issues brought on by inability to work, change in employment status or cut backs from corporations or personal businesses may have contributed to the increase in anxiety and depression. The mental health disorders may have been further propelled by the increase in substance use observed in this age group, including an increase in self-harm behaviors and suicide.

Although our study utilized clinician diagnosis of mental health disorders there are limi-

tations associated with the use of electronic health records. Primarily that all diagnoses were recorded and captured accurately, and that the data reported in the record only reflect that of individuals who have sought care. Furthermore, the UC-CORDs data set consists of record solely from academic health science centers which may not reflect trends in smaller communities or in areas where there are no academic health science centers.

Our observations demonstrate the impact of the lockdown and COVID-19 pandemic on mental health disorders in California. The use of public health measures, including the lock down, and the availability of vaccines has resulted in California having a COVID-19 death rate of 228 per 100,000, making it one of the lowest in the United States. However, we are just beginning to understand the impact of the totality of the pandemic and the measures instituted to stop the spread on mental and physical health. More studies are desperately needed to understand the factors that contribute to these issues so that interventions and policies can be instituted when another global pandemic emerges.

Chapter 4

Cardiovascular Biomarkers Monitoring

Continuous monitoring of blood pressure (BP) can help individuals manage their chronic diseases such as hypertension, requiring non-invasive measurement methods in free-living conditions. Recent approaches fuse Photoplethysmograph (PPG) and electrocardiographic (ECG) signals using different machine and deep learning approaches to non-invasively estimate BP; however, they fail to reconstruct the complete signal, leading to less accurate models. In the first part of this chapter, we propose a cycle generative adversarial network (CycleGAN) based approach to extract a BP signal known as ambulatory blood pressure (ABP) from a clean PPG signal. Our approach uses a cycle generative adversarial network that extends the GAN architecture for domain translation, and outperforms state-of-the-art approaches by up to 2× in BP estimation.

To further investigate the importance of BP and heart rate, we leverage the importance of such biomarkers in COVID-19 context. In the second part of this chapter, we show the power of BP and HR in COVID-19 detection. The world has been affected by COVID-19

coronavirus. At the time of this study, the number of infected people in the United States is the highest globally (31.2 million infections). Within the infected population, patients diagnosed with acute respiratory distress syndrome (ARDS) are in more life-threatening circumstances, resulting in severe respiratory system failure.

Various studies have investigated the infections to COVID-19 and ARDS by monitoring laboratory metrics and symptoms. Unfortunately, these methods are merely limited to clinical settings, and symptom-based methods are shown to be ineffective. In contrast, vital signs (e.g., heart rate) have been utilized to early-detect different respiratory diseases in ubiquitous health monitoring. We posit that such biomarkers are informative in identifying ARDS patients infected with COVID-19.

In this study, we investigate the behavior of COVID-19 on ARDS patients by utilizing simple vital signs. We analyze the long-term daily logs of blood pressure (BP) and heart rate (HR) associated with 150 ARDS patients admitted to five University of California academic health centers (containing 77,972 samples for each vital sign) to distinguish subjects with COVID-19 positive and negative test results. In addition to the statistical analysis, we develop a deep neural network model to extract features from the longitudinal data. Our deep learning model is able to achieve 0.81 area under the curve (AUC) to classify the vital signs of ARDS patients infected with COVID-19 versus other ARDS diagnosed patients. Since our proposed model uses only the BP and HR, it would be possible to review data prior to the first reported cases in the U.S. to validate the presence or absence of COVID-19 in our communities prior to January 2020. In addition, by utilizing wearable devices, and monitoring vital signs of subjects in everyday settings it is possible to early-detect COVID-19 without visiting a hospital or a care site.

4.1 Blood Pressure Waveform Reconstruction from PPG using Cycle Generative Adversarial Networks

4.1.1 Background

Blood pressure (BP) monitoring is critical for early detection of cardiovascular and respiratory diseases [46, 22]. High blood pressure can be the source of mortality and morbidity for the aging population [32]. Hence, continuous monitoring of BP – via the bio-markers systolic (SBP) and diastolic blood pressure (DBP) – can help diagnose chronic severe conditions. Currently the established method for measuring SBP and DBP (usually in mmHg) is to use a medical-grade cuff-based instrument, which is neither comfortable nor feasible for continuous BP monitoring in everyday settings.

Thanks to recent advances in Internet-of-Things (IoT), it is now possible to record vital signs in everyday settings. For instance, smartwatches can record heart rate and blood oxygen using photoplethysmograph (PPG) and electrocardiogram (ECG) sensors. PPG sensors emit and reflect light into/from blood vessels, whose measurement form a signal that is proportional to continuous blood volume in the unit of time. This signal has been shown to correlate strongly with BP [80] as shown in Fig. 4.1.

State-of-the-art approaches have investigated different methods ranging from feature-based statistical estimations [62, 134, 80] to deep learning-based [132, 133] approaches to estimate BP (i.e., SBP and DBP). The former approaches are limited to a pre-defined set of features that can diminish the information of the entire input signal. Furthermore, this needs an expert to define such features. However, the second approach utilizes convolutional layers to embed the segmented signal and absorb more information into a training model. Although these approaches accurately estimate SBP and DBP values, they fail to reconstruct the entire ambulatory blood pressure (ABP) signal. It has been shown that the waveform itself

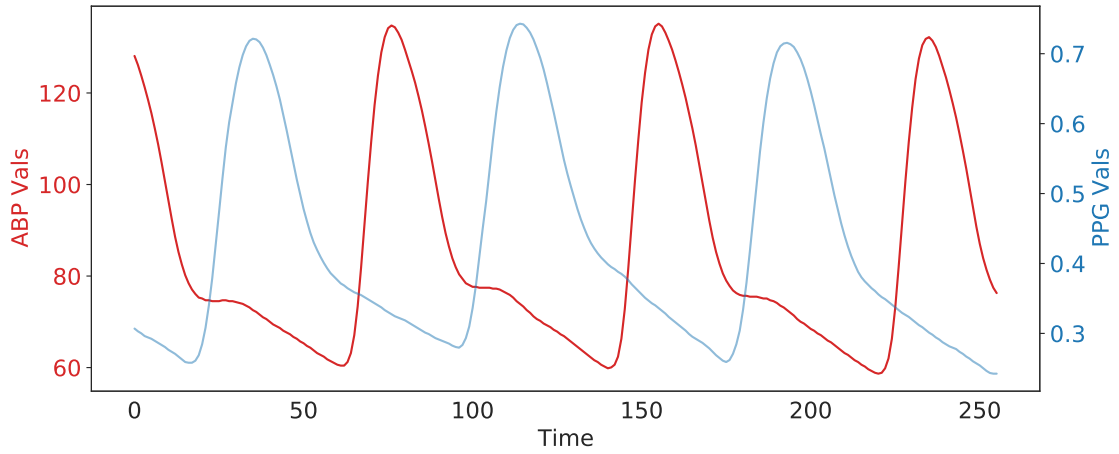


Figure 4.1: Correlation between BP and PPG signal.

contains a rich set of information on the underlying causes of cardiovascular diseases [76, 139]. Moreover, to measure other cardiovascular features, such as cardiac volume, there is a need for the ABP waveform [104].

In this section, we propose a deep learning method based on cycle generative adversarial network (CycleGAN) [147] to reconstruct the entire ABP waveform using a PPG signal. CycleGAN has been widely used in unsupervised learning for domain transformation. We train and test our model on Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC) II online waveform database [53] using 5-fold cross-validation with more than 90 subjects. In addition, since state-of-the-art approaches train and test on the same subjects' data, we perform a similar mechanism to compare our proposed model against the related work in the literature. On cross-subject evaluation, our proposed model achieves prediction error ($MAE \pm \sigma$) of 2.89 ± 4.52 mmHg and 3.22 ± 4.67 mmHg for SBP and DBP, respectively. Furthermore, per-subject evaluation outperforms the results of the state-of-the-art methods (2.29 ± 0.88 mmHg for SBP and 1.93 ± 2.61 mmHg for DBP). In summary, our CycleGAN method outperforms the state-of-the-art approaches for ABP waveform construction, as well as SBP and DBP estimations improving the estimation accuracy by up to $2\times$.

4.1.2 Related Work

Blood pressure estimation has been investigated in prior work [62, 132, 116, 80]. The majority of the related works attempt to extract SBP and DBP using ECG and/or PPG signals [133, 132]. Recent methods mainly rely on machine learning and deep learning by extracting features from the input signals [103]. The limitation of such studies would be the the under-utilization of the information in the signal. On the other hand, thanks to the progress in deep learning methods, researchers have focused on building deep neural networks that can generate embeddings of a given signal [133] to address the limitation of feature engineering.

On the other hand, synthesizing the entire ABP signal instead of extracting numerical SBP and DBP values can lead to a valuable source of information. There exists a few studies on BP waveform reconstruction. The state-of-the-art uses statistical methods as well as a wavelet neural network to reconstruct the ABP signal [80, 76]. While these studies minimally fulfill the standards, we show the estimation accuracy can be significantly enhanced (up to 2×) by proposing a CycleGAN-based model that reconstructs the ABP signals.

4.1.3 Material and Methods

4.1.4 Dataset

We employed the MIMIC-II online waveform database [53], which contains different bio-signals of thousands of subjects hospitalized between 2001 and 2012. This dataset contains PPG and corresponding ABP signal with the sampling frequency of $f_s = 125Hz$. We use 5 minutes of recording for 92 randomly selected subjects to evaluate our model. For this dataset, we randomly select 75 subjects for training and 17 subjects for testing. We repeat this procedure using 5-fold cross-validation while keeping each subject's data in only one

fold. Furthermore, to compare our results with literature, we randomly select 20 subjects to train and test for each subject separate from others (the first 80% for training and the rest for testing).

Pre-processing

The obtained signals had minor noises; hence, we apply the traditional Fourier Transform (FFT) approach to eliminate unwanted information. We use a band-pass filter with cutoff frequencies of 0.1 and $8Hz$ to remove noises from the PPG signal. On the contrary, we utilize a low-pass filter with a cutoff frequency of $5Hz$ to clean the ABP signal. For both PPG and ABP, we normalize the values of the signals for each subject. Afterward, each signal is divided into windows of 256 samples with 25% overlap for the downstream learning task using CycleGAN.

Evaluation metrics

We utilize the mean absolute error (MAE), and root mean square error ($RMSE$) to evaluate the performance of ABP construction for both SBP and DBP. These metrics have been widely used in the literature and show the difference between predicted and the true value. MAE and $RMSE$ can be calculated as follows:

$$MAE_{S/DBP} = \frac{1}{N}\sum_{i=1}^{N}|T^i_{S/DBP} - P^i_{S/DBP}| \tag{4.1}$$

$$RMSE_{S/DBP} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T^i_{S/DBP} - P^i_{S/DBP})^2} \tag{4.2}$$

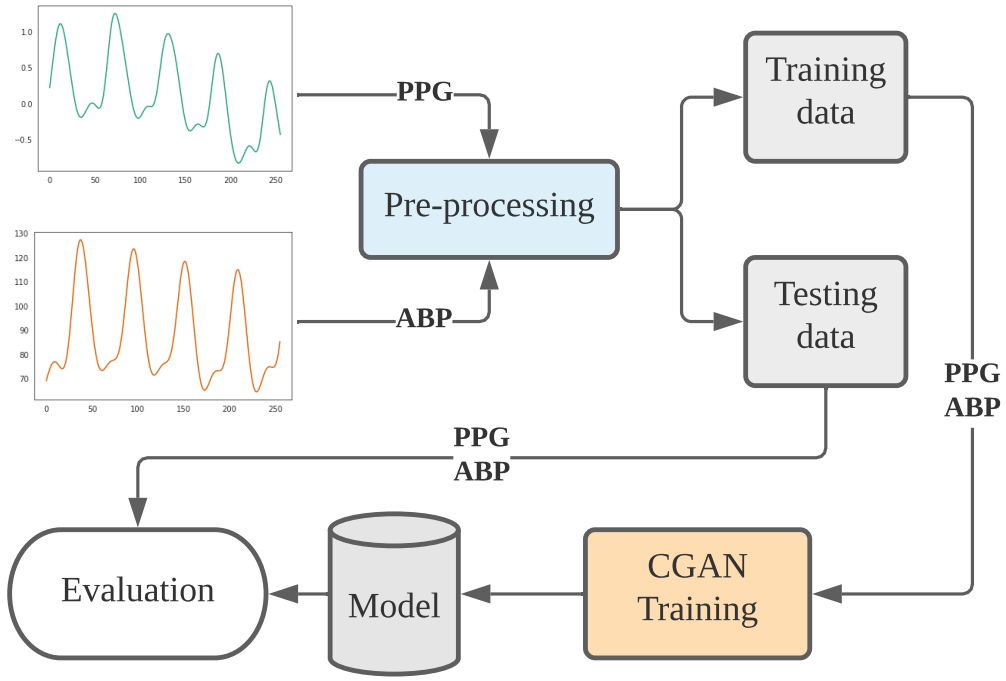


Figure 4.2: BP estimation pipeline.

As these formulas show, MAE and $RMSE$ are proportional to the averages of absolute and square differences between true (T) and predicted (P) values of all samples (N).

In addition, we compare our results with British Hypertension Society (BHS) [107] guidelines. This guideline divides the accuracy of blood pressure measurement into three groups based on the different ranges of estimations. Table 4.1 summarizes these ranges with their corresponding fraction of data.

Furthermore, Bland-Altman plots have been utilized to illustrate the agreement between the true and predicted values. The X-axis of this plot shows the mean, and the Y-axis represents the differences between the estimated and the true value and provides 95% confidence intervals.

Table 4.1: BHS standard ranges.

	Percentage Error		
	≤ 5 mmHg	≤ 10 mmHg	≤ 15 mmHg
Grade A	60%	85%	95%
Grade B	50%	75%	90%
Grade C	40%	65%	85%

PPG to ABP Translator (PAT)

In this work, we use the Cycle Generative Adversarial Networks (CycleGAN) to reconstruct ABP signals from raw PPG signals. Recent studies have already shown that the CycleGAN can be employed as one of the most powerful tools for signal-to-signal translation [19, 145].

The CycleGAN proposed by Jun-Yan Zhu et al. [148] is an extension of the GAN architecture. The GANs are composed of a generator network and a discriminator network. The generator network is trained to learn the real data distribution. It starts from a latent space as input and attempts to generate new data similar to the original domain. The discriminator network aims to take the generated data as an input and predict whether it is from the dataset (real) or the generated one (fake). After each epoch, the generator is updated to better fool the discriminator, while the discriminator is updated to accurately detect the generator's fake data.

The CycleGAN consists of two generators and two discriminator networks working in pairs. The idea behind the CycleGAN is to take data from the first domain as an input and generate data for the second domain as an output, and vice versa. In the PAT module, the goal of CycleGAN is to learn the mapping between PPG signals (domain X) and ABP signals (domain Y).

Each domain contains a set of training samples $\{x_i\}_{i=1}^{N} \in X$ and $\{y_i\}_{i=1}^{N} \in Y$ used directly from MIMIC-II dataset. There are two generators in this module with mapping functions as $G : X \rightarrow Y$ and $F : Y \rightarrow X$. The two discriminators are named D_X and D_Y. D_X aims to

distinguish between the real PPG signals (x_i) and the generated PPG signals ($F(y)$), while D_Y aims to discriminate between the real ABP signals (y_i) and the generated ABP signals ($G(x)$).

The adversarial losses [54] are used to match the distribution of the synthetic signals to the data distribution of the original signals. They are applied to both mapping functions ($G : X \rightarrow Y$ and $F : Y \rightarrow X$). The objective of the mapping function G as a generator and its discriminator D_Y is expressed as below: (We indicate the distributions of our data as $x \sim p_{data}(x)$ and $y \sim p_{data}(y)$.)

$$\begin{aligned} L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data}(y)}[log D_Y(y)] \\ + E_{x \sim p_{data}(x)}[log(1 - D_Y(G(x)))] \end{aligned} \tag{4.3}$$

where G attempts to generate ABP signals ($G(x)$) that look similar to original ABP collected from MIMIC-II dataset (domain Y), while D_Y aims to discriminate between generated ABP signals ($G(x)$) and real samples (y). Similarly, adversarial loss for the mapping function F is expressed as $L_{GAN}(F, D_X, Y, X)$.

The adversarial losses as the final objective loss function are not sufficient enough to guarantee that the learned functions can translate an individual input from the first domain into a desired output in the second domain. Therefore, cycle consistency losses are added to the final objective loss function. The cycle consistency losses guarantee the mapping from an individual input (x_i) to a desired output (y_i) by considering learned mapping functions to be cycle consistent. Cycle consistency means for each PPG signal x from domain X we must have $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ while for each ABP signals y we have $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$. The cycle consistency behavior is indicated as:

$$
\begin{aligned}
L_{cyc}(G, F) = E_{x \sim p_{data}(x)}[||F(G(x)) - x||_1] \\
+E_{y \sim p_{data}(y)}[||G(F(y)) - y||_1]
\end{aligned}
\tag{4.4}
$$

The final objective is the weighted sum of the above loss functions:

$$
\begin{aligned}
L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) \\
+L_{GAN}(F, D_X, Y, X) \\
+\lambda L_{cyc}(G, F)
\end{aligned}
\tag{4.5}
$$

where λ controls the relative importance of the two objective functions and is set to 10 in our work.

G aims to minimize the objective while an adversary D attempts to maximize it. Therefore, our model aims to solve:

$$
G^*, F^* = \arg \min_{G,F} \max_{D_X, D_Y} L(G, F, D_X, D_Y) \tag{4.6}
$$

The overview of CycleGAN is shown in Fig. 4.3 which is directly from [148].

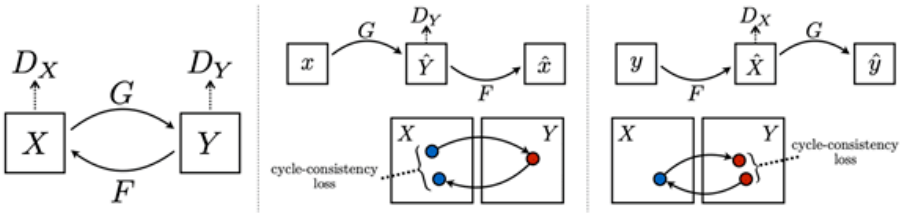


Figure 4.3: CycleGAN overview from [148].

We use the CycleGAN architecture proposed by [148] in our work. The architecture of generative networks contains two stride-2 convolutions, nine residual blocks [61], and two fractionally-strided convolutions with stride 0.5. This network is adopted from Johnson et al.

[68]: `c7s1-64,d128,d256,R256,R256,R256,R256,R256,R256,R256,R256,R256,u128,u64,c7s1-3`

Fig. 4.4 shows an overview of the generator network in CycleGAN.

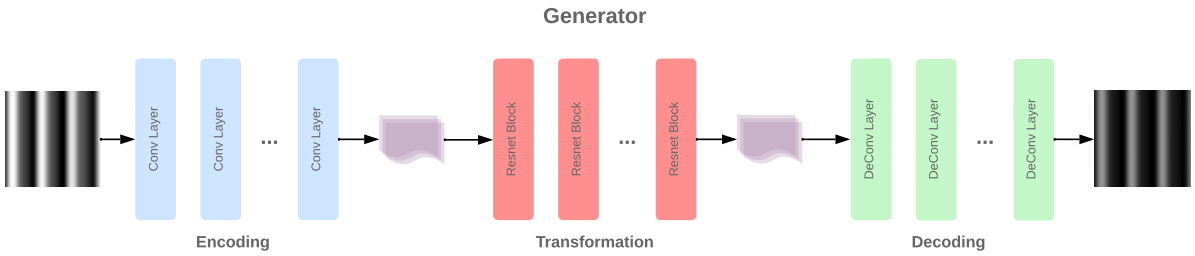


Figure 4.4: Generator network of CycleGAN.

The discriminator networks use 70×70 PathGANs [65] aiming to classify whether the signals are fake or real. The discriminator architecture is as follows: `C64-C128-C256-C512`

Fig. 4.5 shows an overview of the discriminator network in CycleGAN.

