Biopsychosocial Data Analytics and Modeling

By

MEENA SANTHANAGOPALAN

School of Engineering, Information Technology and Physical Sciences
FEDERATION UNIVERSITY

A dissertation submitted to the Federation University in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Science, Engineering and Information Technology.

SUPERVISOR: DR. MADHU CHETTY
ASSOCIATE SUPERVISOR: DR. CAMERON FOALE
ASSOCIATE SUPERVISOR: DR. BRITT KLEIN

APRIL 2021

AUTHOR'S DECLARATION

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution, and to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

SIGNED: DATE: 26 APRIL 2021

DEDICATED TO

SHRI SHARADAMBAL

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors. Associate Prof Madhu Chetty, my principal supervisor, who has provided me with valuable guidance and support. He helped me not only in academics to improve my presentation skills but has also supported me when I faced personal problems during my candidature. His encouragement and critical suggestions are of great importance to me.

Dr. Cameron Foale, my associate supervisor, has always mentored me during my candidature and supported me to enhance my technical knowledge. His constructive advice and practical suggestions have always helped me deliver better.

Prof Brit Klein, my associate supervisor, has provided me with valuable feedback and suggestions in my research work in her busy schedule.

I am fortunate that I had the best team of supervisors whose guidance, patience and availability has helped me throughout the development and the writing of this thesis.

I thank the School of Science, Psychology and Sport, Federation University Australia, for supplying the necessary data and information from the conducted 'Wellness study' Digital Health Intervention program. I also thank all the staff members who have supported me in this research.

I acknowledge and thank the support provided by the Australian Government Research Training Program (RTP) Fee-Offset Scholarship through Federation University Australia. I am also grateful for some seeding financial support from the School of Science, Psychology and Sport and the Collaborative Research Network (Federation University Australia).

The information in this thesis was edited by Elite Editing, and editorial intervention was restricted to Standards D and E of the Australian Standards for Editing Practice.

I thank my parents and family for their continuous and unparalleled love, help, sacrifice and support. My husband has stood by me like a rock and constantly encouraged me and my little daughter has filled me with positive energy to achieve my goals despite the challenges.

Abstract

Sustained customisation of digital health intervention (DHI) programs, in the context of community health engagement, requires strong integration of multi-sourced interdisciplinary biopsychosocial health data. The biopsychosocial model is built upon the idea that biological, psychological and social processes are integrally and interactively involved in physical health and illness.

One of the longstanding challenges of dealing with healthcare data is the wide variety of data generated from different sources and the increasing need to learn actionable insights that drive performance improvement. The growth of information and communication technology has led to the increased use of DHI programs. These programs use an observational methodology that helps researchers to study the everyday behaviour of participants during the course of the program by analysing data generated from digital tools such as wearables, online surveys and ecological momentary assessment (EMA). Combined with data reported from biological and psychological tests, this provides rich and unique biopsychosocial data.

There is a strong need to review and apply novel approaches to combining biopsychosocial data from a methodological perspective. Although some studies have used data analytics in research on clinical trial data generated from digital interventions, data analytics on biopsychosocial data generated from DHI programs is limited. The study in this thesis develops and implements innovative approaches for analysing the existing unique and rich biopsychosocial data generated from the wellness study, a DHI program conducted by the School of Science, Psychology and Sport at Federation University. The characteristics of variety, value and veracity that usually describe big data are also relevant to the biopsychosocial data handled in this thesis. These historical, retrospective real-life biopsychosocial data provide fertile ground for research through the use of data analytics to discover patterns hidden in the data and to obtain new knowledge.

This thesis presents the studies carried out on three aspects of biopsychosocial research. First, we present the salient traits of the three components - biological, psychological and social - of biopsychosocial research. Next, we investigate the challenges of pre-processing biopsychosocial data, placing special emphasis on the time-series data generated from wearable sensor devices. Finally, we present the application of statistical and machine learning (ML) tools to integrate variables from the biopsychosocial disciplines to build a

predictive model. The first chapter presents the salient features of the biopsychosocial data for each discipline. The second chapter presents the challenges of pre-processing biopsychosocial data, focusing on the time-series data generated from wearable sensor devices. The third chapter uses statistical and ML tools to integrate variables from the biopsychosocial disciplines to build a predictive model. Among its other important analyses and results, the key contributions of the research described in this thesis include the following:

1. using gamma distribution to model neurocognitive reaction time data that presents interesting properties (skewness and kurtosis for the data distribution)
2. using novel 'peak heart-rate' count metric to quantify 'biological' stress
3. using the ML approach to evaluate DHIs
4. using a recurrent neural network (RNN) and long short-term memory (LSTM) data prediction model to predict Difficulties in Emotion Regulation Scale (DERS) and primary emotion (PE) using wearable sensor data.

ABBREVIATIONS

Short-form	Expanded form
AI	Artificial Intelligence
AIC	Akaike information criterion
ANS	Autonomous Nervous System
BDNF	Brain-derived neurotrophic factor
BPS	Biopsychosocial
CART	Classification and regression tree
CRISP-DM	The Cross-Industry Standard Process for Data Mining
DASS	Depression Anxiety Stress scale
DERS	Difficulty in Emotion Regulation scale
DHI	Digital Health Intervention
ECG	Electrocardiogram
ELISA	Enzyme-Linked Immuno Assay
EMA	Ecological Momentary Assessment
FGF2	Fibro-blast Growth Factor
FOH	Faculty of Health
GAM	Generalised Additive Model
GLM	Generalised Linear Models
HRV	Heart Rate Variability
ICT	Information and Communication Technology
IL-6	Interleukin-6
K-6	Kessler Psychological Distress Scale
KS-test	Kolmogorov-Smirnov test
LOT-R	Life Orientation Test - Revised scale
LSTM	Long Short Term Memory
mHealth	mobile Health
MAAS	Mindfulness Attention Awareness scale
MAR	Missing at Random
MHC	Mental Health Continuum scale
MI	Multiple Imputation
ML	Machine Learning
MLE	Maximum Likelihood Estimation
NLP	Natural Language Processing
OSEMN	Obtain, Scrub, Explore, Model, iNterpret
PE	Primary Emotion
qPCR	quantitative Polymerase Chain Reaction

ABBREVIATIONS CONTINUED..

Short-form	Expanded form
RCT	Randomised Controlled Trials
RNN	Recurrent Neural Network
SEMMA	Sample, Explore, Modify, Model, and Assess
TDSP	Team Data Science Process

Related Publications

Published Articles

Conference Articles

- Meena Santhanagopalan, Madhu Chetty, Cameron Foale, Britt Klein: Modeling neurocognitive reaction time with gamma distribution Publication: ACSW '18: Proceedings of the Australasian Computer Science Week MulticonferenceJanuary 2018 Article No.: 28 Pages 1–10 https://doi.org/10.1145/3167918.3167941

- Meena Santhanagopalan, Madhu Chetty, Cameron Foale, Britt Klein: Relevance of frequency of heart-rate peaks as indicator of 'biological' stress level ICONIP 2018: International Conference on Neural Information Processing, Springer, Cham, Switzerland, pp. 598-609, doi: 10.1007/978-3-030-04239-4_54.

Journal Article

- Meena Santhanagopalan, Madhu Chetty, Cameron Foale, Britt Klein: Towards Machine Learning approach for Digital-Health intervention program. Aust. J. Intell. Inf. Process. Syst. 15(3): pages 16-24 (2019) http://ajiips.com.au/papers/V15.4/v15n4_20-28.pdf

Articles awaiting a decision

Journal Article

- Meena Santhanagopalan, Madhu Chetty, Cameron Foale, Britt Klein, 'Artificial neural network based classification of physiological time series data for digital health intervention', Elsevier Journal - Internet Interventions, submitted on 23 April 2021.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

Holistic health is an approach to life. It considers a person as a whole being and how they interact with the environment. It accentuates the connection between body, mind and mood. Its goal is to achieve maximum wellbeing. 'High-level wellness is defined as an integrated method of functioning which is oriented towards maximising the potential of which the individual is capable, within the environment where the individual is functioning' [1]. This approach encourages individuals to recognise their responsibility for their wellbeing and the everyday decisions that affect their health.

1.1 Biopsychosocial model - a holistic approach

In the broad domain of healthcare, for a very long time, wellbeing conditions have been explained in accordance with normative models that are standards for evaluation [2]. Such normative models have given rise to health models that are consistent with the biomedical model and are very reductive in nature. Such models contend that all diseases can be attributed to biological causes in the body. Hence, their treatments are typically biological in nature (such as surgery or medications) [3]. In recent times, health psychology researchers have started implementing the holistic approach using the biopsychosocial model [4]. The biopsychosocial model is an expanded view that attributes disease outcomes to the intricate, variable interactions between different biological factors (genetic, biometric, biochemical), psychological factors (mood, experience, neurocognitive) and social factors (cultural, familial, socio-economic). In this way, the biopsychosocial model advocates a

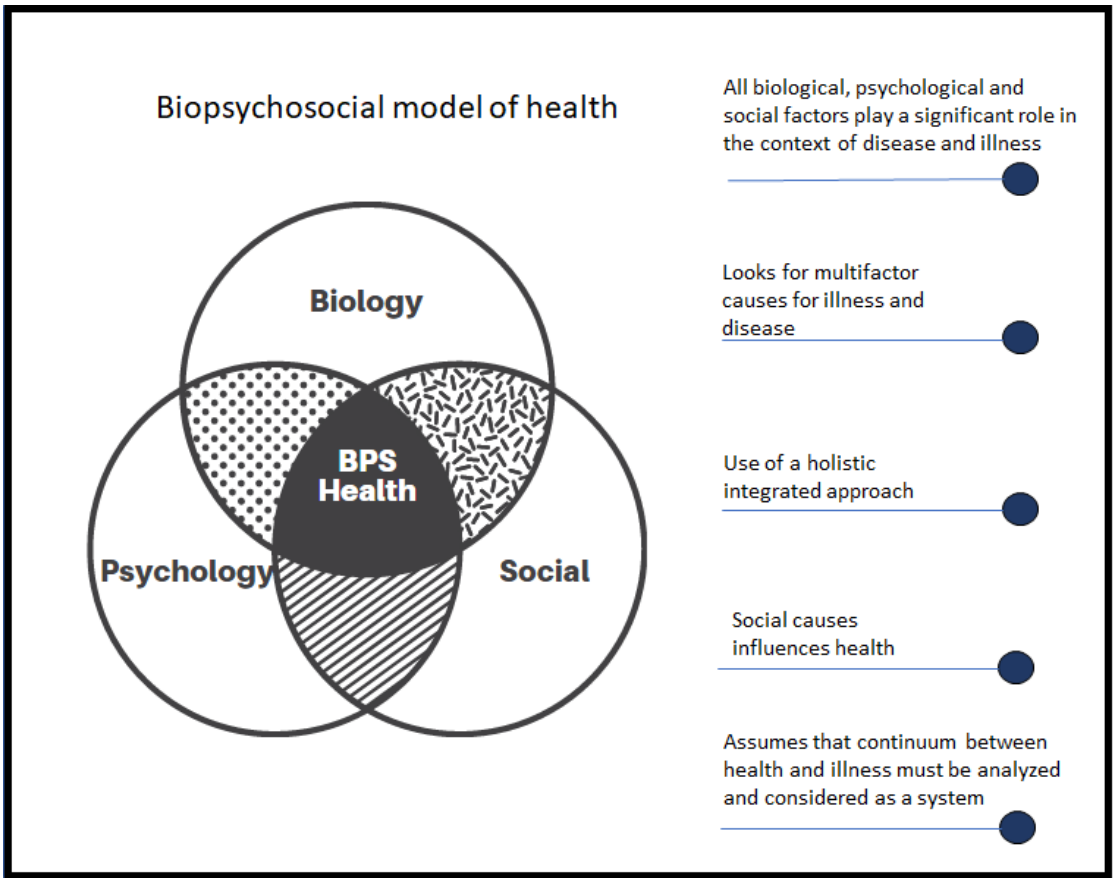


Figure 1.1: Biopsychosocial model of health

more comprehensive investigation of the health relationships between mind and body [5]. Since this model is based on the basic premise that any health or disease condition results from the interaction between biological, psychological, and social factors, it has helped us move beyond the old dualism that separated the body from the mind [2]. This approach towards the biopsychosocial model has marked a shift from a traditional biomedical model centred only on the body to a medicine centred on the person [2, 5, 6]. Figure 1.1 depicts the interdisciplinary biopsychosocial model of health.

In healthcare, there is growing awareness that biopsychosocial screening of patients can assist in planning for more effective treatment in case of illness [7]. Biopsychosocial screening has already proven effective for supportive care and research related to cancer [8]. Overall, there is a wider understanding in the supportive care community that greater levels of interdisciplinary integration between the biological, psychological and

social disciplines are needed. Biopsychosocial screening programs have the potential to bridge the gap between these patient-centred and family-centred systems. In the health domain, biopsychosocial screenings take the form of biopsychosocial interviews. These are assessments typically conducted by therapists and counsellors at the beginning of therapy or a health intervention to assess for biological, psychological and social factors that can contribute to a patient's problem. In this way, patient-centred interview methods transform the general biopsychosocial model, initially formulated by Engel [5], into a specific model for each clinical encounter, translating it into an evidence-based, consistently defined model. The biopsychosocial model implies that for optimal management of conditions, healthcare providers must recognise that patients' participation in health and disease management is affected by their existing medical conditions, psychological factors and socio-cultural barriers in the environment [9]. A significant number of published papers [10–12] have raised concerns that the existing biopsychosocial model has unclear boundaries between the biology, psychology and social disciplines. In the light of these unclear boundaries, the biopsychosocial model vaguely tends to suggest that biology and psychology are two separate fields in medicine and this has led to errors in identifying and categorising patient symptoms as biological or psychological, masking the underlying biomedical approach [13]. In more recent years, digital health technologies have been transforming the healthcare landscape [14–16]. These technologies have facilitated, increased access to data and provided scalable and comprehensive data from varied sources [17]. In the context of the biopsychosocial model of health, these advancements in various digital health technologies help capture biological, psychological and social information; they thus provide a non-dichotomous understanding of an individual's medical condition and contribute to unmasking the biomedical focus of the original model [18].

1.2 Integrating healthcare using digital health data

As discussed earlier, digital health makes use of information and communication technology (ICT) to advance human health, healthcare services, and wellness for individuals and across populations [19–23]. The positive impact of technology on the health and wellbeing of individuals, communities and populations is unprecedented [14–16]. Recent technological achievements [24, 25] have revolutionised clinical practice, from prevention to diagnosis [26, 27] and from monitoring to disease management [28–32], and have enabled unprecedented public interest and engagement in self-management and wellbeing [33–36]. Successfully developing, integrating and implementing new technologies and methods in digital health requires a shift from the single disciplinary approach followed by healthcare

professionals, medical scientists and computer scientists/engineers to a joint multidisciplinary approach between the various groups [36]. Digital health is a multidisciplinary domain and spans disciplines including computer science, engineering, information science, journalism, economy, clinical medicine, public health, epidemiology, and others [33, 36, 37]. A joint multidisciplinary collaboration between technology specialists and healthcare experts to conduct collaborative research at all levels, will accelerate access to the best evidence and healthcare services [36]. There have been rapid changes in healthcare caused by embracing digital health technologies in different aspects of health and disease [18]. These changes have been driven by several factors, including the following:

- the increased availability of digital biological data generated on different platforms and its easy accessibility to healthcare providers and patients. For example, deoxyribonucleic acid as a digital molecule has become the cornerstone of digital genomics services [38, 39]
- the usefulness of portable biosensors in providing information to support and manage their own health and to enable affordable access to people living in low socio-economic conditions and for people in remote geographical environments [40, 41]
- the growing trend of increased personal health surveillance and monitoring by integrating personal self-tracking data with electronic medical records [42]
- the increased use of artificial intelligence (AI) in healthcare to provide better patient outcomes and reduce costs
- the growing use of health-related social media data, at both the individual and group levels, to improve quality of care [43].

With advancements in digital health, healthcare screening that aligns with the biopsychosocial model [44–46] has helped health practitioners to evaluate the potential benefits of a health solution at the individual, population or organisational level using digital tools. Treatments and health solutions are driven by health information which is a crucial factor in improving health conditions, wellbeing and quality of life. Digital health programs have used a range of ICT healthcare strategies, such as creating websites, portals, social networks, and synchronous and asynchronous communication, among others, to facilitate participation from individuals seeking healthcare solutions [47, 48].

The use of mobile and wireless digital health interventions (DHI) has allowed healthcare practitioners and researchers in the healthcare domain to develop and implement different kinds of DHI programs. Some of these programs focus on preventing ill health, while others focus on delivering early intervention and mental health treatment [49, 50]. The broad range of DHI programs addresses mental and physical health concerns and allows wider choice of healthcare management for both the healthcare practitioners and the patients involved. Broad types of digital health programs include but are not limited to:

1. self-help DHI programs
2. therapist assisted digital programs.

All digital health programs generally involve working through many interactive intervention modules over a period. DHI program frameworks in clinical and non-clinical practice increasingly use patient-centric interviews, such that clinicians and healthcare practitioners can identify a scientific biopsychosocial model specific to each patient [51]. The openness to and demand for biopsychosocial interventions is also driven by developments in behavioural health and medicine that focus on behaviour and health, which contribute significantly to the maintenance of overall wellbeing and quality of life [52].

The influence of digital, electronic and mobile technologies (such as mobile health [mHealth]) is improving physical and mental health behaviours and outcomes [53]. The implications of ubiquitous and pervasive digital technologies for healthcare and public health are profound. Many such technologies are now explicitly designed for medical and health purposes, contributing to the digital health phenomenon that has recently emerged. Mobile digital devices, applications ('apps'), websites and platforms to which they connect offer not only ready access to medical and health information on the internet, but also offer new ways of monitoring, measuring and visualising the human body and sharing personal information and experiences with others. Randomised controlled trials (RCT) are used to evaluate DHI programs, which are carried out in clinical settings or various other everyday contexts—for example, healthcare facilities, communities, neighbourhoods, workplaces, schools and universities. Many interventions of the 'Wellness study' DHI program mentioned in this research (see Section 2.9) were accessed in the 'real world'; however, they were tested artificially via research protocols. The participants provided saliva and blood samples and undertook three computerised cognitive tests when they met with the researcher on site. They wore the BASIS wristwatch during the course of the DHI program that monitored their biometrics. It must be noted that the design for

data collection is not part of the research study presented in this thesis. The design for data collection was predetermined prior to the rollout and implementation of the 'Wellness study'. All aspects of the 'Wellness study' and its interventions were digital, as all the data captured, used and stored took the form of digital signals involving the use of computer technology. The digital nature of health intervention programs offers the potential to leverage technological innovations to develop proactive connected health [54].

1.3 Extracting knowledge from digital health data

More than a decade ago, Sir Muir Gray [55] forecast that *'Knowledge is the enemy of disease – the application of what we know already will have a bigger impact on health and disease than any drug or technology likely to be introduced in the next decade.'* Using technology efficiently to collect, understand and disseminate knowledge is of the utmost importance to fulfilling this vision. Many recent research achievements in the areas of web science, data mining and analytics, medical ontologies and recommender systems have created immense opportunities for better evidence dissemination, medical advice and development of personalised persuasive intelligent systems [36]. As the different fields of medicine continue to leverage digital tools to improve the reach and effectiveness of digital health programs, a huge opportunity arises to mine the large quantity of big data produced by these digital tools. Digital health data encapsulate the richness and granularity of individuals' behaviour. The health data record the confluence of elements that affect an individual's behaviour in the current moment and the within-individual evolution of behaviour over time [56].

Digital health uses data captured via digital technology to measure individuals' health behaviour in everyday life and provides access to digital therapeutic tools anytime and anywhere [25, 57]. Smartphones and wearable devices have a range of inbuilt sensors, such as accelerometers, wearable electrodes, temperature sensors, Bluetooth, GPS, light sensors, microphones and proximity sensors, as well as systems logs of calls and short-message service use [58]. These smart devices enable passive, ecological sensing of behavioural and physiological features, such as one's sleep, physical activity, social interactions, electrodermal activity and cardiac activity [59]. Ecological momentary assessment (EMA) surveys facilitated by mobile apps capture an individual's responses to questions and provide snapshots of social interactions, stress, pain, mood, eating, physical activity, mental health symptoms and substance use. Social media data provide information about individuals' behaviour, preferences and social networks. This kind of digital data is also referred to as

'digital exhaust' [60] and enables the continuous measurement of individuals' behaviour and physiology in naturalistic settings [58].

Digital health research is important and must be carried out by building upon what is already known. This ensures that the investments in digital health in our healthcare systems are informed by the strongest possible evidence base [17, 61]. Although the aim of digital health research is to improve or advance health, this depends heavily on the research methods and tools that are available in ICT disciplines such as computer science and information systems [62, 63]. Data mining models and machine learning (ML) methods are increasingly being used for analytics on healthcare data generated from different sources [64–66]. Data mining explores a large existing dataset to unearth previously unknown patterns, relationships and anomalies that are present in the data. ML, a subset of AI, is used to examine large datasets and then 'learn' patterns that can enable predictions to be made about new datasets [67]. ML has attracted considerable research interest in developing smart DHIs [68, 69]. ML is increasingly used in data analytics to advance the pace of digital health evidence generation [70] and to establish a data-rich research infrastructure that enables continuous learning and evaluation [68].

1.4 Motivation

Applying analytics to data produced by digital health technologies has emerged as an efficient approach for evaluating DHIs. DHIs incorporating ML algorithms in real-life studies are useful and effective. However, given the low number of identified studies and the low rigour of ML evaluation methodology used therein, there is a vast need for the research community to conduct further studies in intervention settings to demonstrate the potential of ML and AI in clinical practice as well as in evaluating DHIs in real-life settings [68].

Most prior studies have evaluated interventions in clinical settings; these studies are retrospective validations of ML algorithms and models wherein one or more datasets are available [68]. Moreover, these studies, which have evaluated RCT interventions in clinical settings, have focused primarily on data from a single discipline centred only on the body, thus treating illness as a purely biological event [71–73]. To date, limited research has been conducted to evaluate or analyse DHIs by following the biopsychosocial model in a real-life setting [74]. The research reported in this thesis analyses the retrospective wellness dataset generated from the 'Wellness study' DHI program. The framework of this program

is described in Section 2.9. The DHI program produces a complex set of biopsychosocial data collected from disparate data sources. The existing real-life biopsychosocial data provide fertile ground from which to apply the analytics process model and thereby discover patterns hidden in the data. Different statistical methods and ML methods are used to pre-process, analyse and discover new knowledge from the data analysis.

1.5 Aims and objectives

The primary aim of this research is to explore and discover relevant insights by conducting a retrospective study and analysis of the biopsychosocial data generated from the DHI program. The focus is on developing new approaches and computational tools for the various phases: Obtain, Scrub, Explore, Model and iNterpret (OSEMN) of the data analytics process. The data analytics work in this thesis lies at the intersection of the information technology (IT), statistics and healthcare domains. The research on biopsychosocial data analysis has been broken down into several key objectives:

- explore the salient features of the transdisciplinary biopsychosocial data generated from the 'Wellness study' DHI program
- examine the challenges involved in data pre-processing, with a focus on missing data and on imputation methods
- explore the use of predictive analytics to predict aspects of biopsychosocial wellbeing by analysing patterns in the biopsychosocial data.

1.6 Contributions

The key contributions of the thesis are summarised below.

1. modelled neurocognitive reaction time using gamma distribution (*Chapter - 3*). Analysis of participants' reaction time in the Psychology Experiment Building Language (PEBL) Go/No-Go test showed that the reaction time data are more compatible with a gamma distribution. This is because gamma distribution considers higher order moments of data, such as skewness and kurtosis, using shape and scale parameters rather than the commonly applied Gaussian distribution, which is defined using mean and variance.
2. used novel 'peak heart-rate' count metric to quantify level of 'biological' stress (*Chapter - 3*). Heart-rate data are one key biophysiological indicator of stress. A new metric,

'peak heart-rate count', was proposed to measure 'biological' stress. Further, a new approach using the median and interquartile range (IQR) of daily heart-rate data was used to define *maximum threshold*. The proposed new metric, 'heart-rate peak count', and the associated new MedIQR method was used to analyse the distribution of the heart-rate peaks between the two interventions, mindfulness and physical activity, to determine which intervention was more beneficial to participants during the DHI program.

3. repurposed the formula AUC_G for cortisol score computation (*Chapter - 3*). The formula AUC_G was used on the biological saliva data to compute a single cortisol score from four cortisol readings that were recorded on a single day and changed over the course of that day.

4. comparatively evaluated nonlinear classifiers to identify features that can predict *mental ill-health* represented by the Depression Anxiety Stress Scale (DASS) (*Chapter - 5*). ML and nonlinear classifiers were explored in existing biopsychosocial data. Real-life biopsychosocial features were input to nonlinear classifiers to predict the mental illness state (*mental ill-health*) represented by the DASS score.

5. conducted wearable data analysis using a recurrent neural network (RNN) and long short-term memory (LSTM) data sequence classification model (*Chapter - 5*). Developed an RNN-LSTM model using a sequence classification algorithm to predict negative affective variable Difficulties in Emotion Regulation Scale (DERS) and to predict primary emotion (PE) based input time-series wearable data.

1.7 Organisation of thesis

The research reported here was conducted to achieve the stated research objectives, and the thesis chapters outline the various phases of the data analytics process model that were undertaken to achieve the overall objective. Chapters 3 and 4 reports on the work carried out as part of the Obtain, Scrub and Explore data phases of the OSEMN data mining process model [75]. Chapter 5 reports the work carried out as part of the Explore, Model and iNterpret data phases of the OSEMN data mining process model. Since different phases of the data analytics process model overlap, the chapters reflect this nature when subsets of biopsychosocial data are investigated in chapters 3, 4 and 5. A brief outline of the chapters is given below.

This chapter details the preparatory concepts of this research, including motivation, aims and objectives, contributions, and the thesis outline. Chapter 2 presents a detailed review of the literature on the DHI programs that align with the biopsychosocial model of health and discusses the use of analytics on digital health evidence data using supportive data mining and ML methods. The chapter concludes by examining the characteristics of biopsychosocial data.

Chapter 3 explores the salient features of the transdisciplinary biopsychosocial data as a first step. This chapter is divided into different sections for the three disciplines of the biopsychosocial model: a. analysis of biological data (3.1) and analysis of physiological data (3.2), b. analysis of psychological data (3.3) and c. analysis of social data (3.4). The biological and physiological data are separated within the subsection, as the biological data strictly refer to the blood and cortisol data, whereas the physiological data refer to the time-series wearable data, also known as the biometric data. The work done on the subsets of biopsychosocial data and the contents of this chapter follow the Obtain, Scrub and Explore data stages of the OSEMN data mining framework.

Chapter 4 focuses on the challenges of using real-time biopsychosocial data to prepare it for statistical analysis using statistical procedures or ML methods. It addresses the challenge of handling missing wearable data and investigates the missingness patterns in the data to choose an appropriate imputation strategy. This chapter follows the Obtain, Scrub and Explore data stages for the time-series wearable sensor data collected using BASIS wristwatches.

Chapter 5 details the statistical and ML methods to examine possible contributory connections between participants' different biopsychosocial features of participants in the DHI program. The first part of the chapter (5.1), explores building the prediction model using nonlinear classifiers to compare and evaluate the two DHIs. The second part of the chapter (5.2), further explores building the prediction model using the ML method.

Chapter 6.1 summarises the study conclusions and proposes directions for further work.

CHAPTER 2

LITERATURE REVIEW

The biopsychosocial model in healthcare and research is gaining momentum. Healthcare providers and researchers have started appreciating the psychological and social factors that can mediate health outcomes, as opposed to using the biomedical model alone [76]. Despite recurrent criticisms and uneven uptake, the biopsychosocial model has influenced core aspects of medical practice, education and research across many areas of health and science. Over time, emerging evidence has been associated with adopting the biopsychosocial model of health in relation to specific conditions, such as chronic pain management [77] or mental health problems [78, 79]. Slowly but steadily, the biopsychosocial model is being incorporated into the design of research programs that involve analysing major aspects of biopsychosocial data [80].

This chapter reviews the research carried out in the broad domain of biopsychosocial data analysis and DHIs. It reviews various papers that apply analytics to digital health evidence data and examine ways to apply data analytics to biopsychosocial data generated from DHI programs. The rest of this chapter is organised as follows:

- Section 2.1 briefly discusses the perspective of the biopsychosocial model of health;
- Section 2.2 provides a synopsis of using DHI to promote biopsychosocial health;
- Section 2.3 studies the benefits and scope of digital health analytics for the data generated from DHI programs.

- Section 2.4 discusses different approaches to data mining;
- Section 2.5 outlines the different ML methods that can be used as part of data mining;
- Section 2.6 examines the need for predictive models to integrate biopsychosocial data;
- Sections 2.7 and 2.8 describe the characteristics of the digital health interventions and the typical characteristics of highly dimensional biopsychosocial data produced as part of the 'Wellness study' DHI program;
- Section 2.9 describes the framework of the 'Wellness study' DHI program; and
- Section 2.10 summarises the chapter.

2.1 Perspective of the biopsychosocial model

The perspective of the biopsychosocial model holds to the idea that biological, psychological and social processes are integrally and interactively involved in health and wellbeing [5, 81–84]. The premise that these subsystems are nested and inextricably connected has stimulated innovations in the design and implementation of interventions to promote health. As a guiding framework, the biopsychosocial model has proven remarkably successful, as it has enabled healthcare providers to forge a multi-level, multi-systems approach to human functioning [85].

Over the years, applied research across a range of substantive areas has affirmed the values of the biopsychosocial perspective and demonstrated how biological, psychological and social processes operate together to affect physical and mental health outcomes. Kusnanto et al. [86] review the implementation of the biopsychosocial model in clinical practice and its contribution to clinical outcomes. They describe the biopsychosocial model as particularly helpful for addressing chronic diseases and ill-defined illnesses to which patients experience unique responses. Wijma et al. [87] recommend using the clinical biopsychosocial assessment prior to providing physical therapy to patients with chronic pain. The use of the biopsychosocial approach in clinical practice, allows healthcare providers to provide proper explanations of the neurophysiology of pain and biopsychosocial interactions in an interactive and patient-centred manner. The biopsychosocial approach has also facilitated the recovery of people with chronic illness [88–90].

Apart from chronic illness and pain management, the biopsychosocial model has also been successful in other fields of healthcare, such as psychiatric rehabilitation and workplace injury management. In their recent study, Tong et al. [91], discuss the importance of sense of community (SOC) for people with serious mental illness. The study provides a detailed understanding of the predictors of SOC by using the biopsychosocial model as a conceptual framework. Managing emotions, social support and stress have been shown to play important roles in the progression and management of cardiac disease and cancer [88, 92]. By introducing behavioural interventions, physicians have demonstrated success in promoting smoking cessation [93]. In Australia, the Queensland Government has successfully applied the biopsychosocial approach to workplace injury management [94]. A range of rehabilitation interventions have been developed that align with the biopsychosocial model; these are based on the principle that effective injury management relies on understanding the whole person, including their biological, psychological and social needs and the challenges that arise therein.

Studies in the domain of health psychology have considered social factors in the form of self-report measures and marital satisfaction [95–98] in an effort to embrace and examine the multiple systems [76] that underlie the biopsychosocial model. Other studies [99, 100] in this domain have assessed biological variables, including physiological reactivity to proposed health interventions. Some studies have assessed macro-variables, also known as demographic variables, such as ethnicity, income and age [4, 101]. Overall, the review of literature on adopting the biopsychosocial model in the domain of health psychology shows that investigators focus on the interplay between psychological and social factors or between biological and psychological factors. This shows that immense opportunity exists to explore interconnections between the biological and social disciplines within the domain of research and practice in health psychology.

Although advances have been made in specifying connections between biological, psychological and social processes, the full potential of the biopsychosocial model remains untapped in many domains of medicine. In their review paper, Farre et al. [102], study the interplay between the biopsychosocial disciplines and note that the understanding of this interplay can shape the nature of medical work both philosophically and in practice. The paper outlines some key controversies and criticisms raised by other authors related to the biopsychosocial model in medicine and includes solutions to overcome these shortcomings. A major criticism it raises is that the biopsychosocial model does not include a method to identify relevant biopsychosocial data [103]. There is no methodological guidance on the

level of analysis needed in each discipline of the biopsychosocial model, since it is often not known which factor might be responsible for a given condition [10, 12, 104]. Suls et al. [85], explain the challenges of integrating the biopsychosocial model into health research and recommend ways to fix this. For instance, the authors note that in the field of health psychology, only 26% of studies have examined measures from all the disciplines of the biopsychosocial model. This view can be extended to health researchers in other fields, who most assuredly have given much less consideration to the psychological and social influences that affect wellbeing.

The biopsychosocial model is the basis of World Health Organisation's (WHO) International Classification of Functioning [105]. This has provided an impetus for using this model clinically and structuring clinical guidelines. Smith et al. [103], suggest using patient-centred interview methods in practice so that clinicians can identify a scientific biopsychosocial model specific to each patient. Hence, there is a strong focus on psychological interviews, which are the most important source of biopsychosocial data during each clinical encounter. These interviews can generate highly relevant biopsychosocial information for healthcare providers. In an effort to simplify the adoption of the biopsychosocial model, Karunamuni et al. [106] introduce an updated theoretical model to examine the interrelationship between the different disciplines of the biopsychosocial model of health and present a framework that could have important implications for clinical practice and research.

Despite the criticisms, the biopsychosocial model has influenced core aspects of medical practice, education and research. More optimal use of existing bodies of evidence, bringing together evidence-based methodological advances of the biopsychosocial model and existing evidence on the psychosocial needs associated with specific conditions or populations can help bridge the gap between philosophy and practice. Digital health technologies can address the limitations of the biopsychosocial model by providing more accessible, scalable and comprehensive data from various aspects of health and disease. Further, the availability of various technologies that can continue to capture biological, psychological and social information provide a non-dichotomised understanding of an individual's medical condition and help unmask the biomedical focus of the original model. With digitisation, the adoption of the biopsychosocial model has increased in the areas of chronic illness and monitoring general wellbeing and affective experiences [107–112]. In the next section, we review the role of DHIs in promoting biopsychosocial health.

2.2 Digital health intervention

Digital health is an advancing phenomenon in modern healthcare systems [113]. Currently, numerous stakeholders across the world are continuously exploring and evaluating the potential benefits of digital health solutions at the individual, population and organisational levels [114]. The notion of 'the digitally engaged patient' is emerging across healthcare policy, service provision and research domains [33, 108]. Using IT to deliver health interventions is a promising approach, as it has wide reach and low cost, is readily accessible and provides for scalable interventions for large populations, thus addressing time and resource constraints [107, 115, 116].

DHI is closely related to the notion of mHealth, telemedicine, telecare and health IT [117, 118]. The use of technology has the potential to transform the face of health service delivery across the globe, and the driving factors for this are the rapid advances in mobile technologies, the rise in new opportunities to integrate mHealth into existing eHealth services and the continued growth in coverage of mobile cellular networks. DHI describes the action of intervening with 'tools and services that use information and communication technologies (ICT) to improve prevention, diagnosis, treatment, monitoring and management of health and lifestyle' [118]. DHIs are programs that are delivered via digital technologies such as smartphones, websites and text messages [61, 119]. DHI programs [120] have an enormous potential as scalable tools to enhance health and healthcare delivery by improving effectiveness, accessibility, safety and personalisation [121].