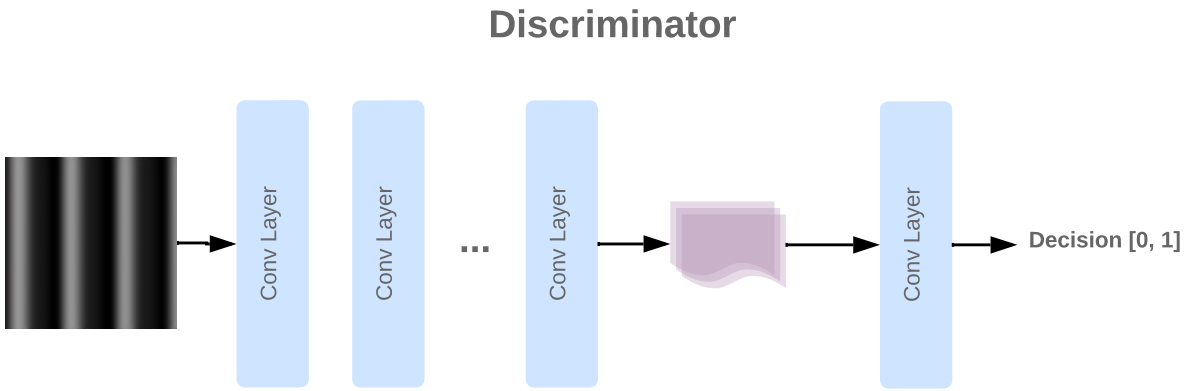


Figure 4.5: Discriminator network of CycleGAN.

To put everything together, Fig. 4.6 shows the entire system for PPG to ABP conversion.

This model aims to convert PPG to ABP and ABP to PPG at the same time. To be able to perform such a conversion, we transformed our signals to a 256×256 image but copying the entire signal window 256 times. Fig. 4.7 displays such images.

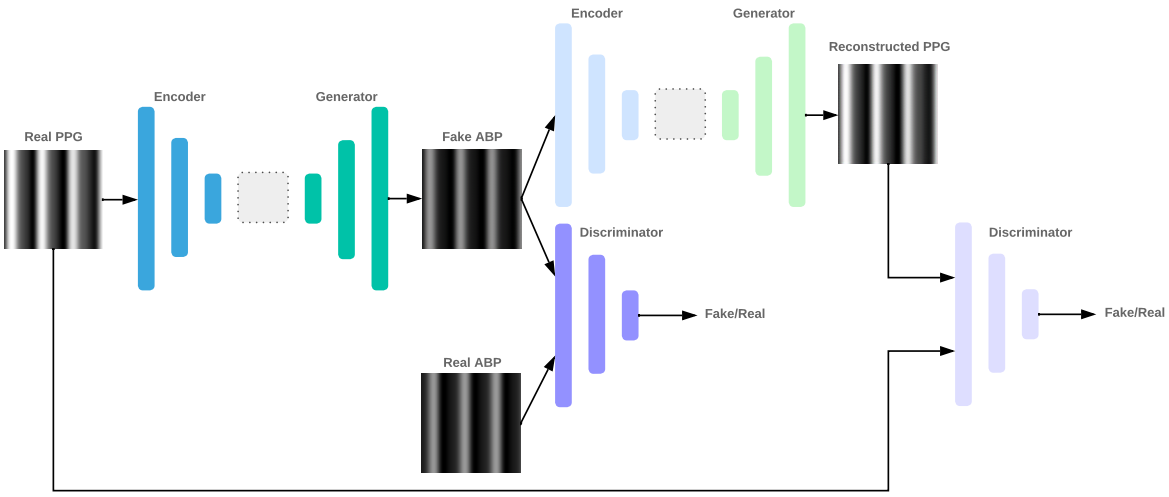


Figure 4.6: PPG to ABP overview using CycleGAN.

Table 4.2: Average performance of our proposed model.

	MAE	$RMSE$	μ	σ	r	P
SBP	2.89	5.18	0.67	4.52	0.97	$< .001$
DBP	3.22	4.82	1.78	4.67	0.94	$< .001$

4.1.5 Experimental Results

After transforming back the signal values to the original range (due to normalization), we extract SBP and DBP and compare them with the actual ABP signal. State-of-the-art approaches split each subject's data into train and test and build a separate model for each user. Hence, it is challenging to generalize their proposed models. Our study trains the entire CycleGAN on one set of subjects and tests it on new subjects given their corresponding PPG signal.

Contrary to the conventional methods, we perform 5-fold cross-validation on the entire data such that each subject will be categorized to be either in the training subset or the testing subset. Table 4.2 summarizes the results for the MIMIC-II. r shows Pearson correlation with the corresponding P-value. μ and σ represent the mean and standard deviation of the estimation error, respectively. Fig. 4.8 illustrates the prediction error distribution of SBP

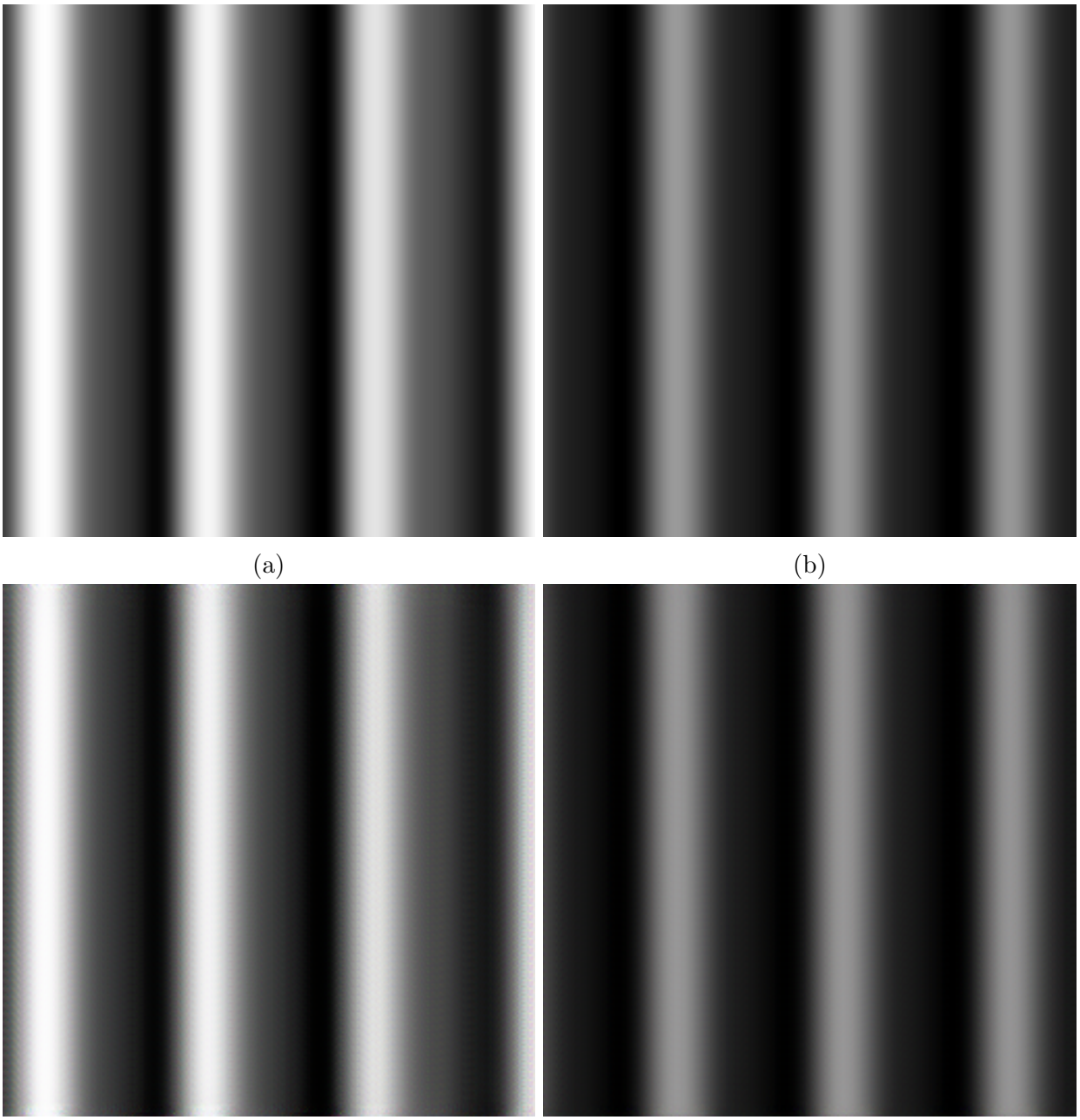
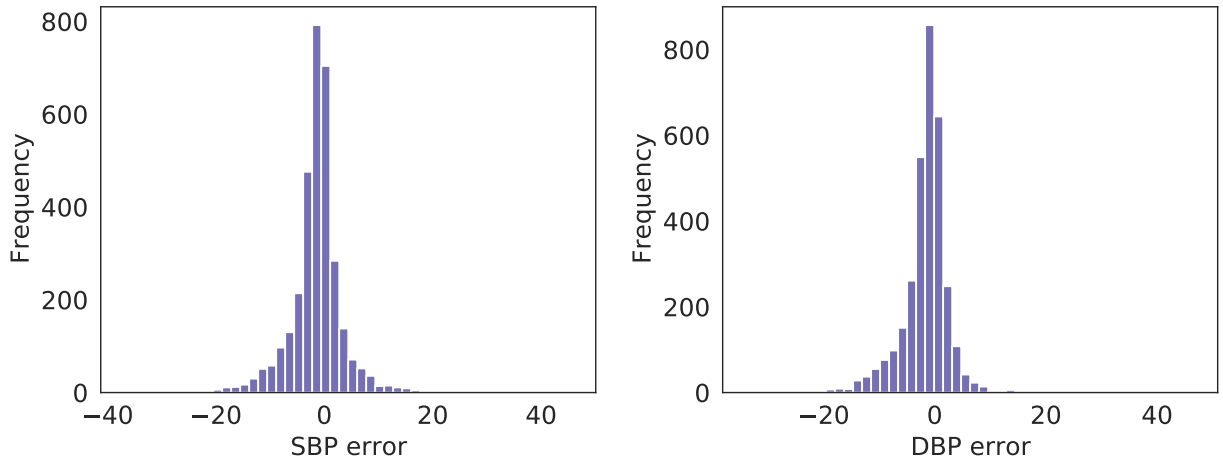


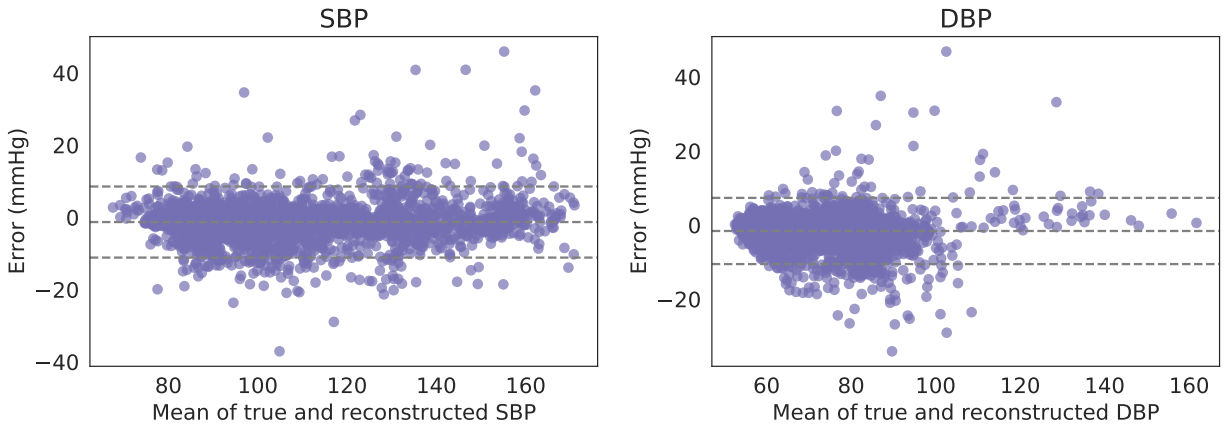
Figure 4.7: Transformed PPG signal (a) and ABP signal (b) to 256 by 256 images for a sample user and their corresponding reconstructed images using CycleGAN (c) and (d).

and DBP, respectively. Table 4.3 summarizes our results with regards to the BHS standard. Our method passes the *Grade A* requirements for all the criteria. In addition, to show how the error is distributed across different values of blood pressure, Bland-Altman plots have been utilized for both SBP and DBP (Fig. 4.9). In addition, the distribution of the true



(a) (b)

Figure 4.8: SBP (a) and DBP (b) prediction error.



(a) (b)

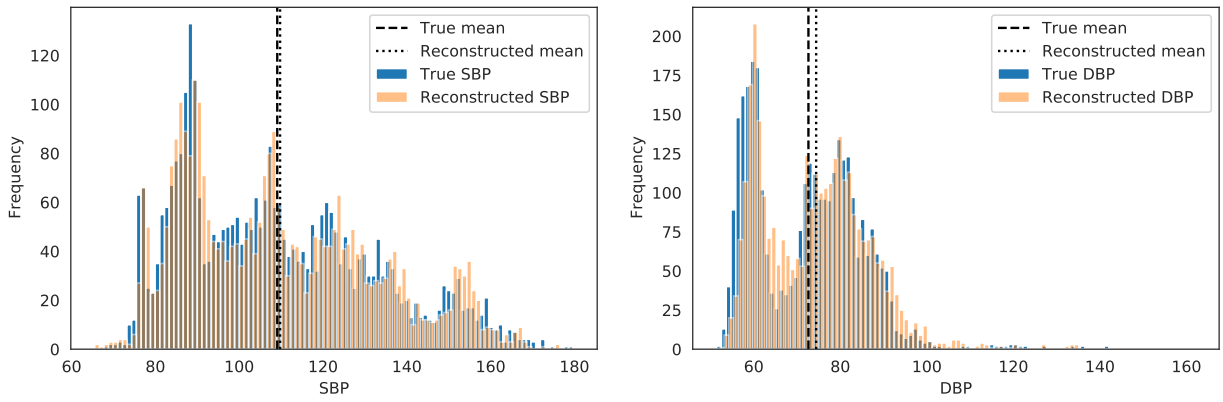
Figure 4.9: SBP (a) and DBP (b) Bland-Altman plots.

Table 4.3: Our results in regards to BHS standard ranges.

	Percentage Error		
	≤ 5 mmHg	≤ 10 mmHg	≤ 15 mmHg
SBP	85% (A)	95% (A)	98% (A)
DBP	81% (A)	94% (A)	98% (A)

values and the reconstructed values as well as their corresponding mean values is illustrated in Figs. 4.10a and 4.10b for SBP and DBP, respectively.

To be able to compare our results against the related work in the literature, we perform



(a) (b)

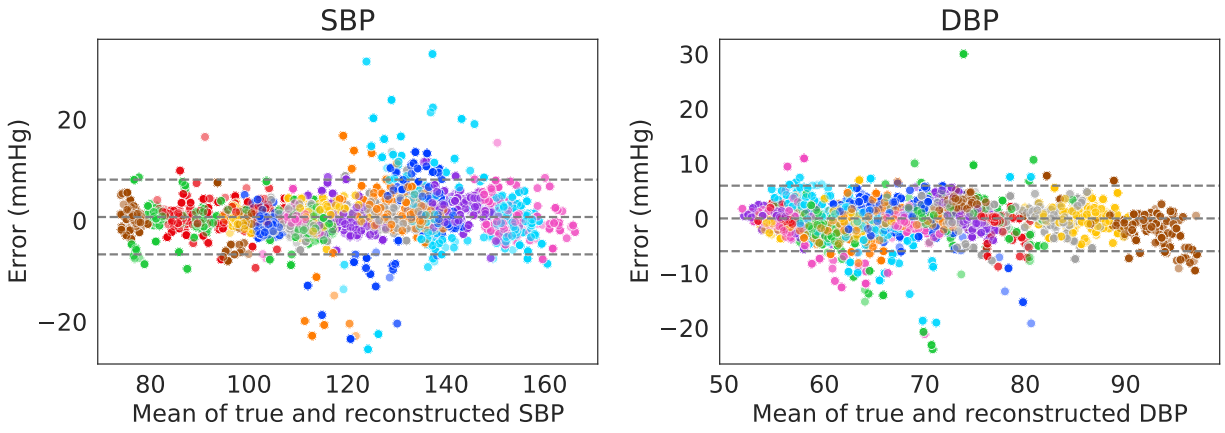
Figure 4.10: SBP (a) and DBP (b) distributions and their corresponding mean values.

Table 4.4: Comparison of our CycleGAN-based model performance with prior works.

		SBP				DBP			
		MAE	$RMSE$	μ	σ	MAE	$RMSE$	μ	σ
Waveform	**Our work**	**2.29**	**3.22**	0.88	2.99	**1.93**	**2.61**	0.91	2.32
	[76]	5.9	-	0.9	5	3.5	-	0.9	3.5
	[80]	-	-	2.32	2.91	-	-	1.92	2.47
SBP-DBP estimation	[64]	6.32	8.78	0.69	8.75	3.89	5.48	1.23	5.34
	[103]	3.97	8.9	0.050	7.99	2.43	4.18	0.187	3.37
	[133]	3.70	-	0.21	6.27	2.02	-	0.24	3.40
	[84]	3.42	5.42	0.06	4.19	2.21	3.29	0.18	2.65
	[116]	3.36	-	-	4.48	5.59	-	-	7.25
	[73]	3.8	-	-	3.46	2.21	-	-	2.09

per-subject train and test procedures for a fair comparison. For this purpose, we select 20 random subjects from the MIMIC-II database whose ABP and PPG signals are retained. Table 4.4 shows the results of our proposed model as well as the recent related work for per-subject evaluation. Our model outperforms the studies with waveform reconstruction as well as those with only SBP-DBP values estimation. Moreover, Fig. 4.11 presents the Bland-Altman plots for these subjects (the colors show individuals) within the agreement limits (95% confidence intervals).

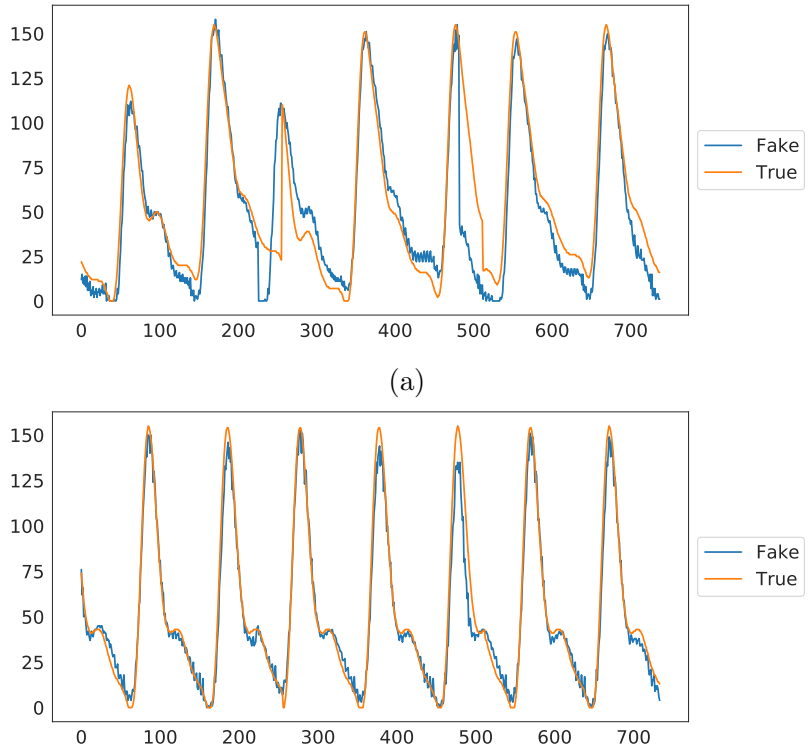
Furthermore, to show how the reconstruction performs for the whole signal, we randomly selected 2 subjects and plotted the estimated signal as well as the original one. These training



(a) (b)

Figure 4.11: SBP (a) and DBP (b) per-subject Bland-Altman plots.

are based on per-subject evaluation. Fig. 4.12 visualizes these two signals.



(a)

(b)

Figure 4.12: True and reconstructed (fake) ABP signal for 2 subjects.