DHI programs are increasingly being used to promote mindfulness and physical activity based approaches in health and lifestyle as part of public health campaigns [122]. In the practice of psychological therapies, standard mindfulness-based interventions use digital technology tools to engage participants in mindfulness assignments [123]. Digital technologies deliver mindfulness-based interventions to support older, informal carers with managing stress and sleep [124]. Computer scientists, interaction designers and other inventive practitioners and researchers have undertaken a wide variety of experiments using digital tools to support social wellbeing. In this space, the innovations in technologies range from responsive architecture, installations and dynamic furniture to handheld gadgets and wearables [125].

Researchers have focused on using DHI in a therapeutic capacity for musculoskeletal conditions, bringing moderate benefits to the area of reducing pain and stronger benefits to the area of improving functional disability [107]. Findings show that chronic musculoskeletal conditions, particularly lower back pain, are optimally managed using a biopsychosocial approach [126] that incorporates both physical and psychosocial elements in the rehabilitation program. Studies in which a DHI includes an exercise/physical activity component alongside a psychosocial approach have achieved statistically significant improvements in functional disability [127–129]. DHI using the biopsychosocial approach can contribute positively towards diminishing the personal, societal and economic impact of musculoskeletal conditions, which, as our population ages, is only set to grow.

Other areas in which significant achievements have been made using eHealth/mHealth are for tuberculosis treatment control, smoking cessation [130]], remote pulmonary rehabilitation, physical activity telecoaching [131] and self-management support [132]. Implementing DHIs in respiratory medicine is another area of great focus, as it promotes eHealth research on novel methodologies to evaluate digital interventions in respiratory diseases [133].

Although there is increasing evidence that digital interventions can successfully effect meaningful changes in health-related behaviour, there are key challenges in the delivery of DHI programs [134]. These challenges relate to participation in DHI programs namely: low usage, high attrition and small effect sizes [135–137]. The impact of the challenge regarding effect size varies depending on what condition or illness is being treated. Effect sizes are typically large for DHIs that measure conditions such as anxiety, depression and insomnia. In the prevention space, effect sizes are far smaller. In the context of comparing interventions, there is less room for change between pre- and post-intervention results and people show lower illness scores. Therefore, when evaluating DHIs in the prevention space, it is necessary to work with a large number of samples (N) to detect small changes over time. The issues of low samples and small effect sizes limit the delivery and evaluation of DHIs. Morrison [135] recommends using psychological theories and models alongside conceptual frameworks that guide intervention planning and development to address the challenges of low usage, high attrition and small effect sizes to a large extent.

The framework of DHI programs enables the easy collection and consolidation of biopsychosocial data. DHIs provide opportunities to tackle health system challenges and offer the potential to enhance the quality and sustainability of general wellbeing. The WHO has

published guidelines that classify DHIs in an attempt to standardise the vocabulary used within the diverse communities working in digital health [138]. Arigo et al. [139] highlight key areas of opportunity and recommend steps to advance intervention development, evaluation and commercialisation using digital technologies such as wearable devices and mobile apps. Online programs to advance DHI development are popular compared to mobile apps, as they provide for a larger evidence base. With recent advancements in technology and the just-in-time rule becoming more commonplace, mobile apps should also become more effective [140]. In the field of behavioural medicine, the future of digital health behavioural science research lies in finding ways to promote and strengthen more robust academic industry partnerships [139].

The WHO is addressing digital health as a global strategy in 2020–24 [141]. It defines digital health as 'the field of knowledge and practice associated with the development and use of digital technologies to improve health' [141]. Digital health is a rapidly growing industry, and demand is increasing for remote monitoring services caused by rising incidences of chronic diseases. This is a major factor propelling global market growth. Digital health is an accessible and affordable solution for people who do not have access to the traditional health system.

Advances in IT have enabled easy access to varied forms of healthcare data. Advances in data sensing and acquisition technologies have led hospitals and healthcare institutions to collect enormous amounts of healthcare data about their patients. Advanced analytical techniques are required to effectively understand and build knowledge from the collected healthcare data so that these data can be effectively transformed into meaningful and actionable information.

2.3 Digital health analytics

Digital health analytics can be defined as the discovery and communication of patterns in health data [142]. It has emerged as an efficient approach to evaluating DHIs [143]. Pham et al. [142]] discuss in detail the practice of analytics as an efficient and effective means of supporting digital health evidence generation. Their study aimed to generate transferable insights into the process of implementing analytics to evaluate DHIs. The study concluded that continued implementation of analytics can help accelerate the rate of digital health evidence generation to build a data-rich research infrastructure that, in turn, can enable continuous learning and evaluation.

It has become easier to extend the benefits of healthcare to the wider community because of the increasing use of new technologies such as big data analytics, AI and ML. These digital technologies are effective tools to improve clinical practices and operations and support public health management and medical research. Compared to other industries, the healthcare industry's adoption of digital technology lags behind, and this is due to technological challenges, data privacy obstacles, lack of healthcare system integration and health data quality issues. To overcome these challenges, research should focus on using AI and analytics to explore and harness digital health data and support evidence-based medicine. Further, research must investigate ways in which healthcare information systems can work together in a cohesive way to deliver effective care to patients and address privacy issues regarding the use of digital technologies in healthcare.

The broad availability of data has led to growing interest in methods for extracting useful information and knowledge from data [144]. One of the longstanding challenges in healthcare informatics has been the ability to deal with healthcare data and the increasing need to glean actionable insights that drive performance improvement [145]. Some studies have mined the rich log data generated by users engaged in digital interventions and have successfully generated evidence of their impact on health outcomes [146]. Although many studies have used analytics to examine clinical trial data generated from interventions, the use of analytics on biopsychosocial data generated from DHI programs in a real-life setting is not available.

The work done in this thesis endeavours to analyse real-world complex biopsychosocial data generated from a well structured DHI program rolled out by the School of Science, Psychology and Sport at Federation University. The unique biopsychosocial data are

analysed using statistical methods and ML tools. The different studies and experiments in this thesis use the existing biopsychosocial data to measure the participants' wellbeing using different representative features. The digital technologies used in the intervention programs helped to collect the participants' neurocognitive, physiological, biological and social data. Data analytics on the transdisciplinary biopsychosocial data uncovered the challenges of data quality and data integration along with the challenges exposed by the interplay of the underlying complex data's variety, value and veracity. A data mining approach was used to efficiently analyse the large complex data that were gathered from different data sources. In the next section, we review some well-known models for data mining used in different industry domains and describe the suitability of the OSEMN process model for the study conducted in this thesis.

2.4 Models for data mining

Data mining is the process of learning from data, whereas ML is learning from the data. That is, data mining is the process of extracting useful information from a vast quantity of data; ML is the process of becoming trained on a 'training' dataset. At the end of the training, ML teaches the computer how to make sense of data to make predictions about new datasets [67]. There are different approaches to analysing data, and the methods used can vary for different domains.

2.4.1 Cross-industry standard process for data mining

The cross-industry standard process for data mining (CRISP-DM) [147] is noted to be the default time-tested methodology for data science. It has been widely used for a long time, not only for data mining but also for predictive analytics and big data projects [148]. CRISP-DM breaks down the process of data mining into six different phases [149], as shown in Figure 2.1.

Figure 2.1: A CRISP-DM process model

With using CRISP-DM data mining mode, there are no strict ways of moving between the different phases of the processes. Moving back and forth between the phases is encouraged and may be required. The outcome of every phase determines if one should move to the next step or instead repeat the previous step(s). The outer circle (see Figure 2.1) describes the cyclic nature of data mining. This implies that even when a solution has been deployed, the process can continue to create a better version [150]. CRISP-DM is built on an analytics-focused process framework called Knowledge Discovery in Databases and is predominantly used for enterprise data science projects [151]. A notable limitation of this method is that it lacks a standardised framework to guide the customisation process that would allow the method to be adopted by organisations whose capabilities are not mature enough [152].

2.4.2 Sample, Explore, Modify, Model and Assess

Another popular methodology, SEMMA [153], takes its name from the different stages that comprise the process: Information Exploitation, Sample, Explore, Modify, Model and Assess. SAS Institute, which developed the model, describes it as a toolset for carrying out the core tasks of data mining [149]. The methodology is itself a cycle (see Figure 2.2), the internal steps of which can be performed iteratively according to needs [154].

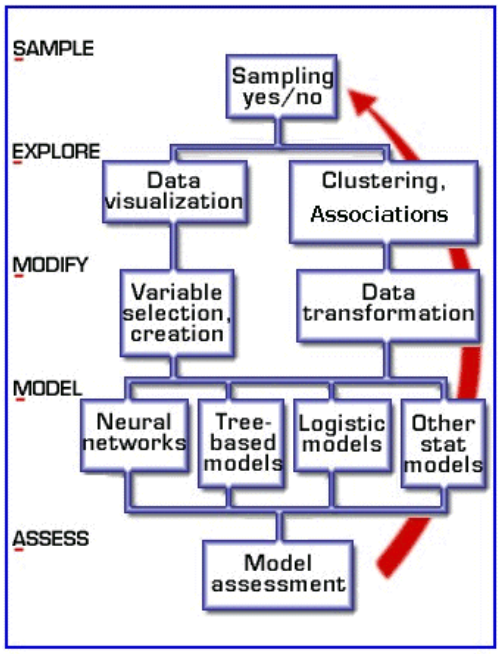


Figure 2.2: SEMMA process model

The SEMMA model processes are easy to understand, and this model is often used to develop and maintain information exploitation projects [154]. Structurally, both CRISP-DM and SEMMA methodologies have similarities, but they differ in several aspects, including tasks, activities and phases. The CRISP-DM model has a phase dedicated to problem understanding, with reference to business, defining objectives, resources, roles and other factors to set data mining goals. Conversely, SEMMA proposes to access the data to be analysed directly. It defines a sample and applies data mining techniques without considering the business objectives [154].

2.4.3 Agile methods

Agile methods for data analysis, such as Agile Knowledge Discovery Database [155] and Analytics Solutions Unified Method for Data Mining [152], have been used for projects with smaller scope, less complexity and fewer security issues. Agile methods are more often used within organisations that encourage more freedom [156].

2.4.4 Technology driven method

Technology used for data science has a great impact on performance, and one such technology-driven methodology is Microsoft Team Data Science Process (TDSP), which was introduced in 2016. It incorporates specific tools to make exploratory data analysis more efficient. TDSP can be customised for ML or AI projects. These tools are free and can be easily integrated into any workflow that uses Python or R. However, TDSP uses a specific set of Microsoft tools and infrastructure and is platform-dependent. Foundational Methodology for Data Science is similar to TDSP but has more detailed steps and is platform-independent [157].

2.4.5 Obtain, Scrub, Explore, Model and iNterpret

Hilary Mason and Chris Wiggins [75] introduced the OSEMN process framework. This framework represents the general workflow that data scientists typically perform, as shown in Figure 2.3. OSEMN was developed as a process for independent research [158], and, therefore, business-focused project phases, such as business understanding, deployment, operations and optimisation, are omitted. This makes the OSEMN model ideal for use in exploratory research projects that prioritise data analysis. The OSEMN process has five phases: *Obtain data*, *Scrub data*, *Explore data*, *Model data*, and *iNterpret data*. The first two phases, *Obtain data* and *Scrub data*, often consume between 50% to 80% of the time in a complex data analysis pipeline. This is justified, as any analytical model is only as good as input data. In the third phase, *Explore data*, users try to understand the patterns and values of their data and use different types of visualisations and statistical testings to support their findings. The next phase is *Model data*, which involves generating a model from the acquired data. The last and most important phase in the OSEMN model is *iNterpret data*, which interprets the data. In this phase, users primarily answer the questions that provoked the data modelling processes and needs. The phases in the OSEMN model can iterate until a satisfactory condition is met. Dineva et al. [159] discuss the OSEMN approach, which is used for working with data acquired by Internet of Things devices and their integration into beekeeping. The paper by Guhr et al. [160] describe the many tools that can be used for the different steps of the OSEMN process. The paper implements a proof of concept to demonstrate the feasibility of IT operations by combining common data science tools with enterprise architecture. CRISP-DM and OSEMN methodologies are process-focused, as they work regardless of the technology they use.

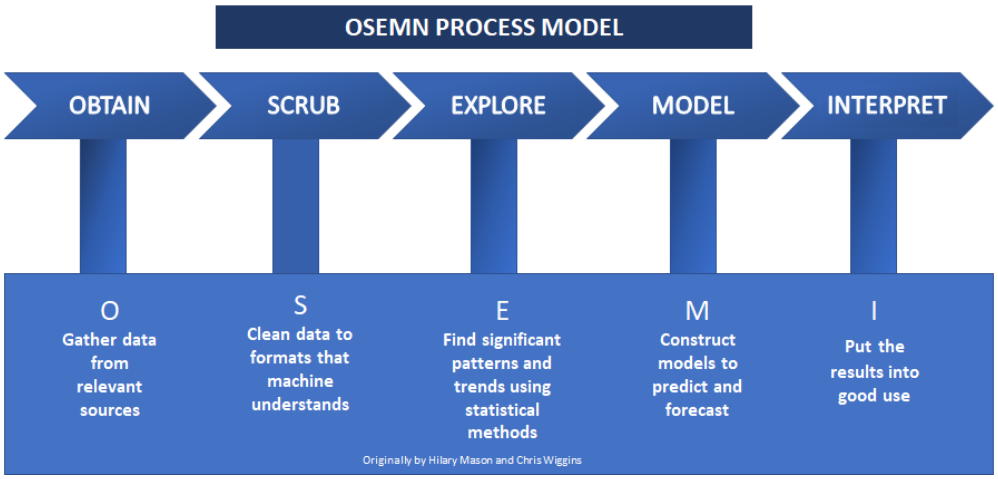


Figure 2.3: An OSEMN process model

In analysing the biopsychosocial data in this study, we closely followed the OSEMN process model because it can be represented as a scientific process that comprises different steps in a data science pipeline and is well suited to research focused on data analytics. Python and R predominantly served as query languages for the Obtain and Scrub data steps. Application suites such as SPSS were used to Explore and transform the data. ML methods supported by specialised Python libraries were used to build the Model, iNterpret the data and evaluate the generalisation of the algorithm. In the next section, we review some broad categories of ML methods and their applications.

2.5 Machine learning methods

ML arose as a subfield of AI, which dealt with methods for improving the knowledge or performance of an intelligent system over time based on the system's learning experience. The system gains its learning experience by analysing data from the environment and making predictions about unknown quantities. Over the years, this data analysis aspect of ML has come to play a very large role in the field of AI. ML examines patterns and can predict future behaviour by learning from those patterns. Data mining is often used as an information source on which ML is based. As ML algorithms are data-driven, the data determine the system's response. For example, when developing an application to detect photos, a traditional algorithm would first define what a face is. By contrast, an ML algorithm would not have such a predetermined definition but would instead learn by examples by getting trained via several photos containing faces (i.e., learn by example).

Most data mining frameworks discussed in Section 2.4 have a Model step that can be mapped to the ML stage in the data mining process. In the Model step, the predictive models are applied and the models' predictive accuracy is evaluated. In the iNterpret step, the model is evaluated to ensure that it satisfies the original business goals. Broadly ML algorithms can be of three types.

2.5.1 Supervised learning

Supervised learning finds an unknown function that connects known inputs to unknown outputs [161]. A given set of predictors predict the target variable in a supervised ML algorithm. We generate a function that maps inputs to desired outputs using this set of predictors In this way, the training process continues until the model achieves a desired level of accuracy for the training data. The supervised ML algorithms use the features and annotations in the training set to induce a model to predict the annotations of the instances in the testing set [162]. Biologists have used supervised ML algorithms such as classification and regression to understand the complex aging process that affects nearly all animal species [162]. Decision trees, Naive Bayes and Support Vector Machines (SVM) are commonly used algorithms in supervised learning [163–167]. Classification and regression algorithms are two kinds of supervised learning algorithms [168].

2.5.2 Unsupervised learning

Unsupervised learning involves pattern recognition without the use of an output label [162]. The unsupervised learning algorithm uses all input variables to predict the outcome. Clustering and association mining techniques are suitable for unsupervised learning. Clustering algorithms used in unsupervised learning can identify inherent groupings within unlabelled data. Labels can be subsequently be assigned to each data value [163, 169, 170]. Association mining algorithms in unsupervised learning are used to identify association rules that describe the relationships between the features that are suitable for clustering [162]. K-means, hierarchical clustering and principal component analysis are the most commonly used unsupervised learning algorithms, among others [171–173]. Even though unsupervised learning uses all input variables, it extracts limited attributes from the data and relies on previously learned patterns to recognise likely classes within the dataset [174, 175]. Unsupervised learning is well suited to feature reduction problems in large datasets and clustering tasks that lead to the creation of new classes in the unlabelled data [176, 177].

2.5.3 Reinforcement learning

Reinforcement learning occurs when the machine trains itself continually using trial and error and when it is exposed to an environment [178]. In such a condition, the machine learns from past experience and tries to capture the best possible knowledge to make accurate decisions. The application of reinforcement learning is limited to areas where a vast amount of data can be generated. Games offer an excellent environment for a reinforcement learning agent wherein the agent can explore different trials in a virtual world, since the cost of exploration is affordable [179, 180]. This type of learning is less useful for business applications compared to supervised or unsupervised learning.

Supervised ML algorithms for predictive modelling are described as learning a target function f that best maps input variables X to an output variable Y; $Y = f(X)$. The widely used ML method is to learn the mapping $Y = f(X)$ to make predictions of Y for new X. This form of ML method is referred to as predictive modelling or predictive analytics. ML algorithms are ever evolving and have been used in various domains, such as data mining, image recognition [181], speech recognition [182], self-driving cars [183], email and spam filtering [184], online fraud detection [185] and many more. Schmidt et al. [161] provide a comprehensive overview and analysis of the most recent research that has used ML. They critique the interpretability of ML algorithms in detail and introduce different concepts of interpretability, such as simulatability, decomposability, algorithmic transparency and post-hoc knowledge extraction. In this context, the authors describe the ML algorithms as black-box algorithms that do not explicitly represent the connection between features and prediction. They are not transparent, so they cannot describe the causality between the model inputs and the outputs. Despite these interpretability challenges, ML algorithms are successful in several different fields.

Linear ML algorithms are represented by linear functions that describe the relationship between two variables by producing a straight line when graphed. Nonlinear function has a shape that is not a straight line when graphed. Linear regression, logistic regression and linear discriminant analysis are examples of linear algorithms. Nonlinear algorithms include classification and regression trees (CARTs), Naive Bayes, k-nearest neighbours, learning vector quantisation and SVM. Ensemble methods create multiple models and then combine them to produce improved results. Algorithms that use ensemble methods include Random Forest and AdaBoost. There are many different ways to represent the patterns that can ML can discover; some are described below.

1. Dimensionality reduction. This refers to methods that reduce the number of input variables in a dataset.

2. Decision trees. This method applies a 'divide and conquer' approach to the problem of learning from a set of independent instances.

3. Classification rules. The precondition of a classification rule is a series of tests similar to the tests on nodes in decision trees, and the conclusion gives the class or classes that apply to the observations covered by that rule or the probability distribution over the classes. Classification rules are popular because each rule represents an independent chunk of knowledge. With classification rules, it is easy to add new rules to an existing rule set without disturbing ones already there, whereas adding to a tree structure may require reshaping the whole tree [186].

4. Association rules. Association rules can predict probability of relationships between attributes. Different association rules express different regularities that underlie the dataset, and they generally predict different things [187]. The methods for decision trees and rules work most naturally with nominal attributes. They can also be used for numeric attributes by using numeric value tests in the decision tree or rule induction algorithm. Alternatively, the numeric attributes can be changed into nominal attributes.

5. Linear regression. This regression technique is used when the outcome or class is numeric and all the attributes are numeric. The class for the attributes is represented as a linear combination of attributes with predetermined weights. The membership function is 1 for instances that belong to that class and 0 for other instances [188]. Although linear regression achieves good results in practice, the membership values it produces are not proper probabilities because they can fall outside the range of 0 to 1 [189].

6. Logistic regression. This is a predictive analysis algorithm based on the concept of probability. It is a linear model that draws linear decision boundaries between data points and is used for classification problems.

7. Clustering. Clustering techniques are used when the instances are to be divided into natural groups. No classes are predicted in this technique. The classic clustering technique is called k-means.

The growth of advanced analytics and ML and AI techniques has opened up numerous possibilities for transforming complex data into meaningful and actionable insights to support decision-making [190, 191]. All healthcare stakeholders can harness the power of data, using analytical techniques not only to analyse existing data (descriptive analytics), but also to predict future outcomes (predictive analytics) and determine the best action for the current situation (prescriptive analytics) [192, 193].

2.6 Predictive models for integrating biopsychosocial data

With the advent of the digital age, many experts have advocated for greater adaptation of digital health solutions [194]. Digital transformation creates numerous opportunities to improve health outcomes and increase the efficiency of healthcare systems [195]. The advancement of digital technology in healthcare has led to an abundance of information collected for each individual at a micromolecular level. DHI programs enable collecting a myriad of biological, psychological and social data that shed light on an individual's state of wellbeing. Recent studies [196, 197], have hypothesised that most complex diseases are caused by the combined effects of many diverse factors, including different genetic, genomic, behavioural and environmental factors. Therefore, it is essential to build integrative models that consider variables from multiple disciplines simultaneously and combine the information present for an integrative study for biopsychosocial marker discovery. To achieve this, information present in each dataset is assessed by its capabilities to predict a disease's endpoint. Such integrative studies aim to build a predictive model by combining datasets.

Prediction models are widely used in the analysis of healthcare data. Basic statistical prediction models use linear regression, the generalised additive model (GAM), logistic regression, Bayesian models and Markov random fields. The dependent variable in linear regression is assumed to be a linear combination of the features with corresponding estimated regression parameters [198]. Linear regression is widely used for clinical cost predictions [199, 200] and estimation of medical inspections [201] in the area of clinical data analysis. The GAM is used to model continuous outcomes in regression [202] and can be described as a linear combination of smooth functions [203]. Logistic regression is another popular binary classification method widely used for prediction tasks [204–206]. Logistic regression predicts the output based on the assumption that there is a linear relationship between the features and the log odds of the probabilities [203]. Implementations of

the Bayesian model, such as Naive Bayes and Bayesian networks, are used for clinical predictions [207, 208].

In addition to basic prediction models, recent developments in ML and data mining literature show other prediction models being used by biomedical researchers for clinical applications. These models include decision trees and artificial neural networks. Decision trees are the most widely used clinical prediction models [209], wherein predictions are made by asking a series of well-defined questions (splitting criteria) about a test record. Depending on the answers to these questions, the test record hierarchically falls into a smaller subgroup where the contained individuals are similar to one another with respect to the predicted outcome. The splitting criteria help find the locally optimum decisions that can minimise the within-node homogeneity or maximise the between-node heterogeneity [203]. Other tree-based models, such as C4.5 [210] and ID3 [211], use information entropy to determine the best splits. A CART [212] can only produce binary splits, and the best split is selected where the *Gini* index is minimised. Compared to other methods, decision trees are straightforward and constructing them does not require knowledge of the underlying distribution. Moreover, decision trees can easily handle all kinds of data types for the input data [203]. The challenge lies in finding an optimal decision tree. Usually, a tree induction algorithm is a heuristic-based approach that can make decision tree unstable [213]. A wide range of applications have used decision trees for medical decision-making problems [214, 215]. Another widely used prediction model is an artificial neural network (ANN). ANNs are inspired by biological neural systems [216], in which simple artificial nodes called 'neurons' are combined via a weighted link to form a network that stimulates a biological neural network. As a result of the ANN's ability to model complex mapping functions and because of the rich literature in the ML community, ANNs have been widely used in various biomedical applications [217, 218]. Some noteworthy applications include decision support systems [219], medical intervention [220] and medical decision-making [221].

Integration of diverse biopsychosocial data using predictive models is a vast research topic. Most of the studies scattered throughout the literature were developed from a clinical perspective and were not based on real-life biopsychosocial data collected for the person. To the best of our knowledge, there are currently no studies that aim to review integrative approaches, combining biopsychosocial data from a methodological perspective. There are many technical challenges in integrating biopsychosocial data generated from DHI programs. Some of these challenges are listed below.

1. Difference in nature of data
   As the data being integrated are collected from multiple disciplines, the nature of the data from each discipline can vary significantly. Physiological data collected from wearable devices have a strong temporal aspect, whereas biological data collected as part blood and saliva tests are not. Self-reported social data collected in the form of surveys can be unstructured and have missing values. Integrating variables with such different formats, types, structures, dimensionalities and missing values is challenging in the data mining and ML domain.

2. Statistical significance
   The high dimensionality of biopsychosocial data generated from DHI programs combined with low sample size poses challenges for finding statistically significant biopsychosocial markers (classical statistical n < p problem) [222]. Selecting the most useful features for building the prediction model is challenging and requires support from domain experts.

3. Different biases and assumptions
   Each discipline of the biopsychosocial data can have its own set of biases and assumptions. Depending on the design framework of the DHI program, the data being integrated may be different because of the difference in experimental designs and protocols.

4. Interpretability
   Another generic challenge of integrating the biopsychosocial data is that the obtained model has to be interpretable. This means that the effect of individual biopsychosocial markers on the wellbeing of the individual must be clearly identified by the prediction model; otherwise, the prediction model from the integrative study cannot be used efficiently.

Though these challenges exist, the growing availability of systems analytics, statistical tools, big data, ML and mixed-model trajectory analyses have facilitated research on integrative study of the biopsychosocial data in both clinical and non-clinical environments.

2.7 Characteristics of digital health interventions

The biopsychosocial data generated from the 'Wellness study' DHI program were used for this research. The DHI program sought to compare the effectiveness of the mindfulness and the physical activity interventions. Mindfulness is widely described as purposefully paying attention to unfolding moment-by-moment experiences with an open, non-judging and accepting attitude [223]. Mindfulness training enhances self-regulation abilities, including attention mechanisms, behaviour flexibility and emotion regulation [224, 225]. Part of the 'Wellness study', the mindfulness intervention measured dispositional mindfulness, which is often defined as a 'basic human quality' [226] that everyone possesses to a varying extent. Self-report questionnaires, such as the Difficulty in Emotion Regulation Scale [227, 228], the DASS [229, 230], the Mental Health Continuum (MHC) [231, 232], the Revised Life Orientation Test (LOT-R) [233, 234] and the six-item Kessler Psychological Distress Scale (K-6) [235, 236], are widely used to measure human conditions and traits such as depression, anxiety, stress and optimism. The Mindful Attention Awareness Scale (MAAS) measures mindfulness. The other intervention as part of the 'Wellness study' DHI program was physical activity. Physical activity refers to any movement made by the body or produced by skeletal muscles that resulting in energy expenditure [237]. Physical activity broadly encompasses exercise, sports and activities performed as part of daily living, occupation, leisure and active transportation [238]. The health benefits of regular physical activity are well established and convincing [239, 240]. The physical activity intervention program as part of the 'Wellness study' helped participants make small changes to the amount of physical activity they did, helped them find an activity they enjoyed, identified barriers and set achievable goals.

The selection criteria and sample size of participants in a DHI program strongly influence the power of the statistical tests conducted on the data. A precise statistical preparation of a DHI program framework must consider the selection criteria and the power needed to obtain valuable results. The participant selection criteria (inclusion and exclusion) must be carefully chosen to avoid possible confounding factors and to exclude participants for whom the intervention is useless or dangerous [241]. Moreover, the selection criteria should be not too severe; the risk is to conduct the DHI in an overly selected population and to obtain results not generalisable for actual clinical practice. In the context of the DHI program 'Wellness study' conducted by the School of Science, Psychology and Sport at Federation University, participants excluded from the study were pregnant women and people with a documented life-threatening illness (e.g., cancer), traumatic brain injury, debilitating

physical mobility issues, cardiovascular disease or current untreated severe (psychotic) mental illness. Another recurrent problem in most DHI programs is a low recruitment rate. This is due to recruitment difficulties, inadequate selection criteria or a small number of participants willing to join. Moreover, some participants leave the DHI program in the middle, and to define the outcome for such participants is not possible.

The biopsychosocial data are quite unique. Many questions can be explored with these data. They present a very interesting source for various data mining, pre-processing and visualisation techniques which can be applied to this dataset.

2.8 Characteristics of the biopsychosocial data

The characteristics of variety, value and veracity that usually describe big data are also relevant to the biopsychosocial data handled in this thesis. We now examine this interplay of variety, value and veracity [242, 243] and the challenges they pose in processing data.

2.8.1 Heterogeneity

The biopsychosocial data used in this research were heterogeneous in nature due to the diversity of data sources from which these data were collected and the diverse data types. The biopsychosocial data is divided into the following main categories:

1. *Demographic data* comprised features such as age and gender that were obtained during screening and demographic survey. The features of employment area and income were excluded from the study.

2. *Self-reported levels* included features such as social support, religiousness, physical health, mental health and quality of life. Each of these features was a self-reported value supplied by the participants using a Likert scale prior to the start of the DHI. The self-reported measure of medication was excluded from study.

3. *Test scale levels* consisted of features such as the following:

    a) The DASS level measured distress along three axes, depression, anxiety and stress, and equated them to clinical levels. DASS describes the measure of the negative emotional states of depression, anxiety and stress (*mental ill-health*).

    b) The K-6 level measured general psychological distress.

    c) The DERS level measured how participants managed their emotions.

d) The MAAS level measured how mindful the participants were.

e) The MHC and LOT-R levels measured wellbeing and optimism.

4. *Ecological momentary assessment data (EMA)* consisted of emotion and experience data such as primary and secondary emotions and intensity of emotions experienced.

5. *Psychological test levels* consisted of features such as PEBL [244] Go/No-Go test mean accuracy, trail test total time and Stroop test total errors for each participant. These features were derived from the computerised neurocognitive tests that were conducted to measure cognitive flexibility, selective attention and response inhibition.

6. *Biological test levels* comprised the cortisol levels (cortisol is a hormone that is mainly released at times of stress and manages many essential bodily functions) and blood proteins level (BDNF, iL-6) recorded at the start of the intervention during Week 1. Blood proteins data were not used because of missing data values.

Increasing digitisation in healthcare had led to more variety than the transactional sources used within the four walls of a clinical facility. The emergence of new data sources, such as smartphones and wearable devices, coupled with the need to integrate data within essential business transactional systems for example, electronic health record which creates an urgent need to recognise and address variety as one of the most critical factors in healthcare [245]. The biopsychosocial data in this study were generated from different data sources, including biological, neurocognitive tests, self-reported assessment by participants in the form of online questionnaires, EMA survey results and time-series data generated from wearable sensor devices during the DHI program. The data types were a mix of structured, semi-structured and unstructured data, including Likert values, text data and wearable reported time-series data. This heterogeneity in data sources and data types accounted for the large variety and required extensive data analysis and profiling (see Figure 2.4).

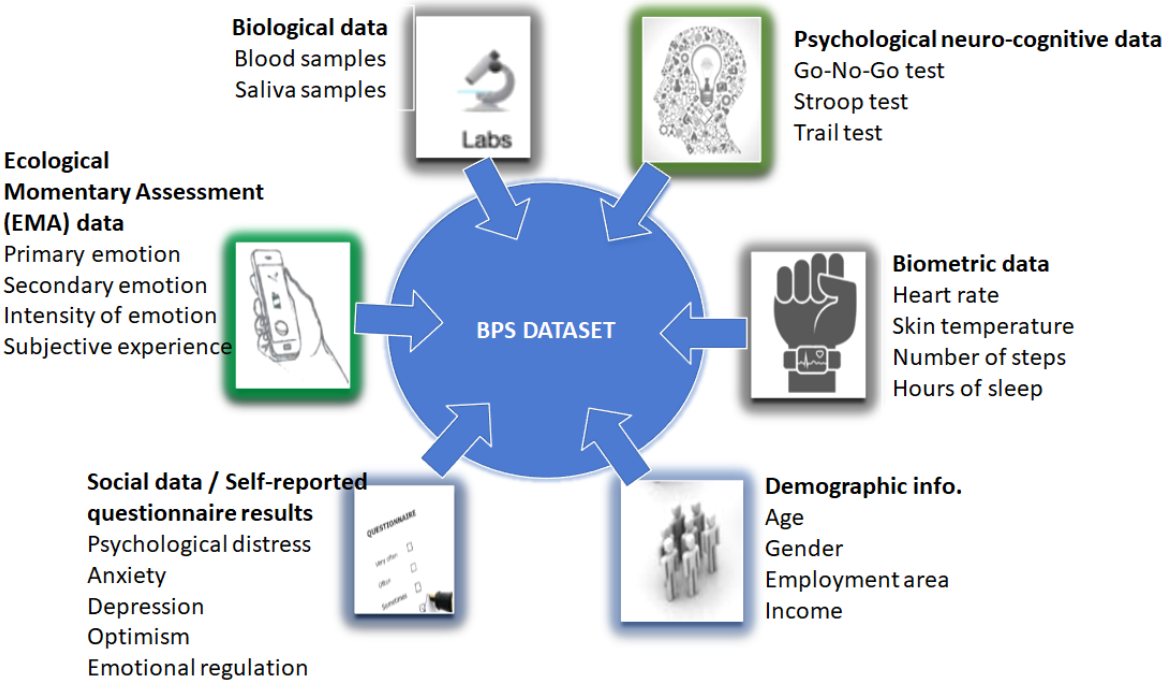


Figure 2.4: Biopsychosocial data mapped to the biopsychosocial model of health

2.8.2 Accuracy

The accuracy or value of data refers to their usefulness for the intended purpose. The end goal of the data analytic system is to extract value from the data. The value of the data is also related to the veracity of the data. The biopsychosocial data used in this research were unique, as they were real-world data collected from participants in the DHI program under real-world conditions. Blonde et al. [246] discuss the interpretation and impact of real-world clinical data and their limitations. They note that real-world data can be subject to bias and confounding factors. Missing data values in inconsistently collected real-world data can reduce the statistical validity of tests conducted. The biopsychosocial data in this study had missing values for different measures of interest - this was particularly the case for the physiological data collected via the wearable device. Since participation in the DHI program was voluntary, participants were not obligated to hold on to the wearable all the time during the experiment. Different imputation methods were reviewed and selected to deal with missing values to improve the value and veracity of the data.

2.8.3 Fitness of use

The data collected must be fit for use. Fitness of use or data veracity count for more than data quality and it must be examined in light of data fidelity. Data fidelity is defined as data quality in the context and appropriateness of its use. In practice, veracity depends on the participants' ability to report high-quality data, especially in the context of data collected using surveys and self-reported assessments of affective experiences [247–249]. Various data preparation tasks, such as data wrangling, normalisation, sampling and filtering, were used on the biopsychosocial data in this study to improve their veracity.

2.9 The Wellness study program framework

The DHI program was called the 'Wellness study'. For the remainder of this thesis, the resultant biopsychosocial dataset is referred to as the wellness dataset. This program was open to Australian residents aged 18 years and older who had access to the internet, owned a mobile phone and were willing to travel to one of Federation University's campuses on two separate occasions for the pre- and post-biological and psychological tests. The goal of the 'Wellness study' was to compare the effectiveness of the mindfulness and the physical activity interventions. These were online programs designed to improve wellbeing. To evaluate this, participants were randomly allocated to one of the three conditions: either the mindfulness program or the physical activity program for an immediate start (these participants belonged to the immediate start intervention group) or a waitlist control group. Those randomly allocated to the waitlist control group were required to wait eight weeks before commencing one of the interventions; however, they were allowed to choose the program they wanted to participate in. The mindfulness intervention was about learning how to become more aware of moment-to-moment experiences (e.g., thoughts, feelings, images and sensations). The physical activity interventions was about making a small change to the amount of physical activity the person did, helping them find an activity they would enjoy, identifying barriers and setting achievable goals. Both the intervention approaches were found to be mentally and physically beneficial and, to the best of our knowledge, no other online study has compared these two approaches as part of a randomised control trial in a real-life setting.