4.2 Detection of COVID-19 Using Heart Rate and Blood Pressure: Lessons Learned from Patients with ARDS

4.2.1 Background

Acute respiratory distress syndrome (ARDS) is a potential life-threatening consequence of infection with SARS-CoV-2, the novel coronavirus that causes COVID-19 [81]. ARDS is characterized by an overwhelming immune response and non-cardiogenic pulmonary edema that compromise gas exchange, resulting in severe respiratory failure. ARDS mortality ranges from 40%-60%; however, it is unclear if the mortality rate is substantially higher if associated with COVID-19 infection, as it varies from 28.8%-62% [81, 131]. Currently, more than 136 million people worldwide have been infected with SARS-CoV-2 [5]. In the United States, 31.2 million people have been infected with over 562,000 deaths [5]. The impact of the COVID-19 pandemic is considerable and efforts to mitigate its spread through early detection cannot be over-emphasized.

Infections to COVID-19 have been conventionally investigated in clinical settings by monitoring laboratory metrics and symptoms [66, 33]. These studies have focused on a large amount of subjective questionnaires and invasive laboratory test results. For example, Jehi *et al.* [66] use a large number of features extracted from demographics, comorbidities, immunization history, symptoms, travel history, laboratory vairables, and medications to predict the infection with COVID-19. Li *et al.* [81] show that the oxygenation index and respiratory system compliance could be leveraged to study ARDS patients infected with COVID-19. Force *et al.* [49] propose that ARDS caused by factors rather than COVID-19 results in reduced lung compliance. However, reduced lung compliance in ARDS is typical of the disease [81].

Such diagnostics are the gold standard methods to investigate COVID-19 and ARDS pa-

tients; however, they are limited to hospitals and clinical settings. Moreover, subjective symptom-based analyses were shown to be an ineffective strategy to qualify an individual's likelihood of contracting COVID-19 [33]. In contrast, various studies showed that vital signs such as heart rate and blood pressure could be exploited for early detection of infections and respiratory diseases [125, 90, 51]. We posit that such biomarkers are informative in identifying ARDS patients infected with COVID-19. These biomarkers can be collected continuously and remotely due to the recent advancements in wearable electronics and Internet-of-Things-based devices (e.g., Omron® HeartGuide wrist-band [10]). Therefore, the effectiveness of these biomarkers in early COVID-19 detection extends the monitoring services to remote settings.

Recognition of COVID-19 infections using big sensory data necessitates novel modeling and analysis techniques. The state-of-the-art studies often use traditional statistical models to predict COVID-19 infections. These studies have mostly studied the linear statistical relationship and association between the health parameters or extracted features from the subject's demographics, symptoms, laboratory tests, and medications [66, 33]. For example, a full multi-variate logistic model is constructed in [66] to predict COVID-19 using extracted features. However, such data with complex intensive longitudinal structure and temporal characteristics need to be investigated using nonlinear and advanced methods. Machine learning algorithms, including Artificial Neural Networks, can be tailored in this regard to extract linear/nonlinear correlations in the data throughout the health monitoring. In this study, we investigate the behavior of COVID-19 on ARDS patients by proposing a deep neural network (DNN) model which utilizes three longitudinal features from the University of California COVID Research Data Set (UC-CORDS) [14]: systolic and diastolic blood pressure and heart rate. We compare individuals who developed ARDS with and without COVID-19 to assess potential markers that could be used in early detection and prevention strategies. Moreover, we utilize statistical features and neural networks to distinguish between ARDS caused by COVID-19 and other factors.

Table 4.5: Age distribution of subject with different COVID-19 test results.

	Number of Participants		**Number of Samples**	
Age range	**Negative**	**Positive**	**Negative**	**Positive**
20 - 40	10	10	5724	5291
41 - 60	27	30	13955	16068
61 - 80	27	24	16395	15020
80 - 100	11	11	2096	3423

4.2.2 Methods

Data Set

UC-CORDS data set provides comprehensive, structured information of patients admitted to the hospital at the University of California's five academic health centers (i.e., UC Davis Health, UC San Diego Health, UC Irvine Health, UCLA Health, and UCSF Health). This data set provides a wide range of information, including different observations, measurements and COVID-19 test results of patients.

Notably, the vital signs, including heart rate, systolic and diastolic blood pressure, are recorded continuously every 30 minute. Since the data set is fully anonymized, it is not possible to access actual dates. However, we only considered *hospitalized* patients who were diagnosed with ARDS (IDs 4195694 and 4191650 from SNOMED vocabulary [8]) and we included data after their first COVID-19 test. Since the number of observations with negative COVID-19 test results is more than positives, we randomly selected fewer patients with negative test results based on the age distribution. This re-sampling resulted in a more balanced data set (i.e., 39,802 data points for each feature in the positive group and 38,170 samples in the negative group). As of April 1^{st} 2021, this led to 150 participants for the positive and negative test groups (i.e., 75 participants for each group). Table 4.5 shows the age distribution of the patients per each COVID-19 test result.

In addition, another valuable aspect of this data set is the longitudinal monitoring of the

vital signs. The data set contains, on average, 136.6 and 57.4 days for the negative and positive test groups, respectively.

Ethics

The data was jointly reviewed by the Institutional Review Boards of all UC Health campuses and was determined to be non-human subjects research. Moreover, UC-CORDS does not contain any patient identifier such as name and phone number. As such, UC-CORDS is a HIPAA limited data set.

Statistical Analyses

To show the correlation of features (i.e., blood pressures and heart rates) and COVID-19 test results, statistical features have been extracted. We measure basic features, including mean, minimum (min), maximum (max), and standard deviation (std) of DBP, SBP, and HR for each subject. Besides, we utilize the Point Biserial correlation between the proposed features and COVID-19 test results. This correlation, which is similar to Pearson's correlation, is used when one of the variables is binary, and the other variable is a continuous number [126]. In other words, this measurement indicates the difference between the categorical groups' distribution.

Neural Networks

In this study, we are interested in the COVID-19 detection using longitudinal heart rate and blood pressure monitoring. To perform the detection, we propose a DNN architecture combining convolutional neural network (CNN), and a long short-term memory (LSTM). Such a model is utilized to leverage the embedded structure of longitudinal data. We have

Table 4.6: The architecture of the proposed neural network.

Layer	**Output Shape**
1D CNN	(None, 4, 14, 64)
MaxPooling	(None, 4, 7, 64)
1D CNN	(None, 4, 5, 32)
MaxPooling	(None, 4, 2, 32)
Flatten	(None, 4, 64)
LSTM	(None, 64)
Flatten	(None, 64)
Dense	(None, 100)
Dense	(None, 1)

considered three channels of vital signs, i.e., heart rate (HR), systolic and diastolic blood pressure (SBP and DBP), as the inputs and the COVID-19 test result for the network's output. Table 4.6 summarizes the detailed structure of the proposed network. It consists of CNNs (capturing the spatial information), followed by a max-pooling layer, a LSTM layer (capturing the order in time series data) [63], and finally two fully connected layers (extracting the embeddings). We use grid search to tune the hyperparameters (e.g., number of filters and neurons) of the DNN. We randomly select 75% (112) of patients as train, the rest as the test data, and accuracy, f1 score and AUC metrics were chosen for the performance evaluation.

We label positive COVID-19 test results with '1' (26,984 samples in the train, 46.66%, and 8,018 samples in the test data, 75.97%) and the negative ones with '0'. For the learning task, TensorFlow package has been utilized.

Besides, to assess the detection's effectiveness, we test our model on different time intervals on test subjects. In other words, we are interested in the possibility of COVID-19 test result detection by only using a limited number of samples (in days). We evaluate the model with different interval sizes, which is extracted including N days ($N \in \{2, 4, ..., 60\}$) of subject's data. This evaluates the model's performance by looking only at a limited number of days.

Finally, for visualization purposes, the t-SNE method [86] was used over the dense layer's

Table 4.7: Point Biserial correlation of statistical features and COVID-19 test results (* shows significant correlation).

		Correlation	***P*-value**
HR	mean	0.20	.012*
	std	0.21	.009*
	min	-0.10	.22
	max	0.18	.020*
DBP	mean	-0.19	.015*
	std	-0.19	.016*
	min	-0.10	.20
	max	-0.14	.075
SBP	mean	-0.15	.063
	std	0.08	.32
	min	-0.17	.031*
	max	0.13	.10
Age		-0.05	0.51

Table 4.8: 95% confidence intervals of HR, DBP and SBP for each COVID-19 test result group.

Biomarker	**95% CI (Positive)**	**95% CI (Negative)**
HR	95.97-96.35	84.87-85.24
DBP	64.95-65.20	70.83-71.13
SBP	117.66-118.09	127.20-127.72

output to reduce the feature space's dimension to two.

4.2.3 Results

Statistical Observations

We measure basic statistical features of BP and HR and compare them with COVID-19 test results. Table 4.7 shows the Point Biserial correlation between these features and age with COVID-19 test results, and Table 4.8 represents 95% confidence interval (CI) of these biomarkers for each COVID-19 test group.

Table 4.7 shows significant positive correlations between the mean of HR/DBP, std of

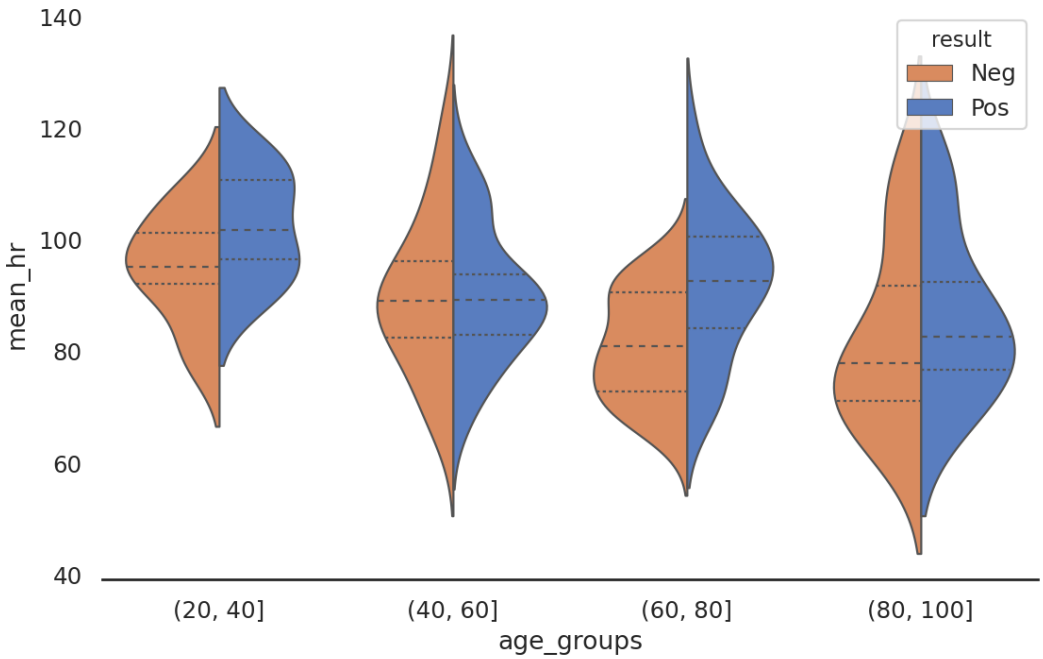
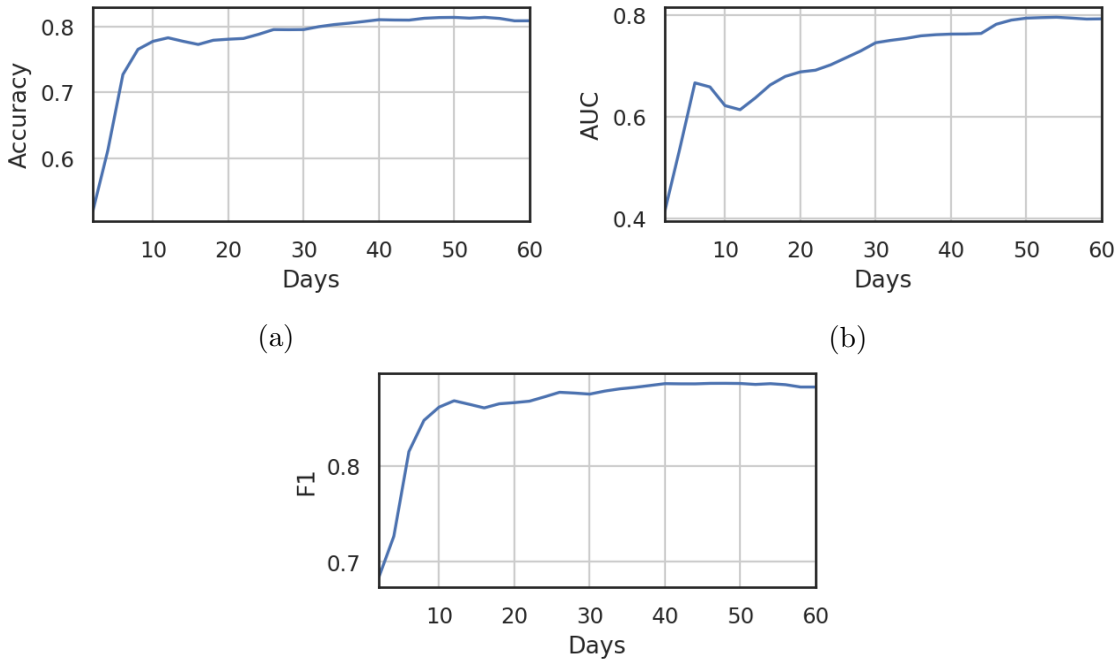


Figure 4.13: The distribution comparison of average HR between the positive and negative test results for each age group. The dashed lines represent the quartiles.

HR/DBP, maximum value of HR, minimum value of SBP and COVID-19 test results. Fig. 4.13 illustrates the difference in the distribution of average HR *(mean_hr)* between each age group for the positive and negative test results. Although *(mean_hr)* shows a significant correlation with COVID-19 test results, there is an overlap in the distribution of such a feature between positive and negative results. This visualization further supports the fact that using only statistical features to detect people infected with COVID-19 is challenging.

Deep Learning

Due to the longitudinal aspect of the data, we consider a DNN architecture to detect COVID-19 test results by only looking at BP and HR. The accuracy of this model reached as high as 0.79, 0.87 precision, 0.84 recall, 0.85 f1 score and 0.81 area under the curve (AUC) for the entire test data. Besides, Fig. 4.14a shows the accuracy of our model with respect to the first 60 days of data. Fig. 4.14b illustrates the corresponding area under the curve (AUC) with



(a) (b)

(c)

Figure 4.14: The performance of the model in terms of accuracy (a), AUC (b) and f1 score (c) using test data with respect to the number of included days.

given days while Fig. 4.14c shows the f1 score. Fig. 4.14a shows an increase in the model's accuracy at the beginning, starting from 0.52 and reaching as high as 0.78% on day 12^{th}. The small drop in the measurements after day 12^{th} is because of the increase in the number of false negatives compared to true positives. By observing more data for both groups, the performance of the model constantly increases (day 20^{th}).

To visualize the extracted features using DNN, we used t-SNE method [86] to reduce the feature space dimension to two. We performed this method on the output of the dense layer with 100 neurons. Fig. 4.15 shows that using extracted features by the DNN, the positive and negative cases are almost separated.

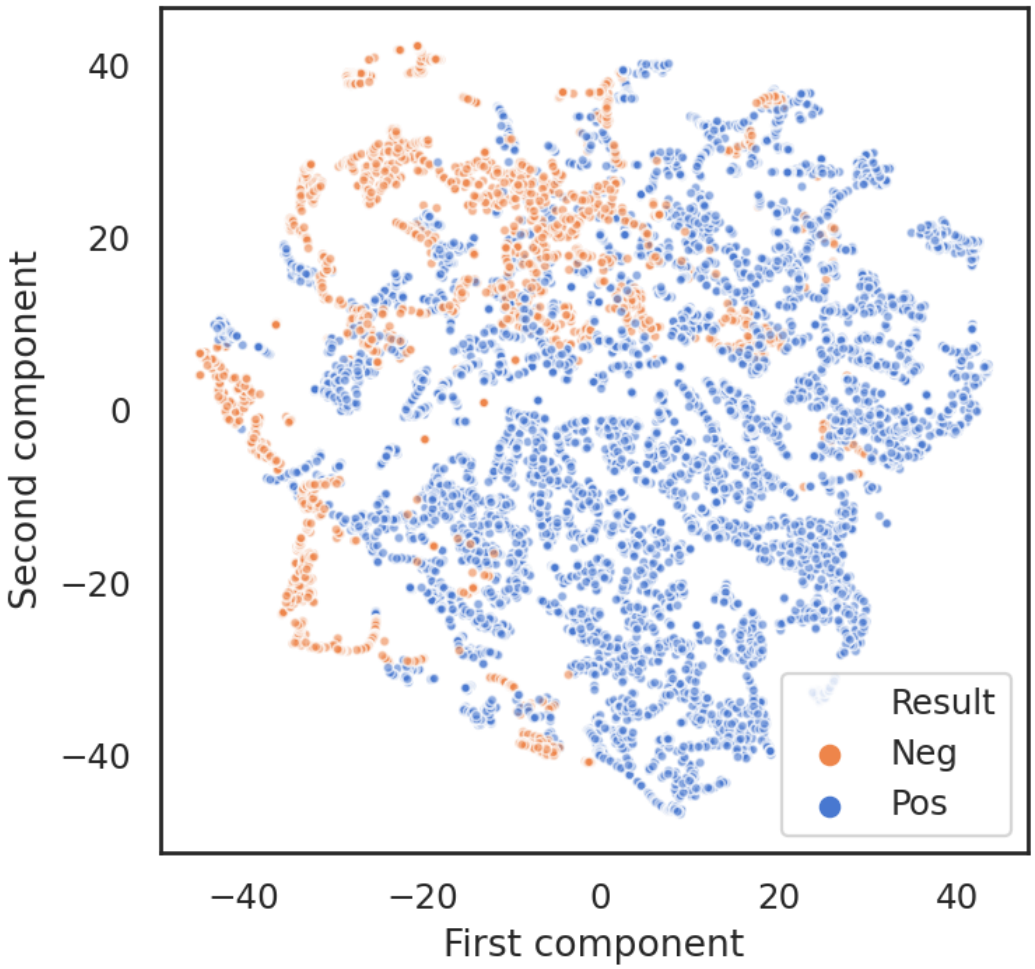


Figure 4.15: The 2-dimensional representation of test data using t-SNE considering the entire test data.

4.2.4 Discussion

A few of our observations warrant additional discussion. First, monitoring of blood pressure and heart rate may provide a useful strategy for individuals living in collective communities, such as nursing homes or rehabilitation facilities, as well as for healthy community-dwelling adults. The potential impact could be to mitigate the spread of COVID-19, as well as allowing early detection of complications associated with infection, such as those at greater risk for ARDS.

Second, we assessed for the presence of comorbidities in COVID-19 positive patient with

ARDS, and reported that comorbid diagnoses such as type 2 Diabetes Mellitus, hyperglycemia, chronic obstructive pulmonary disease, elevated transaminase, and lactic acid dehydrogenase, bradycardia, acute ST segment elevation myocardial infarction, and metabolic derangements were more prevalent (data not shown). This observation is in-line with other reports [81, 131, 143] demonstrating increased vulnerability among those with chronic health conditions, as well as reported metabolic derangements observed with COVID-19 infection, especially among adults over 60 years of age.

Third, there are other potential applications in modeling COVID-19. Specifically, there has been a discussion of how early COVID-19 arrived in the United States; the first cases were reported in California. It would be possible to review data prior to the first reported cases in the U.S. to validate the presence or absence of COVID-19 in our communities prior to January 2020. This is of importance as the viral genome sequence was confirmed in late January 2020, which allowed for the use of polymerase chain reaction to detect viral genetic material [15]. Antibody testing, which has been shown to be inconsistent, was used in the preceding months, raising the question of how early was COVID-19 in the United States.

Moreover, the related works focus on laboratory measurements and symptoms to detect the infected patients or severe positive cases of COVID-19 using statistical methods [66, 33]. In contrast, we considered two easily accessible features as well as utilizing a deep learning method to capture the short- and long-term dependencies in the time series data. There is a correlation between the simple statistical features and COVID-19 test results. However, simple logistic regression models are insufficient due to the overlap in the feature space. Leveraging the nonlinear features extracted from our proposed neural network, we distinguished negative and positive COVID-19 test results with the AUC as high as 0.81 by using only blood pressure and heart rate values.

Although our findings are only based on ARDS population, these achievements could potentially lead future directions of our research to investigate the aforementioned vital signs for

COVID-19 detection tasks with other populations.

4.3 Conclusion

Monitoring blood pressure is essential for early detection and treatment of cardiovascular disease. Conventional methods utilize statistical and machine learning models to estimate SBP and DBP using PPG, a low-cost and straightforward signal; however, they fail to generate the entire signal or synthesize an accurate waveform. In this work, we leveraged the cycle generative adversarial network (CycleGAN) for the first time in the blood pressure estimation domain, and our model achieved MAE of 2.89 mmHg and 3.22 mmHg for SBP and DBP in a cross-subject setting. The per-subject evaluation's performance was 2.29 mmHg for SBP and 1.93 mmHg for DBP, outperforming state-of-the-art approaches by up to 2$\times$ for improving the BP estimation accuracy.

In addition, we proposed a DNN-based model to investigate the non-linear patterns in simple vital signs, namely, blood pressure and heart rate, which can be easily and reliably measured without the need for skilled medical professionals, in ARDS patients with positive and negative COVID-19 test results. Our proposed model achieved 0.79 accuracy, 0.85 f1 score and 0.81 AUC. Using wearable devices, it is possible to monitor vital signs of subjects in everyday settings without visiting a hospital or a care site. Utilizing the proposed model allows early detection of COVID-19 cases in free-living conditions.

Chapter 5

Physical Activity Recommendation

5.1 Background

Physical activity (PA) has undeniable health advantages. The World Health Organization (WHO) and a growing number of national governments worldwide have developed public health-oriented PA recommendations in response to the relevance of physical inactivity as a risk factor for chronic illnesses and early death [50].

The Internet of Things (IoT) and mobile health (mHealth) have made it feasible to link users' physiological data and daily exercise information with their fitness demands by collecting and visualizing individual exercise activities via wearable trackers [47].

Despite the great potential of mHealth services in free-living conditions, the evaluation of exercise recommendations and optimization is challenging. The majority of these mHealth services suffer from non-personalization and having fixed activity suggestions for all the users, which may result in fixed physical activity [146]. Such an inefficiency might have different underlying causes, as described by [47].

Several systematic studies [16, 98, 141] investigated the effectiveness of mHealth in physical fitness and interventions, as well as the need for mHealth technology to promote physical health. Just-in-time adaptive models [47, 17], reinforcement learning-based model [144, 120, 83], neural network-based models [87] are such instances of mHealth designs. Although these studies claim personalization, they suffer from the inability to monitoring real-time heart rate as well as the intensity during the exercise.

Machine learning models can be practical for extracting the patterns of entire data sets, and they work best when there is enough data; however, in a real-time system at the beginning of the studies, the sparsity of the data makes the prediction task challenging (cold start problem). Therefore, Reinforcement Learning (RL) gained enormous attention as they learn the policies by observing the rewards of different user actions. These RL algorithms have been applied in a wide range of healthcare applications [144, 111, 120].

In this study, we designed the first end-to-end closed-loop physical activity recommendation system in the wild using personalized real-time HR monitoring with a proof-of-concept implementation. We designed and implemented a contextual bandit framework as an adaptive (active) learning approach to recommend exercises based on personalized biofeedbacks, exercise intensity, and overcome the cold start problem at the beginning of the study. Our result showed an increase in daily exercise duration ($P < .001$) and walking and recommendation system components had average satisfaction scores of 4.31 (0.60) and 3.69 (0.95), respectively, on a scale of 1 to 5. In addition, participants' confidence in their capacity to do the suggested walking exercises safely and the study's ability to satisfy their needs for physical activity both received average scores of over 4.

To summarize, our main contributions include:

1. We designed and implemented a closed-loop mHealth system with personalized exercise recommendations. The users can interact with different interfaces using smartphone

and smartwatch applications.

2. The system leverages the user's real-time heart rate and exercise intensity as inputs for the system.

3. We conducted a 12-week study with 12 active participants as a proof of concept to demonstrate the functionality and effectiveness of the system.

5.2 Methods

5.2.1 Participant recruitment

Twenty female college participants were recruited for the study between Nov. 2021 and Feb. 2022 via flyers. Participants were eligible to enroll in the study based on the following criteria: at least 18 years old, registered UCI undergraduate and graduate students, in general, good physical health, and without any medical conditions preventing them from engaging in exercise. They also used a smartphone (iPhone or Android compatible with our app), spoke English, agreed to wear the smart devices as well as complete the recommended exercise activities and submit surveys related to their experiences.

5.2.2 Study Procedures

The study procedures were approved by the University Institutional Review Board. All participants received a study information sheet, and completed baseline demographic and exercise history questionnaires. Instructions on how to properly connect the wearable smart devices to the phone were provided by our technical team. In addition, a virtual group introductory session was held by a research staff/coordinator, who provided an overview/de-

scription of the study, expectations, and anticipated timeline of activities, as well as answered any questions participants had. Participants were first assigned to one of two exercise activity levels depending on their exercise history. Subsequently, participants completed weekly walking activities based on the assigned frequency and duration determined by the recommendation system and reported their experiences via our app surveys.

5.2.3 Exercise organization

Ideally, human subjects should get at least 150 minutes of moderate-intensity aerobic activity every week. This moderate-intensity activity can be relative to an individual's fitness level [129]; such movements can raise the heart rate to a certain level. Aerobic activity includes a large portion of muscle movement in a rhythmic way [60]. For this study, individuals were assigned to two different levels based on their current level of physical activity (Level 1 for subjects with low level of physical activity, and Level 2 for the individuals who had exercise routines in their daily life).

Borg test, as well as talk test, are two classical methods for exercise intensity detection. The Borg test contains a numbered list of values that a subject reports in response to the intensity level of an exercise. The drawback of such tests could be generalizability. In addition, such tests are subjective and cannot accurately capture an exercise's intensity. Similarly, as long as an individual can follow up on a conversation while exercising, she is not in the moderate-intensity zone [129].

Another exercise intensity detection method relies on heart rate (HR) values during exercise activity. This method is more objective and robust. Maximum heart rate (HRmax) is commonly used to identify different HR zones. The exact value of HRmax can be identified while performing a VO2max test for individuals; however, this method is not feasible in free-living conditions and daily life [114, 122]. Hence, we used the most common equation for

HRmax, which is 220 – age [114, 122]. Using HRmax, different guidelines suggest HR ranges for moderate-intensity activity. We used the American College of Sports Medicine (ACSM) recommendation for this study. According to ACSM, any physical activity that increases the HR to 63% of the HRmax is labeled as light intensity exercise. Activities ranging from 64% to 76% can be classified as moderate-intensity exercises, and anything above 77% would be considered vigorous exercise [114, 122].

Literature shows that the absence of time is the main reason for not performing physical activities [50]. Because one of the most significant limitations stated for not exercising PA is a lack of time, studies indicate that even 30 minutes of movement per week can improve health [50]. Hence, since we had a healthy younger population cohort, we chose 15 and 20 minutes as the baseline for Level 1 and 2 groups, respectively.

In addition, for inactive people, rather than meeting the physical activity guidelines (150 minutes of moderate-intensity), the global action plan is to start with a baseline and perform small incremental increases in physical activity [40, 99]. Hence, we designed an incremental plan for our exercise recommendation. Fig. 5.1 illustrates such a plan.

5.2.4 Recommendation System Design

The designed exercise recommendation system consist of different components. Fig. 5.2 summarizes the entire recommendation system design. Different data sources record the subject's data and constantly send it to the servers using Wifi/Bluetooth/Mobile network connection. The ZotCare system manages the different modalities of data, and the recommendation system pulls different data using its cloud interface. This system decides about different alerts and/or updates on exercises. In the following sections, we will introduce different components.

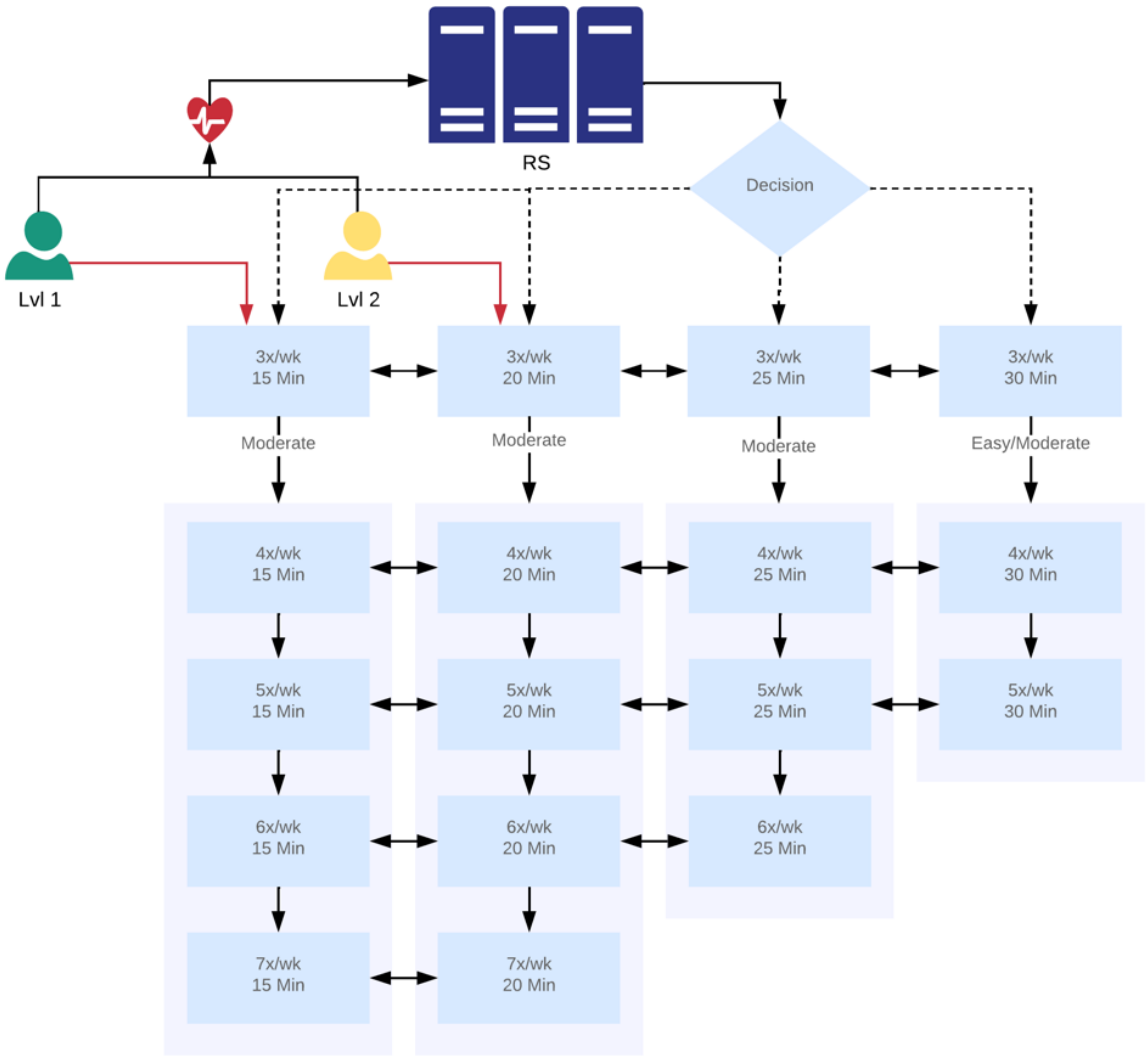


Figure 5.1: Exercise duration setup.

Mobile and Samsung watch applications

For this study, we developed two sets of applications and services. The first group contained the applications for users' mobile devices (i.e., iPhone and Android platforms), which were used to record questionnaires and show the recommendations. The second group consisted of applications developed for the user's smartwatch (i.e., Samsung active 2) to record the user's biomarkers and real-time activities. Fig. 5.3 shows an overview of the user interface of these applications.

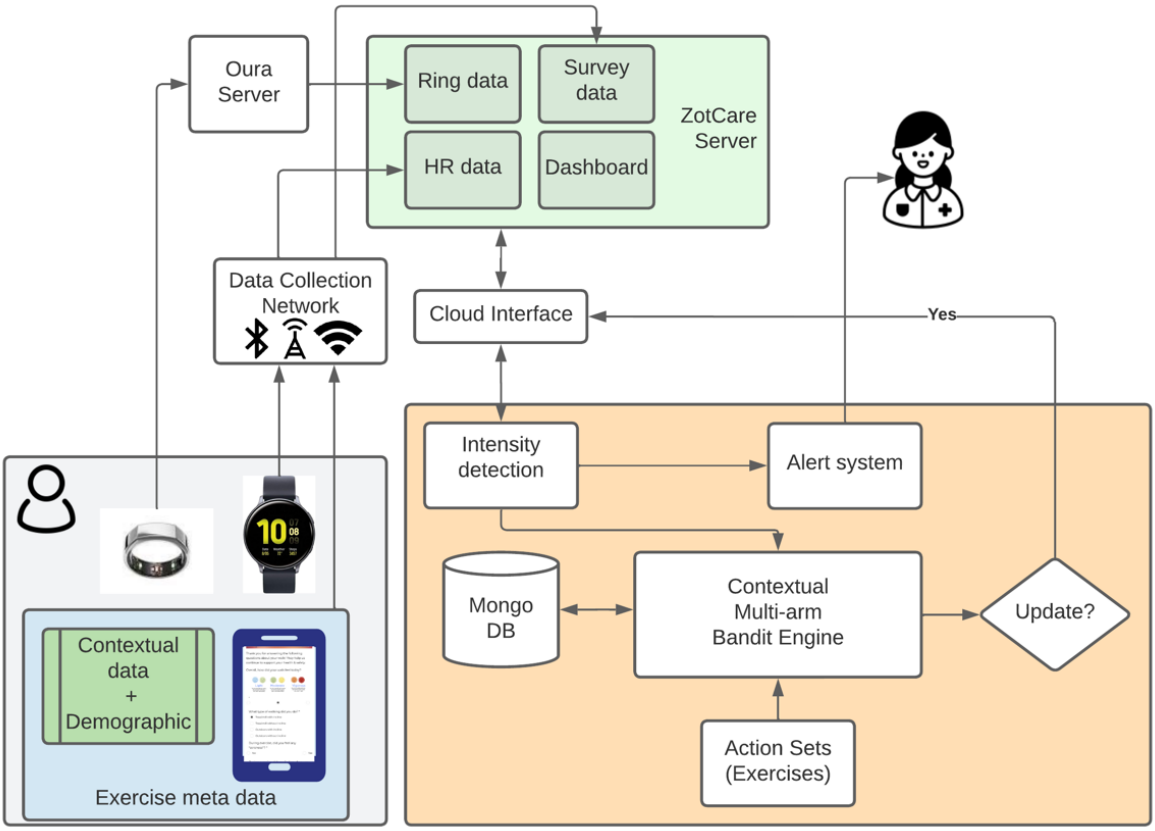


Figure 5.2: The overview of the system.

ZotCare Dashboard

The ZotCare platform [13] has been developed by our group at the University of California Irvine and can provide services for real-time data collection. This platform's dashboard is designed so study coordinators or health experts without programming knowledge can create and manage study parameters, including questionnaires, mobile notifications, user sensor data, and user profiles, in a privacy-preserved form. Fig. 5.4 demonstrates a sample page of such a dashboard for our study.

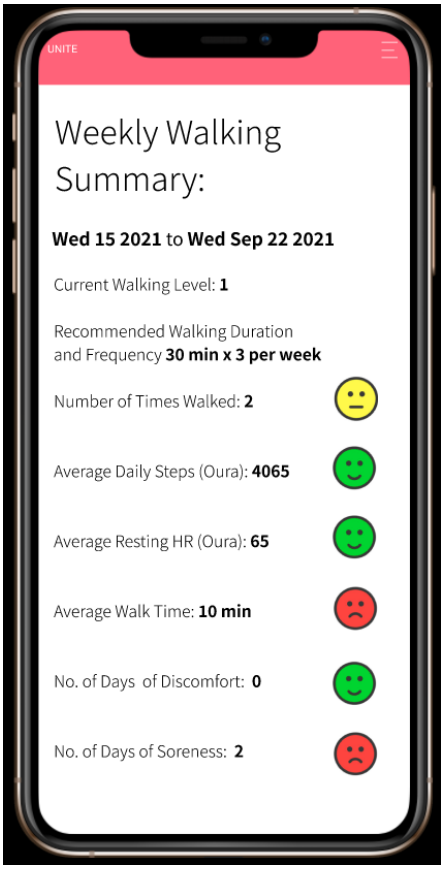


Figure 5.3: An overview of the developed applications for users' cellphones and Samsung smartwatches.

Contextual-bandit Problem

The multi-armed bandit (MAB) algorithms have recently become popular in different domains, including healthcare and recommendation systems. These models are used to maximize the total payout of an agent who is interacting with an environment. At each timestamp t, the model selects an action based on a policy, delivers that to the patient and monitors the reward regardless of the user's context [85].

In exercise recommendations, it is possible to provide a list of suggestions according to their intensity and caloric benefit. Currently, these personalizations are provided only by human health coaches. However, such a method does not capture each individual's dynamic characteristics, which can change person-to-person or/and over time. Furthermore, the limitation

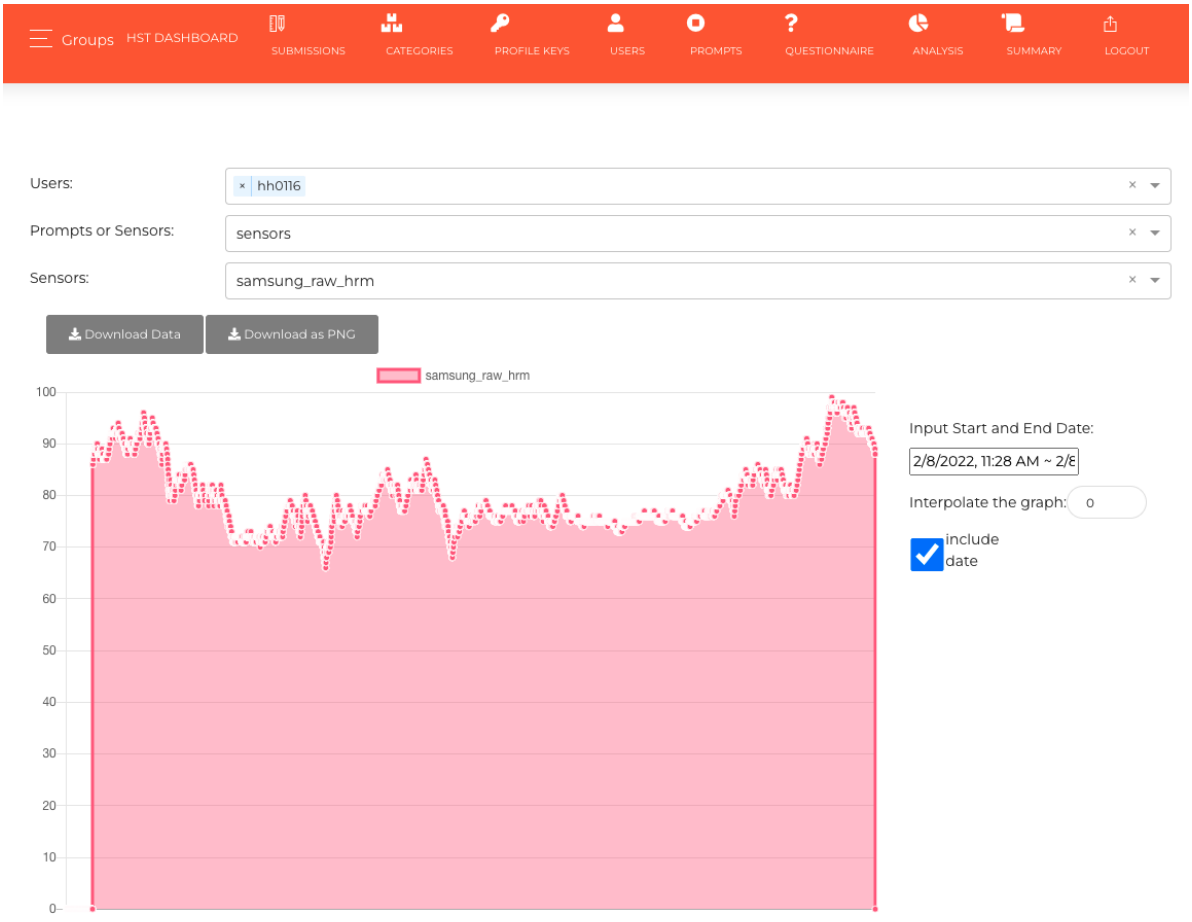


Figure 5.4: An example page of the ZotCare dashboard.

of the data makes it difficult to have perfect predictions, and the cold-start problem is a challenge for such a task. In many applications, there is rich information regarding each action, so having features for every (context, action) pair rather than features associated only with context and shared across all actions would be the more profitable approach.