There were four main assessment phases as well as some 'during' intervention assessments/monitoring activities that involved collecting psychological (online questionnaires), biological (i.e., salivary cortisol samples, 15 ml blood sample), physiological (e.g., resting

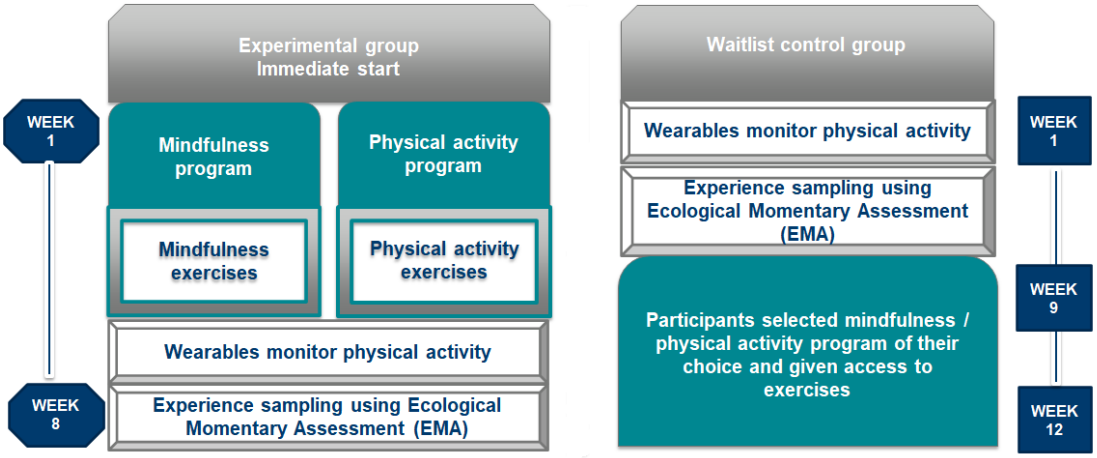


Figure 2.5: Structure of DHI program rolled out by School of Science, Psychology and Sport at Federation University

heart-rate variability) and cognitive (three short computerised cognitive tests measuring attention, inhibition and cognitive flexibility) wellbeing measures. All participants were asked to wear a BASIS wristwatch for seven weeks. The BASIS watches monitored the participants' heart-rate, skin temperature, number of steps taken and sleep. During this time, the participants also had access to BASIS biometrics data to track their measures. The framework of the DHI program is described in Figure 2.5.

2.10 Summary

This chapter has discussed the perspective of the biopsychosocial model and its success in clinical applications in the areas of chronic illness, pain management, psychiatric rehabilitation and workplace injury management. The efforts of health researchers and investigators to specify interconnections between classes of variables in the biology, psychology and social domains are optimistic. It demonstrates these researchers' and investigators' motivation to design studies that embrace all the variables of the holistic model and to analyse how these variables, including feedback loops, reciprocal influences, correlated variables and others, interplay. Although advances have been made in specifying connections between biological, psychological and social processes, the full potential of the biopsychosocial model remains untapped in many domains of medicine. A key reason for this is the insufficient emphasis on transdisciplinary collaboration that strives for theoretical and research developments to cultivate the multi-level, multi-system and multivariate nature of health processes.

With digitisation, adoption of the biopsychosocial model to deliver DHIs has increased. DHI programs generate large quantities of rich granular data, which is also called digital health evidence data. Mining digital health evidence data has led to a growing interest in investigating methods to extract useful information from these data and make data-driven decisions [144]. Data mining models provide the process framework for discovering actionable information from large digital health evidence data sets. As mentioned earlier, many studies have used analytics to examine clinical trial data generated from interventions. The use of analytics on the wellness dataset generated from the 'Wellness study' DHI program in a real-life setting, as presented in this thesis, is new and has not been done before. Data mining models provide the process framework for discovering actionable information from large digital health evidence data sets. The data mining process model makes use of a plethora of ML methods that help build predictive models. Predictive models are needed to integrate the diverse data generated from the 'Wellness study' DHIs. The technical challenges of integrating diverse data are examined in this thesis. Though these challenges exist, the use of advanced computational tools and methods facilitate the integrative study of the wellness dataset. The first step towards the integrative study of the transdisciplinary data was examining the characteristics of the unique wellness dataset in light of the intervention that produced these data. In the following chapter, we study the salient features of the wellness dataset in more detail using exploratory data analysis.

CHAPTER 3

SALIENT FEATURES OF WELLNESS DATASET

The current research project uses the wellness dataset generated from the 'digital health' intervention programs to engage in new knowledge-building discoveries. Biological, physiological (biometric tests), psychological (mood state, neuro-cognitive tests) and social/environmental (Ecological Momentary Assessment[EMA]) tests, were conducted and results collected for participants within the experimental group (the group that was exposed to the conditions of the experiment; mindfulness or physical activity digital health interventions) and the control group (group not exposed to the conditions of the experiment; no mindfulness or physical activity digital health interventions) during the program run period. All the experiment results were together labelled as wellness dataset. The design for data collection is not included as part of the work presented in this thesis. The data collection design was predetermined before implementing and rolling out the 'Wellness study' DHI program. In addition to the 'Wellness dataset', the participants' demographic data were also collected such as age, sex, and profession. These activities carried out in this chapter can be mapped to the Obtain, Scrub and Explore phases of the OSEMN data mining process model.

The study of the 'Wellness dataset' for the proposed research is mapped to the biopsychosocial model of health as illustrated in Figure 2.4. In this chapter, we will explore the salient features of the wellness dataset.

- Section 3.1 explores the data produced from the biological tests that were conducted

twice, once during pre-intervention and next during post-intervention;

- Section 3.2 explores the salient features of the biometric data that was generated from the wearable device;
- Section 3.3 explores the neurocognitive reaction time data produced from the PEBL Go/No Go tests;
- Section 3.4 explores the different kinds of social data collected as part of the EMA surveys and surveys using online questionnaires during the DHI program; and
- Section 3.6 summarises the chapter.

3.1 Analysing the biological data

The biological domain data consists of biological data and biometric data (also known as physiological data), as described in Figure 2.4. The biological data was collected from the blood and saliva tests conducted at the beginning and the end of the DHI program. The biometric data, henceforth referred to as physiological data, was collected from the wearable device worn by the participants during the DHI program. In this section, we will focus on the blood and saliva data and examine their salient features.

The blood tests measured the levels of Brain-derived neurotrophic factor (BDNF), Fibroblast Growth Factor (FGF2) and IL-6 (Interleukin-6). BDNF is a protein found in humans and is also called abrineurin. The BDNF gene encodes this protein. BDNF protein is related to brain development and promotes the survival of nerve cells (neurons) by playing a role in the growth, maturation (differentiation), and maintenance of these cells. FGF2 is related to fibroblast growth and is a growth factor as well as a signaling protein, encoded by the FGF2 gene in humans. FGF2 helps in the 'promotion of endothelial cell proliferation and the physical organisation of endothelial cells into the tubelike structure and thus promote angiogenesis, the growth of new blood vessels from the pre-existing vasculature' [250]. IL-6 is a protein produced by various cells in humans and helps regulate immune responses, making the IL-6 test potentially useful as a marker of immune system activation. There were two types of experiments conducted to test for BDNF:

1. qPCR (quantitative Polymerase Chain Reaction) test was used in the type-1 experiment to measure gene expression in total blood

2. ELISA (Enzyme-Linked ImmunoAssay) test was used in the type-2 experiment to measure protein levels in plasma (mean concentration data).

The 'Wellness study' DHI program predicted improvements after engaging in interventions such as mindfulness. BDNF and FGF2 levels are expected to be high after the health interventions, whereas IL-6, an inflammation marker, is expected to be low after the DHI program. Lower BDNF and FGF2 levels at the end of the DHI program ideally predict higher negative affective variables and vice versa, although not necessary. The negative affective variables described earlier are represented by the DERS (Difficulty in Emotion Regulation Scale) and DASS (Depression Anxiety Stress Scale) scores. These scores are deduced from participants' answers in response to self-report questionnaires during the start, middle and end of the DHI program.

The DHI program's goal was to improve the levels of BDNF and FGF2, and it was hypothesised that the mindfulness intervention would lead to stronger changes in ANS (Autonomous Nervous System) activity markers like cortisol and IL-6. The physical activity intervention would show stronger changes in BDNF. The goal of the DHI program was to improve the levels of BDNF and FGF2 and it was hypothesised that the mindfulness intervention would lead to stronger changes in ANS (Autonomous Nervous System) activity markers like cortisol and IL-6 and physical activity intervention would show stronger changes in BDNF. BDNF readings from qPCR experiments are analysed for the participants to determine which intervention is likely to show improvement in BDNF values at the end of the DHI program. The findings of this study are detailed in Section 3.1.1. There are many missing values for FGF2 and IL-6 in the post interventions tests carried out at the end of the DHI program. This reduced the number of samples, and hence no significant statistical tests could be carried out on this set of data.

The saliva tests were conducted on the participants to measure the cortisol levels. Cortisol is a hormone that is mainly released at times of stress and manages many important body functions alongside regulating a wide range of processes throughout the body, including metabolism, immune response, and the body's response to stress. The participants were provided with a saliva sample testing kit before the start of the DHI. They were asked to drool into a small tube marked with the date and time, and samples were frozen. These samples were then forwarded to the health-scope pathology by the Federation University biomedical lab for further analysis. Section 3.1.2, examines and investigates methods to pre-process the raw saliva sample data and adopt a suitable method to reduce dimensionality prior to data analysis.

3.1.1 Correlation of interventions and BDNF readings

Exploratory data analysis is carried out on the BDNF values for the participants in the DHI program. The BDNF readings at the end of the DHI program between the mindfulness and physical activity sub-groups determine which intervention helped participants increase BDNF values. A total of 12 participants were allocated to the Mindfulness intervention and 6 participants were allocated to the physical activity intervention. Table 3.1 describes the pre and post-intervention BDNF readings for the mindfulness intervention (condition-1) and physical activity intervention (condition-2). Static analysis of the BDNF values pre and post interventions reveals that 83% of participants have increased BDNF readings at the end of the physical activity intervention compared to a 33% increase in BDNF for the mindfulness intervention. The physical activity intervention shows stronger changes in BDNF compared to mindfulness intervention at the end of the DHI program, as was hypothesised earlier. The findings also reveal that, within the physical activity sub-group, for 66.67 % of participants, an increase in levels of BDNF shows a strong correlation with higher LOT-R (measures optimism) test scores recorded during week eight at the end of the DHI program. This is depicted in Table 3.2.

Intervention condition	BDNF Abundance extra Pre	BDNF Abundance extra Post	Delta change in BDNF
1	9.910787	3.255774	*decrease*
1	4.525258	0.525559	*decrease*
1	4.366568	0.024248	*decrease*
1	3.35195	2.009728	*decrease*
1	2.738715	0.0277	*decrease*
1	1.481441	0.115023	*decrease*
1	0.748461	0.798575	***increase***
1	0.72749	1.36462	***increase***
1	0.409802	10.50133	***increase***
1	0.013471	0.205613	***increase***
1	3.036692	1.642053	*decrease*
1	2.407441	0.028637	*decrease*
2	0.992748	2.955702	***increase***
2	0.197784	1.515717	***increase***
2	0.091379	1.631275	***increase***
2	6.200997	2.441047	*decrease*
2	0.299889	1.326846	***increase***
2	0.636839	3.936738	***increase***

Table 3.1: Exploratory data analysis of BDNF readings and correlation to LOT-R levels for the intervention groups 1 and 2 (1-mindfulness sub-group, 2-physical activity sub-group). (Text in **bold** indicate significant changes)

Delta change in BDNF	Change in level of optimism at the end of DHI
increase	not improved
increase	**improved**
increase	no change
decrease	**not improved**
increase	**improved**
increase	**improved**

Table 3.2: Strong correlation of BDNF readings to improved LOT-R levels at end of the DHI program. (Text in **bold** indicate significant changes)

3.1.2 Computation of single cortisol score

The drool method used to collect oral fluid for biological testing was used to collect cortisol levels. Four cortisol level readings were recorded during the saliva test conducted at the start and the end of the DHI. The four cortisol readings were recorded on a single day at spaced-out time intervals. The four readings were denoted by S1 (20 mins after wake), S2 (60 mins after wake), S3 (midday) and S4 (evening). Two methods [251], namely AUC_G and AUC_I are compared to compute a single cortisol score from the set of four readings.

1. AUC_G: This method describes the area under the curve with respect to the ground. The following formula is used for the calculation of AUC_G:
   $AUC_G = \sum_{i=1}^{n-1} \frac{(m_{i+1}+mi)t_i}{2}$

2. AUC_I: This method describes the area under the curve with respect to increase and is calculated from repeated measurements. The area under the curve is calculated with reference to the first value and ignores the distance from zero for all measurements, thereby emphasizing the changes over time.

The comparative study of both these methods shows that the AUC_G method calculates the total area under the curve of all the four cortisol readings; as the area of interest. It thus considers the difference between the single measurements from each other (i.e., the change over time) and the distance of these measures from ground to zero. In contrast to AUC_G, AUC_I ignores the distance from zero for all measurements, thereby emphasizing the changes over time. The AUC_G [251] method is proposed to compute the single cortisol score from the set of 4 readings at the start and the DHI end.

Table 3.3 describes the values chosen for *t0*, *t1*, *t2*, *t3* and *t4* based on the AUC_G method. The readings are named *S1*, *S2*, *S3* and *S4*. Table 3.4 lists the cortisol level computed for a sample of 10 participants based on this method.

Time variable	Minutes after wake-up when cortisol reading was recorded	Time and description	Time variable to calculate AUC_G	Interval between measures (AUC_G)	Cortisol reading to calculate AUC_G
t0	0	At wakeup (e.g., 6 AM)	*t0*	0	-
t1	20	20 mins after first reading	*t1*	0	*S1*
t2	40	40 mins after first reading	*t2*	0	*S2*
t3	360	At mid-day	*t3*	0	*S3*
t4	720	In evening	*t4*	0	*S4*

Table 3.3: Description of time and cortisol measure variables used in the AUC_G method.

S1	*S2*	*S3*	*S4*	*t0*	*t1*	*t2*	*t3*	*t4*	AUC_G computed cortisol level
34	39	15	7	0	20	20	320	360	4790
12	21	4	2	0	20	20	320	360	1540
27	28	14	6	0	20	20	320	360	4170
7	4	2	2	0	20	20	320	360	810
17	20	12	19	0	20	20	320	360	5650
9	17	4	2	0	20	20	320	360	1430
41	32	16	7	0	20	20	320	360	4910
21	14	11	7	0	20	20	320	360	3480
13	13	4	2	0	20	20	320	360	1390
7	10	2	2	0	20	20	320	360	930

Table 3.4: Computation of cortisol level for 10 sample participants from 4 readings using the AUC_G method.

Using the AUC_G method, we can apply a single biological score to a set of four cortisol level readings recorded on a single day that change over time. Cortisol and blood protein data collected as part of the biological tests were cross-sectional, and they were collected once at the start of the DHI and once at the end of the DHI program. On the other hand, the physiological data collected from wearable sensor devices were continuous. In the next section, we will explore the salient features of the physiological data, a subset of the 'Wellness dataset'.

3.2 Analysing the physiological data

As part of our broader effort to examine possible connections between the different biopsychosocial features, we study the bio-physiological factors revealed by the biometric data. The physiological data comprises features: *heart-rate*, *accelerator magnitude*, *galvanic skin response*, *skin temperature*, *steps* and is classified by *activity type*. In this study, we aim to measure the bio-physiological indicator of *stress* using heart-rate data. To achieve this, it is necessary to identify and utilise an effective measure of heart-rate data that can capture the magnitude of *stress*. We propose using 'peak heart-rate count' to measure heart-rate data to indicate *stress*.

Stress [252] can be defined as a 'response to change to maintain the state of stability or homology that the body has maintained against the stimulus to break the mental and physical balance and the stability of the body'. *Stress* can be used to indicate different things, depending on the context in which it is examined. We measure 'biological' *stress*, which refers to the participants' ability to cope with the frequency of high *heart-rate* values marked by *peak(s)*.

In a clinical setting, heart-rate variability (HRV) data collected from electrocardiogram (ECG) records have been a very popular measure to indicate *stress* [253–257]. The gold standard for clinical experiments has been to use HRV, and the best practice is to use ECG machines to capture HRV. Some prior studies [253, 256] show the effectiveness of using *peak(s)* in ECG as a measure to indicate *stress*. However, the use of ECG limits the analysis to controlled clinical settings. It cannot be easily used in uncontrolled 'real-world' settings.

With the digital revolution in wearable technology, users can now monitor their biometrics daily in a 'real-world' uncontrolled setting. Such wearable technologies generate biometric big data. Wearables such as *heart-rate* monitors offer the potential to leverage the biopsychosocial model's strengths [258] of health. Prior work, reported so far, has focused predominantly on well-controlled clinical environments. Reviewing available literature [259–261], it is noticed that there is work that examines the heart-rate variability data, collected from wearable devices, to indicate *stress*. However, these studies have been done in an experimental setting and are validated using self-reported *stress* levels. Hao et al. [259] present the practical challenges associated with measuring *stress* in real-life situations. To the best of our knowledge, there has been no study that has previously examined

the heart-rate data from wearable devices in a 'real-world' setting to indicate 'biological' *stress*.

We propose using the novel 'peak heart-rate count' metric to quantify the level of 'biological' stress to compare the stress level between the experimental and control groups and between the mindfulness and physical activity sub-groups; within the experimental group. For each sample participants, 'heart-rate peak count' is computed for each day. The daily 'heart-rate peak count' is aggregated, and the median 'heart-rate peak count' is computed for each participant in the two groups during the trial period. The spread of 'heart-rate peak count' is compared between the experimental and control groups' participants and between the mindfulness and physical activity subgroups; within the experimental group. This is done using measures such as range and IQR (Inter Quartile Range). Analysing this kind of 'real-world' heart-rate data from wearable sensor devices presents a practical set of limitations and challenges. We also include evaluating the significance and challenges of using *heart-rate* data obtained in a 'real-world' setting from wearable devices; as a measure of 'biological' *stress*. The study proposes a novel 'peak heart-rate count' as a metric to quantify the level of 'biological' *stress*. We apply this metric to investigate real-life *biopsychosocial* data for experimental studies.

3.2.1 Data set-up

The biometric data used for the present study is a subset of the *biopsychosocial* data. This study focuses on the *heart-rate* data collected as part of biometrics, using the *heart-rate monitor* device that the participants wore. The participants were divided into two groups:

1. Experimental group: participants in this group were randomly allocated to one of the two active *'digital health'* intervention conditions, i.e., (physical activity or mindfulness), during the trial period

2. Wait-list control group: participants in this group did not participate in any intervention during the trial period of the first eight weeks.

The participants in both experimental and control groups were asked to wear a BASIS smart-watch during the trial. The biometrics such as *heart-rate*, *skin temperature*, *galvanic skin response* and *steps* were collected. The wearable was equipped with an optical heart-rate sensor that continually tracked beats per second throughout the day, irrespective of what activity the participant was engaged in; i.e., *inactive*, *light activity*, *moderate activity*, *off* or *sleep*. The activities and activity levels were set-up on the participants'

wearable devices aligned to the experiment requirements. However, the control group participants did not receive access to the 'digital health' intervention programs (i.e., access to mindfulness or physical activity exercises) during the trial. The duration of the trial was for eight weeks. The control group had a delayed start in week nine and could then pick one of the two interventions. There was a one-month follow up which the control group participants did not go through. A sample of 20 participants' heart-rate data is used for this study; 10 from the experimental group and 10 from the control group. The experimental group samples include 5 samples from the mindfulness subgroup and 5 samples from the physical activity subgroup.

3.2.2 Heart-rate count as 'biological' *stress* measure

In this section, we describe the limitations of using HRV derived from wearable *'heart-rate'* data, a measure to indicate 'biological' *stress*. We also describe the different biometric data that was collected by the BASIS smart-watch wearable device. While we propose to use *heart-rate* data in our studies, we present a brief discussion on HRV which is commonly used to measure *stress*. We describe the limitations of using HRV when examining 'real-world' heart-rate data collected by wearable sensor devices and the need for a different measure to address this. HRV, also known as the heart period, is defined as the length of the R-R interval. HRV is measured in cardiac time units (beat-by-beat) [262]. On the other hand, *heart-rate* is defined as the inverse of the heart period usually measured by the length of the 'R-R interval' in the ECG. *Heart-rate* is measured in real-time units such as seconds and standardised in beats per minute [262]. While *heart-rate* focuses on average beats per minute, HRV measures the specific changes between successive *heart-rate(s)*. Generally, a low HRV (or less variability in the heartbeats) indicates that the body is under *stress* from exercise, psychological events, or other internal or external stressors. A high *heart-rate* value could indicate that the body is 'biologically' *stressed*. In the context of 'real-world' biometric data collected from wearable devices, we observe the following limitations in applying HRV to determine the level of 'biological' *stress*.

1. Although HRV can be extracted from time-series *heart-rate* data, conditions such as missing lengths of biometric data when the wearable was switched off or removed from the hand diminish data quality. The *heart-rate* data collected for the study was from a real-world setting. Participation in the *'digital health'* intervention program was voluntary. This implies that participants were not obligated to hold on to the wearable all the time during the experiment. This could have compromised the quality of HRV data if extracted from existing *heart-rate* data.

2. Imputing missing *heart-rate* data to compute HRV is not considered favorable for this study due to its various imputation challenges. The wearable device collected the biometric heart-rate data in real-time when the participants were engaged in different activities. Imputing this kind of real-time biometric data is complex and challenging to achieve.

Owing to the above-stated limitations, 'heart-rate' is the choice of measure instead of HRV. Over the last few years, the world has seen wearable technology cover a sizable segment of the technology industry [263]. Smart watches are wrist-worn devices that have 'real-time' fitness and activity tracking capabilities. Wearables accessories such as wrist bands and smart watches are in a unique position to evolve and carry forward the more integrative, holistic, *biopsychosocial* model. The data collected from wearable has the potential to indicate various biometric measures to the user. They can appropriately take actions to regulate them, steering the person towards elevated wellbeing.

The biometric data for the present study was generated from the BASIS smart-watch wearable that was worn by the participants during the DHI program. The technology in this wearable device includes wrist sensors that collect biometric data from human vital signs and activities. The biometrics monitored by the BASIS watch include heart-rate, skin temperature, galvanic skin response (GSR) and other derived statistics such as the number of hours of sleep. The wearable activity sensors measured gross user activity attributes, different from narrowly focused vital signs sensors such as either accelerator magnitude (for the sensor) or acceleration (for the number of steps).

3.2.3 Proposed method to measure 'biological' *stress* levels

We propose to explore the 'heart-rate peak count' as a metric on heart-rate data to measure 'biological' *stress* levels. The following sub-sections provide the details of the proposed method to define and measure heart-rate peaks.

3.2.3.1 Defining peaks

We consider two design parameters to define *peak(s)* in the heart-rate time-series data. The first parameter is the *interval* of *peak*. In our definition, this has been set to a 5 minute limit. By setting the *peak interval* to 5 minutes, we treat the recurring *peak(s)* within a 5-minute time span, as one single *peak* for that period. Considering to set such a time *interval* allows us to eliminate instances of counting separate recurring *peak(s)*, that may arise when a participant removes their wearable device and puts them back within a short

span of 5 minutes. The second parameter used is *maximum threshold* to define a *peak*. A *heart-rate data point* is defined as a *peak* if the magnitude of the data point is >*maximum threshold* and the *interval* of the identified *peak* is >5 minutes. We have the following two approaches for defining the *maximum threshold*.

1. Sum Of Mean and SD (MSD): *maximum threshold* is the sum of mean daily heart-rate and standard deviation computed on this mean

2. Sum of Median and IQR (MedIQR): *maximum threshold* is the *sum* of median heart-rate for the day and IQR of heart-rate data for the day.

MSD approach is not suitable since the computation of mean and standard deviation values assumes that the distribution of heart-rate time-series data is normal, whereas in reality the data distribution is skewed and not continuous. Hence, we propose to use the MedIQR approach, which is robust yet straightforward to the out-lier data.

3.2.3.2 Reference for measuring peak

Resting heart-rate is defined as the number of heart contractions per minute while at rest[264]. In our work, the resting heart-rate is used to define *peak(s)* in the heart-rate distribution. The heart-rate data for the 20 sample participants are filtered for the activities *Sleep* and *Inactive*, to compute the resting heart-rate. Studies indicate that resting heart-rate measurements in a sitting position tend to be 1-2 beats per minute, higher than in a supine position [264]. Physical exercise or any external intervention such as intake of some kinds of medication to lower blood pressure can affect the resting heart-rate [265]. As part of this study's data setup, the task is to investigate if the wearable reported heart-rate for activities *Sleep* and *Inactive*, truly describes the heart-rate when the participant is at rest. If they did, this allows us to use the resting heart-rate as the mean heart-rate. The accelerator magnitude is a good reference measure to validate the resting heart-rate. It is used to check this correlation of data collected by the wearable device during activities. A sample of 5 participants is chosen randomly for whom heart-rate values and accelerator magnitude values are available for different activity conditions.

3.2.4 Analysis of 'heart-rate peak count' as a measure of 'biological' *stress*

In this section, we present the *biopsychosocial* data analysis using 'heart-rate peak count' as the measure of 'biological' *stress*. Participation in the *'digital health* program was voluntary. The participants were not obligated to keep the wearables on throughout the experiment. The heart-rate time-series data collected was not continuous. The interruptions in the heart-rate data ranged from a few seconds to sometimes days when no biometric heart-rate data was collected. This affected the quality of heart-rate data collected and reduced the number of good quality samples that could be used for the study. The nature of the data distribution and the low quality of data limits the statistical tests such as ANOVA (Analysis of variance) that can be applied to the data set. The IQR describes H spread (mid-spread) and is the measure of the statistical dispersion between the 75th percentile (also denoted as quartile 3 or Q3) and the 25th percentile (also denoted as quartile 1 or Q1). It is a measure of variability based on dividing the range of heart-rate data points into quartiles.

3.2.4.1 Variations in accelerator magnitude

Analysing the heart-rate and accelerator magnitude data reveals the following.

1. The maximum value of accelerator magnitude dropped when the activity was *Sleep*
2. Average values for accelerator magnitude drops during activities *Sleep* or *Inactive* compared to conditions of *Light activity* or *Moderate activity*.

Based on the above results, we conclude that heart-rate values during activity conditions *Sleep* and *Inactive* describe heart-rate during rest and can be used as mean heart-rate to define *peak(s)*.

3.2.4.2 'Peak heart-rate count' metric

The distribution of the heart-rate data used for the study is not normal. The distribution of heart-rate data for a sample participant is illustrated in Figure 3.1. In this study, peak heart-rate is defined as the data point, which is greater than the *maximum_threshold* value. The *maximum_threshold* value is the sum of median heart-rate and IQR for the participant for the given day. The magnitude of stress is set to 5 minutes. The interruptions in the heart-rate time-series data and the poor quality of 'real-world' heart-rate data limits the

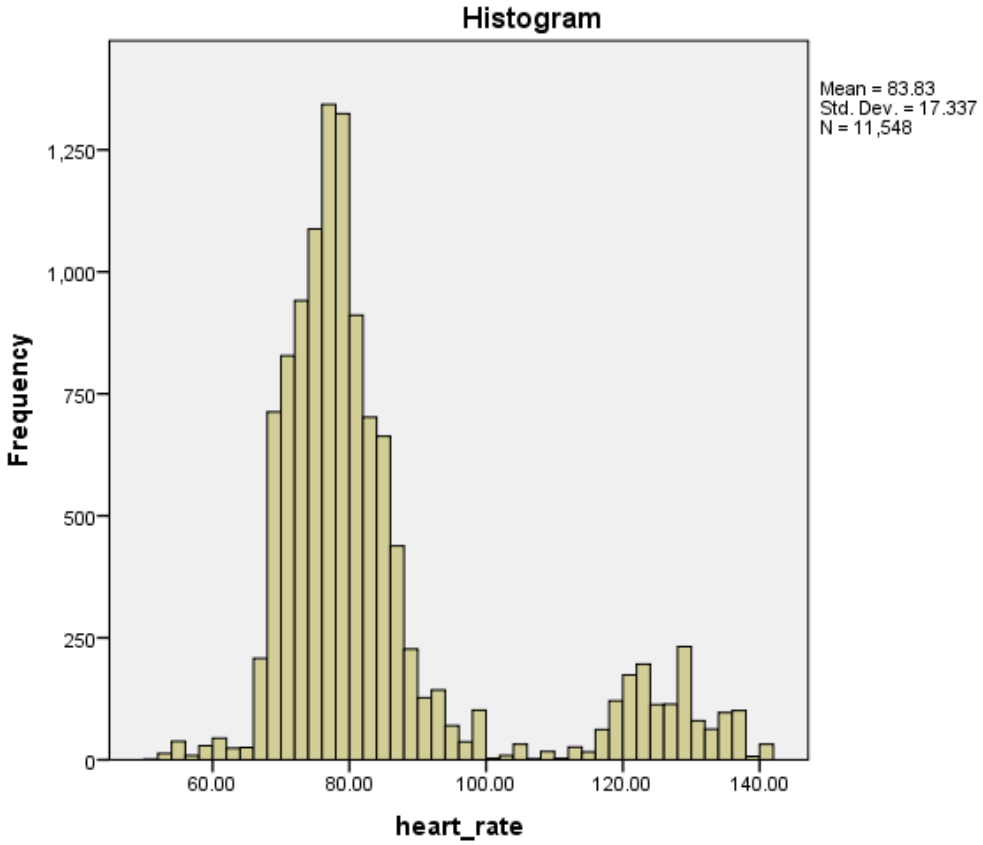


Figure 3.1: Distribution of heart-rate data for sample participant on a random day, during *Sleep* and *Inactive* conditions. X-axis: heart- rate and Y-axis: frequency of the heart-rate.

use of statistical tests to compare the levels of 'biological' *stress* between the experimental and the control groups and between the mindfulness and physical activity subgroups.

We conduct quantitative analysis using other attributes of the distribution, i.e., the IQR and range, to establish the suitability of 'peak heart-rate count' as a measure to indicate 'biological' *stress*. Table 3.5 describes the minimum, maximum, Q1, Q3, mean and range data values computed from the 'peak heart-rate count' distribution. It can be seen that the maximum value of 'peak heart-rate count' (49) causes the control group to have a wider range compared to the experimental group. The H spread of data in the control group is dispersed less consistently due to the wider range, reflecting the high IQR value of 17.25. Within the experimental-physical activity sub-group, we can see that the 'peak heart-rate count' are distributed more consistently within a narrower range of 4 and IQR of 2.5.

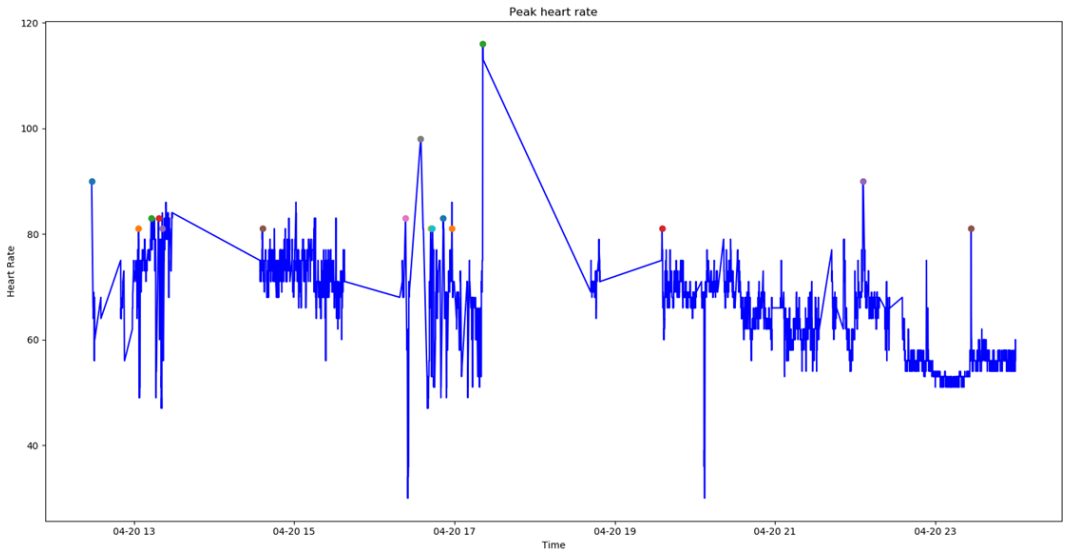


Figure 3.2: Plot of peak heart-rate (generated per second) for a sample participant for a single day. The periods with no fluctuating peaks and valleys represent timestamps where no heart-rate was plotted owing to the absence of activity: *Sleep* and *No Activity*. The x-axis on the graph describes the time date: 2015-04-20, hours range 13:00:00 to 23:00:00. A large number of peak(s) in heart-rate distribution indicates that the participant is experiencing stress.

Table 3.5: Summary of computed Range and IQR values

Variables	**Mindfulness group**	**Physical activity group**	**Control group**
Minimum	16	24	14.5
Q1	20.5	25	19.375
Median	26	27	26.75
Q3	27	27.5	36.625
Maximum	28	28	49
Mean	23.5	26.3	28.35
IQR	6.5	2.5	17.25
Range	12	4	34.5

The experimental-mindfulness subgroup has a larger range of 12 and IQR of 6.5, which reflects the wider spread between the data points. The box and whiskers plot depicts the shape of the data distribution, median and variability. The 'peak heart-rate count' for the experimental-physical activity sub-group participants lies within a smaller range between 24 and 28, whereas the 'peak heart-rate count' for the experimental- mindfulness sub-group participants lies within a wider range of 16 to 28. The IQR is depicted by the shaded area

(see Figure 3.3), which shows a relatively consistent distribution of 'peak heart-rate count' for the experimental-physical activity sub-group as against the experimental-mindfulness sub-group.

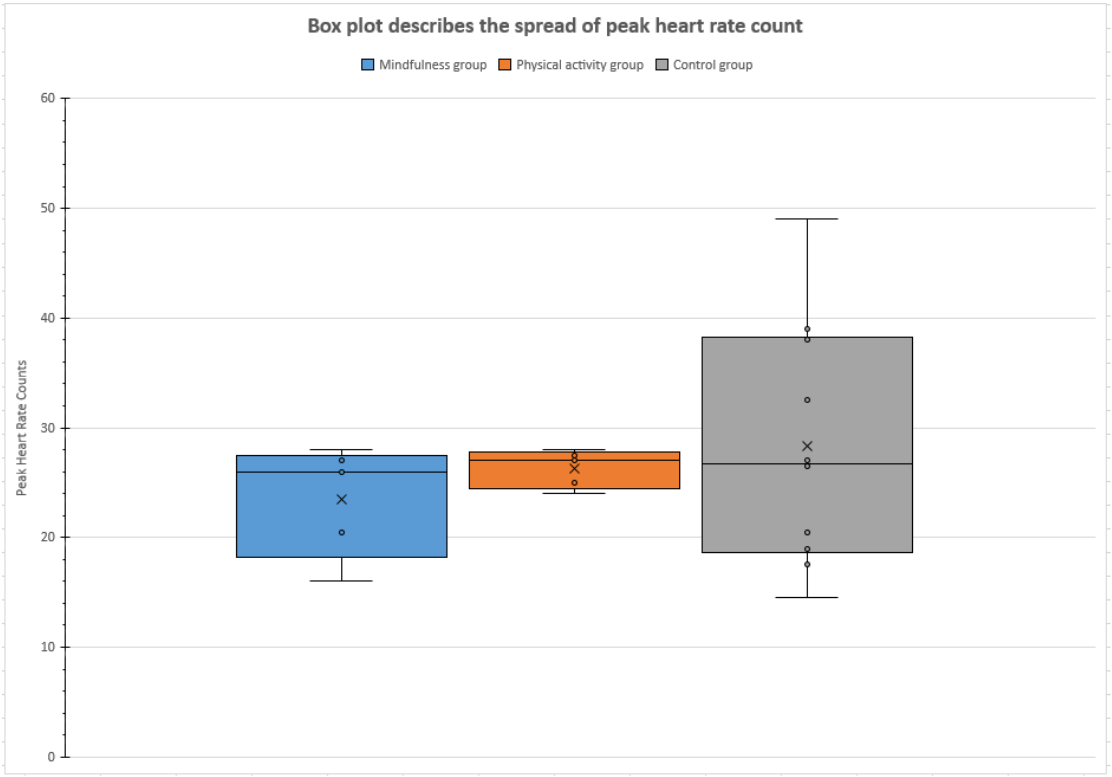


Figure 3.3: Box plot describes spread of median 'peak heart-rate count' within IQR for different groups. Whiskers describe the spread of outlier median 'peak heart-rate count' values.

The proposed new metric, 'heart-rate peak count' and the associated new MedIQR method is seen to be useful in analysing the distribution of the heart-rate peaks, using range and IQR values. Its application on 'real-world' heart-rate data collected from wearable devices presents several challenges related to diminished data quality and limits the application of statistical tests. The observations made from the real-world biometric data are limited by measurement errors. Based on the readings of the IQR measure and range, the participants in the experimental-physical activity sub-group were observed to have 'peak heart-rate count' within a smaller range of 4 with recording a maximum 'peak heart-rate count' of 28 and a minimum 'peak heart-rate count' of 24. On the other hand,

the participants in the experimental-mindfulness activity sub-group were observed to have a 'peak heart-rate count' within a broader range of 12 with recording a maximum 'peak heart-rate count' of 28 and a minimum 'peak heart-rate count' of 16. The IQR and range of the control group participants were larger than the experimental group which indicates that the *'digital health'* intervention has a higher probability of positive impact on the experimental-physical activity sub-group participants. The 'heart-rate peak count' has been used to indicate 'biological' stress, experienced by the participant during the intervention. The results from the study can support the hypothesis that the *'digital health'* intervention program benefited the experimental group participants, particularly the physical activity sub-group.

The biopsychosocial approach to health-care data, emphasises the importance of understanding human health and illness considering the biological, psychological and social factors and their complex interactions in understanding health, illness and health care delivery [266]. The work in this study opens up avenues for further research in the area of real-world data mining from other domains such as psychological and social data on the same longitudinal time line as the biometric data, to provide correlations and validate a composite measure as an indicator of 'biological' *stress*.

3.3 Analysing the psychological data

In this section, we will focus on the psychological discipline of the Biopsychosocial model by analysing the real-life neurocognitive data generated from the neurocognitive tests conducted at the start and end of the DHI program. Three kinds of psychological tests were conducted for participants who participated in the DHI program. They were the PEBL(Psychology Experiment Building Language) based Go/No-Go test, Stroop test and the trail test. We mainly focus on the distribution of reaction time data obtained from participants undertaking the PEBL based Go/No-Go test, using EDA(Exploratory Data Analysis). The neurocognitive tests are used to evaluate a persons' cognitive status such as memory, attention, reaction time, response inhibition along with other cognitive functions.

(PEBL) [244] is an open-source software system for designing and running psychological experiments. In these experiments, participants are presented with a series of binary visual stimuli — the letter R or P. The participants must press a key to indicate when they see the letter P which is configured as the target stimuli for the experiment. The distribution of P versus R is adjusted to be P in 80% of trials for the first round (Round

1) and 20% of trials for the second round (Round 2). The PEBL Go/No-Go test is used to measure a participants' capacity for sustained attention and response control under different experimental conditions or using interventions [267]. The PEBL Go/No-Go test results are analysed in terms of accuracy and reaction time.

'The Go/No-Go task is a response inhibition task where the motor response must either be executed or inhibited' [268]. In this test, the participants are presented with a sequence of letters on a visual tool and are asked to respond to a single target letter by clicking a button. During the test, a 2*2 array is presented which displayed 4 stars one in each square of the array. The single letter P or R is displayed in any one of the squares for a timed duration of 500 milliseconds. The inter-stimulus interval between each letter being displayed is 1500 milliseconds. For the first condition P-Go, participants were asked to respond by clicking a button when they spotted the target letter P and at the same time they were asked to hold back their response to the non-target variable R [267]. The ratio proportion of letter P to R, described under the P-Go condition was 80:20 and the total number of trials for this condition was 160. The second condition was a reversal condition called R-Go, where the participants were asked to respond by clicking a button when they spotted the target letter R and withhold their response to the non-target letter P. The ratio of targets to non-targets stayed the same as in condition P-Go. The P-Go and the R-Go conditions together were made up of 320 trials (160*2). Before taking the Go/No-Go test, the participants were allowed to take a brief practice session to ensure everyone understood how the test worked and what they were required to do. The behavioural performance of the test was evaluated by calculating 4 metrics in each condition: a. Correct responses to target b. Errors of omission c. Errors of commission d. Correct rejections to the No-Go letter. These typically represent sensitivity, specificity, F-score and precision. In addition to these metrics, reaction time (RT) to the Go letter was assessed and computed for the participants. RT stores the reaction time to the stimuli, irrespective of it being a correct response or an incorrect response. Where the reaction time is 1452 ms, it shows that this is the largest response time which identifies participants' inaction (No Response cases). The minimum and maximum reaction of Go/No Go test reaction time data can be analysed with omission and commission errors (see Figure 3.4). In this section, we focus on analysing the distribution of participants' reaction time. We examine some widely used existing methods to describe the distribution of reaction time data. We propose the use of Gamma distribution to describe neurocognitive reaction time data.

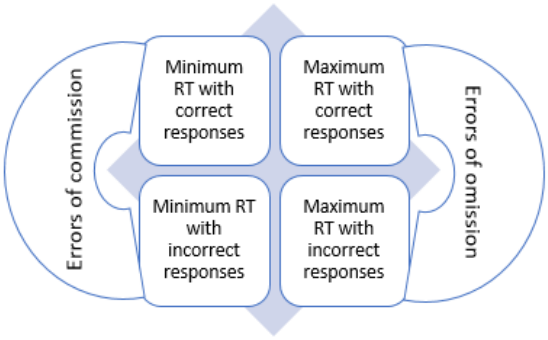


Figure 3.4: Minimum and Maximum Reaction time and relation to error types

3.3.1 Existing methods and their limitations

In this section, we present some of the well-known methods which are widely used to describe the distribution of reaction time data. We also discuss and highlight the limitations of these methods in representing neurocognitive data sets.

3.3.1.1 Mean based ANOVA analysis

Analysis of reaction time data is usually carried out using statistical methods such as Analysis of Variance (ANOVA) on the sample mean which assumes that the distribution of the reaction time data is normal. Recently, Moniz et al. [269]. aimed to present the results of normative data to study Executive Functions (EF), to identify the influence of variable age in reaction time concerning response inhibition and attention, using ANOVA. However, the approach is limited since reaction time data as generated by Psychological Go/No-Go tests usually cannot be considered as normally distributed and hence, using statistical methods such as ANOVA, can produce results that are not truly representative of the problem conditions. Using hypothesis tests on data that are skewed or heteroscedastic, reduces the power of these tests and can fail to detect a real difference between experimental conditions [270].

In our study, the PEBL Go/No-Go tasks' recognition reaction time data, for a typical participant who undertook the 'Mindfulness' program was examined. The Kolmogorov-Smirnov test (KS-test) and Shapiro-Wilk test were done to test the normality of the collected reaction time data for the 137 trials, using SPSS statistical software. The alpha level of significance was set at 0.05. Probabilities > 0.05 mean the data are normal. Probabilities < 0.05 mean the data are not normal. Figure 3.5 illustrates the typical density distribution in the form of a histogram of PEBL Go/No-Go test generated reaction time data for a sample

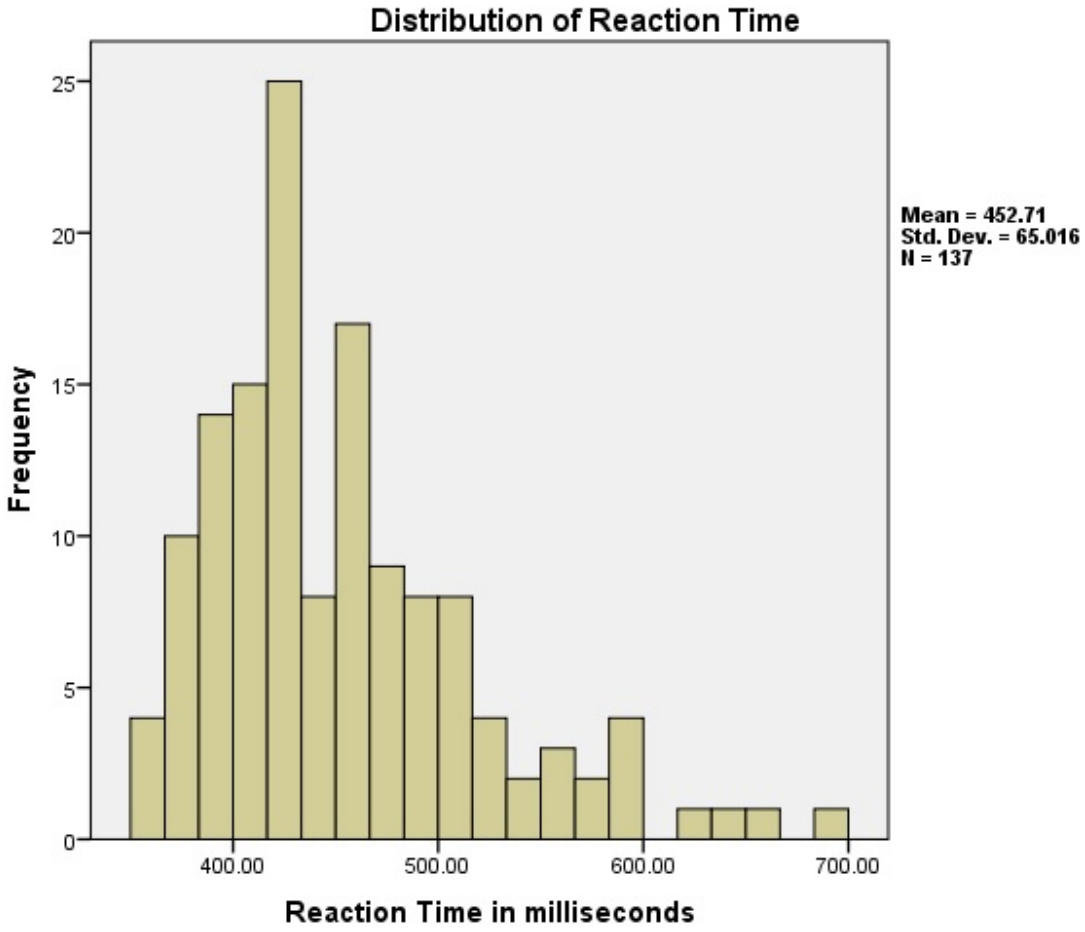


Figure 3.5: Sample density distribution of PEBL Go/No-Go test reaction time data

participant. Figure 3.6 describes the SPSS output for test statistics. Two tests for normality are run. If the dataset is small then we can use the Shapiro-Wilk test. For large datasets, the Kolmogorov-Smirnov test is used. We have used the results of the Shapiro-Wilk test as our data set contains just 137 trials.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Degrees of freedom	Significance	Statistic	Degrees of freedom	Significance
ReactionTime	.121	137	< 0.0	.920	137	< 0.0

a. Lilliefors Significance Correction

Figure 3.6: Results of Normality test

W is the test statistic and is given by $W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

'Statistic' is the *test statistic* W for the Shapiro-Wilk test. 'Significance' is the *significance* for the test, i.e., the p-value. The p value is <0.0 for the Shapiro-Wilk test, which is less than 0.05 (the alpha significance level set prior to the test). Hence we conclude that the sample data are significantly different from a *normal distribution*. From Figure 3.5, it can be seen that reaction time data are not normally distributed but rises rapidly on the left and typically has a long positive tail on the right. Reaction times of other participants also mostly follow this shape, with a highly-skewed uni-modal distribution. This is characteristic of the psychological reaction time task responses, such as in PEBL Go/No-Go tasks, where participant's reaction to a target is gathered within a range, e.g., between 350 MS to 600 MS, after which there are delayed reactions, that stagger the data to have a long right tail. Reaction time can vary across trials for the same participant under the same condition. Small values for reaction time could be a result of fast guesses and very high values of reaction time could be because the participant was inattentive. These reaction times can strongly influence the outcome of hypothesis tests. Reaction time data has been commonly analysed using the mean and standard deviation. The mean and variance in a *normal distribution* are not interdependent. When using ANOVA to compare distributions between two experimental conditions, it would be hard to conclude, whether a statistically significant difference arose from a real change in the means or in the variances. In the context of reaction time generated from psychological tests, the mean and standard deviations are not robust measures for analysing variations of reaction time data of participants between experimental conditions, for the following reasons.

- Mean is not reflective of typical response if the distribution is skewed, because the mean is distorted in the direction of the *skewness*

- Standard deviation can be greatly increased by a relatively low number of slow reaction time(s)
- when analysing reaction time between two or more experimental conditions, the mean reaction time can be the same even if two populations are genuinely different. This has been described in a scenario where the modal part of the distribution can decrease in value while the number of data in the tails can increase, thereby producing a null effect of the condition [270].

Hence, examining the mean alone could obscure interesting details, such as the behaviour of fast and slow reaction times, across the conditions of an experiment.

3.3.1.2 Applying transformations

Where the reaction time data is not normally distributed, statistical methods such as ANOVA cannot be applied directly even though these tools come with their limitations as discussed in the previous section. One approach commonly applied by researchers is to transform the reaction time data to follow a *normal distribution* such that statistical methods such as ANOVA can be applied. With skewed data such as reaction time, data transformations have the potential to lessen the impacts of skew by reducing the larger values to a greater extent than smaller values. Care must be taken when transforming reaction time data because it is possible to eliminate significant effects by transformation and this, in turn, could lead to issues of interpretation after transformation [271]. Lo and Andrews [272] have discussed the theoretical implications of applying transformation on reaction time data generated from psychological studies. The study indicates that by routinely applying transformations to yield the *normal distribution* required for Linear Mixed Models (LMM), the researcher may ultimately fail to test their hypothesis, using the dependent variable that underpinned their theoretical predictions. It also indicates how, to interpret the results and compare with earlier published ANOVA data, the estimates are often 'back transformed' to raw reaction time metric, which can be unreliable.

3.3.1.3 Ex-Gaussian distribution

The exploratory data analysis must be conducted on the whole distribution of reaction time data to identify a parsimonious approximate description [270]. Some of these approaches have been examined earlier with other kinds of data distributions that are not normal but have not been specifically applied to reaction time data generated from the psychological PEBL Go/No-Go tests under an experimental condition. Heathcote et al. [273] examine

how analysis based on mean reaction time does not take distributions' shape into account, leading to obscure aspects of performance. In this study, the author has referenced the work done by Hohle [274], where response time distribution has been described using a 3-parameter model consisting of the convolution of the normal and exponential distributions, the Ex-Gaussian. The use of Ex-Gaussian distribution was supported by the argument that the duration of the decision process might be exponentially distributed and suggested that the residual latency of the decision process might be approximately Gaussian. This explanation appears valid in the given context where the reaction can be described as an output that is generated when a system receives input and grinds it through a series of discrete processing stages. In order to react, one needs to make a decision first which is representative of one or more discrete processing stages. This process of decision making can be viewed as exponential in nature. The residual latency of the decision process is indicative of the output reaction time at the end of the decision-making stage. The studies reported in [275] and [276] have applied Ex-Gaussian distributions and have shown these to fit response time distributions well. The implementation of Ex-Gaussian distribution using maximum likelihood estimation has been described in detail using MATLAB in [277]. However, the reaction times being measured in the studies using the Ex-Gaussian distribution are not necessarily from Psychology based tests but could pertain to the analysis of reaction time in other domains. Sternberg [278] argues that this method of modeling reaction time using Ex-Gaussian distributions may not be suitable for all processes where reaction time is monitored. The author further explains that although such a model can describe waiting times between initiation of independent telephone calls in large parallel networks, or the times between decays among the innumerable atoms in a lump of radioactive material, this model cannot be applied to describe the mental process that accomplishes something using and building upon prior knowledge or information stored in memory, unless all the work is done in an instant. This limitation is attributed to the 'no-memory' property unique to the exponential distribution. Although the Ex-Gaussian distributions cannot model all forms of reaction time, the Ex-Gaussian distribution has two excellent properties; (i) it is well parameterised with location (shift) parameter, a separate skew parameter as well as a *scale* parameter and (ii) it fits some kinds of reaction times very well. However, the method is limited by the need to examine the nature and context of reaction time data that needs to be analysed before applying this distribution method.

3.3.2 Proposed method to model neurocognitive reaction time

In order to analyse different other properties of reaction time distributions, different from arithmetic means, Generalised Linear Models (GLMs) models were examined. The use of GLMs allows statistical assessment on original reaction time metric and also helps to meet the mathematical constraints set by the statistical model. The assumptions for GLM are more relaxed compared to the strict assumptions laid out for LMM, which incorporates ANOVA. For instance, by using GLM, response variables can follow error distribution models other than a *normal distribution*. In a GLM model, each outcome Y of the dependent variable is assumed to be generated from a particular distribution in the exponential family, which includes the Gamma probability distribution.

As stated earlier, although GLM based on a family of exponential distributions including Gamma distribution is used to analyse reaction time in other domains, they are not used in psychology to analyse participants' reaction time on neurocognitive tests. We propose Gamma distribution for representing the distribution of recognition reaction time data, produced from the PEBL Go/No-Go tests. As elaborated earlier, the primary motivation to choose Gamma distribution to describe reaction time data is that it is more favorable to describe reaction time data compare to the traditional Gaussian distribution or transformations to Gaussian because mean and variance are both calculated from the same *shape* and *scale* parameters whereas the mean and variance in Gaussian distribution are not inter-dependent. This implies that as the means of the Gamma random variable increases, its variance also increases. This property of a Gamma distribution provides more confidence and power to use the distribution, to represent probability distributions of reaction time data. We can further extend the analysis to compare the Gamma distribution parameters between experimental conditions to tell whether the statistic test was significant enough. The Gamma distribution is defined as the sum of a series of exponential processes. Each exponential process may have a different scale but the average scale of the processes is captured by the *scale* parameter β. The *shape* parameter α reflects the approximate number of exponentials contributing to the function [279].

A random variable X that is Gamma-distributed with *shape* α and *scale* β is denoted by:

$$X \sim \Gamma(\alpha, \beta)$$

The density function [280] of the Gamma distribution is given by:

$$f(x|\alpha,\beta) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma\alpha} \tag{3.1}$$

for x > 0 and $\alpha, \beta > 0$, where $\Gamma(\alpha)$ is a complete Gamma function. The *shape* and *scale* parameters are related to each other through the mean and variance of the Gamma distribution.

$$Mean = \alpha * \beta \tag{3.2}$$

$$Variance = \alpha * \beta^2 \tag{3.3}$$

3.3.3 PEBL Go/No-Go experiment study

The sample data used for studies here is the reaction time data generated from the PEBL Go/No-Go test. This data was collected for 10 participants, 2 males and 8 females, with a mean age of 36.2 years. The sample data was collected as part of a digital health intervention program by School of Science, Psychology and Sport at Federation University; from a group of Australian participants who subscribed to participate in this program voluntarily. Participants were presented with a series of binary visual stimuli —the letter R or P. The participants were asked to press a key when they saw the letter P which was configured as the target stimuli for the experiment. Each participant underwent a total of 160 trials as part of round 1 (block 1) where target P was shown 80% of the times and 160 trials as part of round 2 (block 3) where target P was shown 20% of the times. The size of PEBL Go/No-Go test trials was initially set to be equal (160 trials for each round) for each participant.

3.3.4 Analysis of reaction time data distribution

The pre-processing of data reduced the size of trials for the participants in each round even though the size of PEBL Go/No-Go test trials was set to be equal for each participant. The pre-processing was necessary to remove timeout data, i.e., reaction time data where the timeout instances exceeded 1450 milliseconds. All analysis was conducted using the Statistical Package for the Social Sciences (SPSS), version 23 and R, version 3.3.1. The level of significance was set at $p < 0.05$. Exploratory data analysis was used to summarise the descriptive statistics. Figure 3.7 illustrates the kernel density plots and their *skewness* and *kurtosis* values for the 9 samples that were analysed for round 1 PEBL Go/No-Go test trials. The density distribution plots clearly show that the participant reaction time data, in most cases, is positively skewed. The data distribution is unimodal having one local maximum except in one instance, for participant ID 679_1, where there are two peaks. The *skewness* of reaction time data in the kernel density plots describes the measure of datasets symmetry while the *kurtosis* of the reaction time distributions refers to the measure of the combined weight of the tails relative to the rest of the distribution. The results of the Anderson Darling normality tests reveal that in 9 out of 10 cases, the sample distributions were not normally distributed (p-value was smaller than the alpha significance level set at 0.05). These 10 sample distributions were analysed for Round 1 of PEBL Go/No-Go test trials.

In order to assess other properties of the reaction time data distribution, such as *skewness* and *kurtosis*, we plotted the Cullen and Frey graph in statistical tool R. All the 10 sample distributions of reaction time data were analysed for round 1 and round 2 of the PEBL Go/No-Go test trials. In total, we analysed the *skewness* and *kurtosis* of the reaction time distributions for 20 samples. The R Cullen and Frey graphs were used to visualise the likely distribution, the *skewness* and *kurtosis* followed. These graphs for the sample data are shown in Figure 3.8 and Figure 5. The observation is depicted by a blue circle that describes the likely distribution followed by the *skewness* and *kurtosis* of the sample reaction time data. From Figure 3.8, we can see that 'Observation' on *skewness* and *kurtosis* are more likely to follow a Gamma or a Weibull distribution for 7 out of 20 samples. Although the 'Observation' in 9 samples fall under the range of a Beta distribution, as described in Figure 3.9; this distribution belongs to a family of continuous probability distributions defined in the interval [0,1], and hence does not suit our study of reaction time data which can be any random positive number. The closest distribution for these samples is the Gamma / Weibull distribution as seen from the Cullen and Frey graphs. Four

samples examined fall outside the range of any descriptive distribution. As the next step, we used the test statistic from the sample Kolmogorov-Smirnov test (KS-test) and Akaike information criterion (AIC), to determine the most compatible distribution for our reaction time data. The goodness-of-fit tests (in this case, KS-test) is used to indicate whether it is reasonable to assume that the random sample of reaction times comes from the specified Gamma / Weibull distribution. The AIC is a measure of the relative quality of statistical models for a given set of data. Figure 3.8, lists the Gamma distribution parameters for the populations' samples, computed using Maximum Likelihood Estimation (MLE). An alternative to the MLE is the Method of Moments estimation which is known to be a poor estimator, owing to inefficiency, for small *shape* values [281–283]. The null and alternate hypothesis for conducting the KS-test is stated as follows: H0: Sample data comes from the stated distribution HA: Sample data do not come from the stated distribution The level of significance was set at: $p < 0.05$.

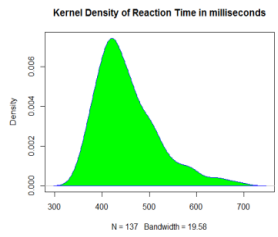
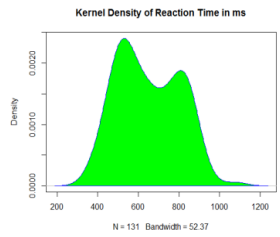
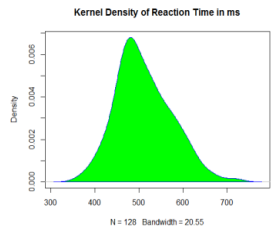
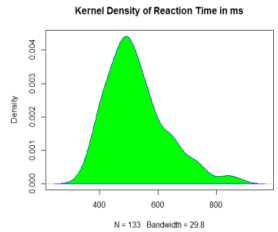
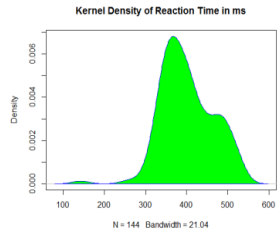
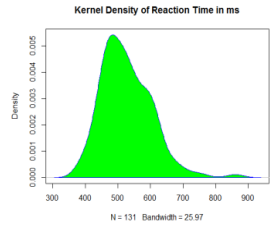
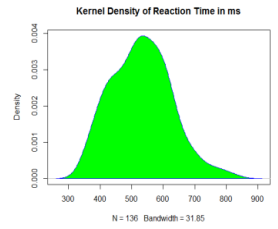
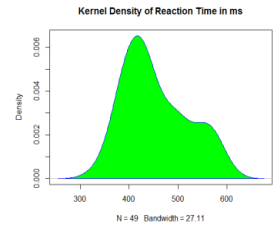
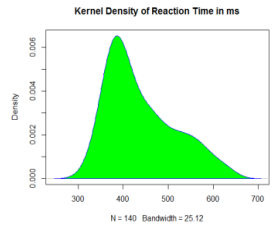
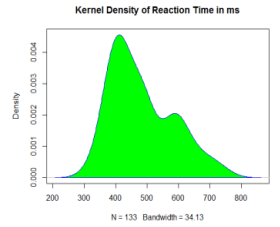
Kernel density distribution plots		
673_1  Skew: 1.1306 Kurtosis: 1.2699 P value of AD-test: 0.0002	679_1  Skew: 0.2183 Kurtosis: -0.9377 P value of AD-test: 0.00012	678_1  Skew: 0.4894 Kurtosis: 0.1204 P value of AD-test: 0.0235
711_1  Skew: 0.9095 Kurtosis: 0.7475 P value of AD-test: < 0.0	712_1  Skew: -0.0815 Kurtosis: 0.7909 P value of AD-test: 0.0002	744_1  Skew: 0.9629 Kurtosis: 1.8890 P value of AD-test: 0.0031
732_1  Skew: 0.3453 Kurtosis: -0.1067 P value of AD-test: 0.425	733_1  Skew: 0.5250 Kurtosis: -0.8474 P value of AD-test: 0.0028	739_1  Skew: 0.7708 Kurtosis: -0.4149 P value of AD-test: < 0.0
	738_1  Skew: 0.6996 Kurtosis: -0.4054 P value of AD-test: < 0.00	

Figure 3.7: Kernel density distribution plot for reaction time data (Round 1)

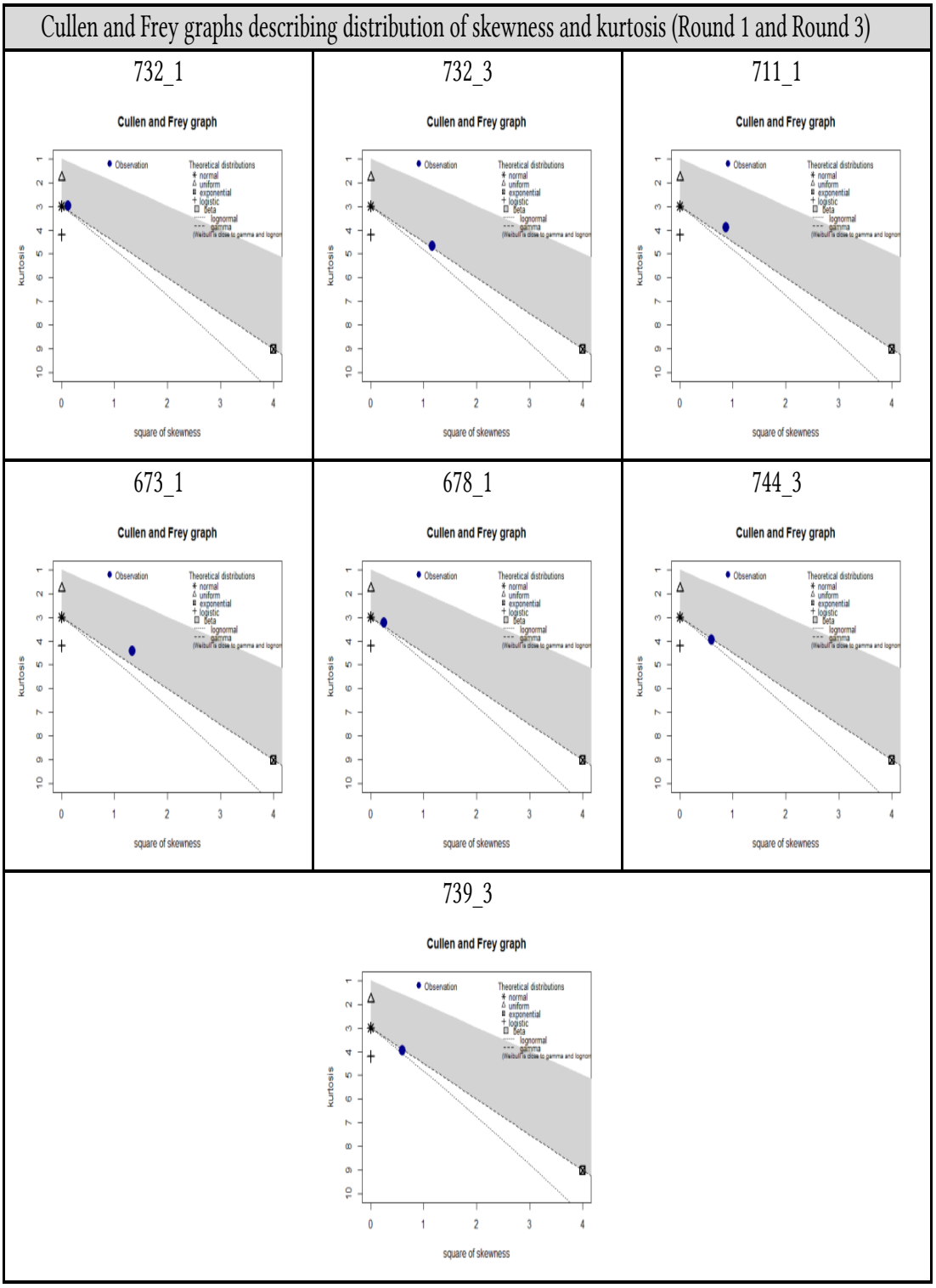


Figure 3.8: Visualisations from Cullen and Frey Graph (Gamma/Weibull)

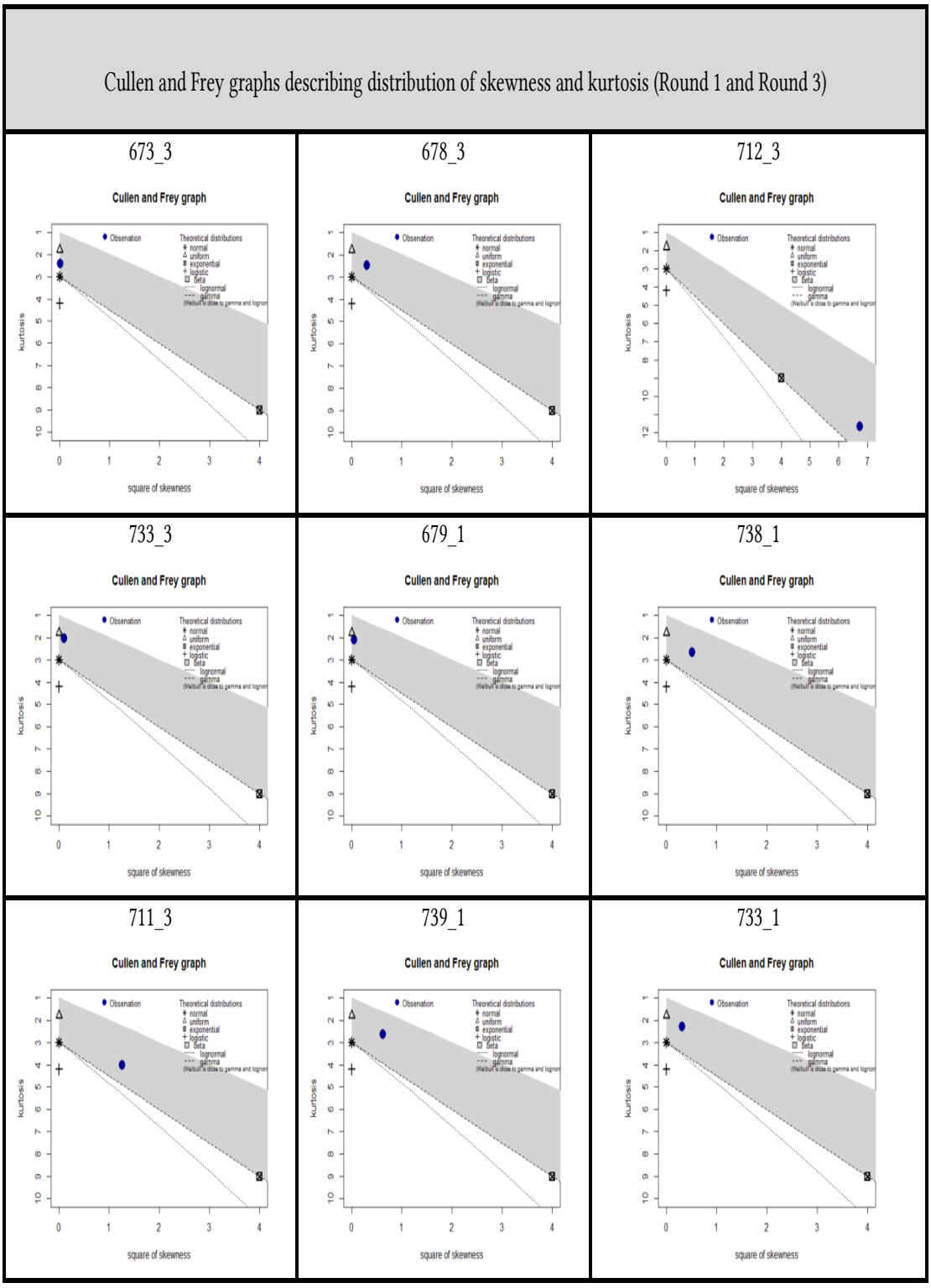


Figure 3.9: Visualisations from Cullen and Frey Graph (Beta distribution)

ID	Gamma				Weibull				Preferred distribution
	Shape	Rate	AIC	One-sample KS test statistic	Shape	Scale	AIC	One-sample KS test statistic	
673_1	52.9126	0.1169	1522.9580	0.0674	6.4652	481.9701	1567.3580	0.0124	Gamma
673_3	93.0714	0.1717	363.2773	< 2.2e-16	10.4457	567.4644	365.6816	0.8798	Gamma
678_1	67.7527	0.1324	1423.1180	0.4206	8.1031	540.0527	1446.8960	0.1511	Gamma
678_3	208.3214	0.3730	328.6786	0.7001	14.2221	577.4763	335.2543	0.6630	Gamma
679_1	17.6205	0.0271	1691.5870	0.2407	4.6517	710.3543	1696.7930	0.1157	Gamma
679_3	49.9474	0.0809	380.4244	< 2.2e-16	6.2800	657.3020	390.4055	0.3427	Gamma
711_1	27.3583	0.0521	1604.4030	0.4204	4.9638	569.2789	1634.8690	0.0367	Gamma
711_3	25.8948	0.0428	399.7757	0.6959	4.7449	656.7764	408.0075	0.5935	Gamma
712_1	37.7439	0.0940	1613.8370	0.4494	6.9533	428.1976	1612.2470	0.0790	Gamma
712_3	16.7876	0.0357	483.0640	0.0051	7.8300	498.6794	453.1952	0.2854	Weibull
732_1	31.8991	0.0601	1622.8390	0.7975	5.8809	570.6043	1637.4350	0.6868	Gamma
732_3	46.1180	0.0737	407.5460	0.8360	6.1565	667.8267	417.4847	0.6055	Gamma
733_1	50.6413	0.1115	549.7175	0.2309	7.2329	483.2539	557.7254	0.0735	Gamma
733_3	22.9027	0.0388	153.2775	0.6219	5.2239	641.5653	154.0669	0.6982	Gamma
738_1	24.5387	0.0508	1596.2200	0.0842	4.9433	525.0606	1619.4600	0.0310	Gamma
738_3	49.9565	0.0894	385.6924	0.9738	6.8326	594.1733	392.0527	0.0000	Gamma
739_1	36.0768	0.0822	1600.3410	0.0112	5.8631	471.8065	1629.8370	0.0091	Neither Gamma nor Weibull
739_3	68.1439	0.1324	358.9912	0.3026	10.4674	540.0222	353.9073	0.8317	Weibull
744_1	50.2101	0.0953	1502.9610	0.5504	6.3390	560.8972	1541.9170	0.0560	Gamma
744_3	136.2196	0.2397	352.8615	< 2.2e-16	12.4648	590.4582	356.5301	0.6442	Weibull

Figure 3.10: Computation of Gamma distribution parameters

The results from the KS-test show that most samples of analysed reaction time data are more compatible with Gamma distribution than with a Weibull distribution. Based on the KS-test results, we can accept the null hypothesis that the sample reaction time data follows Gamma distribution. The AIC values are also lower for a Gamma compared to the Weibull distribution [284]. Based on KS-test and AIC values' combined results, we can infer that the Gamma distribution is well suited to represent recognition reaction time generated from PEBL Go/No-Go test. The Gamma distribution is bounded on the left by zero. This property is important for analysing the PEBL Go/No-Go test reaction times, as negative reaction time is not a valid scenario. Hence, a distribution that excludes negative values is readily applicable [170, 282]. The Gamma distribution offers a tremendous amount of flexibility in the shape of the distribution function [170]. Moreover, this flexibility allows the Gamma distribution to fit any reaction time data with reasonable accuracy, while other distributions may fit only specific reaction time values. These reasons make Gamma distribution useful allowing for a holistic analysis of the PEBL Go/No-Go test, reaction time data.