The Contextual multi-armed bandit can address the challenges mentioned earlier. In the contextual multi-armed bandit (CMAB) problem, a learning model monitors the user's current context, recommends an action (i.e., exercise in our context), and observes the reward of the recommended action. The objective of such a model is to minimize the cumulative sum of losses with respect to the context. The benefit of the contextual bandit learner compared to the regular multi-armed bandit is the context which makes CMAB personalized

and an ideal alternative for dynamic environments. In healthcare, CMAB enables precision medicine to leverage suitable actions for individuals instead of similar actions for multiple patients [85].

There are different algorithms to explore the finite set of actions, and for this study, we used the Epsilon-Greedy approach. In this approach, the model explores the arms (exercises) with a probability of epsilon (0.05 in our case) and with a probability of 1-epsilon exploits the known actions.

There are two main reasons for the exploration trade-off. First, to examine new actions, and second, in mobile health setups, although interventions might tend to have positive feedback on the selected action, they can have a negative impact on future rewards because of the user's habituation [83].

In the contextual bandit problem, the environment generates a pair (x_t, l_t) at time t such that x_t is a context vector and l_t is the loss vector. The learner chooses an action at and observes the loss of the corresponding action. The objective is to sustain a small cumulative regret

$$R_t = \sum_{t=1}^{T} l_t(a_t) - \sum_{t=1}^{T} l_t(\pi^*(x_t))$$

over time [29]. In this equation, $\pi^* \in argmin_{\pi \in \phi} E_{(x,l)}[l(\pi(x))]$ is the optimal policy with ϕ being the set of policies.

Optimization. There are different methods to solve the optimization problem. We used the regression with importance weights $w_t > 0$ introduced by [29].

$$argmin_{f \in F} \sum_{t=1}^{T} w_t (f(x_t, a_t) - y_t)^2$$

The goal is to find a f from a class of regressor functions F to predict a cost y_t having the context c_t and action a_t.

There are different policy evaluation approaches. The direct method is the most straightforward approach, which maps context and actions to the rewards using a regression model [17]. Since most of the directed methods suffer from high bias, in this study, we used Importance Weighted Regression to reduce the variance and bias of the estimator [29]. This method optimizes a regressor f to find the optimal policy which has been used in off-policy learning recommendation scenarios [123].

$$\hat{f} = argmin_{f \in F} \sum_{t=1}^{T} \frac{1}{p_t(a_t)} (f(x_t, a_t) - l_t(a_t))^2$$

$$\hat{\pi} = armin_a \hat{f}(x, a)$$

We considered user identifiers, weight, and the number of days from the beginning of the study as the input features (i.e., context). We had access to the sleep features (e.g., total sleep time, resting hr, sleep quality.) using the Oura ring; however, due to the manual sync issue, we did not have access to the real-time features. Hence, for this study, we discarded those features for real-time analysis and only performed offline analysis.

The duration of the exercises was considered as the actions for this study. In the case of moderate-intensity duration, we increased the exercise frequency. We trained two different models: one for Level 1 subjects and another for the Level 2 subjects.

Feedback Policy (exercise intensity)

To model the intensity feedback of the exercise, we used the subject's HR during each exercise. Thanks to the ZotCare system, we were able to have semi-real-time data recorded by Samsung Active 2 smartwatch. Since the smartwatch is prone to noises during recording and data collection, we removed invalid values and used a 10-second moving average filter over recorded non-zero HR data. Later on, the exercise reward for the subject s was defined as below:

$$r_s = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{1+e^{\Omega_s(HR)-HR_i}} - \frac{1}{1+e^{O_s(HR)-HR_i}}$$

The Ω and O in this equation represent 64% to 76% of HR_{max} (i.e., moderate-intensity range) for each subject s. HR_i reflects the ith HR during the exercise. This reward function is convex which meets the requirement for the convexity of the cost function.

The intuition behind this reward function is that for HR values in the moderate-intensity range, the inner part of the summation tends to be close to 1.0. For values outside this range, this term will be close to 0. The average makes this reward function smoother as well as by increasing the number of HR points in the moderate-intensity range, the reward function increases, consequently. Fig. 5.5, depicts the reward function and the corresponding moderate-intensity range for a sample user with $\Omega = 131.3$ bpm and $O = 151.5$ bpm.

Safety module

To ensure that the exercises are safe, we designed an alert system notifying the nurse about the outcome of the exercise. The alerts are generated in case of the subject's HR exceeds

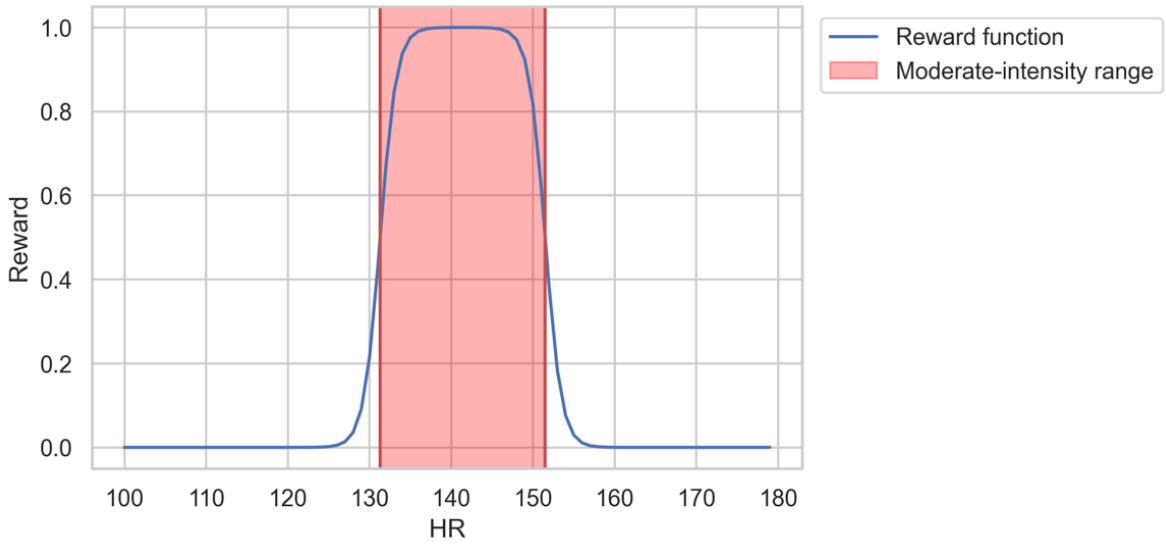


Figure 5.5: Reward function with respect to the HRs during the exercise.

90% of the HRmax, she feels new discomfort during the exercise, and/or has soreness due to the exercise. In case of any of the outcomes above, the recommendation system maintained the same workout for one more session.

5.2.5 Statistical Analysis

Data Exclusion

During our study, one subject did not participate in any of the exercises, and two subjects dropped out due to personal reasons. Out of the rest, we had four subjects who were experiencing technical difficulties and/or restarted their watch by mistake. Hence, we lost a large portion of their data. Since we were programming the watches manually, the turnaround time was causing us to lose data. In addition, a couple of these subjects submitted the reports incorrectly, making it challenging to track their physical activities. Although these subjects contributed to the reinforcement learning framework, we excluded them from the offline analyses.

Mixed-Effect Analysis

Hierarchical linear mixed (HLM) models (also known as mixed effect models) were exploited to analyze the trends in between- and within-subject variances. We did this analysis using the notation defined and recommended by Raudenbush-Bryk and Bolger-Laurenceau [31, 112]. The dependent variables of interest include duration in light/moderate intensity over time, and exercise duration performed by the subjects.

5.3 Results

5.3.1 Improvement in performed exercise duration trends

During the course of 12 weeks of this study, we monitored subjects' exercise behavior recorded by our application. We only considered those exercises that the subject reported and had HR data available.

Fig. 5.6 shows the average exercise duration with standard error bars for each group (i.e., Level 1 and Level 2) over the study period. In general, it shows an increasing trend throughout the study. In addition, we utilized mixed effect models to illustrate the within- and between-subject trends over time.

Fig. 5.7 illustrates the trends for each subject. The blue dots show the actual duration that the subject did, and the red ones demonstrate the recommended duration. Tables 5.1 and 5.2 summarize the fixed and random effects of such a model. To overcome the exploding gradients effect, we normalized the x-axis (time) before applying the mixed effect model. The general trend for these subjects is significantly increasing ($P < .001$), with an initial intercept of 15.52 minutes of exercise on average ($P < .001$). The correlation of fixed effects

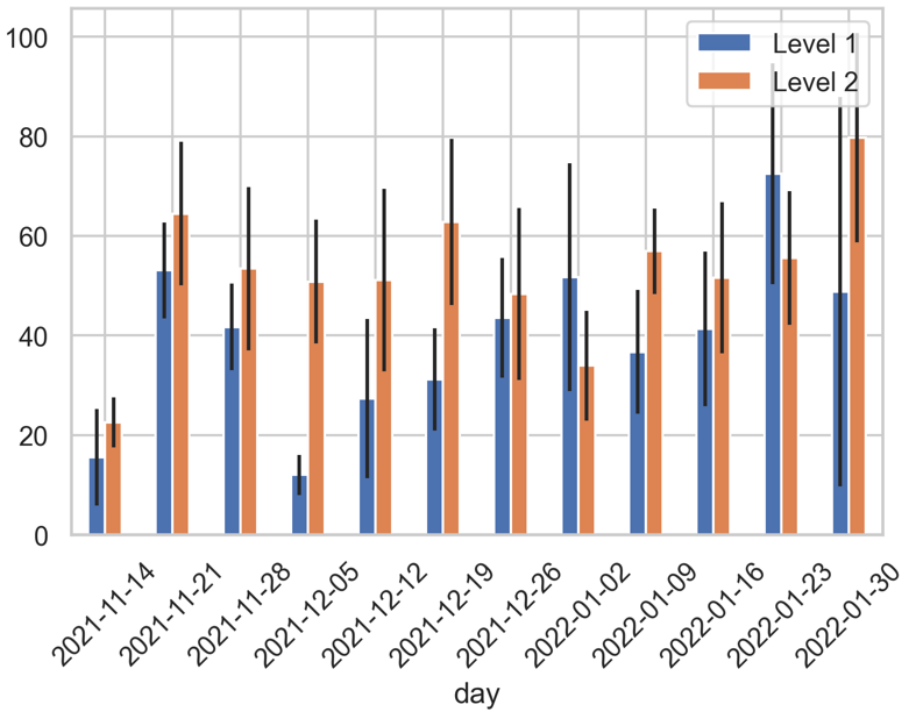


Figure 5.6: Average weekly duration of exercise for each group. The bars represent standard error.

is $r_f = -0.81$, and the correlation of random terms is $r_r = -0.84$. This is not surprising since subjects starting with a lower initial level tend to reach the goal faster than other groups (i.e., Level 2 subjects). Fig. 5.8 illustrates the quantile-quantile plot (Q-Qplot) and normality of the residuals.

Table 5.1: Summary of fixed effects of the HLM fitted to the exercise performed duration.

		Estimates	CI	P
Fixed Effect				
	Intercept	15.52	10.96-20.07	¡.001
	Time	15.24	9.72-20.76	¡.001

Table 5.2: Summary of random effects of the HLM fitted to the exercise performed duration.

		Standard deviation
Random Effect		
	Residual	7.88
	Intercept	6.98
	Time	6.43

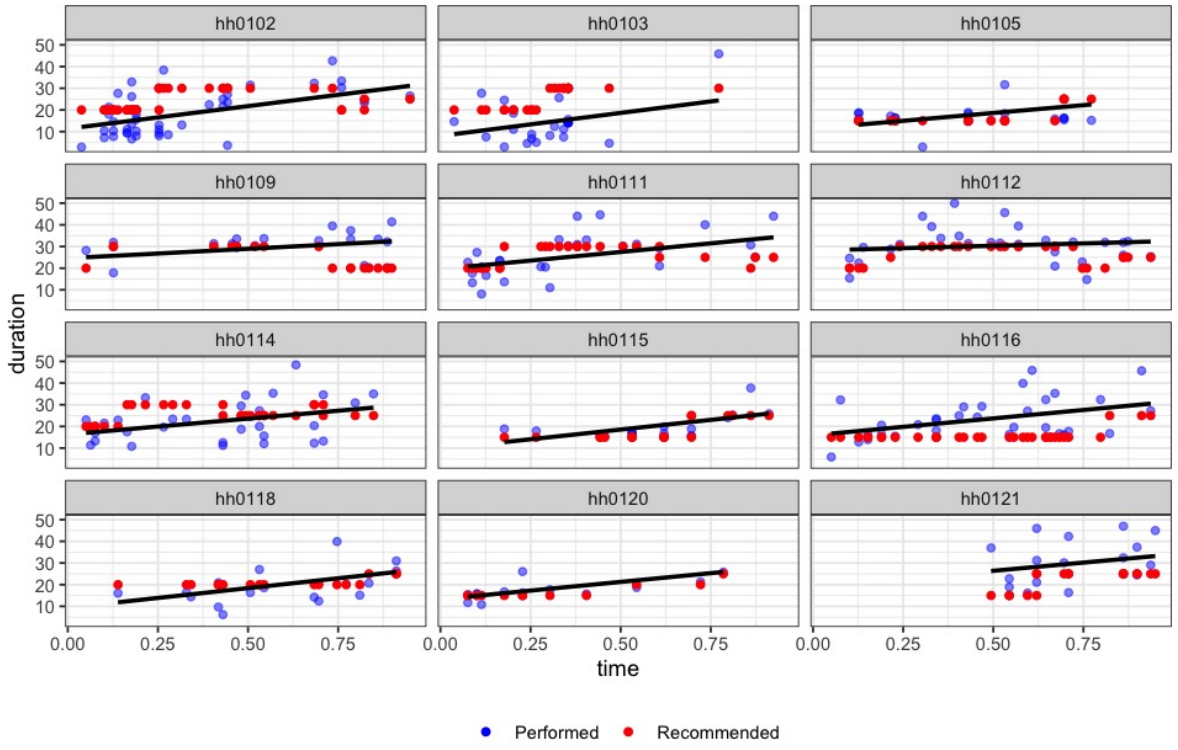


Figure 5.7: Performed vs. recommended duration of exercise per each subject. The line represents the fitted HLM model to each subject.

5.3.2 Minutes in light- and moderate-intensity

In addition to the exercise total duration, the HLM model was utilized to analyze the trend of non-vigorous activity duration (duration in which the HR is below 77% of the HRmax). Figs. 5.9 and 5.10 illustrate per subject trend estimation and dot plot, respectively. Tables 5.3 and 5.4 show the fixed and random effect for light and moderate-intensity exercise duration. The general trend significantly increases the corresponding duration ($P < .001$). The effect correlations are $r_f = -0.84$, and $r_r = -0.93$. Fig. 5.11 shows the Q-Q plot for the residuals. Subjects hh103 and hh112 have the highest and lowest increase, respectively.

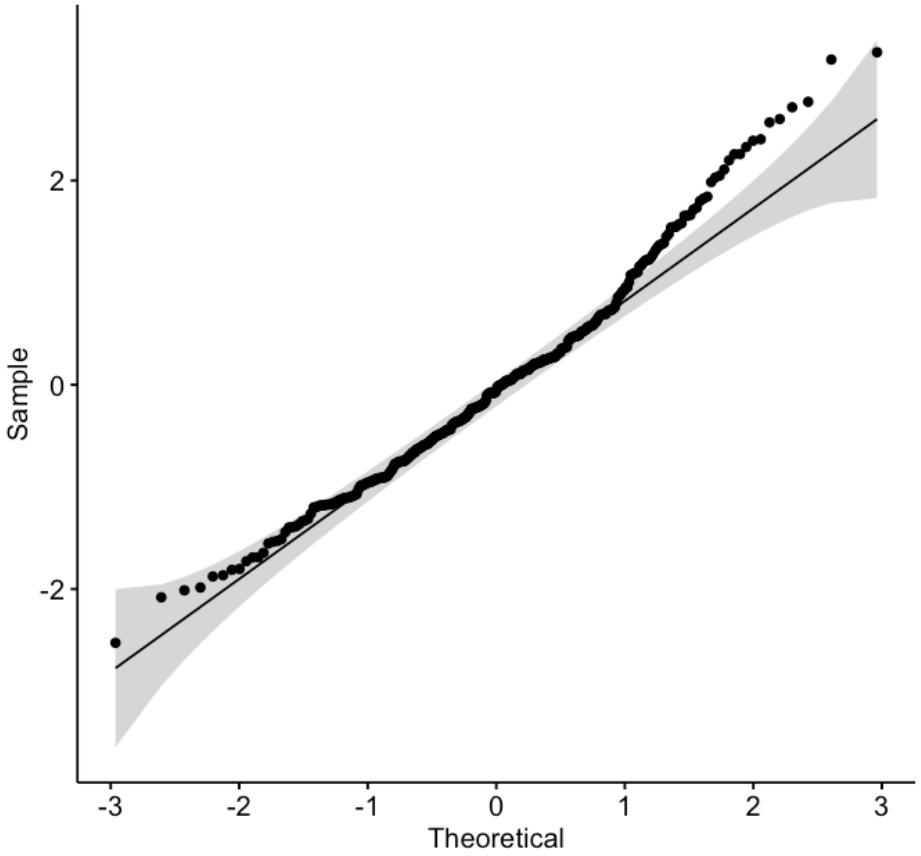


Figure 5.8: Normal Q-Q plot for normality of the residuals for exercise performed duration.

Table 5.3: Summary of fixed effects of the HLM fitted to the minutes in light/moderate-intensity exercise.

		Estimates	CI	P
Fixed Effect				
	Intercept	15.66	11.54-19.77	¡.001
	Time	10.01	5.44-14.57	¡.001

Table 5.4: Summary of random effects of the HLM fitted to the minutes in light/moderate-intensity exercise.

		Standard deviation
Random Effect		
	Residual	8.17
	Intercept	6.25
	Time	4.92

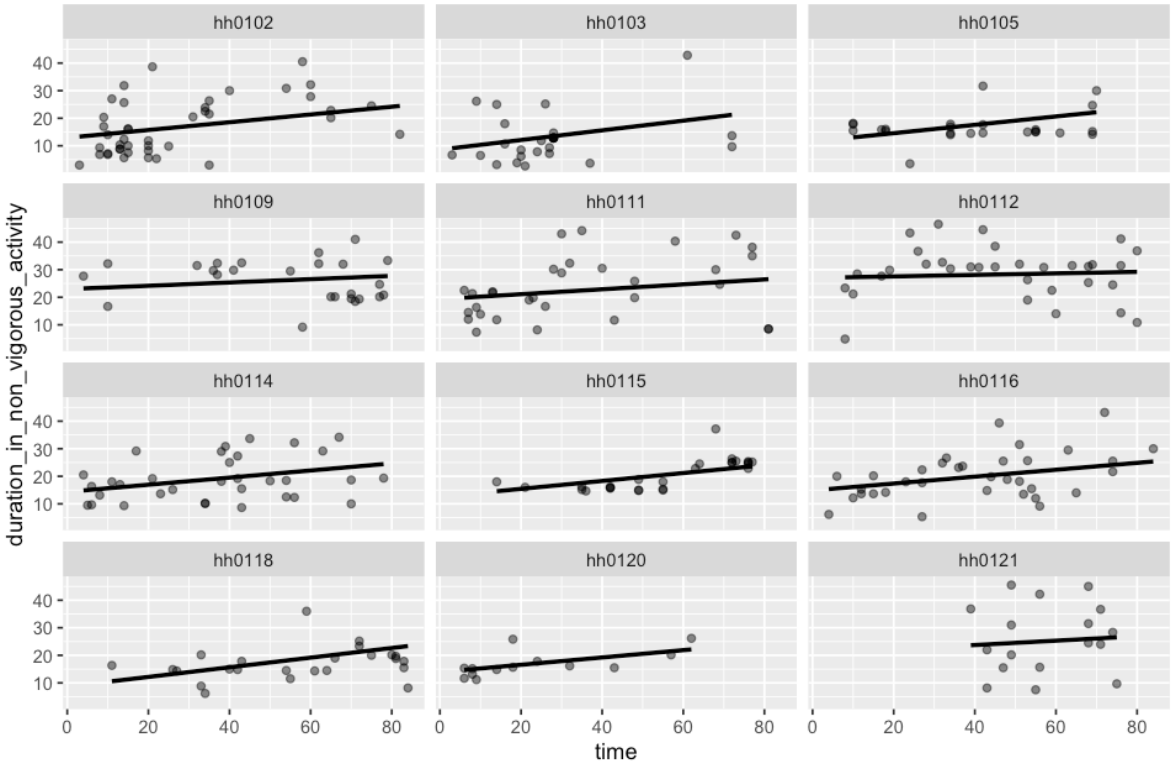


Figure 5.9: Trends in minutes in light and moderate-intensity exercise over time.