Analysing the distribution of reaction time data helps discover hidden patterns that are not visible using traditional approaches, e.g., the statistical methods such as ANOVA, applying transformation, Ex-Gaussian methods or non-parametric testing methods. Based on the research work presented in this study, we see that the recognition reaction time data generated from the PEBL Go/No-Go test is compatible with Gamma distribution. Modeling the data with Gamma distribution enables data to be analysed using interesting distribution properties such as *skewness* and *kurtosis*. When the mean of the Gamma random variable increases, its variance also increases. This property of a Gamma distribution provides more confidence and power to use the distribution, to represent probability distributions of reaction time data. Future work can extend to include Gamma distribution of data to analyse any participant's recognition reaction time changes before and after a digital health intervention. Methods such as generalised likelihood ratio [285] and Likelihood Ratio Tests (LRT) [286], can also be investigated to compare two samples that follow Gamma distribution where the sample sizes are different and where the samples' Gamma distribution parameters are unequal. Building upon the current work, we propose to investigate methods such as parametric bootstrapping with Gamma prior, to transform the data to a Gaussian form with bootstrapping, yet retaining the Gamma *shape* and *scale* estimates to determine the change in mean or variance between the two distributions.

3.4 Analysing the social data

In this section, we will focus on the social discipline of the Biopsychosocial model. Social data was collected during the DHI program via

1. EMA (Ecological Momentary Assessment) surveys
2. Surveys using online questionnaires

EMA surveys collect real-time data based on repeated measures and observations that take place in participant's daily environment [287]. It involves the measurement of behaviours and experiences in naturalistic settings [288]. In this form of a survey, users are repeatedly prompted to report on their experience at fixed times per day [289]. The EMA data collected is less susceptible to recall bias because of a lower reliance on the memory of the participants [290] and this data can provide insights into time-varying dynamics of behaviour and its correlates; and provides more ecologically relevant data [291–293]. The affective mobile app data was a mixture of categorical and numeric data derived from questions from the affective app survey. This data was highly dimensional; this was because the number of records (around 5000) was in the same order as the number of features. There was a strong temporal aspect to the data as all the emotion data was captured 5 times a day for 7 weeks. The exploratory data analysis of the EMA data revealed some properties of this data collected as part of the social discipline of the 'Wellness dataset'. The findings of the exploratory data analysis are listed below.

1. Use of Likert values for responses on the slider scale: Slider scales are similar to Likert scales with much more response categories. Instead of selecting one of the response alternatives, respondents use a slider to position themselves on a certain question. Slider scales were used to capture EMA survey responses during the DHI program. Survey questions asked the participants to indicate their level of agreement e.g., from strongly agree to strongly disagree, using the slider scale. The participants scaled the responses to a 5 or a 10 point format. The measures generated from the slider scales similar to Likert data; were ordinal, discrete and had a limited range. Table 3.6 describes some sample EMA survey questions that used Likert values.

2. Scope for user errors on EMA survey data: The erroneous outliers for numeric type responses were minimal for the affective data. There were only a couple of fields (e.g., sleep duration and duration of time spent talking with people around), that required free text numerical input from the user. All other values in the EMA survey were choices built into the software. The questions requiring categorical responses

allowed participants to skip questions and progress to the next survey question. Non-responses to skipped questions were marked as missing values in the EMA survey data. A few non-numeric, text-based fields were used to allow participants to provide free text data that supported the participants' responses at different points in the EMA survey.

3. Missing values patterns in EMA survey data: The exploratory analysis of missing patterns in the EMA survey data revealed three common missing data patterns as described.

    a) The EMA survey software required that participants submit a value on an active survey session, so there were few missing values for the individual features. Although the participants were prompted by the app five times a day for the repeat survey question - 'What is going on right now'; in most cases, the participants chose to respond only once to this question. Not responding to one or more survey questions was the most common data missing pattern, at the record level.

    b) The exploratory data analysis of the EMA data revealed a gradual declining trend of responses to alerts over the 7 weeks during the DHI program. Most participants actively filled out the surveys in the first week but the responses were sparse over subsequent weeks due to the server being down, so no data got recorded.

    c) There were several missing values for subsequent survey questions in the EMA survey responses, where the participant had reported a negative primary emotion 'sad'. A software error caused this issue.

    d) The default value on the 10 point format slider, was set to 5. The EMA survey software would not advance the page if the user did not move a single slider, but many 5s were indicative of unanswered questions. Figure 3.11 describes an extract of EMA survey data for a sample participant with unanswered questions.

Survey question	Slider bar grid	Likert scale format
I am currently feeling..	0-extremely tired.... 10-extremely energetic	10 point
I try and figure out or try to solve the problem/issue	0-totally disagree.... 10-totally disagree	10 point
Get emotional support from others	0-totally disagree.... 10-totally disagree	10 point
Avoided things wanted/had to do	0-not at all.... 5-a lot	5 point

Table 3.6: Sample EMA survey questions assessed on a Likert scale.

Tired_Energetic	Hungry_Full	PainFree_Pain	Physically_Tense_Relaxed	Mentally_Foggy_Alert	Emotionally_Aroused_Flat	LossControl_InControl	Pessimistic_Optimistic	Insecure_Confident	Threatened_Safe	Worried_WorryFree	BadSelf_GoodSelf
2	5	5	5	5	5	5	5	5	5	5	5
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	6	5	5	5	5	5	5	5	5	5	5
2	6	5	5	5	5	5	5	5	5	5	5
4	4	6	4	4	6	5	5	5	5	5	5
3	4	5	4	4	7	4	5	5	5	5	5
3	3	5	5	3	7	3	5	5	5	5	5
7	7	5	5	5	5	5	5	5	5	5	5
3	7	4	4	3	5	5	5	5	5	5	5
3	3	5	4	3	8	5	5	5	5	5	5
3	3	4	7	5	5	5	6	6	7	7	6
3	8	5	5	5	5	5	5	5	5	5	5
2	7	6	5	5	5	5	5	5	5	5	5
4	5	6	5	5	8	5	5	4	5	5	5
6	9	5	5	5	5	5	5	5	5	5	5
4	8	5	6	4	6	5	5	5	5	5	5
7	4	5	6	7	5	5	5	5	7	6	6
4	3	5	5	5	5	5	5	5	5	5	5
4	3	5	5	5	5	5	5	5	5	5	5
4	8	5	5	5	5	5	5	5	5	5	5
4	3	6	5	4	6	4	5	5	6	5	5
5	4	5	5	5	5	5	5	5	5	5	5
6	2	5	6	5	5	5	5	5	5	5	5
5	3	7	4	6	5	6	5	5	5	5	5
4	3	5	4	4	5	5	5	5	5	5	5
4	6	5	5	6	5	6	6	5	5	5	5
4	6	6	5	5	5	5	5	5	5	5	5
6	7	5	5	5	5	5	5	5	5	5	5
3	5	5	5	5	5	5	5	5	5	5	5
3	6	5	5	5	5	5	5	5	5	5	5
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
6	6	5	5	5	5	5	5	5	5	5	7
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
4	8	7	4	5	7	5	5	5	5	5	5
4	7	5	5	7	5	5	5	5	5	5	8

Figure 3.11: Sample EMA survey data of participant shows slider value unmoved from default value 5 indicating unanswered questions in the EMA survey session.

The participants of the DHI program took up surveys using online questionnaires at three different points in time. During pre-intervention, the participants completed the first online questionnaire that asked participants questions about their mental health and wellbeing, how they deal with their emotions, how mindful and optimistic they are, and a few questions about the demographics and the personality of the participants. During week 4 and week 8, the participants submitted the online questionnaire responses that included the same set of questions about mental health and wellbeing. These online questionnaires enabled the participant to assess and record their levels of emotions, negative feelings, optimism and perceived 'psychological stress' level. Subsection 2.8.1, provides a brief description of these scales. More details for these scales are described here.

- Perceived Stress Scale (PSS): It is the most widely used psychological instrument for measuring the perception of stress [294]. It is a measure of the degree to which situations in one's life are appraised as stressful. The questions in this scale ask participants about their feelings and thoughts during the last month.

- Depression Anxiety and Stress Scale (DASS): The DASS is a set of three self-report scales designed to measure the negative emotional states of depression, anxiety and stress [229, 230]. Each of the three DASS scales contains 14 items, divided into sub-scales of 2-5 items with similar content. Scores for Depression, Anxiety and Stress are calculated by summing the scores for the relevant items.

- Kessler scale (K-6): This scale is a simple measure of psychological distress [235, 236]. The K-6 is a 6-item inventory rated on a 5 point Likert-type scale. It is a truncated version of the K-10 scale.

- Difficulty in Emotion Regulation Scale (DERS): The DERS is a brief, 36-item, self-report questionnaire designed to assess multiple aspects of emotion dysregulation [229, 230]. The measure yields a total score as well as scores on six scales (non-acceptance, goals, impulse, awareness, strategies and clarity) derived through factor analysis.

- Mental Health Continuum scale (MHC): The MHC scale measures emotional wellbeing [231, 232]. The scale contained 14 items related to three dimensions of wellbeing: emotional, social and psychological.

- Revised Life Orientation Test (LOT-R): This scale is a 10-item scale that measured how optimistic or pessimistic people feel about the future [233, 234]. Respondents

use a 5-point rating scale (0 = strongly disagree; 4 = strongly agree) to show how much they agree with 10 statements about positive and negative expectations.

In addition to the responses to the above test scales, the participants also recorded self-reported values for features such as *Social support*, *Religiousness*, *Physical health*, *Mental health* and *Quality of life*. The responses were recorded as Likert values.

3.5 Conclusions from conducted experiments

The concise conclusions and findings of the experiments conducted in Section 3 are summarised below.

1. The static analysis of the BDNF values pre and post interventions reveals that 83% of participants had increased BDNF readings at the end of the physical activity intervention compared to a 33% increase in BDNF for the mindfulness intervention. The physical activity intervention showed stronger changes in BDNF compared to mindfulness intervention at the end of the DHI program, as was hypothesised earlier. The findings also revealed that, within the physical activity sub-group, for 66.67% of participants, an increase in levels of BDNF showed a strong correlation with higher LOT-R (measures optimism) test scores recorded during week eight at the end of the DHI program.

2. The AUC_G method is preferred over the AUC_I method, to compute and apply a single biological score to a set of four cortisol level readings recorded on a single day that change over time.

3. The DHI program benefited the experimental group participants, particularly the physical activity sub-group. This is evident from values of 'peak heart-rate count' which lie within a smaller range compared the control group.

4. The Gamma distribution is well suited to represent recognition reaction time generated from PEBL Go/No-Go test using interesting distribution properties such as skewness and kurtosis.

5. Exploratory analysis of social data revealed properties of Likert values, scope for user errors and missing value patterns in EMA data.

3.6 Summary

The primary aim of the thesis work was to explore and discover relevant insights by conducting a retrospective study and analysis of the existing Wellness dataset. As a first step, to achieve this aim, this chapter has explored the salient features of the transdisciplinary dataset generated by the 'Wellness study' DHI program. The data from three disciplines of the Biopsychosocial model were independently explored and analysed using exploratory and statistical tools. The study of salient features included examining the distribution of different sub-sets of the wellness dataset. We used data aggregation tools to aggregate multiple measures of a single feature collected during the course intervention programs and experiments to study the correlation of features between the biological, psychological and social disciplines, impacted by the mindfulness and physical activity health interventions.

The biological data comprised the blood and cortisol data. The analysis of the biological BDNF data generated pre-intervention and post-intervention revealed high BDNF levels in participants at the end of physical activity intervention compared to the mindfulness intervention. Participants with high BDNF measures also showed a strong correlation with higher LOT-R test scores recorded for these participants during the same time frame. The cortisol data analysis used the established AUC_G method to aggregate the cortisol measures and compute a single cortisol score from the set of 4 readings. Analysis of the physiological data examined the various physiological features captured by the wearable device. In contrast to the cross-sectional biological data, the physiological data was continuous in nature. We analysed the heart-rate data in detail and a new metric, 'heart-rate peak count' was proposed along with the new MedIQR method to analyse the distribution of the heart-rate peaks, using range and IQR values. Analysis of the psychological data examined the data generated from the neurocognitive tests conducted at the start and the end of the DHI program. The three kinds of neurocognitive tests were the PEBL (Psychology Experiment Building Language) based Go/No-Go test, Stroop test and the trail test. The experimental study of psychological data revealed that recognition reaction time data generated from PEBL Go/No-Go test is compatible with Gamma distribution. Modeling the data with Gamma distribution enabled data to be analysed using interesting distribution properties such as skewness and kurtosis. We analysed the social data by examining its properties and their different sources such as EMA surveys and online questionnaires.
In the next chapter - Chapter 4, we examine the challenges in using the retrospective wellness dataset in order to prepare it for statistical analysis.

CHAPTER 4

PRE-PROCESSING CHALLENGES OF WELLNESS DATASET

In the previous chapter, , we studied the salient features of the wellness dataset, which comprised a. biological and physiological data, also known as biometric data, b. psychological data and c. social data. In this chapter, we focus on the challenges of using the real-time wellness dataset to prepare it for statistical analysis using statistical procedures or ML methods. The activities described in this chapter can be mapped to the Scrub and Explore phases of the OSEMN data mining process model. The chapter is organised as follows

- Section 4.1 describes the challenge of missing data and their impact when pre-processing data;
- Section 4.2 examines the missing pattern in physiological time-series data;
- Section 4.3 discusses the choices of imputation approach to impute missing time-series data; and
- Section 4.4 summarises the chapter.

The psychological and social data, the two key components of the wellness dataset, were cross-sectional in nature. These data were collected via biological and neurocognitive tests and self-reported online questionnaires during the course of the DHI program. This set of cross-sectional data comprising the biological and psychological data was collected for all participants at multiple instances (Week 1, Week 4 and Week 8) during the program. The

physiological data were collected in the form of long segments of continuous data and were sourced from wearable sensor devices worn by the participants.

It was necessary to pre-process the wellness dataset to make it fit for downstream analysis. The static analysis of the wellness dataset revealed the issue of missing observations. Missing data can occur at the unit level or the item level [295]. An example of unit-level non-response is when a participant refuses to undertake a test or declines to fill out a survey. In this scenario, no information is collected from the participant, which results in missing values for the entire test or survey. An example of item non-response is when a participant misses completing some portions of a test or skips some questions in a survey. In this scenario, partial information is collected from the participant, which results in missing values for some aspects of the test or survey. At times, technical or system errors also result in missing values, as all data from tests or surveys fail to be recorded. For the purpose of this research, we prioritised analysing missing data due to item non-response.

Missing values in quantitative research lead to biased estimates, loss of information, decreased statistical power and weakened generalisability of findings [295]. Missing data can have a substantial influence on the results produced by statistical analysis. The manner of influence varies depending on the how missing data are handled, but some bias always occurs whenever data are missing. All biases inherent in incomplete data can seldom be removed with any amount of statistical compensation. Paul D Allison [296] concludes that a good solution to the missing data problem is not to have any. On the same subject, he also acknowledges that perfectly complete datasets are highly probable when conducting survey research on human subjects. Dong et al. [297] discuss missing data treatments in their review of quantitative studies published in the *Journal of Educational Psychology* between 2009 and 2010. The findings in this study indicate that several research practices continued using data with missing values and did not explicitly acknowledge the presence of missing data nor detail the approach used to deal with them. As far as possible, the goal should be to minimise missing data; when the problem is inevitable, the dataset with missing data must be closely examined for the missing data mechanism, rate of missing data, missing patterns in data and data distribution before selecting a suitable method to insert the missing data.

The focus of this chapter is to study the challenges of using the real-time wellness dataset to prepare it for statistical analysis using statistical procedures or ML methods. We explicitly evaluate the impact of missing data problems and examine the conditions under which they occur. We also examine the suitability and challenges of several known techniques for handling missing values in the wellness dataset.

4.1 Data pre-processing and impact of missing data

The multidisciplinary wellness dataset used in this study had a large proportion of missing data at the item level. This was particularly prominent in the physiological data collected as long segments of continuous data sourced from wearable sensor devices. The amount of missing data was directly related to the quality of statistical interference [297]. So far, there is no established cut-off in the literature regarding an admissible percentage of missing data in a dataset for valid statistical inferences. Some studies have noted that a missing rate of 5% or less is inconsequential [298], whereas others have maintained that a missing rate of more than 10% can lead to biased statistical analysis [299].

Missing data rate is just one of the many indicators for assessing the missing data problem. Other indicators that have a greater impact on research results include missing data mechanisms and missing data patterns [300].

4.2 Missingness pattern in physiological data

The static analysis of physiological time-series data revealed a large proportion of missing data at the item level. The physiological data, also known as the sensor data, were collected from the BASIS watches worn by participants during the course of the DHI program. This sensor data collected were in real-life settings during the intervention. The physiological data comprised *heart-rate*, *accelerator magnitude*, *galvanic skin response*, *skin temperature* and *steps* and were classified by *activity type*. The presence of missing values in such real-life sensor data is inevitable, as they are collected in real-life settings. The quality controls that can be imposed on data collected in a controlled environment could not be imposed in the real-life setting during the DHI program. Factors such as participant behaviour (keeping the wearable device on during the intervention period), wearable device battery life, device malfunctioning or system glitches preventing the reporting of measures from the wearable device to the server system were the main causes of missing sensor data collected as part of the wellness dataset. Using this sensor data with missing values would

be unfit for research purposes and would severely affect the adoption of statistical or ML methods that could be applied to these data for modelling.

To mitigate this risk, we quantified the impact of the missing data in the time-series physiological data, proposing an imputation strategy that incorporates handling missing data for physiological features in the wellness dataset. We analysed the pattern of missing data for each of the physiological features, and this analysis guided the imputation strategy used for the data. Table 4.1 shows a few entries of records from the physiological data extracted from the biopsychosocial data. Typically, a dataset in analytic form is characterised as a rectangular matrix (with participants shown in rows and participant features shown in columns) and the missing data are the elements not observed in this matrix, which are marked by question marks in Table 4.1.

Table 4.1: Missing data matrix

Date	Participant id	Heart rate	Accelerator magnitude	Skin temp.	Galvanic skin response	Activity type	Steps
17-Jan-2015 13:58:28	rctw111	75	4003.0	85.693	.00040	light activity	.0
17-Jan-2015 13:58:29	rctw111	75	1570	85.693	.00040	light activity	.0
17-Jan-2015 13:58:30	rctw111	?	1060	85.693	.00040	light activity	.0
17-Jan-2015 13:58:31	rctw111	?	124	85.693	.00040	light activity	.0
17-Jan-2015 13:58:32	rctw111	?	?	85.693	.00040	light activity	.0
17-Jan-2015 13:58:33	rctw111	?	?	85.693	.00040	light activity	.0

Missing physiological time-series data, such as *heart-rate*, *accelerator magnitude*, *skin temperature* and *galvanic skin response*, are the consequence of a random process that can be characterised by missingness models. We list three broad types of missingness mechanisms [301], in order to examine the missingness pattern for our physiological time-series data.

1. Missing completely at random (MCAR): A variable is MCAR if the probability of missingness is independent of any subject characteristics. Each participant during the course of the DHI decided to remove or wear the wearable device that captured the physiological data at their own wish, which was completely random. No biometric data were recorded when the wearable device was removed.

2. Missing at random (MAR): A more general assumption, MAR, is that the probability a variable is missing depends only on observed variables. Missing values for the physiological features occurred because the wearable device was turned off by the participant for a certain period. Here the missing values of the *heart-rate* or *skin*

temperature features depended on the observed value of *off*, for the feature *activity level*.

3. Not missing at random: Missingness is no longer 'at random' if its probability depends on variables that are incomplete.

We conducted an experimental study on a sample subset of 10 participants and their physiological data to examine their missingness patterns and determine the missingness mechanism that can be applied to the time-series physiological data.

4.2.1 Experiment - missing data pattern

We used a subset of 10 participants belonging to either the mindfulness or physical activity intervention groups for this experiment. This represents 10% of the physiological time-series data that constitute the wellness dataset. During data collection, the physiological time-series data were captured over an equidistant time interval of one second via the sensor device. The features of *heart-rate*, *accelerator magnitude*, *galvanic skin response*, *skin temperature* and *steps* are numeric and the classifying feature *activity type* is categorical. The configuration of various values for the feature activity level was managed by the company BASIS.

We performed a static exploratory analysis of each of the physiological features against the four different values of the categorical feature *activity type* namely: *inactive*, *light activity*, *moderate activity*, *sleep*. We left out the value *off* for the feature *activity type*, as this refers to the condition in which the BASIS watch was switched off by the participant and therefore did not record any physiological features. The missingness pattern for this condition was invalid for our study. Table 4.2 describes the missing data pattern for the subset of 10 participants used in the experiment.

Table 4.2: Summary of missing data pattern (% of missing values)

Sample	Heart rate	Accelerator magnitude	Galvanic skin response	Skin temp.	Steps	Activity type
112	32.041	0.092	0	0		inactive
112	84.390	0.409	0	0	0.038	light activity
112	95.666	0.073	0	0.023	0	moderate activity
112	99.984	100	100	100	1.038	off
112	2.903	0	0	0	0	sleep
672	12.527	0.100	0	0.004	0.014	inactive
672	75.320	0.399	0	0.024	0.018	light activity
672	95.231	0.348	0.147	0.176	0	moderate activity
672	99.973	100	99.980	99.980	61.077	off
672	0.528	0.033	0.013	0.017	0	sleep
698	34.301	0.019	0	0	0	inactive
698	85.530	0.459	0	0	0	light activity
698	96.896	0.211	0	0	0	moderate activity
698	99.998	100	99.988	99.994	93.534	off
698	2.898	0	0	0	0	sleep
741	12.527	0.100	0	0.004	0.014	inactive
741	75.320	0.399	0	0.024	0.018	light activity
741	95.231	0.348	0.147	0.176	0	moderate activity
741	99.973	100	99.980	99.980	61.077	off
541	0.528	0.033	0.013	0.017	0	sleep
541	12.527	0.100	0	0.004	0.014	inactive
541	75.320	0.399	0	0.024	0.018	light activity
541	95.231	0.348	0.147	0.176	0	moderate activity
541	99.973	100	99.980	99.980	61.077	off
242	0.528	0.033	0.013	0.017	0	sleep
242	12.527	0.100	0	0.004	0.014	inactive
242	75.320	0.399	0	0.024	0.018	light activity
242	95.231	0.348	0.147	0.176	0	moderate activity
242	99.973	100	99.980	99.980	61.077	off
242	0.528	0.033	0.013	0.017	0	sleep

Continued on next page...

Table 4.2 – continued from previous page

Sample	Heart rate	Accelerator magnitude	Galvanic skin response	Skin temp.	Steps	Activity type
673	12.527	0.100	0	0.004	0.014	inactive
673	75.320	0.399	0	0.024	0.018	light activity
673	95.231	0.348	0.147	0.176	0	moderate activity
673	99.973	100	99.980	99.980	61.077	off
673	5.563	0.471	0.199	0.219	0	sleep
698	34.301	0.019	0	0	0	inactive
698	85.530	0.459	0	0.068	0.045	light activity
698	96.896	0.211	0	0	0	moderate activity
698	99.998	100	99.988	99.994	93.534	off
698	2.898	0	0	0	0	sleep
113	29.712	0.214	0	0.014	0.014	inactive
113	76.683	0.808	0	0	0.058	light activity
113	90.118	0.868	0	0.457	0	moderate activity
113	99.991	100	99.989	99.989	83.837	off
113	2.249	0.249	0.120	0.120	0	sleep
739	38.809	0.198	0	0.012	0.015	inactive
739	90.072	0.350	0	0.035	0	light activity
739	95.478	0.452	0.155	0.388	0	moderate activity
739	98.785	100	99.663	99.663	14.478	off
739	4.606	0.122	0.051	0.073	0	sleep

4.2.2 Findings - missing data pattern

The static analysis of the missing data pattern of the physiological feature values reveals the following findings.

1 An average of 2.6% of values is missing for the feature *hear-rate* when the *activity type* is *sleep*. Other physiological features recorded no missing data for this condition. This can be explained by the fact that in these instances, the participant had minimum

activity owing to rest and there were minimum interruptions to the wearable device, allowing it to remain in contact to capture the measurements (see Figure 4.1).

2 An average of 28% values is missing for the feature *heart-rate* when *activity level* is *inactive*. An average of 90% of values is missing for the feature *heart-rate* when the *activity level* is *moderate active* or *light active*. These could pertain to the technical issues with the wearable device that failed to save the feature values to the upstream server system during end-of-day reporting. (see Figure 4.2, 4.3, 4.4).

3 The percentage of missing data for other physiological features when *activity type* is *inactive* or *moderate activity* is minimal.

4 There is a strong correlation between the missing data pattern of the features *heart-rate* and *accelerator magnitude* when *activity level* is *sleep* or *off*. There is potential for correlation between observations with time-series sensor data such as this, where the feature values were collected at adjacent time periods.

5 The missing data pattern shows that the dataset is connected. Any observed data point can be reached by correlating other features through a sequence of horizontal or vertical moves. For example, *heart-rate* can be correlated to the features *activity level* or *accelerator magnitude*. This indicates that the physiological time-series data can be analysed as multivariate.

6 The missing data pattern for the sample set of 10 participants reveals that the timing of missing data for the different variables is informative for the imputation problem. This implies that the probability of a data value missing for a variable is the same only within groups defined by the observed data. For example, the missing values of the variables *heart-rate*, *accelerator magnitude*, *skin temperature* and *galvanic skin response* can, for a period, be explained by the value *off* against the correlated variable *activity level*. Therefore, the missingness pattern in this data is MAR.

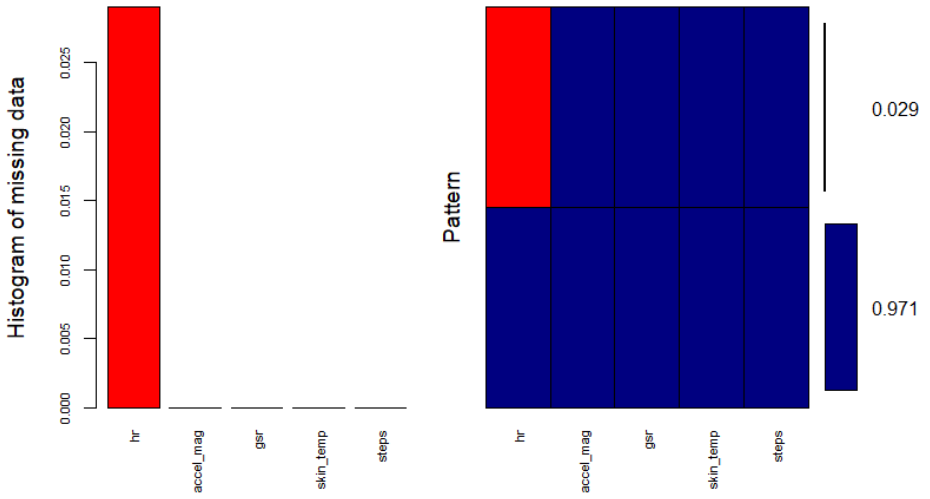


Figure 4.1: Missing data pattern for sample participant where activity level = sleep

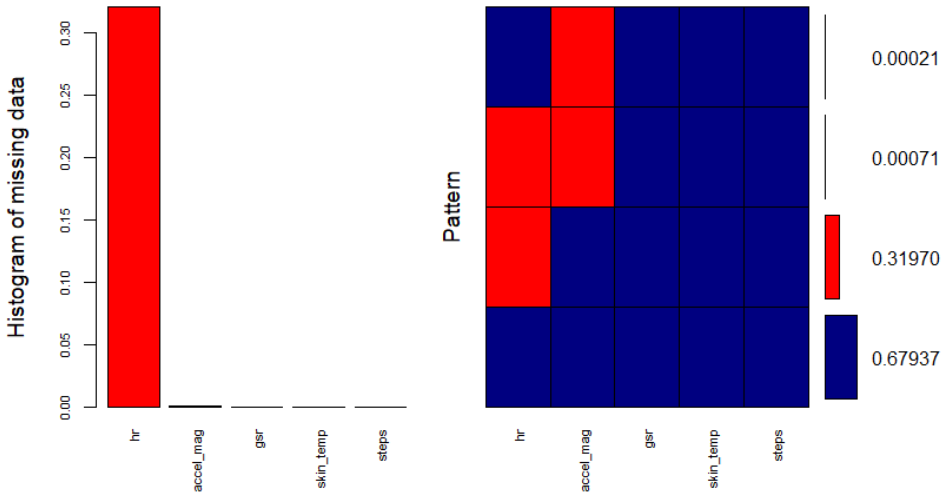


Figure 4.2: Missing data pattern for sample participant where activity level = inactive

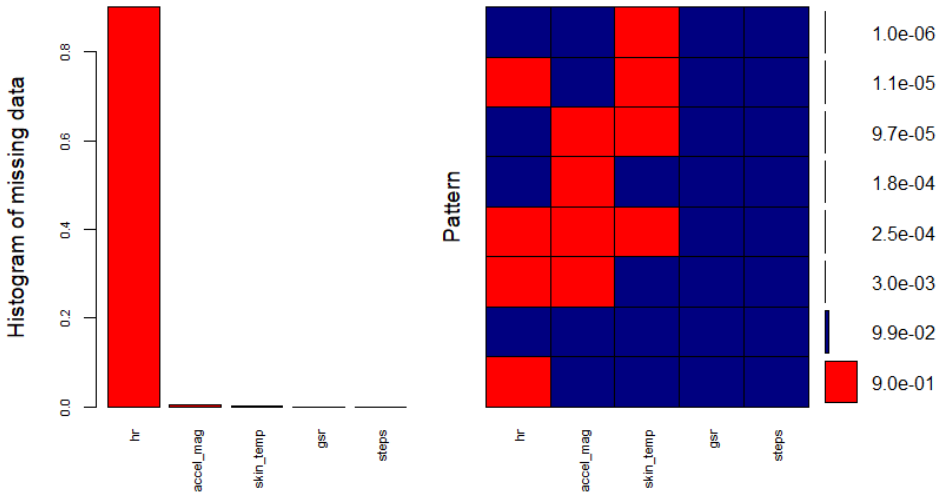


Figure 4.3: Missing data pattern for sample participant where activity level = light activity

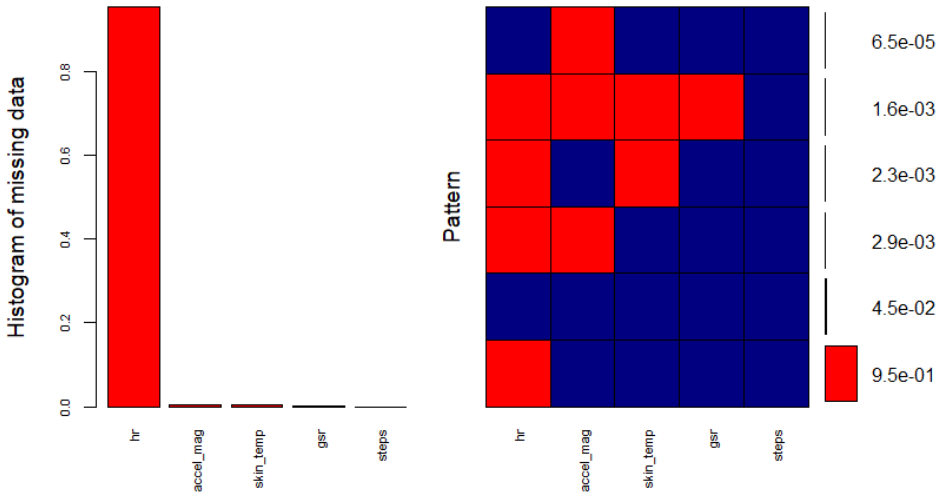


Figure 4.4: Missing data pattern for sample participant where activity level = moderate activity

4.3 Choice of imputation approach

In this section, we first discuss the imputation methodology of some well-known imputation approaches with respect to their application to the wellness dataset. These approaches include the classic multivariate method of multiple imputation (MI) and the univariate method of Kalman filtering. We examine the suitability and discuss the challenges of both these methods with respect to their application to time-series physiological data sourced from wearable devices.