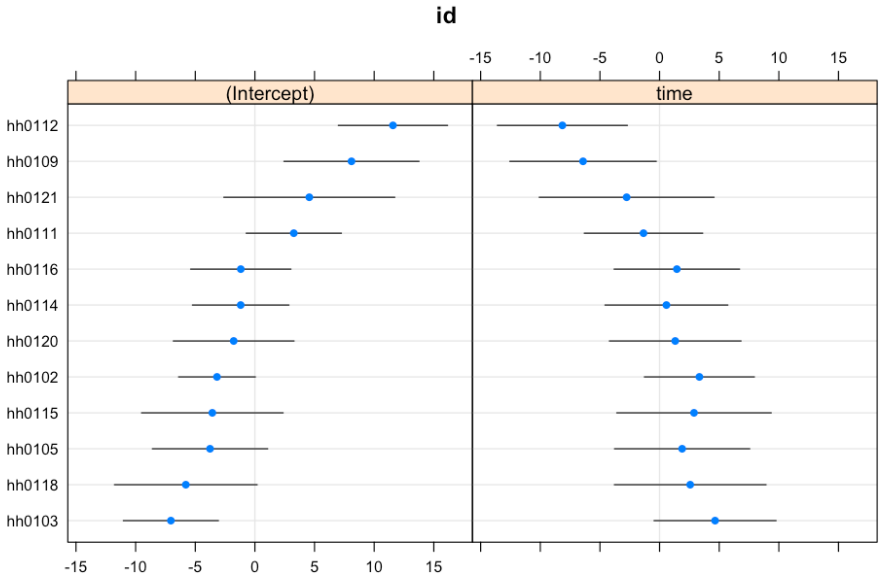


Figure 5.10: Dot plot visualization of between-subject variations compared to the mean trend.

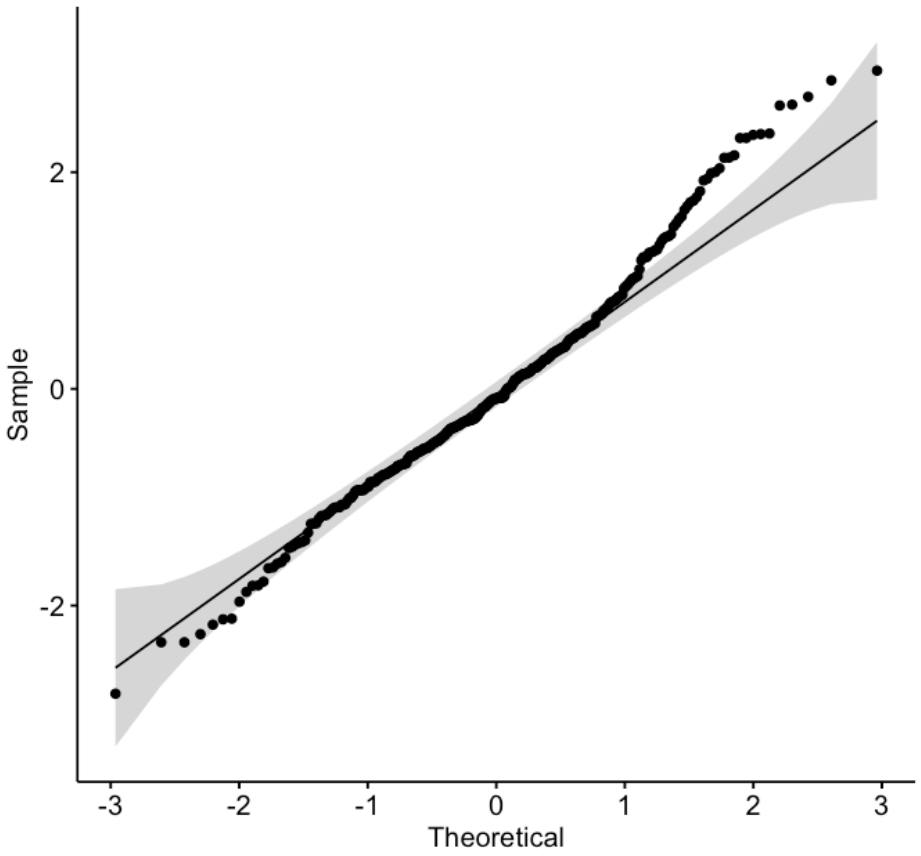


Figure 5.11: Normal Q-Q plot for normality of the residuals for minutes in light/moderate-intensity exercise.

5.3.3 Aggregated weekly performance

To investigate the weekly performance of the users and compare it with the standards and guidelines, we aggregated the exercise execution of the subjects over each week of the study. Fig. 5.12 shows the bar plot of weekly exercise duration for each user regardless of the longitudinal notion of time. Each color represents the exercise minute ranges. Subjects hh116 and hh102 were able to perform more than 150 minutes of weekly exercise.

Furthermore, we investigated the weekly exercise trends of the subjects by applying HLM models to the weekly durations over time. Fig. 5.13 indicates such trends for each user, and Tables 5.5 and 5.6 summarizes the fixed and random effects. The effect correlations are $r_f = -0.86$, and $r_r = -0.85$. According to Figs. 5.7 (daily) and 5.13 (weekly), the

non-aggregated daily results show an increasing trend for subjects; however, for some of the weekly results, the trend is decreasing (e.g., subject hh102). The investigation showed that these subjects tended to submit multiple walking exercises in one day towards the end of the study. There may be several explanations for this behavior, including the decline in user engagement throughout the study discussed by [70]. Since our recommendation system runs daily, only the last exercise performed that day will be considered, which causes the aforementioned issue in the exercise monitoring system.

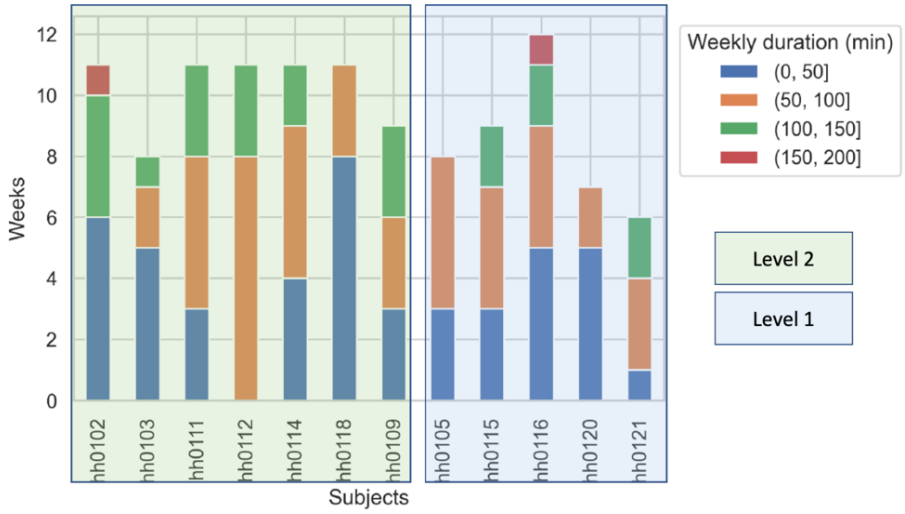


Figure 5.12: Aggregated number of weekly exercise duration performed by the subjects.

Table 5.5: Summary of fixed effects of the HLM fitted to the weekly exercise duration.

		Estimates	CI	P
Fixed Effect				
	Intercept	45.19	23.16-67.21	¡.001
	Time	22.57	-18.19-63.34	.27

Table 5.6: Summary of random effects of the HLM fitted to the weekly exercise duration.

		Standard deviation
Random Effect		
	Residual	36.81
	Intercept	30.15
	Time	56.95

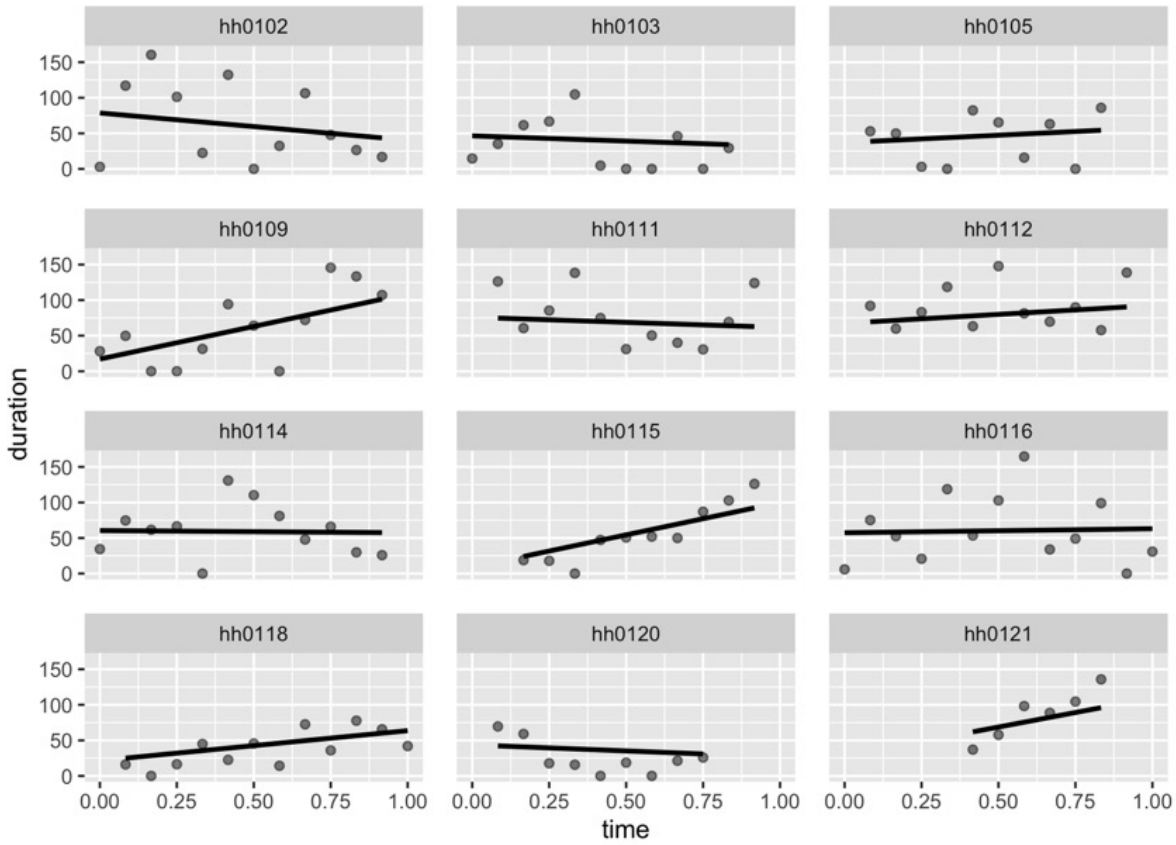


Figure 5.13: Trends in weekly exercise duration.

5.3.4 Recommended exercises

Reward analysis

The average cumulative reward has been monitored to evaluate the recommendation system's performance. In the beginning, since the learner does not know about the subjects, the reward is zero. However, the performance improves as the model explores/exploits different actions given the current context. Figs. 5.15a and 5.15b show the average reward throughout the exercises being done by each group (i.e., Levels 1 and 2). Since two different models were used for the Level 1 group and the Level 2 group, we evaluated them separately. Although the models differed, the overall cumulative average reward follows the same pattern for both groups. Even though the patterns are the same and increasing, the model for Level 2 subjects

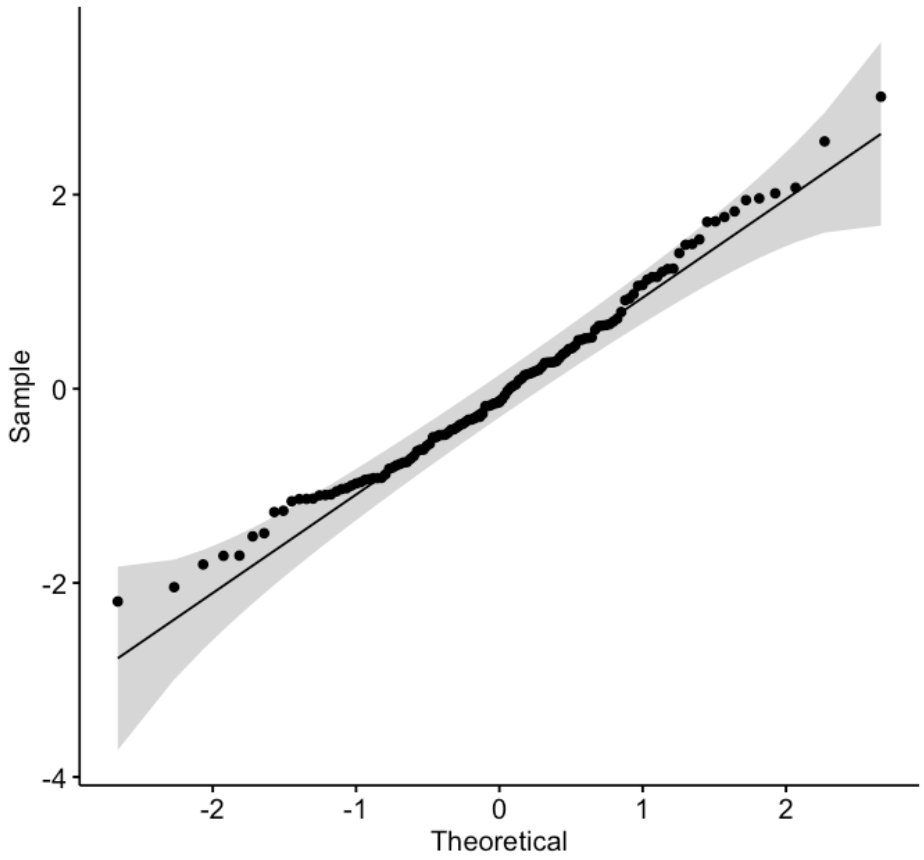
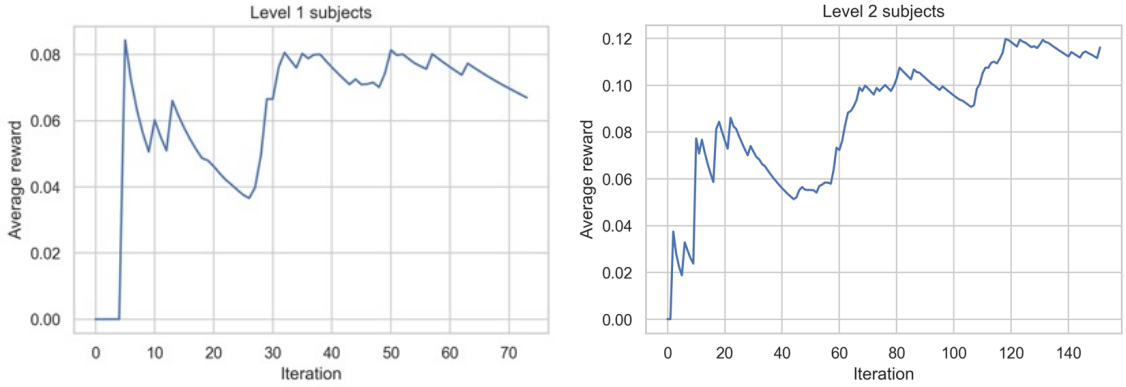


Figure 5.14: Normal Q-Q plot for normality of the residuals for the weekly exercise duration.

performed better due to access to a more significant portion of data (about twice as Level 1 exercise data). Since the nature of our design is similar to active learning techniques, providing more data increases the system's performance.



(a) (b)

Figure 5.15: Average cumulative rewards for subjects in Level 1 and Level 2.

5.3.5 Participants feedback

At the end of the study, exit surveys were administered to participants to better understand their user experiences in order to improve the design and implementation of future studies. Sixteen participants completed the survey questionnaires. The average satisfaction scores of the walking and recommendation system components of the study were 4.31 (0.60) and 3.69 (0.95), respectively, on a scale of 1 to 5. Moreover, an average score of above 4 was reported in relation to the study's ability to meet the physical activity needs of participants, as well as participants' confidence in their ability to safely engage in the recommended walking exercises.

5.4 Discussion

5.4.1 Principal Results

To the best of our knowledge, this study is the first end-to-end closed-loop physical activity recommendation system in the wild using personalized real-time HR monitoring with a proof-of-concept data collection and recommendation. Our proposed model showed an increase in the daily exercise duration of the users which can be useful in getting inactive individuals to perform some physical activity with a comfortable intensity level.

5.4.2 Limitations

In this study, we had multiple apps designed to capture the data. Users had to manually start the HR monitoring app on their smartwatch, which decreased the convenience of our design. Due to this fact, some users forgot to record their physical activity; hence, retrieving

such exercise data was challenging.

In addition, we implemented both the Bluetooth and wifi connection methods in our design. Due to the lower bandwidth of the Bluetooth connection, we experienced some delays in receiving the exercise data.

Moreover, our study had a cold-start problem which caused some initial uncertainty. A pre-trained model over a baseline period could help make more satisfactory predictions. Similarly, the prediction could drastically improve by increasing the duration of the study; however, studies showed that the users' engagement with the system decreases over time [70].

5.4.3 Comparison with Prior Work

Different systematic reviews studied the effectiveness of mHealth in physical fitness and interventions and the need for mHealth technologies to improve physical health [16, 98, 141]. Through mHealth technologies, it is now possible to implement just-in-time adaptive interventions [47, 17]. Yom-Tov et al. [144] used a reinforcement learning approach on patients with diabetes to motivate them to increase their physical activity using encouraging messages. The results showed an improvement in the number of successful intervention messages. Saponaro et al. [120] performed a similar study with healthy subjects, capturing different contextual features to evaluate the effectiveness of nudges in free-living conditions. Rabbi et al. [111] designed a recommendation system to generate health feedback based on physical activity and the log of foods. Liao et al. [83] designed a reinforcement learning algorithm to recommend a treatment policy (activity) to increase the step counts. Although they evaluated their algorithm's performance using different simulations, the algorithm was not tested in the wild. Mahyari et al. [87] designed a model consisting of two inter-connected recurrent neural networks (RNNs) to suggest a new exercise based on the historical sequence

of performed exercises. They extracted features from the name of exercises using word2vec and natural language processing techniques.

Furthermore, other studies used a contextual bandits recommender algorithm to improve emotion regulation, demonstrating that context is essential for effective emotion regulation [17, 26].

Despite the fact that these studies claim personalization, they suffer from monitoring real-time heart rate during physical activity, as well as the intensity of the physical activity.

5.5 Conclusion

Advances in mHealth and Internet-of-Thing have created a new health monitoring and management system era. Nowadays, a general model cannot be used for interventions with different populations, and there is a need for personalized policies and model buildings. Physical activity needs attention since the human body functionally changes person-to-person and over time. Personalization in physical activity recommendations could potentially improve the user's performance as well as his/her engagement. In this study, we proposed a reinforcement learning-based exercise recommendation system that utilizes a person's biomarkers and a context to suggest a new walking exercise that maximizes the user's aerobic capacity. We showed that this system works in an active learning environment.

As a future direction, we are designing such a system for pregnant women, and the reward module works with heart rate reserved since pregnant women's HR norms changes during pregnancy. In addition to the features mentioned in this study, we are utilizing other context features such as sleep quality for the prediction task.

Chapter 6

Summary and Conclusions

Insofar as it is a measure of pleasure, well-being is an essential aspect of our lives and society. We prioritize sleep, physical activity, and mental health as well-being pillars because they could signify body health. Furthermore, the entire world has been substantially affected by a worldwide viral pandemic, which might have a significant impact on societies with sensitive well-being characteristics. As a result, we evaluated the impact of COVID-19 as one of the representations of social well-being threats. Measurement, monitoring, and promotion of well-being may be beneficial to those engaged in disease prevention and health promotion.

The advent of the Internet of Things (IoT) has made it feasible to monitor health outcomes and biomarkers in everyday free-living conditions without having to go to labs or clinical settings. Taking the above into account, this dissertation focused on population and individual-level analyses.

For population-level analysis, we validated two wearable devices with healthy subjects and examined the sleep trends of pregnant women before and during the COVID-19 lockdown. Later on, we investigated the rate of different mental health disorder reports during 2020 compared to 2019, and we showed a significant increase of disorders for the younger popula-

tion.

In individual-level analysis, we first utilized deep learning and machine learning approaches to estimate blood pressure signals from ppg signals collected using wearables; then, we used such biomarkers (i.e., heart rate, systolic and diastolic blood pressure) to detect if COVID-19 or other factors caused ARDS. Finally, we designed and implemented a closed-loop exercise recommendation system in the wild to maximize the user's aerobic capacity.

The takeaway discussion would be the possibility of personalized interventions in the areas that need attention. Using population-level trend analysis, we can observe the vulnerabilities and try to facilitate the situation in well-being using personalized interventions. Furthermore, it would be possible to adjust the general well-being trends by providing optimized and personalized interventions/recommendations to individuals based on their daily needs.

Bibliography

[1] A comprehensive timeline of the coronavirus pandemic at 1 year, from China's first case to the present. https://www.businessinsider.com/coronavirus-pandemic-timeline-history-major-events-2020-3. [Online; accessed 21-July-2020].

[2] A comprehensive timeline of the coronavirus pandemic at 6 months, from China's first case to the present. https://www.businessinsider.com/coronavirus-pandemic-timeline-history-major-events-2020-3. [Online; accessed 21-July-2020].

[3] CAnalysis of the relationship between oil and gold prices. http://www.opf.slu.cz/kfi/icfb/proc2011/pdf/58_simakova.pdf. [Online; accessed 23-Mar-2021].

[4] COVID-19 coronavirus outbreak. Worldometer. https://www.worldometers.info/coronavirus/. [Online; accessed 01-Dec-2020].

[5] COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6. [Online; accessed 31-Aug-2020].

[6] Gear Sport. https://www.samsung.com/us/explore/gear-sport/. [Online; accessed 05-Mar-2020].

[7] Oura Ring. https://ouraring.com/. [Online; accessed 06-Mar-2020].

[8] SNOMED Clinical Terms. https://www.snomed.org/. [Online; accessed 15-Aug-2020].

[9] Survivor Corps. https://www.survivorcorps.com/. [Online; accessed 27-Jun-2022].

[10] Wearable Blood Pressure Monitor and Watch, HeartGuide by OMRON. https://omronhealthcare.com/. [Online; accessed 31-Aug-2020].

[11] Well-Being Concepts. https://www.cdc.gov/hrqol/wellbeing.htm. [Online; accessed 01-Jun-2022].

[12] What does the 'Detect Sleep Periods' button do and how does it work? Actigraph. https://actigraphcorp.force.com/support/s/article/What-does-the-Detect-Sleep-Periods-button-do-and-how-does-it-work. [Online; accessed 18-Oct-2020].

[13] ZotCare. https://healthscitech.nursing.uci.edu/projects. [Online; accessed 01-Aug-2022].

[14] University of California Health creates centralized data set to accelerate COVID-19 research. https://www.universityofcalifornia.edu/press-room/university-california-health-creates-centralized-data-set-accelerate-covid-19-rese 2020. [Online; accessed 15-Aug-2020].

[15] Whole genome of novel coronavirus, 2019-nCoV, sequenced. www.sciencedaily.com/releases/2020/01/200131114748.htm, 31 January 2020. [Online; accessed 31-Aug-2020].

[16] I. Adjerid, G. Loewenstein, R. Purta, and A. Striegel. Gain-loss incentives and physical activity: the role of choice and wearable health tools. *Management Science*, 68(4):2642–2667, 2022.

[17] M. K. Ameko, M. L. Beltzer, L. Cai, M. Boukhechba, B. A. Teachman, and L. E. Barnes. Offline contextual multi-armed bandits for mobile health interventions: A case study on emotion regulation. In *Fourteenth ACM Conference on Recommender Systems*, pages 249–258, 2020.

[18] S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, and C. P. Pollak. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, 26(3):342–392, 2003.

[19] S. A. H. Aqajari et al. An end-to-end and accurate ppg-based respiratory rate estimation approach using cycle generative adversarial networks. *arXiv preprint arXiv:2105.00594*, 2021.

[20] S. A. H. Aqajari et al. Pain assessment tool with electrodermal activity for postoperative patients: Method validation study. *JMIR mHealth and uHealth*, 9(5):e25258, 2021.

[21] S. A. H. Aqajari et al. pyeda: An open-source python toolkit for pre-processing and feature extraction of electrodermal activity. *Procedia Computer Science*, 184:99–106, 2021.

[22] M. Asgari Mehrabadi et al. Detection of covid-19 using heart rate and blood pressure: Lessons learned from patients with ards. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2140–2143. IEEE, 2021.

[23] E. Ash, G. Gauthier, and P. Widmer. Text semantics capture political and economic narratives. *arXiv preprint arXiv:2108.01720*, 2021.

[24] I. Azimi, O. Oti, S. Labbaf, H. Niela-Vilen, A. Axelin, N. Dutt, P. Liljeberg, and A. M. Rahmani. Personalized maternal sleep quality assessment: An objective iot-based longitudinal study. *IEEE Access*, 7:93433–93447, 2019.

[25] I. Azimi, A. M. Rahmani, P. Liljeberg, and H. Tenhunen. Internet of things for remote elderly monitoring: a study from user-centered perspective. *Journal of ambient intelligence and humanized computing*, 8(2):273–289, 2017.

[26] M. L. Beltzer, M. K. Ameko, K. E. Daniel, A. R. Daros, M. Boukhechba, L. E. Barnes, and B. A. Teachman. Building an emotion regulation recommender algorithm for socially anxious individuals using contextual bandits. *British Journal of Clinical Psychology*, 61:51–72, 2022.

[27] R. B. Berry and M. H. Wagner. *Sleep medicine pearls e-book*. Elsevier Health Sciences, 2014.

[28] L. Besedovsky, T. Lange, and J. Born. Sleep and immune function. *Pflügers Archiv-European Journal of Physiology*, 463(1):121–137, 2012.

[29] A. Bietti, A. Agarwal, and J. Langford. A contextual bandit bake-off. *J. Mach. Learn. Res.*, 22:133–1, 2021.

[30] J. M. Bland and D. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476):307–310, 1986.

[31] N. Bolger and J.-P. Laurenceau. *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford press, 2013.

[32] S. Brouwers et al. Arterial hypertension. 2021.

[33] A. Callahan *et al.* Estimating the efficacy of symptom-based screening for covid-19. *NPJ digital medicine*, 3(1):1–3, 2020.

[34] R. Cao, I. Azimi, F. Sarhaddi, H. Niela-Vilen, A. Axelin, P. Liljeberg, A. M. Rahmani, et al. Accuracy assessment of oura ring nocturnal heart rate and heart rate variability in comparison with electrocardiography in time and frequency domains: Comprehensive analysis. *Journal of Medical Internet Research*, 24(1):e27487, 2022.

[35] J. R. Carter, B. M. Gervais, J. L. Adomeit, and I. M. Greenlund. Subjective and objective sleep differ in male and female collegiate athletes. *Sleep Health*, 6(5):623–628, 2020.

[36] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin. Automatic sleep/wake identification from wrist activity. *Sleep*, 15(5):461–469, 1992.

[37] J. D. Cook, M. L. Prairie, and D. T. Plante. Utility of the fitbit flex to evaluate sleep in major depressive disorder: a comparison against polysomnography and wrist-worn actigraphy. *Journal of affective disorders*, 217:299–305, 2017.

[38] J. D. Cook, M. L. Prairie, and D. T. Plante. Ability of the multisensory jawbone up3 to quantify and classify sleep in patients with suspected central disorders of hypersomnolence: a comparison against polysomnography and actigraphy. *Journal of Clinical Sleep Medicine*, 14(5):841–848, 2018.

[39] M. É. Czeisler, R. I. Lane, J. F. Wiley, C. A. Czeisler, M. E. Howard, and S. M. Rajaratnam. Follow-up survey of us adult reports of mental health, substance use, and suicidal ideation during the covid-19 pandemic, september 2020. *JAMA network open*, 4(2):e2037665–e2037665, 2021.

[40] P. de Souto Barreto. Global health agenda on non-communicable diseases: has who set a smart goal for physical activity? *bmj*, 350, 2015.

[41] M. de Zambotti, F. C. Baker, and I. M. Colrain. Validation of sleep-tracking technology compared with polysomnography in adolescents. *Sleep*, 38(9):1461–1468, 2015.

[42] M. de Zambotti, F. C. Baker, A. R. Willoughby, J. G. Godino, D. Wing, K. Patrick, and I. M. Colrain. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. *Physiology & behavior*, 158:143–149, 2016.

[43] M. de Zambotti, S. Claudatos, S. Inkelis, I. M. Colrain, and F. C. Baker. Evaluation of a consumer fitness-tracking device to assess sleep in adults. *Chronobiology international*, 32(7):1024–1028, 2015.

[44] M. de Zambotti, L. Rosas, I. M. Colrain, and F. C. Baker. The sleep of the ring: comparison of the ōura sleep tracker against polysomnography. *Behavioral sleep medicine*, 17(2):124–136, 2019.

[45] H. L. Dunn. High-level wellness for man and society. *American journal of public health and the nations health*, 49(6):786–792, 1959.

[46] D. Ettehad et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. *The Lancet*, 387(10022):957–967, 2016.

[47] J. Fang, V. Lee, and H. Wang. Dynamic physical activity recommendation on personalised mobile health information service: A deep reinforcement learning approach. *arXiv preprint arXiv:2204.00961*, 2022.

[48] T. Ferguson, A. V. Rowlands, T. Olds, and C. Maher. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *International journal of behavioral nutrition and physical activity*, 12(1):1–9, 2015.

[49] A. D. T. Force *et al.* Acute respiratory distress syndrome. *Jama*, 307(23):2526–2533, 2012.

[50] E. Füzéki and W. Banzer. Physical activity recommendations for health and beyond in currently inactive populations. *International journal of environmental research and public health*, 15(5):1042, 2018.

[51] L. Gattinoni *et al.* Covid-19 pneumonia: Ards or not?, 2020.

[52] J. Girschik, L. Fritschi, J. Heyworth, and F. Waters. Validation of self-reported sleep against actigraphy. *Journal of epidemiology*, 22(5):462–468, 2012.

[53] A. L. Goldberger et al. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

[54] I. J. Goodfellow et al. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014.

[55] K. Grym, H. Niela-Vilén, E. Ekholm, L. Hamari, I. Azimi, A. Rahmani, P. Liljeberg, E. Löyttyniemi, and A. Axelin. Feasibility of smart wristbands for continuous monitoring during pregnancy and one month after birth. *BMC pregnancy and childbirth*, 19(1):1–9, 2019.

[56] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami. Internet of things (iot): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7):1645–1660, 2013.

[57] Y.-R. Guo, Q.-D. Cao, Z.-S. Hong, Y.-Y. Tan, S.-D. Chen, H.-J. Jin, K.-S. Tan, D.-Y. Wang, and Y. Yan. The origin, transmission and clinical therapies on coronavirus disease 2019 (covid-19) outbreak–an update on the status. *Military Medical Research*, 7(1):1–10, 2020.

[58] R. Gupta, S. Dahiya, and M. S. Bhatia. Effect of depression on sleep: Qualitative or quantitative? *Indian journal of psychiatry*, 51(2):117, 2009.

[59] S. Haghayegh, S. Khoshnevis, M. H. Smolensky, K. R. Diller, R. J. Castriotta, et al. Accuracy of wristband fitbit models in assessing sleep: systematic review and meta-analysis. *Journal of medical Internet research*, 21(11):e16273, 2019.

[60] W. L. Haskell, H. J. Montoye, and D. Orenstein. Physical activity and exercise to achieve health-related physical fitness components. *Public health reports*, 100(2):202, 1985.

[61] K. He et al. Deep residual learning for image recognition. In *Proceedings of the IEEE CVPR*, 2016.

[62] F. Heydari et al. A chest-based continuous cuffless blood pressure method: Estimation and evaluation using multiple body sensors. *Information Fusion*, 54:119–127, 2020.

[63] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[64] B. Huang et al. Mlp-bp: A novel framework for cuffless blood pressure measurement with ppg and ecg signals based on mlp-mixer neural networks. *Biomedical Signal Processing and Control*, 73:103404, 2022.

[65] P. Isola et al. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on CVPR*, 2017.

[66] L. Jehi *et al.* Individualizing risk prediction for positive covid-19 testing: results from 11,672 patients. *Chest*, 2020.

[67] T. Jimah, H. Borg, P. Kehoe, P. Pimentel, A. Turner, S. Labbaf, M. A. Mehrabadi, A. M. Rahmani, N. Dutt, Y. Guo, et al. A technology-based pregnancy health and wellness intervention (two happy hearts): Case study. *JMIR formative research*, 5(11):e30991, 2021.

[68] J. Johnson et al. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*. Springer, 2016.

[69] E.-J. Kim and J. E. Dimsdale. The effect of psychosocial stress on sleep: a review of polysomnographic evidence. *Behavioral sleep medicine*, 5(4):256–278, 2007.

[70] P. Klasnja, S. Smith, N. J. Seewald, A. Lee, K. Hall, B. Luers, E. B. Hekler, and S. A. Murphy. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of heartsteps. *Annals of Behavioral Medicine*, 53(6):573–582, 2019.

[71] B. P. Kolla, S. Mansukhani, and M. P. Mansukhani. Consumer sleep tracking devices: a review of mechanisms, validity and utility. *Expert review of medical devices*, 13(5):497–506, 2016.

[72] H. Koskimäki, H. Kinnunen, T. Kurppa, and J. Röning. How do we sleep: a case study of sleep duration and quality using data from oura ring. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 714–717, 2018.

[73] Y. Kurylyak et al. A neural network-based method for continuous blood pressure estimation from a ppg signal. In *2013 IEEE International instrumentation and measurement technology conference (I2MTC)*, pages 280–283. IEEE, 2013.

[74] J. Kvedar, M. J. Coye, and W. Everett. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. *Health affairs*, 33(2):194–199, 2014.

[75] N. Lagadec, M. Steinecker, A. Kapassi, A. M. Magnier, J. Chastang, S. Robert, N. Gaouaou, and G. Ibanez. Factors influencing the quality of life of pregnant women: a systematic review. *BMC pregnancy and childbirth*, 18(1):1–14, 2018.

[76] C. Landry et al. Nonlinear dynamic modeling of blood pressure waveform: Towards an accurate cuffless monitoring system. *IEEE Sensors Journal*, 20(10):5368–5378, 2020.

[77] G. J. Landry, J. R. Best, and T. Liu-Ambrose. Measuring sleep quality in older adults: a comparison using subjective and objective methods. *Frontiers in aging neuroscience*, 7:166, 2015.

[78] T. Lappalainen, L. Virtanen, and J. Häkkilä. Experiences with wellness ring and bracelet form factor. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, pages 351–353, 2016.

[79] D. S. Lauderdale, K. L. Knutson, L. L. Yan, K. Liu, and P. J. Rathouz. Self-reported and measured sleep duration: how similar are they? *Epidemiology*, pages 838–845, 2008.

[80] P. Li et al. Novel wavelet neural network algorithm for continuous and noninvasive dynamic estimation of blood pressure from photoplethysmography. *Science China Information Sciences*, 59(4):1–10, 2016.

[81] X. Li *et al.* Acute respiratory failure in covid-19: is it "typical" ards? *Critical Care*, 24:1–5, 2020.

[82] Z. Liang and M. A. Chapa Martell. Validity of consumer activity wristbands and wearable eeg for measuring overall sleep parameters and sleep structure in free-living conditions. *Journal of Healthcare Informatics Research*, 2(1):152–178, 2018.

[83] P. Liao, K. Greenewald, P. Klasnja, and S. Murphy. Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–22, 2020.

[84] W. Lin et al. Energy-efficient blood pressure monitoring based on single-site photoplethysmogram on wearable devices. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 504–507. IEEE, 2021.

[85] Y. Lu, Z. Xu, and A. Tewari. Bandit algorithms for precision medicine. *arXiv preprint arXiv:2108.04782*, 2021.

[86] L. v. d. Maaten *et al.* Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[87] A. Mahyari and P. Pirolli. Physical exercise recommendation and success prediction using interconnected recurrent neural networks. In *2021 IEEE International Conference on Digital Health (ICDH)*, pages 148–153. IEEE, 2021.

[88] J. Mantua, N. Gravel, and R. Spencer. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. *Sensors*, 16(5):646, 2016.

[89] M. Marino, Y. Li, M. N. Rueschman, J. W. Winkelman, J. Ellenbogen, J. M. Solet, H. Dulin, L. F. Berkman, and O. M. Buxton. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep*, 36(11):1747–1755, 2013.

[90] K. Matsumura *et al.* Comparison of the clinical course of covid-19 pneumonia and acute respiratory distress syndrome in 2 passengers from the cruise ship diamond princess in february 2020. *The American journal of case reports*, 21:e926835–1, 2020.

[91] M. A. Mehrabadi, S. A. H. Aqajari, A. H. A. Zargari, N. Dutt, and A. M. Rahmani. Novel blood pressure waveform reconstruction from photoplethysmography using cycle generative adversarial networks. In *2022 44rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022.

[92] M. A. Mehrabadi, I. Azimi, F. Sarhaddi, A. Axelin, H. Niela-Vilén, S. Myllyntausta, S. Stenholm, N. Dutt, P. Liljeberg, A. M. Rahmani, et al. Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study. *JMIR mHealth and uHealth*, 8(11):e20465, 2020.

[93] M. A. Mehrabadi, N. Dutt, A. M. Rahmani, et al. The causality inference of public interest in restaurants and bars on daily covid-19 cases in the united states: Google trends analysis. *JMIR public health and surveillance*, 7(4):e22880, 2021.

[94] M. A. Mehrabadi et al. Detection of covid-19 using heart rate and blood pressure: Lessons learned from patients with ards. *arXiv preprint arXiv:2011.10470*, 2020.

[95] L. J. Meltzer, L. S. Hiruma, K. Avis, H. Montgomery-Downs, and J. Valentin. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. *Sleep*, 38(8):1323–1330, 2015.

[96] L. J. Meltzer, C. M. Walsh, J. Traylor, and A. M. Westin. Direct comparison of two new actigraphs and polysomnography in children and adolescents. *Sleep*, 35(1):159–166, 2012.

[97] R. Mieronkoski, I. Azimi, A. M. Rahmani, R. Aantaa, V. Terävä, P. Liljeberg, and S. Salanterä. The internet of things for basic nursing care—a scoping review. *International journal of nursing studies*, 69:78–90, 2017.

[98] M. Milne-Ives, C. Lam, C. De Cock, M. H. Van Velthoven, E. Meinert, et al. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. *JMIR mHealth and uHealth*, 8(3):e17046, 2020.

[99] M. Miyachi, J. TripeTTe, R. KawaKaMi, and H. MuraKaMi. "+ 10 min of physical activity per day": Japan is looking for efficient but feasible recommendations for its population. *Journal of nutritional science and vitaminology*, 61(Supplement):S7–S9, 2015.

[100] P. Moen, E. L. Kelly, E. Tranby, and Q. Huang. Changing work, changing health: can real work-time flexibility promote health behaviors and well-being? *Journal of health and social behavior*, 52(4):404–429, 2011.

[101] H. E. Montgomery-Downs, S. P. Insana, and J. A. Bond. Movement toward a novel activity monitoring device. *Sleep and Breathing*, 16(3):913–917, 2012.