From the static analysis of the wearable sensor data and the missing data patterns revealed earlier, we can infer that the data are connected and can be treated as multivariate. Further, the missingness pattern of the physiological data, is found to be MAR. Although we see a huge percentage of missing values for *heart-rate* when *activity level* is *inactive* and *moderate activity*, these values can be imputed by correlating the adjacent feature values for the same timestamps using the multivariate data imputation approach or by correlating the historical and future values of the feature on the longitudinal timeline using the univariate data imputation approach if the former approach is not found suitable. The following sections 4.3.1 and 4.3.2 examine the multivariate and univariate imputation algorithms for the different physiological features.

4.3.1 Multivariate data imputation approach

MI for missing data is an attractive method for handling missing data in multivariate analysis [301]. Using MI on the physiological data can provide unbiased estimates for the missing values. The static data analysis of the physiological data of the studied samples in the previous section supports the assumption that the missing data are MAR. Under MAR, the MI approach seeks to retain the advantages of maximum likelihood while also allowing the uncertainty caused by imputation, which is ignored in single imputation, to be incorporated into the completed data analysis. The MI method involves creating more than one set of replacements for the missing values based on plausible models for data, therefore generating multiple completed datasets for analysis. The statistical reasoning behind MI is that the observed data likelihood can be approximated by the average of the completed data likelihood over unknown missing values [302]. However, a multivariate imputation approach such as MI cannot provide an accurate imputation for a variable when missing values for other correlated variables used for imputation occur at exactly the same time intervals.

The physiological data extracted from the wellness dataset contain long periods of interruptions caused by sensor errors (e.g., wearable malfunction) and inconsistent data collection periods (e.g., wearing behaviour and varying compliance by participants) [303]. Such interruptions in physiological time-series data degrade the performance of inference that relies on the complete dataset. Long periods of interruption in the physiological time-series data present a major challenge in the application of multivariate imputation. To overcome this challenge, we used sequences of well-defined lengths of 20 minutes of physiological time-series data for this research.

4.3.1.1 SPSS MI imputation

We applied the SPSS MI function to 20-minute sequences of time-series physiological data to impute values for the feature *heart-rate* for a sample participant. Three imputations sets were generated. The MI imputation method used was Fully Conditional Specification with 10 iterations. The imputation sequence was set in the following order: *galvanic skin response* followed by *skin temperature*, *steps*, *accelerator magnitude* and *heart- rate*. The linear regression imputation model is used by the MI method. The MI method uses the linear regression imputation model. The SPSS MI imputation method imputed implausible missing heart-rate values that were not useful for statistical inferences (see Figure 4.5). The imputed values of *heart-rate* varied drastically over one-second intervals, which is not reasonable compared to previous and future original *heart-rate* values measured over time. This is because the MI method works by imputing the missing values multiple times and then consolidating across imputed datasets to account for variation within and between imputations, reflecting the fact that imputed values are not the known true values. Further, the MI method is based on assumptions relating to the distributional form of the variables and generally assumes a joint multivariate normal model for the continuous variables. Figure 4.5 shows the imputed values generated by the SPSS MI imputation function for the missing values of *heart-rate*. Therefore, the multivariate imputation method of SPSS MI is not found suitable for imputation of physiological data values. In the next section, we examine the univariate data imputation approach to correlate the historical and future values of the feature *heart- rate* on the longitudinal timeline.

1	17-Jan-2015 11:52:00.00	rctw112	62	103.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:01.00	rctw112	62	541.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:02.00	rctw112	62	107.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:03.00	rctw112	62	119.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:04.00	rctw112	62	132.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:05.00	rctw112	62	331.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:06.00	rctw112	62	1152.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:07.00	rctw112	62	1838.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:08.00	rctw112	62	3488.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:09.00	rctw112	62	3835.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:10.00	rctw112	62	812.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:11.00	rctw112	94	2725.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:12.00	rctw112	72	1727.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:13.00	rctw112	81	5236.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:14.00	rctw112	83	1793.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:15.00	rctw112	71	1676.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:16.00	rctw112	63	835.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:17.00	rctw112	79	1729.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:18.00	rctw112	61	504.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:19.00	rctw112	82	498.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:20.00	rctw112	76	256.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:21.00	rctw112	83	3631.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:22.00	rctw112	83	2283.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:23.00	rctw112	83	2161.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:24.00	rctw112	83	2407.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:25.00	rctw112	81	3263.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:26.00	rctw112	81	1640.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:27.00	rctw112	81	1292.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:28.00	rctw112	81	434.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:29.00	rctw112	81	120.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:30.00	rctw112	81	527.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:31.00	rctw112	81	2734.0	.0001441339	91.175	light activity
1	17-Jan-2015 11:52:32.00	rctw112	81	2947.0	.0001441339	91.175	light activity

Figure 4.5: Extract of implausible heart-rate values imputed by SPSS MI method

4.3.2 Univariate data imputation approach

Here, we examine use of the *na_kalman* function to impute missing values on structural time-series model of the physiological data is examined. The *na_kalman* is an imputation algorithm and is available in the imputeTS package in R. Kalman filtering provides estimates of unknown variables given the measurements observed over time. To use the *na_kalman* function, each feature of the physiological sensor dataset is treated as univariate time-series data and the function handles missing values in the sample observations.

4.3.2.1 R-Kalman filter

The static analysis of the physiological time-series data shows a very high percentage of missing values for the feature *heart- rate* under the conditions *light activity* [84%] and *moderate activity* [95%] for the feature *activity type*. Kalman filtering is not considered suitable as an imputation method for feature *heart-rate* because of the large period of missing values. The univariate imputation method of linear trend is suitable for imputing missing values for the feature *heart-rate*. Kalman filtering is suitable for the other physiological features that contain smaller periods of missing values. To examine the suitability of the *na_kalman* function for the remaining physiological variables, we used the time-series structural model fitted by the ML function and passed it to the *na_kalman*

function to impute missing values for the physiological features: *accelerator magnitude*, *galvanic skin response*, *skin temperature* and *steps*. The imputed values for *accelerator magnitude*, *galvanic skin response*, *skin temperature* and *steps* are visualised in Figures 4.6, 4.7, 4.8 and 4.9. Visual inspection of the figures indicates that the imputed values (shown as red dots) fit well in the univariate time-series data for the different physiological features. Therefore, the *na_kalman* function is found suitable for imputing values for the physiological features *accelerator magnitude*, *galvanic skin response*, *skin temperature* and *steps*.

4.3.2.2 Combining univariate imputation methods

The *na_kalman* function, although suitable for most physiological features, is not appropriate for use on the feature *heart-rate*, which has long periods of missing values. We examined other univariate imputation methods for this feature and found that linear trend and linear interpolation are most suitable under this condition:

- The linear interpolation imputation method uses the last complete observation value before the missing data and the first complete observation value after the missing data to impute the missing values. If the first and last observations are missing in the set, imputation fails [304]. For each 20-minute sequence of time-series physiological data, we used linear interpolation to impute missing values for the feature heart-rate where the missing values occurred in between the sequence. Where the missing values occurred at the start or end of the 20-minute sequence, we used the linear trend at point imputation method.

- The linear trend at point imputation method consistently determines the missing values in accordance with the trend of the current structure (e.g., a trend could be that the values tend to increase from the first subject to the last). The missing data are placed into the values decided in an index variable where the sets are scaled from 1 to n [304]. We used the linear trend at point imputation method to impute missing values occurring at the start and end of the 20-minute sequence.

The combination of the univariate imputation methods *na_kalman* function and linear interpolation and linear trend at point were used as the imputation approach to handle missing values for the feature *heart-rate*.

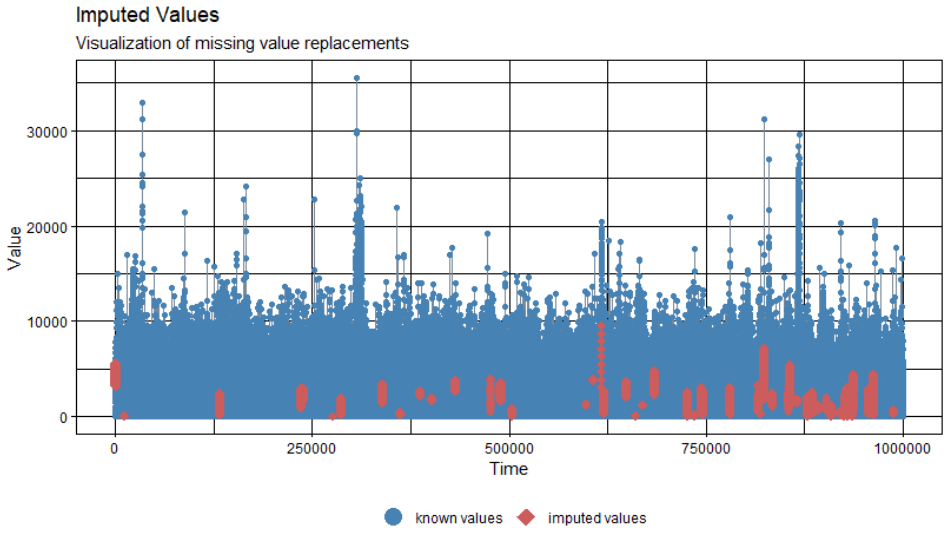


Figure 4.6: Imputation - accelerator magnitude

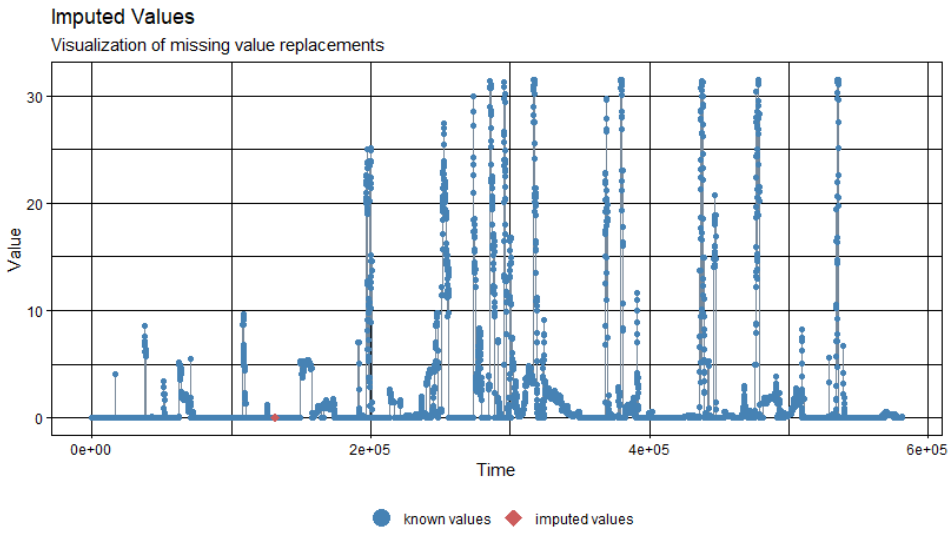


Figure 4.7: Imputation - galvanic skin response

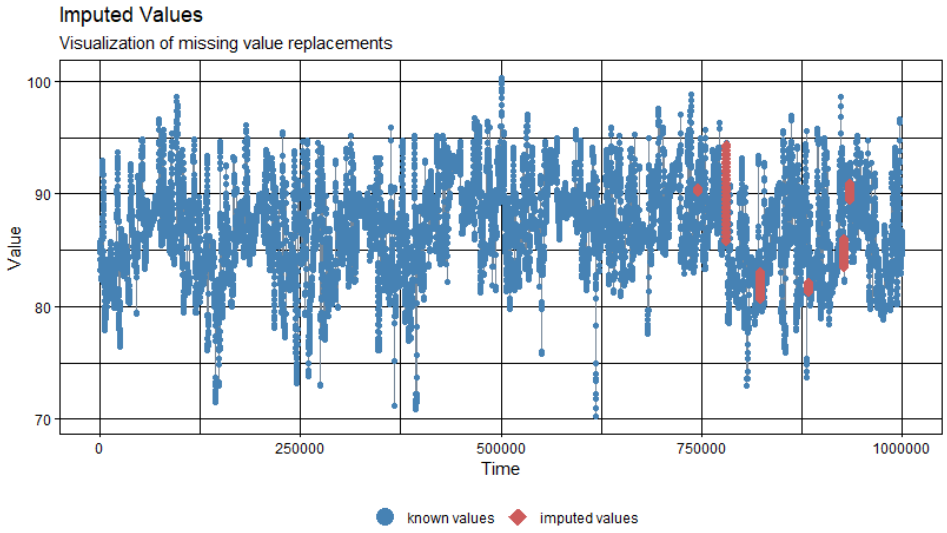


Figure 4.8: Imputation - skin temperature

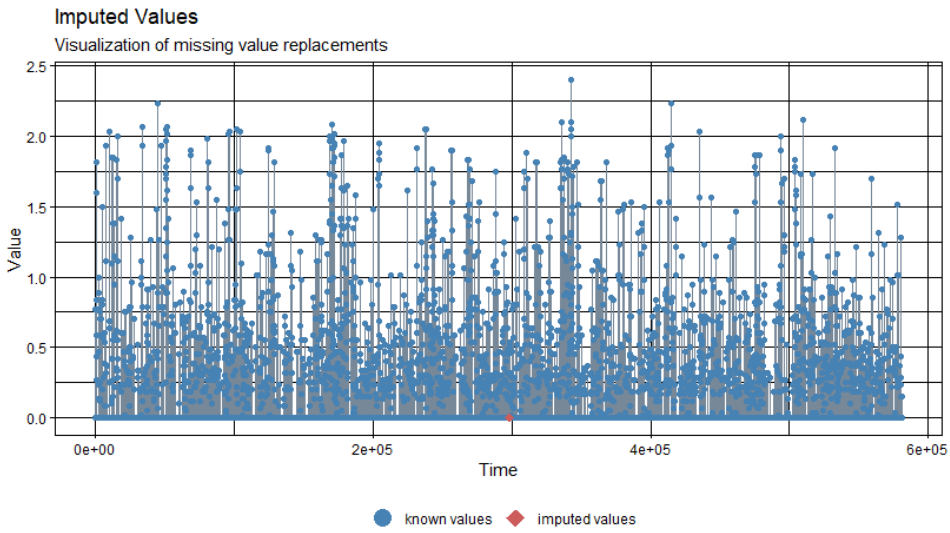


Figure 4.9: Imputation - steps

4.4 Summary

This chapter has highlighted the challenges of data pre-processing. We focused on the problem of missing data, particularly within time-series data, during data pre-processing and their impact on research analysis. We examined the suitability and challenges of multivariate and univariate imputation methods, considering the specific case of the physiological time-series data derived from the wearable device as part of the DHI program. We observed that the multivariate imputation method SPSS MI does not provide accurate imputation for missing features if the correlated variables available for imputation are also missing at exactly the same time intervals. As a result of this limitation, the generated set after imputation is incomplete with missing values. Most real-world time-series data, such as the assessed physiological data obtained from wearable devices, are affected by large intervals of missing values. The findings from the studies conducted reveal that the MI imputation method failed to impute the plausible values of missing *heart-rate* on short, well-defined lengths of uninterrupted data, as the original MI approach assumes a joint multivariate normal for continuous variables. Therefore, a multivariate imputation method such as SPSS MI was found unsuitable for adoption as an imputation method for the physiological time-series data that constituted the complex wellness dataset. Conversely, the univariate method using the *na_kalman* function was found suitable, as it imputes plausible values in conditions with smaller periods of missing values for the continuous variable. This method of imputation is not suitable for data with large quantities of missing values, as seen in this study for the feature *heart-rate*. In this scenario, the *na_kalman* function did not provide a complete imputed set of values, as Kalman filtering relies on neighbouring values observed over time. To overcome this limitation, the univariate methods of linear trend at point and linear interpolation were used to impute the missing values in conditions with longer periods of missing data for the feature heart-rate occurring at the start, end or middle of the 20-minute time-series sequence.

In the next chapter, we exploit the insights developed in this chapter to explore the statistical and ML methods for examining possible contributory connections between the different biopsychosocial features of participants in the DHI program. We use the physiological time-series data that was made ready by imputing missing values using the data imputation approaches discussed in this chapter. We explore building prediction models using nonlinear classifiers on the biopsychosocial data to compare and evaluate the interventions. We also explore building a prediction model using ML and ANN for the physiological time-series data.

CHAPTER 5

A MACHINE LEARNING APPROACH FOR BIOPSYCHOSOCIAL WELLBEING

In the previous chapters, we have studied the salient features of the wellness dataset (produced from the 'Wellness study'), examined ways to pre-process the raw data, and adopted suitable methods to reduce dimensionality prior to data analysis. We further analysed the missing data problems and the conditions under which they occurred, examined the principled methods suitable to handle the missing values in the wellness dataset and assessed the challenges in doing so. As mentioned earlier, the orchestration of this research closely follows the OSEMN process [75, 158] (see Section 2.4). After obtaining, exploring and scrubbing the unique wellness dataset, as a next step, we will model the data for our particular biopsychosocial problem and further evaluate the model using performance metrics. Doing so will enable us to interpret the results from the model and data. In this chapter we present the results of applying ML techniques to perform predictive analytics on the 'Wellness dataset'.

Predictive analytics, a part of data analytics uses data mining and probability to predict results. In predictive analytics, each model is built using a number of predictors and these predictors strongly determine the future decisions made by the predictive model. As part of the data mining process, we pre-processed the 'Wellness dataset' and explored it using statistical tools (see Chapters 3 and 4). In this chapter, we will explore statistical and ML methods on the processed data to examine possible connections between the different

biopsychosocial features for digital and mental health interventions (see Section 2.9), identify trends in data and design models that can estimate future conclusions. We will use supervised learning on the real life 'Wellness dataset', build a prediction model to predict features of interest such as levels of positive or negative affective states as reported by the different scores such as DASS and DERS and predict the primary emotion for a participant based on the participants' wearable sensor data. Finally, we will evaluate the performance of the model using prediction metrics such as training and validation accuracy. The rest of this chapter is organised as follows:

- Section 5.1 investigates the relevance of ML for determining the suitability of a participant into a DHI assignment based on existing biopsychosocial information. We use different non-linear classifiers and compare their prediction accuracy to evaluate their suitability to determine if biopsychosocial features can be used to predict changes DASS scores;
- Section 5.2 investigates the suitability of RNN-LSTM to be applied for a multi-variate time-series classification problem. In Section 5.2.4 we build an RNN-LSTM model to predict DERS improvement affected by one of the two digital health interventions (mindfulness or physical activity). The developed RNN-LSTM model is further modified to predict primary emotion based on physiological features of the wellness dataset (see Section 5.2.5); and
- Section 5.3 summarises the chapter.

5.1 Prediction model using non-linear classifiers

Health care providers use DHI programs to promote engagement within the community. Effective assignment of participants into DHI programs helps increase benefits from the most suitable intervention, such as mindfulness or physical activity. A significant challenge with the roll-out and implementation of DHI is assigning participants into the intervention that is most likely to be successful, based on initially available data. Further, there is limited research evaluating or analysing DHIs using machine learning tools, following the biopsychosocial model occurring in real-life setting. The use of the biopsychosocial model [105] for this purpose is not wide-spread, due to limited personalised interventions formed on evidence-based data-driven models. ML has changed the way data interpretation works by involving methods that have replaced the traditional statistical techniques. In this section, we investigate the suitability of ML for this purpose, study

different non-linear classifiers and compare their prediction accuracy to evaluate their suitability. Further, as a novel contribution, we use real-life biopsychosocial features as input in this study. The results help in developing an appropriate predictive classification model to assign participants to the most suitable DHI. We exploit the various input features of the wellness dataset to evaluate different categories of popular classifiers and compare their performance. This study paves the way for using historical biopsychosocial features to a-priori predict a specific intervention type's suitability for an individual participant.

5.1.1 Summary of study protocol

In this section, we describe the summary of the study protocol. The details of the 'Wellness study' DHI program and its framework are described in Section 2.9. As mentioned earlier (see Section 2.7), the wellness dataset was generated in a research environment, by the School of Science, Psychology and Sport at Federation University, Australia in 2015. Local Australian volunteers, aged between 19 to 59 participated in this experiment. The program's goal was to investigate which of the intervention groups' participants benefited from the interventions conducted. In this study, we analyse the wellness dataset and identify which biopsychosocial features are suitable predictors of *mental ill-health* described by the Depression Anxiety and Stress Scale (DASS) score. The participants were randomly assigned to one of the two interventions in the experimental or control wait-list groups. The participants in mindfulness or physical activity interventions undertook activities relevant to that group, during the experiment period that lasted for 12 weeks. The participants in both the intervention programs were examined for biological, psychological and social conditions as part of various tests. At the start of the intervention, the participants were given biological tests, neuro-cognitive tests, and survey questions that enable them to assess their levels of emotions, negative feelings, optimism, and perceived 'psychological stress' level. Scales, namely, Perceived Stress Scale(PSS), DASS (Depression Anxiety and Stress Scale), Kessler scale (K6), Difficulty in Emotion Regulation Scale (DERS), Mental Health Continuum scale (MHC) and Life Orientation Test (LOT), encapsulate the test level values. At the end of the intervention, all the above mentioned tests were repeated. The test level values were captured again. The biopsychosocial features used in this study are divided into 5 main categories as described in subsection 2.8.1. The levels and scores of all the biopsychosocial features are recorded. The DASS level measures negative affect states, which include stress, depressive and anxiety symptoms (*mental ill-health*). The DASS level is recorded at the start and end of the intervention period. In addition to the biopsychosocial features, an additional feature 'DASS label' (see Figure 5.3)

is computed for each participant based on the change in DASS level at pre-intervention and post-intervention. DASS label *Improvement* is used if the DASS level at the end of the intervention is less than the DASS level recorded at the start of the intervention. DASS label *No Improvement* is used if the DASS level at the end of intervention remains the same or is greater than the DASS level recorded at the start of the intervention.

5.1.2 The method

We propose to apply a score to each biopsychosocial feature during analysis to enable:

1. applying a single biological score to a set of four cortisol level readings recorded on a single day that changed over time. Since the cortisol readings were taken at four different times during the day, we put together the four readings' values to compute one single cortisol reading that would represent the biological score

2. regularizing the scores for the different biopsychosocial features to make them consistent; to be used within classifiers as input features

3. computing correlation coefficient on a 2-dimensional scale. We do this to determine the extent of correlation between the scores from the three disciplines of the wellness dataset generated from the 'Wellness study'.

The output DASS score and the input biopsychosocial features are applied for comparing the ten classifiers mentioned below. Following are the salient factors related to the proposed approach for DHI data analysis.

5.1.2.1 Cortisol readings

Cortisol is a steroid hormone that regulates a wide range of processes throughout the body, including metabolism and immune response, and helps the body respond to stress. The drool method (a method used to collect oral fluid for biological testing) is used for data collection of cortisol levels. As we aim to develop a predictive model using data available prior to the intervention, only the week 1 cortisol readings are used in this study. The 4 readings denoted by S1 (20 mins after wake), S2 (60 mins after wake), S3 (midday) and S4 (evening) are recorded. We propose to use the AUC_G [251] method to compute the single cortisol score from the set of 4 readings.

5.1.2.2 Feature scaling

The wellness dataset contains 18 features. We will normalise each feature value (excluding gender and age) by linearly re-scaling the data values using the observed minimum and maximum values for that feature across the entire dataset, into a new arbitrary range of 0 to 1. The formula for scaling is given by:

$$\frac{(featVal - min(featVal_1 : featVal_n))}{(max(featVal_1 : featVal_n) - min(featVal_1 : featVal_n))}$$

where *featVal* refers to the feature value, *min* is the minimum observed value of the feature and *max* is the maximum observed value of the feature.

5.1.2.3 Trail test reaction time

The trail test reaction time data was generated from PEBL [244] tests. IIn our earlier research [305], we had established that a Gamma distribution best represents neurocognitive reaction time data. In this chapter, we use the Gamma Cumulative Distribution Function (CDF) to analyse trail test reaction time data.

5.1.2.4 Correlation between features

The wellness dataset used for this study is rich in variety, and with high dimensionality (consisting of 18 input features). However, the number of samples used for the study is limited due to missing data. Given the nature of the dataset, imputation of missing data is very challenging, and it can result in unreliable imputed values. We analyse the correlation between the three biopsychosocial disciplines using the Pearson correlation coefficient. The Pearson coefficient, ρ, is computed as follows:

$$\rho = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y}$$

For sampled data, the sample Pearson's correlation coefficient r is given by:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 (y_i - \overline{y})^2}}$$

To explore the relationship between biological, psychological and social measures, the

set of 18 feature scores (see Table 5.1) are reduced to 3 feature scores; one for each discipline (i.e., biology, psychology and social). A single mean score for each discipline of the 'Wellness dataset' is computed using the features scores' simple average. The mean scores for each biopsychosocial dimensions are used to explore the relationship between them and the DASS result at the end of the mindfulness intervention.

Table 5.1: Summary of 18 input feature data used

Demographic features	**Biological features**	**Psychological features**	**Social features**
Age	Cortisol level	Go / No Go test mean accuracy	DERS
Gender		Trail test total time	MAAS
		Stroop test total errors	DAAS
			MHC
			LOT-R
			K-6
			PSS
			Social support
			Religiousness
			Physical health
			Mental health
			Quality of life

5.1.2.5 Classification

We evaluate ten popular classifiers to ascertain which set or subset of biopsychosocial features affect the participants' *mental ill-health* as the DASS score; at the end of the DHI program. These classifiers are Support Vector Machines (SVM), decision trees, random forests, logistic regression, Naïve Bayes, Stochastic Gradient Descent, Linear Discriminant Analysis, Extra Tree classifier, Bagging classifier and Neural net. We choose these classifiers because they are robust to small datasets. The following are key considerations for configuring the classifiers:

1. *Classifier parameter optimisation*: We use the grid search function to scan the dataset to configure optimal classifier parameters. We use the Python GridSearchCV function for classifiers: SVM, decision trees, random forests and neural net.

2. *Number of hidden layers and nodes in neural network*: We propose to use one hidden layer to keep the classifier model simple while still allowing the data points to be separated non-linearly. We use the same number of nodes for the neural net classifier as the number of input features. This decision is made based on test runs carried

out on the neural network classifier with a different number of nodes on the hidden layer. In the test runs, the highest prediction accuracy is seen when the number of nodes is the same as the input features.

5.1.3 Experimental studies

In this section, we first carry out a scatter plot analysis to examine if a correlation exists between the different biopsychosocial features. Subsequently, we apply the ten popular classifiers to the wellness dataset,okp for their comparative evaluation.

5.1.3.1 Dataset

The following table 5.2 describes the participant dataset used for this study. The dataset of 18 participants (11 from the mindfulness program and 7 from the physical activity program) is expanded to 90 samples, using data augmentation by introducing noise around each feature data point.

Table 5.2: Data summary

	Mindfulness program	**Physical activity program**
Number of participants	11	7
Number of features	18	18
Binary classification group 1	Improvement in DASS	Improvement in DASS
Binary classification group 2	No Improvement in DASS	No Improvement in DASS

5.1.3.2 Scatter plot analysis

The correlation between the mean scores in each of the biopsychosocial dimensions is described using scatter plots. The Pearson correlation coefficient value r is computed. The correlation scores for the different dimensions of the wellness dataset,okp[], are summarised in Table 5.3. The input features occurring in each dimension are listed in Table 5.1.

The following conclusions are drawn from the scatter plots:

1. For participants in the mindfulness intervention, who showed an improvement in DASS score, there was a moderate positive correlation between mean biological and mean psychological scores (see Figure 5.1). The correlation coefficient is given

Table 5.3: Correlation scores (Figures in **bold** indicate significant values)

DASS label	**Dimensions compared**	**Pearson correlation r**
Improvement	bio and psycho	**0.55783**
No improvement	bio and psycho	0.3880
Improvement	psycho and social	0.4079
No improvement	psycho and social	0.4580
Improvement	bio and social	0.4079
No improvement	bio and social	0.4580

by r=0.55783 This indicates that both biological and psychological features are good predictors to describe Improvement in DASS score in the case of mindfulness intervention.

2. No linear correlation exists between mean biological, psychological and social scores within the wellness dataset, to indicate any DASS score *Improvement* or *No Improvement* (see Figure 5.2).

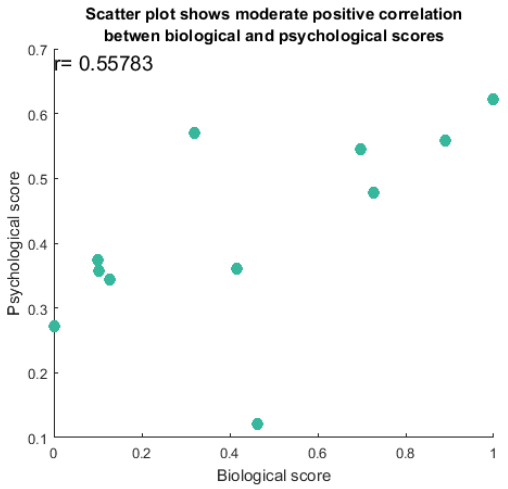


Figure 5.1: The figure describes the correlation between biological and psychological features seen amongst participants in the mindfulness intervention to indicate *Improvement* in the DASS score (moderate positive correlation between mean biological and mean psychological scores)

Although scatter plots are a good tool to visualise linear relationships in a static manner, they cannot reveal the non-linear relationships between features in multiple dimensions. In

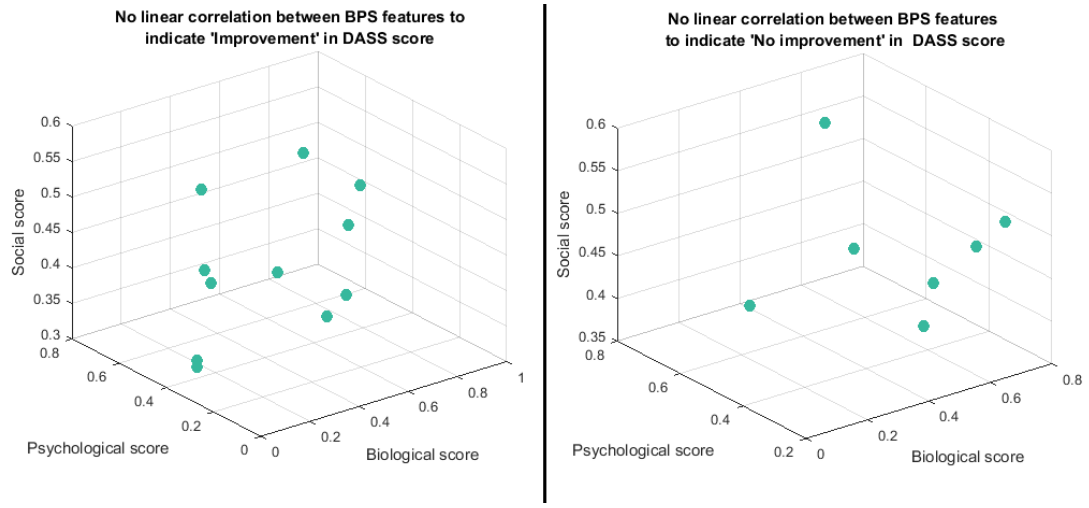


Figure 5.2: The figure indicates no linear correlation between biological, psychological and social features seen amongst participants who showed DASS Improvement, or No Improvement. The values on axes are from 0 to 1 (normalised and re-scaled).

our pursuit to study the influence of the biopsychosocial factors to assign participants into effective DHIs in future, we compare different non-linear classifiers and ascertain which set or subset of these biopsychosocial features affect *mental ill-health* of the participant, measured by the DASS score. The prediction accuracy of these classifiers is compared based on the predicted output DASS score and the input biopsychosocial features.

5.1.3.3 Comparison of classifiers

We compare 10 non-linear classifiers to determine the underlying relationship between the biopsychosocial features. The application of ML methods on the real-life biopsychosocial data (referred to in this thesis as wellness dataset) has not been done before and is unique to this study. The wellness dataset in this study was gathered using a structured DHI program and this adds credibility to this unique data. The present study uses the wellness dataset of 18 participants assigned to the mindfulness program and the target label used is: *Improvement* or *No Improvement* in DASS score. The DASS score describes *wellbeing stress* broadly in the dimensions of depression, anxiety and stress. The DASS score is chosen as the target label as it is more appropriate than other biopsychosocial features available for this study. Figure 5.3 is a snapshot of the table that shows the values of different DASS dimensions that are used to compute the binary target label: *Improvement* or *No Improvement* in DASS score.

Id	Stress_1	Anxiety_1	Depression_1	DASS_Score_Week_1	Stress_8	Anxiety_8	Depression_8	DASS_Score_Week_8	DASS LABEL
nnn1	8	0	6	14	1	0	1	9	IMPROVEMENT
nnn2	13	6	6	25	0	0	0	0	IMPROVEMENT
nnn3	8	3	1	12	12	7	3	22	NO IMPROVEMENT
nnn4	18	4	13	35	20	6	12	38	NO IMPROVEMENT
nnn5	17	4	12	33	17	3	9	29	IMPROVEMENT
nnn6	1	1	1	3	2	2	0	4	NO IMPROVEMENT
nnn7	2	1	0	3	0	0	0	0	IMPROVEMENT
nnn8	9	3	2	14	7	0	0	7	IMPROVEMENT
nnn9	1	1	1	3	5	0	4	9	NO IMPROVEMENT

Figure 5.3: DASS score used as output label for the classifier.

We expand the dataset of 18 participants to 90, using data augmentation by introducing noise around each feature data point. Each data point is surrounded by 4 noise data points by adding and subtracting 10 % and 20% to the original data value. We train the classifiers on the training set of 90 records (known observations for the biopsychosocial feature set of the mindfulness group); using the Leave One Out (LOO) method. In using the LOO method, along with the left out point, we also leave out the augmented data for the left out point. Next, we test the classifiers on the unseen set of 16 records (biopsychosocial feature set of the physical activity group), to determine the classifier's prediction accuracy. The classifiers use different combinations of input features from the wellness dataset to enable a comparative study of their prediction accuracy. A comparative study of the classifier prediction accuracy is provided in the following tables. Classifier prediction accuracy is described as the ratio of the number of correct predictions to the total number of input samples [306]. The classifiers classify the given set of samples into two groups: Binary classification group 1 that describes the improvement in DASS and Binary classification group 2 that describe no DASS improvement. Interesting influences of the biopsychosocial features on the classifiers are described below.