[102] F. Moreno-Pino, A. Porras-Segovia, P. López-Esteban, A. Artés, and E. Baca-García. Validation of fitbit charge 2 and fitbit alta hr against polysomnography for assessing sleep in adults with obstructive sleep apnea. *Journal of Clinical Sleep Medicine*, 15(11):1645–1653, 2019.

[103] S. S. Mousavi et al. Blood pressure estimation from appropriate and inappropriate ppg signals using a whole-based method. *Biomedical Signal Processing and Control*, 47:196–206, 2019.

[104] R. Mukkamala et al. Continuous cardiac output monitoring by peripheral blood pressure waveform analysis. *IEEE Transactions on Biomedical Engineering*, 53(3):459–467, 2006.

[105] H. Niela-Vilén, J. Auxier, E. Ekholm, F. Sarhaddi, M. Asgari Mehrabadi, A. Mahmoudzadeh, I. Azimi, P. Liljeberg, A. M. Rahmani, and A. Axelin. Pregnant women's daily patterns of well-being before and during the covid-19 pandemic in finland: Longitudinal monitoring through smartwatch technology. *PloS one*, 16(2):e0246494, 2021.

[106] H. Niela-Vilen, I. Azimi, K. Suorsa, F. Sarhaddi, S. Stenholm, P. Liljeberg, A. M. Rahmani, and A. Axelin. Comparison of oura smart ring against actigraph accelerometer for measurement of physical activity and sedentary time in a free-living context. *CIN: Computers, Informatics, Nursing*, 2022.

[107] E. O'brien et al. Blood pressure measuring devices: recommendations of the european society of hypertension. *Bmj*, 322(7285):531–536, 2001.

[108] J. M. Peake, G. Kerr, and J. P. Sullivan. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. *Frontiers in physiology*, 9:743, 2018.

[109] I. Perez-Pozuelo, B. Zhai, J. Palotti, R. Mall, M. Aupetit, J. M. Garcia-Gomez, S. Taheri, Y. Guan, and L. Fernandez-Luque. The future of sleep health: a data-driven revolution in sleep science and medicine. *NPJ digital medicine*, 3(1):1–15, 2020.

[110] J. Qi, P. Yang, G. Min, O. Amft, F. Dong, and L. Xu. Advanced internet of things for personalised healthcare systems: A survey. *Pervasive and Mobile Computing*, 41:132–149, 2017.

[111] M. Rabbi, M. H. Aung, M. Zhang, and T. Choudhury. Mybehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 707–718, 2015.

[112] S. W. Raudenbush and A. S. Bryk. *Hierarchical linear models: Applications and data analysis methods*, volume 1. sage, 2002.

[113] M. Reading Turchioe, L. V. Grossman, A. C. Myers, J. Pathak, and R. M. Creber. Correlates of mental health symptoms among us adults during covid-19, march–april 2020. *Public Health Reports*, 136(1):97–106, 2021.

[114] D. Riebe, J. K. Ehrman, G. Liguori, M. Magal, A. C. of Sports Medicine, et al. *ACSM's guidelines for exercise testing and prescription*. Wolters Kluwer, 2018.

[115] B. M. Roane, E. Van Reen, C. N. Hart, R. Wing, and M. A. Carskadon. Estimating sleep from multisensory armband measurements: validity and reliability in teens. *Journal of sleep research*, 24(6):714–721, 2015.

[116] M. Rong et al. A multi-type features fusion neural network for blood pressure prediction based on photoplethysmography. *Biomedical Signal Processing and Control*, 68:102772, 2021.

[117] A. Sadeh. The role and validity of actigraphy in sleep medicine: an update. *Sleep medicine reviews*, 15(4):259–267, 2011.

[118] I. Sadek, J. Biswas, and B. Abdulrazak. Ballistocardiogram signal processing: a review. *Health information science and systems*, 7(1):1–23, 2019.

[119] M. M. Sánchez-Ortuño, J. D. Edinger, M. K. Means, and D. Almirall. Home is where sleep is: an ecological approach to test the validity of actigraphy for the assessment of insomnia. *Journal of Clinical Sleep Medicine*, 6(1):21–29, 2010.

[120] M. Saponaro, A. Vemuri, G. Dominick, and K. Decker. Contextualization and individualization for just-in-time adaptive interventions to reduce sedentary behavior. In *Proceedings of the Conference on Health, Inference, and Learning*, pages 246–256, 2021.

[121] N. Sattari, M. A. Mehrabadi, S. A. H. Aqajari, J. Zhang, K. Simon, E. Alzueta, T. Dulai, M. de Zambotti, F. Baker, A. Rahmani, et al. Sleep quality prediction during the menstrual cycle based on daily sleep diary reports. In *SLEEP*, volume 44, pages A33–A33. OXFORD UNIV PRESS INC JOURNALS DEPT, 2001 EVANS RD, CARY, NC 27513 USA, 2021.

[122] J. L. Scheid and E. O'Donnell. Revisiting heart rate target zones through the lens of wearable technology. *ACSM's Health & Fitness Journal*, 23(3):21–26, 2019.

[123] T. Schnabel, A. Swaminathan, A. Singh, N. Chandak, and T. Joachims. Recommendations as treatments: Debiasing learning and evaluation. In *international conference on machine learning*, pages 1670–1679. PMLR, 2016.

[124] H. Scott, L. Lack, and N. Lovato. A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. *Sleep Medicine Reviews*, 49:101227, 2020.

[125] S. P. Shashikumar *et al.* Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *Journal of electrocardiology*, 50(6):739–743, 2017.

[126] D. J. Sheskin. *Handbook of parametric and nonparametric statistical procedures*. crc Press, 2020.

[127] M. Spielmanns, D. Bost, W. Windisch, P. Alter, T. Greulich, C. Nell, J. H. Storre, A. R. Koczulla, and T. Boeselt. Measuring sleep quality and efficiency with an activity monitoring device in comparison to polysomnography. *Journal of clinical medicine research*, 11(12):825, 2019.

[128] D. Stammbach and E. Ash. Docscan: Unsupervised text classification via learning from neighbors. *arXiv preprint arXiv:2105.04024*, 2021.

[129] S. J. Strath, L. A. Kaminsky, B. E. Ainsworth, U. Ekelund, P. S. Freedson, R. A. Gary, C. R. Richardson, D. T. Smith, and A. M. Swartz. Guide to the assessment of physical activity: clinical and research applications: a scientific statement from the american heart association. *Circulation*, 128(20):2259–2279, 2013.

[130] D. Talevi, V. Socci, M. Carai, G. Carnaghi, S. Faleri, E. Trebbi, A. di Bernardo, F. Capelli, and F. Pacitti. Mental health outcomes of the covid-19 pandemic. *Rivista di psichiatria*, 55(3):137–144, 2020.

[131] X. Tang *et al.* Comparison of hospitalized patients with ards caused by covid-19 and h1n1 [published online ahead of print, 2020 mar 26]. *Chest*, pages 30558–4, 2020.

[132] M. S. Tanveer et al. Cuffless blood pressure estimation from electrocardiogram and photoplethysmogram using waveform based ann-lstm network. *Biomedical Signal Processing and Control*, 51:382–392, 2019.

[133] A. Tazarv et al. A deep learning approach to predict blood pressure from ppg signals. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 5658–5662. IEEE, 2021.

[134] G. Thambiraj et al. Investigation on the effect of womersley number, ecg and ppg features for cuff less blood pressure estimation using machine learning. *Biomedical Signal Processing and Control*, 60:101942, 2020.

[135] E. Tobaldini, E. M. Fiorelli, M. Solbiati, G. Costantino, L. Nobili, and N. Montano. Short sleep duration and cardiometabolic risk: from pathophysiology to clinical evidence. *Nature Reviews Cardiology*, 16(4):213–224, 2019.

[136] E. Toon, M. J. Davey, S. L. Hollis, G. M. Nixon, R. S. Horne, and S. N. Biggs. Comparison of commercial wrist-based and smartphone accelerometers, actigraphy, and psg in a clinical cohort of children and adolescents. *Journal of Clinical Sleep Medicine*, 12(3):343–350, 2016.

[137] R. P. Troiano, D. Berrigan, K. W. Dodd, L. C. Masse, T. Tilert, M. McDowell, et al. Physical activity in the united states measured by accelerometer. *Medicine and science in sports and exercise*, 40(1):181, 2008.

[138] P. Varma, M. Junge, H. Meaklim, and M. L. Jackson. Younger people are more vulnerable to stress, anxiety and depression during covid-19 pandemic: A global cross-sectional survey. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 109:110236, 2021.

[139] R. Velik. An objective review of the technological developments for radial pulse diagnosis in traditional chinese medicine. *European Journal of Integrative Medicine*, 7(4):321–331, 2015.

[140] Q. Wang, Y. Liu, and X. Pan. Atmosphere pollutants and mortality rate of respiratory diseases in beijing. *Science of the Total Environment*, 391(1):143–148, 2008.

[141] S. Wang, K. Sporrel, H. van Hoof, M. Simons, R. D. de Boer, D. Ettema, N. Nibbeling, M. Deutekom, and B. Kröse. Reinforcement learning to send reminders at right moments in smartphone exercise application: A feasibility study. *International Journal of Environmental Research and Public Health*, 18(11):6059, 2021.

[142] P. Watson and A. Petrie. Method agreement analysis: a review of correct methodology. *Theriogenology*, 73(9):1167–1179, 2010.

[143] C. Wu *et al.* Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in wuhan, china. *JAMA internal medicine*, 2020.

[144] E. Yom-Tov, G. Feraru, M. Kozdoba, S. Mannor, M. Tennenholtz, I. Hochberg, et al. Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. *Journal of medical Internet research*, 19(10):e7994, 2017.

[145] A. H. A. Zargari et al. An accurate non-accelerometer-based ppg motion artifact removal technique using cyclegan. *arXiv preprint arXiv:2106.11512*, 2021.

[146] M. Zhou, Y. Fukuoka, Y. Mintz, K. Goldberg, P. Kaminsky, E. Flowers, A. Aswani, et al. Evaluating machine learning–based automated personalized daily step goals delivered through a mobile phone app: Randomized controlled trial. *JMIR mHealth and uHealth*, 6(1):e9117, 2018.

[147] J.-Y. Zhu et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[148] J.-Y. Zhu et al. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In *2017 IEEE ICCV*, 2017.

Appendix A

Mental health disorders

A.1 Different mental health disorders and their corresponding categories.

Acute panic state due to acute stress reaction	Anxiety
Acute post-trauma stress state	Anxiety
Acute Stress disorder	Anxiety
Adjustment disorder with mixed anxiety and depressed mood	Anxiety
Anxiety disorder	Anxiety
Anxiety disorder of childhood OR adolescence	Anxiety
Anxiety State	Anxiety
Dream anxiety disorder	Anxiety
Generalized anxiety disorder	Anxiety
Organic anxiety disorder	Anxiety
Posttraumatic stress disorder	Anxiety

Stress	Anxiety
Social phobia	Anxiety
Obsessive-compulsive disorder	Anxiety
Panic disorder with agoraphobia	Anxiety
Agoraphobia without history of panic disorderwithout limited symptom attacks	Anxiety
Adjustment disorder with anxious mood	Anxiety
Panic disorder without agoraphobia	Anxiety
Agoraphobia without history of panic disorder	Anxiety
Phobia	Anxiety
Claustrophobia	Anxiety
Generalized social phobia	Anxiety
Agoraphobia	Anxiety
Bipolar affective disorder, current episode depression	Bipolar
Bipolar affective disorder, current episode manic	Bipolar
Bipolar affective disorder, current episode mixed	Bipolar
Bipolar affective disorder, currently depressed, in full remission	Bipolar
Bipolar affective disorder, currently depressed, mild	Bipolar
Bipolar affective disorder, currently depressed, moderate	Bipolar
Bipolar affective disorder, currently manic, in full remission	Bipolar
Bipolar affective disorder, currently manic, moderate	Bipolar
Bipolar affective disorder, currently manic, severe, with psychosis	Bipolar

Bipolar disorder	Bipolar
Bipolar disorder in partial remission	Bipolar
Bipolar disorder in remission	Bipolar
Bipolar I disorder	Bipolar
Bipolar I disorder, most recent episode hypomanic	Bipolar
Bipolar I disorder, single manic episode	Bipolar
Bipolar II disorder	Bipolar
Depressed bipolar I disorder	Bipolar
Depressed bipolar I disorder in full remission	Bipolar
Depressed bipolar I disorder in partial remission	Bipolar
manic bipolar I disorder in full remission	Bipolar
Manic bipolar I disorder in partial remission	Bipolar
mixed bipolar affective disorder, mild	Bipolar
mixed bipolar affective disorder, moderate	Bipolar
mixed bipolar I disorder	Bipolar
mixed bipolar I disorder in partial remission	Bipolar
moderate mixed bipolar I disorder	Bipolar
Psychosis and severe depression co-occurrent and due to bipolar affective disorder	Bipolar
severe bipolar I disorder, single manic episode with psychotic features	Bipolar
severe bipolar II disorder, most recent episode major depressive, in partial remission	Bipolar
severe depressed bipolar I disorder with psychotic features	Bipolar

severe depressed bipolar I disorder without psychotic features	Bipolar
severe manic bipolar I disorder without psychotic features	Bipolar
severe mixed bipolar I disorder without psychotic features	Bipolar
Mixed bipolar affective disorder, severe, with psychosis	Bipolar
Psychosis and severe depression co-occurrent and due to bipolar affective disorder	Bipolar
Bipolar affective disorder, currently manic, severe, with psychosis	Bipolar
Adjustment disorder with disturbance of conduct	child/adolescent
Adjustment disorder with mixed disturbance of emotions AND conduct	child/adolescent
Attention deficit hyperactivity disorder	child/adolescent
Nonaggressive unsocial conduct disorder	child/adolescent
Impulse control disorder	child/adolescent
Physical aggression	child/adolescent
Chronic motor tic disorder	child/adolescent
Autism spectrum disorder	child/adolescent
Attention deficit hyperactivity disorder, combined type	child/adolescent
Conduct disorder, childhood-onset type	child/adolescent
Transient tic disorder	child/adolescent
Conduct disorder	child/adolescent
Tic disorder	child/adolescent
Conduct disorder, adolescent-onset type	child/adolescent

Disruptive mood dysregulation disorder	child/adolescent
Aggressive unsocial conduct disorder	child/adolescent
Attention deficit hyperactivity disorder, predominantly hyperactive impulsive type	child/adolescent
Attention deficit hyperactivity disorder, predominantly inattentive type	child/adolescent
Hperkinetic conduct disorder	child/adolescent
Intermittent explosive disorder	child/adolescent
Major depression single episode	depression
Major depression single episode, in partial remission	depression
Mild major depression, single episode	depression
Mild recurrent major depression	depression
Moderate major depression, single episode	depression
Moderate recurrent major depression	depression
Postpartum depression	depression
Recurrent Major depression	depression
Recurrent major depression in full remission	depression
Recurrent major depression in partial remission	depression
Recurrent major depression in remission	depression
Severe major depression, single episode, with psychotic features	depression
Severe major depression, single episode, without psychotic features	depression
Severe recurrent major depression without psychotic features	depression
Single episode of major depression in full remission	depression

recurrent depression	depression
Adjustment disorder with depressed mood	depression
Reactive depressive psychosis	depression
Recurrent major depressive episodes, severe, with psychosis	depression
dysthymia	depression
Adjustment disorder with mixed anxiety and depressed mood	mood
Mood disorder due to a general medical condition	mood
Mood disorderwith manic features due to general medical condition	mood
Mood disorder with mixed features due to general medical condition	mood
Mood swings	mood
Episodic mood disorder	mood
Mood disorder with major depressive-like episode due to general medical condition	mood
Mood disorderwith depressive features due to general medical condition	mood
Mood disorder	mood
Borderline personality disorder	personality
schizoaffective disorder, bipolar type	psychosis
Brief reactive psychosis	psychosis
Psychotic disorder	psychosis
Chronic disorganized schizophrenia	psychosis
Chroniclatent schizophrenia	psychosis

Residual schizophrenia	psychosis
Schizoaffective disorder, bipolar type	psychosis
Schizophrenia in remission	psychosis
Schizoaffective disorder, depressive type	psychosis
Schizophrenia	psychosis
Schizophreniform disorder	psychosis
Acute exacerbation of chronic schizophrenia	psychosis
Chronic paranoid schizophrenia	psychosis
Delusional disorder	psychosis
Latent schizophrenia	psychosis
Paranoid schizophrenia	psychosis
Schizoaffective disorder	psychosis
Schizoaffective schizophrenia in remission	psychosis
Chronic disorganized schizophrenia with acute exacerbation	psychosis
Subchronic schizoaffective schizophrenia	psychosis
Subchronic schizophrenia	psychosis
Chronic schizoaffective schizophrenia	psychosis
Disor anized schizophrenia	psychosis
Simple schizophrenia	psychosis
Acute exacerbation of chronic catatonic schizophrenia	psychosis
Catatonic schizophrenia	psychosis
Chronic undifferentiated schizophrenia	psychosis
Undifferentiated schizophrenia	psychosis
Acute exacerbation of chronic schizoaffective schizophrenia	psychosis

Chronic residual schizophrenia	psychosis
Organic delusional disorder	psychosis
Chronic post-traumatic stress disorder	PTSD
Psychoactive substance-induced withdrawal syndrome	substance use
Psychoactive substance use disorder	substance use
Alcohol amnestic disorder	substance use
Alcohol-induced anxiety disorder	substance-induced mental disorder
Cannabis-induced anxiety disorder	substance-induced mental disorder
Psychoactive substance-induced organic anxiety disorder	substance-induced mental disorder
Psychoactive substance-induced organic intoxication	substance-induced mental disorder
Cocaine-induced mood disorder	substance-induced mental disorder
Alcohol-induced mood disorder	substance-induced mental disorder
Drug-induced delusional disorder	substance-induced mental disorder
Psychoactive substance abuse	substance-induced mental disorder
Sedative, hypnoticAND/OR anxiolytic-induced mood disorder	substance-induced mental disorder
Drug-induced mood disorder	substance-induced mental disorder
Sedative, hypnotic AND/OR anxiolytic-induced psychotic disorder with delusions	substance-induced mental disorder
Opioid-induced psychotic disorder with hallucinations	substance-induced mental disorder
Psychoactive substance-induced organic anxiety disorder	substance-induced mental disorder
Cocaine-induced psychotic disorder with hallucinations	substance-induced mental disorder
Opioid-induced psychotic disorder with delusions	substance-induced mental disorder

Psychoactive substance-induced organic delusional disorder	substance-induced mental disorder
Psychoactive substance-induced organic mental disorder	substance-induced mental disorder
Alcohol-induced psychotic disorder with delusions	substance-induced mental disorder
Cannabis delusional disorder	substance-induced mental disorder
Drug-induced psychosis	substance-induced mental disorder
Induced psychotic disorder	substance-induced mental disorder
Psychoactive substance-induced organic hallucinosis	substance-induced mental disorder
Psychoactive substance-induced organic mood disorder	substance-induced mental disorder
Alcohol-induced psychosis	substance-induced mental disorder
Cannabis-induced psychotic disorder with hallucinations	substance-induced mental disorder
Cocaine delusional disorder	substance-induced mental disorder
Opioid-induced mood disorder	substance-induced mental disorder
Psychoactive substance-induced organic amnestic disorder	substance-induced mental disorder
Psychoactive substance-induced organic delirium	substance-induced mental disorder
Suicidal thoughts	suicide

Table A.1

Appendix B

Abbreviations

ABP	Ambulatory blood pressure
ACSM	American College of Sports Medicine
ARDS	Acute respiratory distress syndrome
AUC	Area under the curve
BHS	British Hypertension Society
BP	Blood pressure
CMAB	Contextual multi-armed bandit
CNN	Convolutional neural network
CycleGAN	Cycle generative adversarial network
DBP	Diastolic blood pressure
ECG	Electrocardiographic
GAN	Generative adversarial network
HR	Heart rate
HRV	Heart rate variability
IoT	Internet-of-Things
LSTM	Long Short-Term Memory
MAE	Mean absolute error
MIMIC	Multi-parameter Intelligent Monitoring in Intensive Care
ML	Machine learning
PPG	Photoplethysmogram
PSG	Polysomnography
PTSD	Post-traumatic stress disorder
RL	Reinforcement learning
RMSE	Root mean square error
SBP	Systolic blood pressure
SE	Sleep efficiency
SOL	Sleep onset latency
TST	Total sleep time
WASO	Wake after sleep onset