Table 5.4: Classifiers using 18 biopsychosocial features as input and their prediction accuracy. (Figures in **bold** indicate significant values)

Classifier	**Feature count**	**Prediction accuracy Physical activity group**	**Prediction accuracy Mindfulness group**
Neural net	18	0.4375	0.38889
Bagging classifier	18	0.4375	**0.55556**
Extra test classifier	18	**0.625**	**0.55556**
Linear discriminant analysis	18	**0.625**	0.44444
Stochastic gradient descent	18	**0.5625**	0.5
Naïve Bayes	18	0.4375	0.38889
Logistic regression	18	**0.5625**	**0.55556**
Random forest	18	0.375	**0.5**
Decision trees	18	**0.5625**	**0.61111**
SVM	18	**0.5625**	**0.61111**

Table 5.5: Classifiers using 3 biological features (includes demographic features) as input and their prediction accuracy. (Figures in **bold** indicate significant values)

Classifier	**Feature count**	**Prediction accuracy Physical activity group**
Neural net	3	**0.5625**
Bagging classifier	3	**0.5625**
Extra test classifier	3	0.4375
Linear discriminant analysis	3	0.5
Stochastic gradient descent	3	0.3125
Naïve Bayes	3	0.4375
Logistic regression	3	0.5
Random forest	3	0.5
Decision trees	3	0.4375
SVM	3	0.4375

Table 5.6: Classifiers using 5 psychological features (includes demographic features) as input and their prediction accuracy. (Figures in **bold** indicate significant values)

Classifier	**Feature count**	**Prediction accuracy Physical activity group**
Neural net	5	0.5
Bagging classifier	5	0.5
Extra test classifier	5	**0.5625**
Linear discriminant analysis	5	0.5
Stochastic gradient descent	5	**0.5625**
Naïve Bayes	5	0.5
Logistic regression	5	0.5
Random forest	5	**0.5625**
Decision trees	5	**0.5625**
SVM	5	**0.5625**

Table 5.7: Classifiers using 14 social features (includes demographic features) as input and their prediction accuracy. (Figures in **bold** indicate significant values)

Classifier	**Feature count**	**Prediction accuracy Physical activity group**
Neural net	14	**0.5625**
Bagging classifier	14	**0.5625**
Extra test classifier	14	0.5
Linear discriminant analysis	14	0.4375
Stochastic gradient descent	14	0.375
Naïve Bayes	14	0.375
Logistic regression	14	0.4375
Random forest	14	0.5
Decision trees	14	0.5
SVM	14	**0.5625**

Table 5.8: Classifiers using combinations of biopsychosocial features (includes demographic features) as input and their prediction accuracy on the unseen test set (physical activity group). (Figures in **bold** indicate significant values)

Classifier	**Bio and Psycho**	**Psycho and Social**	**Social and Bio**
Neural net	**0.5625**	**0.5625**	**0.5625**
Bagging classifier	**0.5625**	0.375	0.3125
Extra test classifier	0.4375	**0.5625**	0.4375
Linear discriminant analysis	**0.5625**	0.5	0.4375
Stochastic gradient descent	0.5	0.5	0.4375
Naïve Bayes	0.375	0.4375	0.4375
Logistic regression	0.5	0.3125	0.375
Random forest	0.375	0.43755	0.4375
Decision trees	0.375	0.4375	0.4375
SVM	**0.5625**	**0.5625**	**0.5625**

1. Table 5.4 shows that the majority of classifiers perform better than a random prediction accuracy of 50%; when all the biopsychosocial features are used for training the classifier. Extra tree classifier and linear discriminant analysis show high prediction accuracy for mindfulness intervention (at 62.5%). Decision trees and SVM show high prediction accuracy for the physical activity intervention (at 61.1%).

2. Table 5.5 and Table 5.7 show that fewer classifiers perform reasonably when only the biological features or only social features are used for training the classifier.

3. Table 5.6 shows that more number of classifiers (5 out of 10) perform fairly when only psychological features are used for training. However, the prediction accuracy of these classifiers is lower (at 56.25%), compared to the case where all the biopsychosocial features are used to train the classifier.

4. Table 5.8 shows that other combinations of biopsychosocial features used for training the classifiers, do not result in higher prediction accuracy than what is achieved by using all the biopsychosocial features for training the classifier.

5.1.4 Analysis and discussion

The addition of data points using noise enabled us to effectively apply the classifiers, reducing over-fitting and avoiding poor performance during testing. The additional data points make the input space smoother and easier for training. Although the biological data such as cortisol level was clinically obtained and was of good quality, the feature values of social data, in the form of self-reported scores is subjective and only as good as the interest level of the participant. The prediction accuracy of the classification model affected by the sparsity of the wellness dataset and the quality of the feature values in the dataset, can be further improved by improvements in field data collection techniques. The wellness dataset used for this study, being from an earlier conducted DHI program, was collected with a different objective to this comparative study. In this study, we have used the DASS to measure *mental ill-health* health and examined the ML methods' prediction accuracy to predict which intervention is more suitable, based on the underlying biopsychosocial features. Realizing that most/all the participants were not clinical, it is possible that little changes can occur during the intervention period. The framework of the DHI program that conducted the biopsychosocial tests partly focused on the DASS variables as part of the social data collected in the form of *Test scale levels*. We can achieve the productive use of the wellness dataset with more integrated tests spanning all dimensions of the biopsychosocial model within the framework of DHI program.

This study used ML and non-linear classifiers on real-life biopsychosocial features, compared their prediction accuracy and evaluated their suitability to a DHI. The experimental studies indicated that while the methods such as scatter plots were unable to reveal the linear relationship between the features of the 'Wellness dataset', the ML approach was successful in identifying which features are appropriate as predictors of *mental ill-health*. Prediction accuracy of classification models improved when all biopsychosocial features are considered input features compared to using features from just one or two domains. Classifiers such as extra tree classifier and linear discriminant analysis showed high prediction accuracy (at 62.5%) in the mindfulness group. Decision trees and SVM showed high prediction accuracy (at 61.1%) in the physical activity group. While the research work focused on establishing the relevance of ML approach for DHI, the prediction accuracy can be improved further by improved classifier design and data collection.

5.2 Prediction model using RNN-LSTM

In the previous section, we recognised the power of ML methods using non-linear classifiers to identify which features are appropriate as predictors of *mental ill-health* represented by the DASS score. We chose the classifiers for the experiment based on their robustness to small datasets. Most datasets generated from DHI programs have a small sample size because the number of participants enrolled in these programs is limited. Small sample sizes pose a challenge for finding statistically significant biopsychosocial markers. The wellness dataset generated by the 'Wellness study' DHI program in this research, is also marked by small sample size as they were a limited number of participants who volunteered to participate in this program. Even though the wellness dataset is highly dimensional, the small sample size of participants limits the application of statistical tests at the participant cohort level for the two different types of interventions.

In contrast to this, the large sample size of time-series physiological data, collected through the BASIS watch wearable sensor device, provides ample scope to examine ML methods on this data. The physiological data used in this research is long segments of continuous data over an equal time interval of one second, collected during the 'Wellness study' DHI program. We will use multivariate time-series physiological data, a subset of the existing wellness dataset, in the experiments described in this section. We will feed this multivariate physiological time-series data as inputs to the ML algorithm to build the analytical prediction model. To handle the large volume of time-series physiological data, where data is intrinsically linked to the notion of time, the ML algorithm must understand sequences (of time) instead of isolated samples.

In this study, we apply ANNs to analyse, model and interpret the time-series physiological data. In the past, ANNs have widely been applied to real-world problems in areas of business, education, economics and many aspects of life problems [307]. Most researchers have used ANNs and ML successfully to solve classification problems [308, 309]. ANNs are excellent identifiers of trends and patterns in data [310], and they are suited for forecasting and prediction needs. The choice of using ANN for our experiments was driven the physiological time-series data representation and the nature of the problem being solved. In the case of time-series data, the order of inputs is an important attribute as it is for most sequential data.

In ANNs such as feed forward neural networks (FFNN), information is transmitted only in one direction from the input nodes, through any hidden nodes, and finally to output nodes. Some examples of FFNNs are single-layer perceptron and multilayer perceptron. Although FFNNs can acquire high predictive power with sequence data, using the overlapping windowed dataset method, they are incapable of producing the same accuracy when the input sequence is rearranged. This is because FFNNs do not distinguish the order within the data and therefore miss some information. Therefore a neural network that can remember the order of inputs is required. Recurrent Neural Networks (RNNs), a class of ANNs, has a short term memory and supports the output of the previous state to be fed as the input to the next state. These networks are called recurrent as they carry out repeated tasks for every observation of the input sequence and the output relies on the preceding computations. We can think of RNN to have 2 inputs namely the current and the recent past and these inputs represented as a sequence, contain important information to predict the future observation. This property of RNNs makes it powerful for modeling sequence data such as time-series or natural language [311]. In our experimental studies we are interested in using an internal state memory to process the sequential nature of the time-series physiological data to predict features of interest that describe biopsychosocial wellbeing such as *Improvement* or *No Improvement* on DERS and primary emotion. Hence we will choose RNN for the experiments in our study.

Although RNNs can learn to recognise short-term time-dependent patterns and relationships, it is impractical to use standard RNNs to learn long-term patterns in data sequences. This is because of the 'vanishing gradient problem' that occurs when, during back-propagation, the loss function gradient becomes negligibly small after only a few recurrences [312]. The nature of the physiological time-series data which is represented in the form of long segments of continuous observations over a one second time interval, demands storing of data observations for a long time and learning from long-term dependencies. A type of RNN called RNN-LSTM is well suited for this requirement. RNN-LSTM networks are a type of RNN that uses special units called 'memory cell' that can store and maintain information in memory for long periods. This is in addition to standard units of the RNN model architecture. The RNN-LSTM network uses gates to control when information enters the memory, when information is output, and forgotten. The input gate determines whether or not to let the new input in, the forget gate determines which information to delete that is not important and the output gate decides what information to output [313]. These three gates are analog and are based on the sigmoid function whose range is between 0 to 1. This capability of RNN-LSTM enables the neural network to learn

longer time dependencies compared to vanilla RNNs [312]. LSTM networks [312] have been widely used in Natural Language Processing (NLP) and have achieved state-of-art performance in sequence tagging tasks including Part-of-Speech tagging [314], Named Entity Recognition [314] and Chunking. [315].

The RNN-LSTM network is well-suited to construct the predictive model for our experimental studies as it not only allows the model to treat the time-series physiological data as sequential data but is also capable of remembering information for a long period. The length of the physiological time-series data is 900 to 2520 times steps long and the RNN-LSTM architecture is suitable to handle this length of time in a sequential manner.

We will conduct two experiments using the RNN-LSTM model on time-series physiological data as input features. The first experiment will use the RNN-LSTM model to a priori predict the likelihood of improvement of the negative affective variable DERS, prior to participation in a DHI program. The outcome of this experiment can be used as guidance to determine the group of participants, who are more likely to benefit from a similar DHI program in the future, based on their physiological features before enrolment. The second experiment will use the RNN-LSTM model to predict participants' primary emotion at the point in time, based on the wearable sensor device's physiological features twenty minutes preceding the reported time-stamp. Both the experiments will use classification models to solve the time-series sequence classification problems. We will use supervised learning for the experiments that operate on labelled data for training so that the model can predict the class label for the unseen sequences in the test set.

5.2.1 Sequence classification and time-series data

The physiological time-series data collected from the BASIS watch wearable device is sequential because the data is represented as an ordered list of events. There are two approaches to analyse time-series data namely, sequence prediction and sequence classification.

1. Sequence prediction: A problem in which historical sequence information is used to predict the succeeding value or set of values in sequence, is called sequence prediction. A sequence prediction problem differs from a supervised learning problem. It imposes that the observations follow a well-defined sequence and this sequence be preserved when training the model to make predictions. Sequence prediction is not a relevant approach in the context of our experimental study, as we are not interested in

predicting the future value of a physiological feature based on its historical sequence information.

2. Sequence classification: This refers to a predictive modeling problem where a sequence of inputs is distributed over time or space. The aim is to predict a category or class label for an input sequence. In this way, we build a classification model which uses a labelled dataset so that the model can predict the class label of an unseen sequence [316].

Our experiments described in this section will analyse the physiological time-series data treating it as a sequence classification problem. This is because we are interested in predicting a class label for the input physiological time-series data. As mentioned earlier, we are also interested in storing the historical sequence information in the input time-series data to make the class predictions.

5.2.1.1 Sequence classification

Sequence classification algorithms are broadly divided into two categories. They are binary sequence classification algorithms (see Figure 5.4) and multi-class sequence classification algorithms (see Figure 5.5). The two experiments described in this section are sequence classification experiments as it predicts a class label for a given input sequence of vectors. The first experiment (Subsection 5.2.4) is a binary sequence classification experiment as the output class is either 0 which represents DERS *Improvement*, or 1 which represents DERS *No improvement*. The second experiment (Subsection 5.2.5) is a multi-class sequence classification experiment, as the output class can be any one of the eight primary emotions.

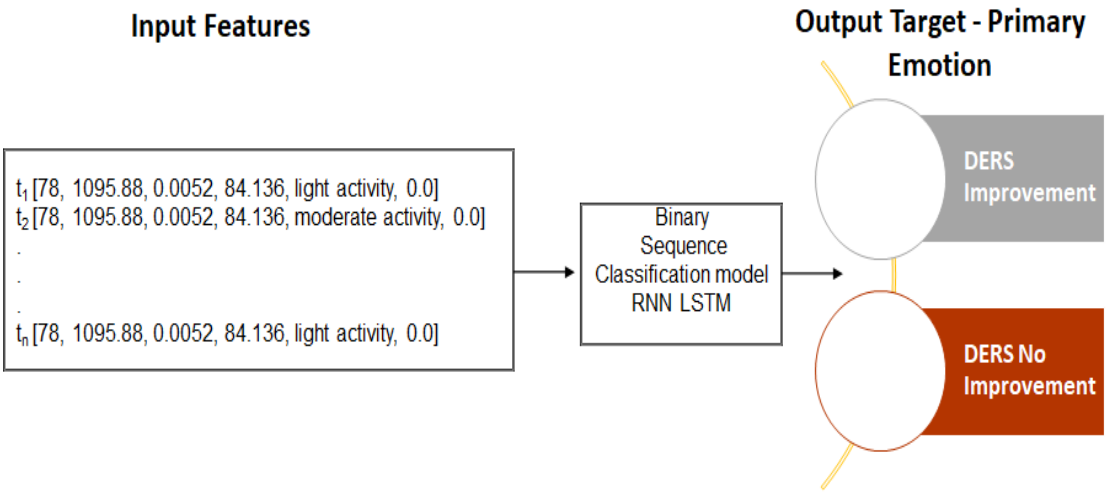


Figure 5.4: Binary sequence classification model
(The input features represent the physiological features: heart-rate, accelerator magnitude, skin temperature, galvanic skin response, activity level and steps).

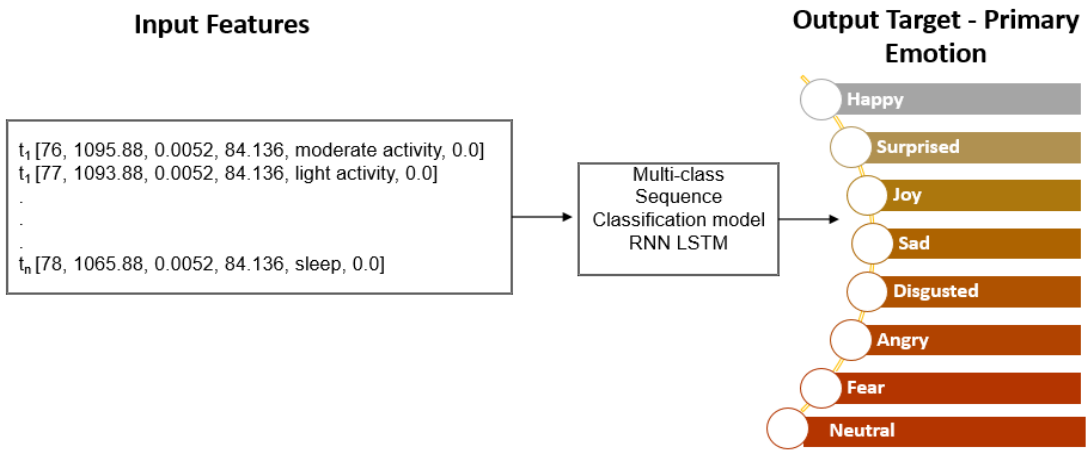


Figure 5.5: Multi-class sequence classification model
(The input features represent the physiological features: heart-rate, accelerator magnitude, skin temperature, galvanic skin response, activity level and steps).

Sequence classification has been used in a broad range of real-world applications. Xing et al. [317] have discussed the various challenges in sequence classification on different data

types including time-series data. In health-informatics, classifying ECG time-series can predict if the data is sourced from a healthy person or a person with heart disease [318]. In the area of information retrieval, classifying documents into different topic categories has been very successful [319]. Other interesting areas where classification algorithms have been used successfully, include classifying query log sequences to distinguish web-robots from human users [320, 321] and classifying transaction sequence data in a bank to combat money laundering [322]. In the area of genomic research, the functions of a new protein are learned from classifying protein sequences into existing categories [323]. Other examples of sequence classification problems include DNA sequence classification [324], anomaly detection [325, 326] and sentiment analysis to predict whether the sentiment of the text is positive or negative for a given set of sequences.

Time-series sequence classification has been actively researched for a very long time [327, 328]. Existing research works have used multivariate time-series classification for gesture recognition [329] and motion recognition [330]. In the area of Human Activity Recognition (HAR), Norgaard et al. [331], describe an RNN-LSTM framework to capture the dependency between consecutive sensor data samples and generate labels for every micro-segment. The time-series ECG dataset was used to demonstrate the performance of a semi-supervised learning classifier [318]. The recent work by Yang et al. [332] describes a Convolutional Neural Network(CNN) framework, to perform sensor classification by using multivariate time-series sensors' data, as inputs.

To the best of our knowledge, RNN-LSTM has not been used for time-series sequence classification, in the context of physiological time-series data from the DHI program in a real-world setting, to predict any aspect of biopsychosocial wellbeing. The work presented in the following experiments is new. The outcome of the experiment can be used as guidance to determine the group of participants, that are more likely to benefit from a similar kind of DHI program in the future, prior to their enrolment into the interventions. In this way, we build upon existing physiological data to enable enrolment into future DHI programs, informed by the classification and prediction models' results.

5.2.1.2 Time-series data

The physiological time-series data can be approached as either univariate time-series or multivariate time-series. When each feature of the physiological time-series data is treated in isolation, it represents the time-series data as univariate (e.g., using only the feature *heart-rate* as input to the sequence classification model). In this case, the univariate time-series is a sequence of real values ordered in timestamp ascending order.
E.g.,

$$(t_1, 76)\ (t_2, 76) \ldots (t_n, 78)$$

is a simple time-series recording the *heart-rate* data from timestamp

$$t_1 \text{ to } t_n$$

On the other hand, a multi-variate time-series is a sequence of numerical vectors. For e.g.,

$$(t_1,\ (76,\ 1095,\ lightactivity))\ (t_2,\ (76,\ 1093,\ lightactivity))$$

$$\ldots (t_n,\ (65,\ 1065,\ moderateactivity))$$

is a time-series recording a vector of features for *heart-rate*, *accelerator magnitude* and *activity level*. Copious amounts of unlabelled time-series data are often readily available however, in most instances, supervised time-series data is very difficult or expensive to obtain. Our experiments will use supervised multi-variate time-series physiological data derived from the existing wellness dataset. The time-series data is supervised because it is labelled with the output target label (DERS *Improvement* or DERS *No Improvement* for experiment 1 (section 5.2.4) and with one of eight primary emotion values for experiment 2 (see Section 5.2.5). As described in the previous chapter (see Section 4.3.2.2), the raw physiological time-series data were pre-processed using data imputation. The univariate imputation methods such as *na_kalman* function, linear interpolation and linear trend at point was used to impute missing values for *heart-rate*. The *na_kalman* function was used to impute missing values for *accelerator magnitude*, *skin temperature*, *galvanic skin response*, and *steps*.

The experimental problems in this section are expressed as time-series sequence classification problems and will be solved using a feature-based classification approach. This approach will transform the input sequence (multivariate time-series physiological data), into a vector of features. The sequences of feature vectors will then be fed as input to the RNN-LSTM model to predict the output class label. The performance of the RNN-LSTM model will depend vastly on the choice of hyper-parameters, hence they must be

chosen carefully to get good results. Being a relatively new model, there are no established guidelines for configuring RNN-LSTM hyper-parameters [313]. The next section will review some commonly used approaches, methods, and considerations for hyper-parameter optimisation.

5.2.2 Hyper-parameter optimisation approach

Achieving a good LSTM [312] network requires selection and optimisation of hyper-parameters such as tuning the number of recurrent units, selecting the depth of the network, selecting the dropout rate and many more [333]. Hutter et al. highlight the significance of hyper-parameter selection that often makes the difference between mediocre and state-of-art performance for ANN models [334]. Nakisa et al. [335] evaluate and compare different hyper-parameter optimisation methods using a dataset collected from wearable sensors to classify emotions. Most studies use grid search and manual search to configure neural networks. A widely used strategy is the combination of manual search and grid search [336] and the use of ML software packages such as *libsvm* and *scikit.learn* [337].

The grid search method of hyper-parameter search is also called parameter sweeping. It is an exhaustive search through a manually defined subset of possible hyper-parameters [333]. The disadvantage with using grid search is that it is computationally expensive. Randomly chosen trials are more efficient for hyper-parameter optimisation than trials on a grid [338]. When using randomised search, the options and ranges for hyper-parameters must be applied carefully. When using random search, sample parameters are set at random for a fixed number of times. This method of setting hyper-parameters is reasonably efficient and also provides the advantages of implementation simplicity and reproduceability of pure grid search. The problem with random search is that it does not adapt its behaviour based on previous outcomes. In some cases, a single poorly chosen hyper-parameter (e.g., a very high learning rate), can prevent the model from learning efficiently.

Manual tuning of hyper-parameter is an effective approach. The designer iteratively selects different architectures and hyper-parameters and hones it to a high-performance region of the hyper-parameter space. Using a manual search approach to configuring a neural network such as RNN-LSTM, there is no 'correct answer' on the number of hidden neurons or the number of layers one should choose for the model. A drawback of using manual search is the difficulty in reproducing results [338]. We will use the manual search method for the experiments to determine the values of the hyper-parameters. The choice of manual search is preferred as it is flexible to assign different values, change a part of

them, train the model again and check the difference in the performance scores without using automation to select or change values of parameters.

The experiments will use and assign different values to the hyper-parameters a. dropout layer b. activation function c. learning rate. The dropout layer is used to prevent over-fitting since it ignores the randomly selected neurons during training and in this way reduces the sensitivity to the specific weights of individual neurons [339]. The choice of activation function on the output layer depends on the kind of problem we want the model to solve [340, 341]. Common choices are linear functions, sigmoid function and softmax functions. We will use the sigmoid and softmax activation functions to solve the sequence classification problems. For the binary sequence classification problem described in experiment 1 (see Section 5.2.4), we will use the sigmoid activation to make clear predictions for a two-class problem. The probabilities produced by the sigmoid activation are independent and are not constrained to sum to 1. This is because the sigmoid activation function views each raw output value separately and hence the predictions are clear. For the multi-class sequence classification problem described in experiment 2 (see Section 5.2.5), we will use the softmax activation function since it can handle multiple output classes. In using this function, the outputs are interrelated. This is because the softmax probabilities will always sum to 1 by design. Hence in a multi-class sequence classification problem, to increase the likelihood of one class, the likelihood of the other class must decrease by an equal amount.

The learning rate hyper-parameter will be used to control the extent of changes to be carried out on the RNN-LSTM model based on the estimated error, each time the weights of the model are updated. Selecting an optimum value for the learning rate is a challenging task. A value too small may lengthen the training process and may also lead training to get stuck. A very large value may result in learning a sub-optimal set of weights quickly or may lead to an unstable training process. We will configure the value of this parameter to be within the range 0.0 and 1.0. Leslie N. Smith [342] describes a powerful method to select the learning rates and optimise the learning schedule for a neural network by training with cyclical learning rates instead of fixed values. However, for our experiments, we will follow a trial and error approach to tune and optimise the learning rate and the values of other hyper-parameters such as the loss function number of hidden layers and number of units in each hidden layer. The manual tuning of hyper-parameters allows us to compare and check the difference in performance scores when applied to the unseen test data.

5.2.3 Output target label

In this section, we describe the choice for the output target variables used in the two experiments.

5.2.3.1 Difficulty in Emotion Regulation Scale

DERS is widely used to measure of emotion (dys-) regulation ability in clinical and nonclinical populations [343]. It will be used as the output target variable for experiment 1 (see Section 5.2.4). DERS is designed to measure multiple dimensions of emotion regulation ability in a comprehensive way. The dimensions of DERS described below are adapted from the work reported by Williams et al. [257]. Levels of emotion regulation difficulties as measured by DERS are associated with different kinds of psychopathology such as personality disorders, post-traumatic stress disorders, depression and anxiety [344–348]. Higher scores for the negative affective variable DERS indicate poorer emotion regulation by the participant [343]. The DERS score measures emotion (dys-) regulation ability broadly in the dimensions of:

1. Nonacceptance: The items in this dimension reflect an inclination to experience negative secondary emotional responses or a non-accepting reaction to distress, mainly shame, guilt, or self-blame regarding one's own (negative) emotions.

2. Goals: The items in this dimension refer to troubles following goal-directed behaviour and reflect difficulties associated with concentration or carrying out tasks when upset.

3. Impulse: Items in this dimension describe difficulties controlling emotions or behaviours when in distress. It also includes items reflecting the perception of emotions as overwhelming.

4. Awareness: These items refer to the absence of emotional awareness and are made up of reverse-scored items. These items reflect the tendency to pay attention and to acknowledge emotions.

5. Strategies: This refers to limited access to emotion regulation strategies. The items in this dimension reflect the belief that the person's negative emotions cannot be regulated. It suggests a feeling of hopelessness when confronted with one's negative feelings.

6. Clarity: These items describe a lack of emotional clarity. They refer to the ability of a person to understand emotions. A high score in these dimensions indicates a high degree of confusion regarding emotions.

We find that DERS as a measure of emotion (dys-) regulation ability is closely linked with physiological variables such as *heart-rate*, *galvanic skin response* and level of physical activity. Researchers have examined the relation between emotion regulation and heart rate in the past. Williams et al. [257] highlighted two distinct facets of emotion regulation: impulse control and emotional clarity. These facets of emotion regulation were items of interest when investigating the link between emotion regulation, resting vmHRV (Vagally mediated Heart-Rate Variability), and related health outcomes including morbidity and mortality. Thayer and Lane [349] propose that the characteristic beat-to-beat variability in the heart-rate time-series and heart-rate variability (HRV) - serves as a readily available index and measure of emotion regulation capacity. GSR, physical activity and sleep are also closely related to emotion regulation. Existing research has shown that an increase in GSR is related to emotional intensity; the more an individual experiences an emotion, the more likely a detectable change in GSR measure is seen. Low levels of GSR are linked to emotions or experiences of relief [350] or contentment [351]. In the context of physical activity and sleep, the work by Zang et al. [352] examines how the mind-body exercise intervention improves implicit emotion regulation ability. Sleep deprivation was a contributing factor that diminished the capacity to regulate emotion [353]. As described above, DERS as a negative affective variable that measures emotion (dys-) regulation, is closely linked to the physiological variables of *heart-rate*, *galvanic skin response* and level of physical activity. A lower DERS score at the end of the DHI program in week 8 that at the start of the program at week 1 indicates *Improvement* in emotion regulation ability that is brought about by the mindfulness or physical activity intervention. A same or higher DERS score at the end of the DHI program compared to the start of the program indicates *No Improvement* in emotion regulation ability. Hence DERS score is a good choice to compute the output target label for experiment 1 (see Section 5.2.4).

5.2.3.2 Primary emotion

Primary emotion is used as the output target variable for experiment 2 (see Section 5.2.5). The 'Wellness study' DHI program described primary emotion as any one of a limited set of emotions: *Angry*, *Sad*, *Happy*, *Disgusted*, *Surprised*, *Joy*, *Neutral* and *Fear*. The primary emotion is the reaction to an external event and secondary emotion is the emotion felt about the primary emotion itself. For example, one may feel angry about being hurt or shame about feeling anxious. In this case, the feeling of hurt and anxiousness are primary emotions, and anger and shame are secondary emotions. The primary and secondary emotions were captured as part of the EMA surveys that captured emotion data 5 times a day during the DHI program. In experiment 2 (see Section 5.2.5), we draw associations between emotion experiences and physiological variables. For this purpose, primary emotion is best suited as the output target variable since it records the direct reaction to an external event.

In the following sections, we will describe the RNN-LSTM model designed to predict the output target labels, using the physiological time-series data. Both experiments use Keras 2.2.4 GPU and Python 3.6.10 as the programming environment.

5.2.4 RNN-LSTM model to predict DERS using wearable data

This section will describe the RNN-LSTM model designed to predict DERS *Improvement* or *No Improvement* using the wearable data (physiological time-series data).

5.2.4.1 Encoding

The physiological time-series data consists of numerical features and categorical features. The categorical variable *activity type* contains four categories namely *inactive*, *light activity*, *moderate activity* and *sleep*. Since the LSTM-RNN network is not suited to operate on labelled data, it is necessary to encode the categorical variable for the efficient implementation of the algorithm. One hot encoding is chosen as the preferred encoding method for categorical variable *activity type*. This is because the categories in the feature variable *activity type* do not possess any ordinal relationships. Using one hot encoding, we will represent the variable *activity* type as an integer. We will then replace the encoded integer variable using a new binary variable for each unique integer value.

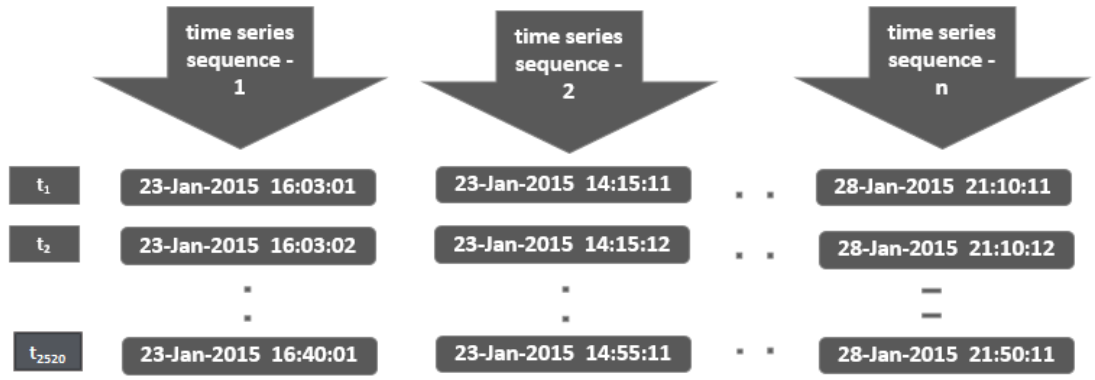


Figure 5.6: Representation of physiological time-series sequence. (Each time-series sequence is 2520 time steps long.)

5.2.4.2 Length of time-series data

As mentioned earlier, the physiological time-series data is represented as a sequence of numerical vectors that contains the values for the features *heart-rate*, *accelerator magnitude*, *skin temperature*, *galvanic skin response*, *activity level* and *steps*. The time-series sequences will be created with an identical length of 42 minutes. The decision to keep time-series sequences of length set to 42 minutes is for two reasons.

1. 42 minutes is sufficiently long to capture significant trends in data collected for any of the physiological variables

2. The exploratory analysis of the physiological data reveals that 42 minutes of sequential data is easier to extract, and is marked by fewer data imputations for missing values.

Figure 5.6 describes the physiological data represented as multivariate time-series sequences. The data in each sequence consists of 5 numerical features namely *heart-rate*, *accelerator magnitude*, *skin temperature*, *galvanic skin response*, *steps* and 1 categorical feature *activity level*.

$$(t_1, (80.4, 2925.0, 0.000154, 91.17, lightactivity, 0.0))$$

5.2.4.3 Examining class balance

Using multivariate time-series data for sequence classification, we will require the output classes to be represented equally. The training data set collected for experiment 1 is imbalanced as the number of samples in the training set for each class label is not the same. This is a common problem with most real-life datasets. The primary cause for the class imbalance occurs because the input features are not synchronous with the output target label. DERS is recorded at the participant level and in contrast, the physiological data is recorded for each second of activity engaged by the participant for the length of the time-series sequence. The RNN-LSTM model being developed is oblivious to participants' distribution and is based on the physiological time-series data distribution. Hence it is necessary to replicate the DERS output target label for each time-step in the time-series sequence pertaining to the same participant. To mitigate the class imbalance problem in the training set, we will include more time-series data using data imputation and re-arrange the input data, such that there are balanced classes for each target label (DERS *Improvement* and DERS *No Improvement*). During the re-arrangement of input data, the sequential structure of the time-series data will be kept intact.

5.2.4.4 Training and test data set-up

Table 5.9 describes the distribution of the training and test data used for experiment 1. To mitigate the class imbalance problem, we will use a modified balanced training and test set whose participants will belong to a mixed cohort of experimental and control groups.

Table 5.9: Experiment 1: Training and test data distribution

	Training set	**Test set**
Number of participants	10	9
Participants in class label *Improvement*	5	5
Participants in class label *No improvement*	5	4
Number of samples	14212	11832

5.2.4.5 Loss function

Different loss functions such as cross-entropy loss, hinge loss and squared hinge loss are examined for experiment 1 that predicts a binary value: DERS *Improvement* or *No Improvement* (0 or 1). Hinge loss and Hinge squared loss are primarily developed for Support Vector Machine (SVM) models. In using the hinge loss function, the margins between decision boundaries and data points are maximised. This ensures that each point is correctly classified. The points that are farther away from the decision margins have a more significant loss value and so these points are penalised. The binary cross-entropy loss function comes from a maximum likelihood estimate of our model's parameters. We will use binary cross-entropy loss for experiment 1 since it is suitable for solving binary classification problems where the target values are within the set {0, 1}.

5.2.4.6 Performance metrics - predicting DERS using wearable data

The input training data is fed to the RNN-LSTM model and the model is fit on the training set. As mentioned earlier, the manual search approach for hyper-parameter tuning is used which involves several rounds of training using different values for the various hyper-parameters. The hyper-parameters namely: dropout layer, optimiser learning rate, number of units, and number of hidden layers are tuned on the model's training set for several iterations. The performance metrics for each iteration are noted and compared to adjust to hyper-parameter values to find the best fitting model. Table 5.10 compares the training results for training accuracy and training loss for different hyper-parameter values during the different trials. The best performing model is observed in Trial 1, which generates a training accuracy of 0.69 and training loss of 0.60. The hyper-parameter values for this best-performing trained model is highlighted in bold in Table 5.10.

A manual 3-fold cross validation was performed on the training set to boost confidence in the training metrics using the best fitting RNN-LSTM model (see Section 5.11). The results of the 3-fold cross validation on training data revealed that the training accuracy recorded (at 66%) for Fold 1 were very close to the training accuracy recorded for remaining folds (Fold 2 and Fold 3). The results of the training accuracy for the manual 3-fold cross validation (average accuracy across folds at 66%) provided confidence in the training accuracy reported as part of Trial 1 (at 69%) on the original training set.

Table 5.10: Experiment 1: Trial-wise performance metrics (Figures in **bold** indicate significant values)

Parameters	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7
Training accuracy	**0.697**	0.63	0.68	0.57	0.68	0.65	0.65
Training loss	**0.60**	0.65	0.53	0.68	0.58	0.60	0.61
Dropout	0.65	0.65	0.67	0.65	0.66	0.65	0.65
Adam learning rate	0.0001	0.01	0.001	0.00001	0.0001	0.0001	0.0001
Units in each hidden layer	18	18	18	18	18	18	18
No. of hidden layers	1	1	1	1	1	2	1

Table 5.11: Training metrics reported on manual 3 fold cross-validation

Fold number	Training loss	Training accuracy
Fold 1	0.5995	0.6659
Fold 2	0.5994	0.66600
Fold 3	0.5996	0.6665

Table 5.12: Optimised Model parameters

Parameter	Value
Dropout value	0.65
No. of epochs	3
Loss function	binary cross-entropy
Optimiser	Adam Learning rate for optimiser 0.0001
No. of units in hidden layer	18
No. of dense layer	1
No. of units in dense layer	18
No. of epochs	3

Table 5.12 lists the optimised model parameters for the best-performing model on the training set (seen during Trial 1) which recorded a training accuracy of 69%. As a next step, the associated hyper-parameter values and weights were validated on 5 variants of unseen test sets (Test set 1 to Test set 5). The performance metrics (training accuracy and training loss) of the RNN-LSTM model for experiment 1 are described in Table 5.13. The number of epochs for each unseen test set was set to 3. The unseen test set 5 reported the best test accuracy of 63% (see Table 5.13). Validating the best performing RNN-LSTM network model on 5 variants of unseen test set, provides us with increased confidence in

the performance metrics reported by the RNN-LSTM model for this experiment.

Table 5.13: Best performing RNN-LSTM model applied on 5 variants of unseen test set. (Figures in **bold** indicate significant values)

Test set	Epoch number	Test loss	Test accuracy
1	1	0.68	0.62
1	2	0.68	0.62
1	3	0.68	0.62
2	1	0.65	0.62
2	2	0.65	0.62
2	3	0.65	0.62
3	1	0.67	0.60
3	2	0.67	0.60
3	3	0.67	0.60
4	1	0.67	0.57
4	2	0.67	0.57
4	3	0.67	0.57
5	1	0.66	**0.63**
5	2	0.66	**0.63**
5	3	0.66	**0.63**

The results of experiment 1 show that the RNN-LSTM sequence classification model can be used successfully on physiological time-series data generated from the DHI program in a real-world setting. The developed RNN-LSTM model has learned from long sequences of vector information to predict the binary class of the output target variable (DERS *Improvement* or *No improvement*) with a testing accuracy of 63 %. Table 5.14 lists the values of some hyper-parameters and reports the performance metrics for the developed model.

Table 5.14: Experiment 1: summary of performance metrics. (RNN-LSTM Model to predict DERS using wearable data.)

Parameter	Value
Size of the training set	35530000 records
No. of sequences in training set	14212
Size of 1 sequence in training set	2520 seconds or 42 minutes
Training loss	0.60
Training accuracy	0.697
Size of an unseen test set	29580000 records
No. of sequences in an unseen test set	11832
Size of 1 sequence in the test set	2500 seconds or 42 minutes
Testing loss	0.65
Testing accuracy	0.63

The black-box nature of ML algorithm limits the transparency of the RNN-LSTM model to describe the causality between inputs and outputs of the model. To overcome this problem, we repeated the experiment with each physiological feature to determine the feature that most influenced the prediction accuracy. The outcome of this experiment revealed that no single physiological feature had an overbearing influence on the model prediction metrics. Further, no single features' prediction accuracy matched the prediction accuracy of 63% as reported by the interplay of all the physiological features used as input for the experiment. Therefore, all the physiological features played an equal role in determining the prediction accuracy. Table 5.15 describes the training and test metrics reported on individual physiological features on the RNN-LSTM model.

The next section describes the RNN-LSTM model designed to predict primary emotion using wearable data (physiological time-series data).

Table 5.15: Experiment 1: summary of performance metrics when an individual physiological feature is input to RNN-LSTM network.

Feature	Training loss	Training accuracy	Testing loss	Testing accuracy
heart-rate	0.587	0.637	0.678	0.603
accelerator magnitude	0.651	0.637	0.674	0.6035
galvanic skin response	0.620	0.6835	0.7547	0.5105
skin temperature	0.6547	0.6374	0.673	0.603
activity type	0.649	0.637	0.674	0.603

5.2.5 RNN-LSTM classification model for emotion prediction

This section describes the RNN-LSTM model that predicts the primary emotion using the wearable data (physiological time-series data). During the 'Wellness study' DHI program, wearable devices used passive sensing technologies during activity tracker to detect sleep or heart-rate monitoring to capture the physiological data [354]. In contrast to this, the EMA surveys captured more detailed time courses of individual's psycho-social and subjective experiences and how they related to other phenomena such as their primary and secondary emotions and their physiological states and experience data. The EMA surveys were facilitated via a smartphone application and were successful in prompting the participants to report particular behaviours, subjective experiences, and various contextual variables. Data collected by the EMA surveys comprised the participant's primary emotion, along with other values for secondary emotion, experiences, and interactions with others.

The experiment 2 in this section uses the existing EMA data generated from the 'Wellness study' DHI program. As part of data pre-processing activities, the EMA data was synchronised with physiological time-series data collected from the wearable sensor device, by matching both these datasets' timestamps. The goal of the experiment is to build a multi-class sequence classification model that can predict the primary emotion at a particular point in time, based on the physiological time-series data reported by the participant; 5 minutes preceding the time when the primary emotion was recorded. The test environment for this experiment is set up using Keras 2.2.4 GPU and Python 3.6.10 was used as the programming environment.

5.2.5.1 Training and test data set-up

Experiment 2 solves a multi-class sequence classification problem because the output class *primary emotion* comprises eight values namely: *Angry*, *Sad*, *Happy*, *Disgusted*, *Surprised*, *Neutral*, *Joy* and *Fear*. RNN-LSTM network will be used to build the classification model. The experiment will use labelled data for training so that the model can be used to predict the class label for the test set that contains unseen sequences. *Primary emotion* is set as the target output label. The labelled training set uses the physiological and EMA survey data for 9 participants. The input features to the RNN-LSTM model are the physiological time-series data that comprise features: *heart-rate*, *skin temperature*, *galvanic skin response* and *steps*. The time-stamp of the output target label *primary emotion* will be used to synchronise the output target label with the physiological time-series data collected from the wearable device. The last 5 minute (300 seconds) lag of physiological data was discarded. This was

done because we wanted to exclude participant's physiological data that was collected while the participant answered the EMA survey questions, which approximately took about 5 minutes. Discarding the last 5 minutes physiological data reduced the length of each sequence from 20 minutes to 15 minutes (900 seconds), in both the training set and the test set. These data sequences are represented as numeric vectors. These activities were done as part of data pre-processing, prior to the start of the experiment. Table 5.19 describes the time-series dataset used for this study. The training set is reasonably balanced (see Table 5.16), whereas the test set is imbalanced (see Table 5.17), as there is a limited amount of EMA data for primary emotion values 4, 5, 7 and 8.

Table 5.16: A Balanced training set

		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	1	7200	12.9	12.9	12.9
	2	7200	12.9	12.9	25.8
	3	7200	12.9	12.9	38.7
	4	5400	9.7	9.7	48.4
	5	7200	12.9	12.9	61.3
	6	7200	12.9	12.9	74.2
	7	7200	12.9	12.9	87.1
	8	7200	12.9	12.9	100.0
	Total	55800	100	100	

Table 5.17: An imbalanced test set

		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	1	2700	20.0	20.0	20.0
	2	2700	20.0	20.0	40.0
	3	2700	20.0	20.0	60.0
	4	900	6.7	6.7	66.7
	5	900	6.7	6.7	73.3
	6	1800	13.3	13.3	86.7
	7	900	6.7	6.7	93.3
	8	900	6.7	6.7	100.0
	Total	13500	100	100	

The RNN-LSTM model was executed on the training set and reported a high training accuracy of 96.77%. The test accuracy on the unseen test set was low at 47%. The performance metrics did not improve with any changes to the hyper-parameter design and

Table 5.18: A balanced test set

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	900	16.7	16.7	16.7
	2	900	16.7	16.7	33.3
	3	900	16.7	16.7	50.0
	5	900	16.7	16.7	66.7
	6	900	16.7	16.7	83.3
	8	900	16.7	16.7	100.0
	Total	5400	100.0	100.0	

iterative runs on the training and test set data (see Table 5.20). The low accuracy score on the test set is due to the class imbalance in the test set data.

Table 5.19: Experiment 2: Training and test data distribution

	Training set	**Test set**
Count of participants	60	6
No. of samples in training set	55800	13500
Ratio of samples in each set	91.08%	8.82% %

Table 5.20: Experiment 2: Trial-wise performance metrics for the balanced training set and unbalanced test set (Figures in **bold** indicate significant values)

Parameters	**Trial 1**	**Trial 2**	**Trial 3**	**Trial 4**	**Trial 5**
Training accuracy	0.8225	0.8870	0.8548	**0.9677**	0.9677
Training loss	0.4080	0.4303	0.3538	0.1863	0.1577
Test accuracy	0.400	0.400	0.400	**0.470**	0.400
Test loss	5.45	1.7	7.04	3.69	3.48
Dropout	0	0	0	0	0
Adam learning rate	0.01	0.001	0.01	0.01	0.01
Model loss function	categorical cross-entropy	kullback leibler divergence	kullback leibler divergence	kullback leibler divergence	categorical cross-entropy
Units in each hidden layer	hidden layer 1 - 45 hidden layer 2 - 8	hidden layer 1 - 45 hidden layer 2 - 8	hidden layer 1 - 45 hidden layer 2 - 8	hidden layer 1 - 45 hidden layer 2 - 8	hidden layer 1 - 45 hidden layer 2 - 8
No. of hidden layers	2	2	2	2	2

As a next step, the class imbalance of the test set is investigated and corrected such that samples for each output label (i.e., each class of primary emotion) are balanced. The revised test set, described in Table 5.18 is balanced by restructuring the participant data. The temporal aspect of the sequential time-series data is kept intact.

5.2.5.2 Performance metrics - predicting PE using wearable data

The RNN-LSTM model is executed on the training data for several iterations with different values for model hyper-parameters. The best performing RNN-LSTM model on the training set reported a training accuracy of 72.5% (see Figure 5.8) and a low training loss of 1.0478 (see Figure 5.7). The best performing RNN-LSTM model was re-run on the unseen balanced test set. The test accuracy improved and was reported at 67% with a test loss of 1.5598 (see Table 5.21).

To improve the performance metrics, the output classes for this experiment were reduced from 8 classes to 3 classes. This was done by reclassifying the 8 output primary emotions to 3 emotions namely positive emotions comprising (*Happy*, *Joy*, *Surprised*); negative emotions comprising (*Sad*, *Depressed*, *Fear*, *Anxious*) and neutral emotions comprising (*Neutral*). The test accuracy of the RNN-LSTM model with three output classes, on the test set deteriorated (50%) compared to the earlier seen test accuracy of 67% for eight output classes (see Table 5.21). Table 5.22 describes the training and test metrics for the RNN-LSTM three-class model with and without the dropout layer. The test accuracy's poor performance can be explained by the interesting phenomena, often seen in a multi-class problem. As described by Abramovich et al [355], 'in a multi-class problem, the precision of classification improves as the number of classes grows. This is because, more accurate feature selection is possible since even weaker significant features, which are not sufficiently strong can be manifested in a coarse classification, being shared across the classes'. This has a stronger impact as the number of classes increases.

Table 5.21: Experiment 2: Training and Test metrics (Balanced training and test data sets) (Figures in **bold** indicate significant values)

Dense layer	Activation output layer	Model loss	Optimiser	Learning rate	Epochs	Training accuracy	Training loss	Test accuracy	Test loss
1	softmax	kullback_leibler_divergence	adam	0.001	600	0.96774191	0.127969174	0.4	3.38
1	softmax	kullback_leibler_divergence	adam	0.001	600	0.96774191	0.127969174	0.666666687	2.03837
1	softmax	kullback_leibler_divergence	adam	0.001	200	0.88709676	0.430371608	0.5	1.28852
2	softmax	kullback_leibler_divergence	adam	0.01	200	0.69354838	0.879156711	0.333333343	7.25998
2	softmax	categorical_cross_entropy	adam	0.01	200	0.93548387	0.180161358	0.5	3.66896
2	softmax	categorical_cross_entropy	adam	0.01	600	0.93548387	0.105320831	0.333333343	7.39356
1	softmax	kullback_leibler_divergence	adam	0.001	800	0.98387098	0.034867851	0.333333343	3.04841
1	softmax	kullback_leibler_divergence	adam	0.001	400	0.98387098	0.089431563	0.333333343	1.88943
1	softmax	kullback_leibler_divergence	adam	0.0001	400	0.56451613	1.3029569	0.5	1.6633
1	softmax	kullback_leibler_divergence	adam	0.0001	600	**0.72580647**	**1.047843291**	**0.666666687**	**1.55985**
1	softmax	kullback_leibler_divergence	adam	0.01	600	0.87096775	0.323096675	0.333333343	6.41299

Table 5.22: Experiment 2: Training and Test metrics for a three-class model (Balanced training and test data sets)

Dropout layer	Dense layer	Activation output layer	Model loss	Optimiser	Learning rate	Epochs	Training accuracy	Training loss	Test accuracy	Test loss
0.5	1	softmax	kullback_leibler_divergence	adam	0.001	600	0.91935486	0.423808637	0.5	1.08629
no	1	softmax	kullback_leibler_divergence	adam	0.001	600	0.77419353	0.318217117	0.5	0.42848

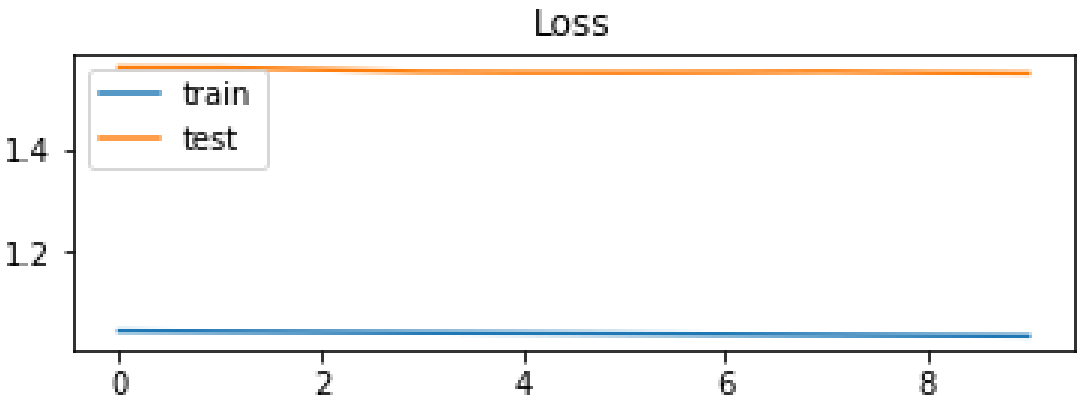


Figure 5.7: Training versus test loss

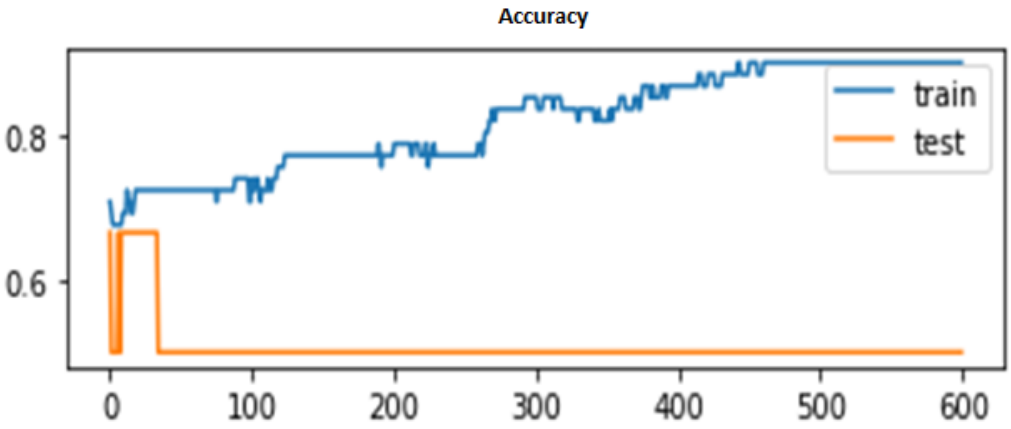


Figure 5.8: Training versus test accuracy (X-axis shows the number of epochs (600) and Y-axis shows the accuracy percentage)

The results from experiment 2 reveal that the RNN-LSTM model accuracy can be further improved by increasing the sample size of participants used for training. The low sample size of participants' EMA data had limited the size of the training and test sets used for the study. The imputation of EMA data is complex and hence this strategy was not considered as a means to increase the sample size. Small sample sizes of participants are a typical challenge in handling real-world DHI program data. The voluntary nature of participation in EMA surveys lead to non-responses or skipped questions which further resulted in

missing values in EMA survey data. The work in this experiment supports the choice of using the RNN-LSTM model for modeling physiological time-series data to predict primary emotion.

5.2.6 Analysis of results

The analysis of the prediction accuracy of the experiments conducted in this section reveal that the results can be improved further using following approaches.

1. The highest prediction accuracy set was reported to be 63% for the experiment RNN-LSTM model to predict DERS using wearable data 5.2.4. Although the prediction accuracy is reasonable it can be improved by performing the experiment using RNN-LSTM architecture that implements the computationally expensive grid search method of hyper-parameter optimisation.

2. The highest prediction accuracy was reported to be 67% for the experiment RNN-LSTM model to predict DERS using wearable data 5.2.5. The prediction accuracy of this experiment can be increased by making enhancements to data collection methods, to improve the size and quality of collected data and reduce class imbalance problems. Although data imputation can be used to overcome the class imbalance problem in some cases, the imputation of real-world EMA data as seen in this experiment is complex and challenging.

5.3 Summary

In this chapter, we explored the use of predictive analytics to predict aspects of biopsychosocial wellbeing by analysing patterns in the wellness dataset aligned to the biopsychosocial model of health. We used statistical and ML methods to examine possible connections between the different biopsychosocial features of participants who undertook one of the two interventions of the DHI program, namely the mindfulness intervention and the physical activity intervention. We investigated the relevance of ML for determining the suitability of a participant into a DHI assignment based on existing biopsychosocial information. We used different non-linear classifiers and compared their prediction accuracy to evaluate their suitability to determine if biopsychosocial features can be used to predict DASS. The experimental studies indicated that while the methods such as scatter plots were unable to reveal the linear relationship between the features of the wellness dataset, the ML approach successfully identified which features are appropriate as predictors of *mental ill-health*.

We further investigated the suitability of RNN-LSTM to predict DERS improvement, affected by one of the two DHI (mindfulness or physical activity). We carried out two experiments to solve the time-series sequence classification problem and we reviewed various approaches for model hyper-parameter optimisation. As part of experiment 1, we developed an RNN-LSTM model to predict DERS *Improvement* or *No Improvement* based on wearable data. Further, we adapted this model in experiment 2 to predict primary emotion at a particular point in time, based on the participant's physiological time-series data. The RNN-LSTM model is aptly suited for sequence classification problems described in experiments 1 and 2. The model is efficient in dealing with sequential time-series data and can remember information for a long period. Prediction accuracy of the RNN-LSTM model used in this study can be further improved by increasing participants' sample size within the training and test sets.

In the following chapter (Chapter- 6), we summarise the work reported in this thesis and future directions for research.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

Biopsychosocial wellbeing is a consequence of the interaction between the biological, psychological and social factors as described within the biopsychosocial model of health. One of the many limitations of the biopsychosocial model is that it has unclear boundaries between the biology, psychology and social disciplines. In recent years, digital health technologies have addressed these limitations by providing more accessible, scalable and comprehensive data from various aspects of health and disease. Using digital health allows us to capture biological, psychological and social information, offer a non-dichotomous understanding of an individual's medical condition and help unmask the biomedical focus of the original biomedical model.

This increased research into digital health has paved the way for integrating healthcare data from different platforms. Advancements in digital healthcare technologies have advanced clinical practice, from prevention to diagnosis and from monitoring to disease management [36]. It has led to unprecedented public interest and engagement in self-management and wellbeing. With advancements in digital health, healthcare screening aligned with the biopsychosocial model has helped health practitioners evaluate the potential benefits of a health solution at the individual, population or organisational level using digital tools. The use of mobile and wireless DHI has allowed healthcare practitioners and researchers in the healthcare domain to develop and implement different kinds of DHI programs. Digital health program frameworks in clinical and non-clinical practice increasingly use patient-centric interviews, enabling clinicians and healthcare practitioners

to identify a scientific biopsychosocial model specific to each patient. The openness to and demand for biopsychosocial interventions is also driven by developments in behavioural health and medicine that focus on behaviour and health, which are significant contributors to maintaining overall wellbeing and quality of life.

Digital tools are being used to improve the reach and effectiveness of DHI programs, leading to immense opportunity for mining the large amount of data produced by digital health programs and applying analytics to evaluate these programs' effectiveness [139]. Thus, digital health research grows in importance and is based upon what is already known, so that investments in the digital health space can be informed by the strongest possible evidence base [17]. Applying digital health analytics using data mining and ML has emerged as an efficient approach to evaluating data-rich interventions. The small number of identified studies and the low rigour of ML evaluation methodology used therein indicates a vast need for the research community to conduct further studies in the area of DHI programs to demonstrate the potential of ML to evaluate DHIs in real-life settings.

Several studies have evaluated interventions in a clinical setting, and these are retrospective validations of ML algorithms and models given the availability of one or more datasets. Further, these studies have focused primarily on data from a single discipline centred only on the body, treating illness as a purely biological event. To date, few researchers have evaluated or analysed DHI by following the biopsychosocial model in a real-life setting. The research in this thesis analyses the existing wellness dataset generated from the 'Wellness study' DHI program conducted by the School of Science, Psychology and Sport at Federation University. The DHI program produced a complex set of biopsychosocial data (referred to as the wellness dataset). This real-life wellness dataset provided fertile ground for applying the data analytics process model and discovering patterns hidden in the data. Different statistical methods and ML methods were used to pre-process, analyse and discover new knowledge from the data analysis.

6.1 Conclusions

As mentioned earlier, this thesis has explored the existing wellness dataset generated from the DHI program. We used data analytics on the transdisciplinary wellness dataset to discover patterns and relevant insights. The data analytics work in this thesis was carried out following the OSEMN data mining process framework. This work is novel research undertaken in the biopsychosocial domain, as it systematically dissects unique and complex real-world biopsychosocial data. The data analytics work in this thesis was at the intersection of IT, statistics and healthcare domains. We examined the transdisciplinary wellness dataset and uncovered the challenges of data quality and data integration along with the challenges exposed by the interplay of the underlying complex data's variety, value and veracity. The key achievements and significance of the research reported in this thesis are summarised below.

- We modelled neurocognitive reaction time using gamma distribution. We examined reaction time data obtained from participants undertaking neurocognitive tests using electrodermal activity to assess their distribution. Analysis of participants' reaction time in the PEBL Go/No-Go test showed that the reaction time data are more compatible with a gamma distribution. It also clearly demonstrated that these can be better modelled by gamma distribution that considers higher order moments of data, such as skewness and kurtosis, using shape and scale parameters rather than the commonly applied Gaussian distribution, which is defined using mean and variance.

- We developed a novel 'peak heart-rate' count metric to quantify level of 'biological' stress. We analysed and focused on heart-rate data as one key biophysiological indicator of stress. A new metric, 'peak heart-rate count', was proposed to measure 'biological' stress. Further, a new approach using the median and IQR of daily heart-rate data was used to define *maximum threshold*. The proposed new metric, 'heart-rate peak count', and the associated new MedIQR method were used to analyse the distribution of the heart-rate peaks between the two interventions, mindfulness and physical activity, to determine which was more beneficial to the participants during the DHI program. With this, we can apply the new metric, 'heart-rate peak count', and the associated new MedIQR method for studies where HRV data are not available.

- We conducted a comparative evaluation of nonlinear classifiers to identify features that can predict *mental ill-health* represented by the DASS—ML and nonlinear

classifiers were explored on the existing wellness dataset. Real-life biopsychosocial features were input to nonlinear classifiers to predict mental ill health represented by DASS score. The work carried out showed that the prediction accuracy of the classification model improved (by 62.5% for mindfulness intervention and by 61.1% for physical activity intervention) when all features from the three disciplines of the biopsychosocial model were used. This finding underpins the need to develop future DHI programs aligned with the biopsychosocial model of health. Well-performing nonlinear classifiers such as extra tree classifier and linear discriminant analysis (with prediction accuracy of 62%) and decision trees and SVM (with prediction accuracy of 62%) can be applied to build prediction models to evaluate similar interventions in the mindfulness and physical activity spaces.

- We undertook wearable data analysis using the RNN-LSTM sequence classification model. We developed an RNN-LSTM model using a sequence classification algorithm to predict negative affective variable DERS and PE based on input time-series wearable data. This work supports the choice to use the RNN-LSTM network to a build prediction model that can predict positive and negative affective variables and PE sourced from EMA surveys. The results from this study can be adapted and applied by healthcare specialists to correlate the information gained from wearables with that of EMA-reported findings. The study findings can also help accelerate and improve the design of DHI programs in the mindfulness and physical activity spaces, which use sensor devices and EMA surveys in parallel to collect data.

- We used the AUC_G formula for cortisol score computation. The formula was used on the biological saliva data to compute a single cortisol score from four cortisol readings that were recorded on a single day and changed over the course of that day.

6.2 Future work

Through this research, we were able to comprehensively study and implement the data analytics techniques, using statistical and ML methods, on the existing wellness dataset generated from the 'Wellness study' DHI program. To extend the research presented in this thesis, the following future research studies can be explored.

- Future research work can focus on further enhancing the use of ML methods on DHI programs. Currently, limited research provides evidence of the effectiveness of ML learning applications in DHI programs. This is a promising area for future work, as ML techniques in digital health can enable healthcare solutions to achieve their full potential via their application within the healthcare and scientific community.
- Further, future research work focused on collecting richer biopsychosocial data from all three disciplines (biology, psychology and social) may be designed to use all the variables of the holistic biopsychosocial model. It can help analyse how these variables interplay in addition to investigating feedback loops, reciprocal influences and correlated variables, thus allowing for a multi-level and multivariate analysis.
- Another research area that can be explored is the generation of synthetic biopsychosocial data using real-world data and ML methods. Such a dataset can serve as a benchmark dataset for future researchers.

BIBLIOGRAPHY

[1] D. Hl, "What high-level wellness means," *Canadian Journal of Public Health-revue Canadienne De Sante Publique*, vol. 50, pp. 447–457, 1959.

[2] F. Irtelli and F. Durbano, "Quality of Life and Biopsychosocial Paradigm: A Narrative Review of the Concept and Specific Insights," 2020. [Online]. Available: https://doi.org/10.5772/intechopen.91877

[3] T. Schramme and S. Edwards, "Handbook of the Philosophy of Medicine," 2017. [Online]. Available: https://doi.org/10.1007/978-94-017-8688-1

[4] J. Suls and A. Rothman, "Evolution of the biopsychosocial model: prospects and challenges for health psychology," *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, vol. 23 2, pp. 119–25, 2004. [Online]. Available: https://doi.org/10.1037/0278-6133.23.2.119

[5] G. L. Engel, "The need for a new medical model: a challenge for biomedicine," *Science*, vol. 196, pp. 129–36, 1977. [Online]. Available: https://doi.org/10.1037/H0089260

[6] B. D. Thomas, B. C. Bultz, "How much longer must medicine's science be bound by a seventeenth century world view?" *Psychother Psychosom*, vol. 57, pp. 3–16, 1992.

[7] B. C. Thomas and B. Bultz, "The future in psychosocial oncology: screening for emotional distress–the sixth vital sign," *Practical Assessment, Research and Evaluation*, vol. 4, pp. 779–84, 2008. [Online]. Available: https://doi.org/10.2217/14796694.4.6.779

[8] M. Loscalzo, K. Clark, S. Pal, and W. F. Pirl, "Role of biopsychosocial screening in cancer care," *Cancer J*, vol. 19, pp. 414–20, 2013. [Online]. Available: https://doi.org/10.1097/PPO.0b013e3182a5bce2

[9] G. Engel, “The biopsychosocial model and the education of health professionals,” *Annals of the New York Academy of Sciences*, vol. 310, pp. 169–87, 1978. [Online]. Available: https://doi.org/10.1016/0163-8343(79)90062-8

[10] S. N. Ghaemi, “The rise and fall of the biopsychosocial model,” *Br J Psychiatry*, vol. 195, pp. 3–4, 2009. [Online]. Available: https://doi.org/10.1192/bjp.bp.109.063859

[11] K. Kendler, “The Rise and Fall of the Biopsychosocial Model: Reconciling Art and Science in Psychiatry,” *American Journal of Psychiatry*, vol. 167, pp. 999–1000, 2010.

[12] N. McLaren, “A Critical Review of the Biopsychosocial Model,” *Australian and New Zealand Journal of Psychiatry*, vol. 32, pp. 86 – 92, 1998. [Online]. Available: https://doi.org/10.3109/00048679809062712

[13] T. B. Benning, “Limitations of the biopsychosocial model in psychiatry,” *Advances in medical education and practice*, vol. 6, pp. 347–352, 2015. [Online]. Available: https://doi.org/10.2147/AMEP.S82937

[14] H. Thimbleby, “Technology and the Future of Healthcare,” *Journal of Public Health Research*, vol. 2, 2013. [Online]. Available: https://doi.org/10.4081/jphr.2013.e28

[15] J. A. Figueiredo and V. Eiriz, “Analysis of the impact of technological innovation on healthcare services,” *International Journal of Behavioural and Healthcare Research*, vol. 1, p. 234, 2009. [Online]. Available: https://doi.org/10.1504/IJBHR.2009.028458

[16] N. Rajkumari, “Impact of Technology in Better Dispersal of Health Care Methods and Improving Health Care Systems,” *Occupational medicine and health affairs*, vol. 2014, pp. 1–3, 2014. [Online]. Available: https://doi.org/10.4172/2329-6879.1000162

[17] K. Gray and C. Gilbert, “Digital health research methods and tools: Suggestions and selected resources for researchers,” 2018. [Online]. Available: https://doi.org/10.1007/978-3-319-67513-8_2

[18] A. Ahmadvand, R. Gatchel, J. Brownstein, and L. Nissen, “The Biopsychosocial-Digital Approach to Health and Disease: Call for a Paradigm Expansion,” *Journal of Medical Internet Research*, vol. 20, 2018. [Online]. Available: https://doi.org/10.2196/jmir.9732

[19] D. Lupton, “Digital health now and in the future: Findings from a participatory design stakeholder workshop,” *Digital health*, vol. 3, 2017.

[20] M. Mitchell and L. Kan, “Digital Technology and the Future of Health Systems,” *Health Systems & Reform*, vol. 5, pp. 113 – 120, 2019. [Online]. Available: https://doi.org/10.1080/23288604.2019.1583040

[21] T. L. D. Health, “Reflecting on a future ready for digital health,” *The Lancet. Digital Health*, vol. 2, pp. e209 – e209, 2020. [Online]. Available: https://doi.org/10.1016/s2589-7500(20)30087-x

[22] C. Almario, “The Effect of Digital Health Technology on Patient Care and Research,” *Gastroenterology & hepatology*, vol. 13 7, pp. 437–439, 2017.

[23] I. Clay, “The Future of Digital Health,” *Digital Biomarkers*, vol. 4, pp. 1 – 2, 2020. [Online]. Available: https://doi.org/10.1159/000511705

[24] J. Greiwe and S. Nyenhuis, “Wearable Technology and How This Can Be Implemented into Clinical Practice,” *Current Allergy and Asthma Reports*, vol. 20, 2020. [Online]. Available: https://doi.org/10.1007/s11882-020-00927-3

[25] S. Bhavnani, J. Narula, and P. Sengupta, “Mobile technology and the digitization of healthcare,” *European heart journal*, vol. 37 18, pp. 1428–38, 2016. [Online]. Available: https://doi.org/10.1093/eurheartj/ehv770

[26] T. Davenport and R. Kalakota, “The potential for artificial intelligence in healthcare,” *Future Healthcare Journal*, vol. 6, pp. 94–98, 2019. [Online]. Available: https://doi.org/10.7861/futurehosp.6-2-94

[27] R. El-Kareh, O. Hasan, and G. Schiff, “Use of health information technology to reduce diagnostic errors,” *BMJ Quality & Safety*, vol. 22, pp. ii40 – ii51, 2013. [Online]. Available: https://doi.org/10.1136/bmjqs-2013-001884

[28] M. Georgeff, “Digital technologies and chronic disease management,” *Australian family physician*, vol. 43 12, pp. 842–6, 2014.

[29] D. Hudson and M. Cohen, “The role of information technology in disease management,” *4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, 2003.*, pp. 62–65, 2003. [Online]. Available: https://doi.org/10.1109/ITAB.2003.1222561

[30] D. Weatherall, B. Greenwood, H. Chee, and P. Wasi, “Science and Technology for Disease Control: Past, Present, and Future,” 2006.

[31] G. Demiris, L. Afrin, S. Speedie, K. Courtney, M. Sondhi, V. Vimarlund, C. Lovis, W. Goossen, and C. Lynch, “White Paper: Patient-centered Applications: Use of Information Technology to Promote Disease Management and Wellness. A White Paper by the AMIA Knowledge in Motion Working Group,” *Journal of the American Medical Informatics Association : JAMIA*, vol. 15 1, pp. 8–13, 2008. [Online]. Available: https://doi.org/10.1197/jamia.M2492

[32] B. Celler, N. Lovell, and J. Basilakis, “Using information technology to improve the management of chronic disease,” *Medical Journal of Australia*, vol. 179, 2003. [Online]. Available: https://doi.org/10.5694/j.1326-5377.2003.tb05529.x

[33] D. Lupton, “The digitally engaged patient: Self-monitoring and self-care in the digital health era,” *Social Theory & Health*, vol. 11, pp. 256–270, 2013. [Online]. Available: https://doi.org/10.1057/sth.2013.10

[34] K. Morton, L. Dennison, C. May, E. Murray, P. Little, R. McManus, and L. Yardley, “Using digital interventions for self-management of chronic physical health conditions: A meta-ethnography review of published studies,” *Patient education and counseling*, vol. 100 4, pp. 616–635, 2017. [Online]. Available: https://doi.org/10.1016/j.pec.2016.10.019

[35] B. Riegel, S. Dunbar, D. Fitzsimons, K. E. Freedland, C. Lee, S. Middleton, A. Stromberg, E. Vellone, D. E. Webber, and T. Jaarsma, “Self-care research: Where are we now? Where are we going?” *International journal of nursing studies*, p. 103402, 2019. [Online]. Available: https://doi.org/10.1016/j.ijnurstu.2019.103402

[36] P. Kostkova, “Grand Challenges in Digital Health,” *Frontiers in Public Health*, vol. 3, 2015. [Online]. Available: https://doi.org/10.3389/fpubh.2015.00134

[37] I. Chouvarda, N. Mountford, V. Trajkovik, T. Loncar-Turukalo, and T. Cusack, “Leveraging Interdisciplinary Education Toward Securing the Future of Connected Health Research in Europe: Qualitative Study,” *Journal of Medical Internet Research*, vol. 21, 2019. [Online]. Available: https://doi.org/10.2196/14020

[38] D. Ho, S. R. Quake, E. McCabe, W. Chng, E. Chow, X. Ding, B. Gelb, G. Ginsburg, J. Hassenstab, C.-M. Ho, W. Mobley, G. Nolan, S. Rosen,

P. Tan, Y. Yen, and A. Zarrinpar, “Enabling Technologies for Personalized and Precision Medicine,” *Trends in biotechnology*, 2020. [Online]. Available: https://doi.org/10.1016/j.tibtech.2019.12.021

[39] D. Hansen, M. Dinger, O. Hofmann, N. Thorne, and T. F. Boughtwood, “Preparing Australia for genomic medicine: data, computing and digital health,” *Medical Journal of Australia*, vol. 210, 2019. [Online]. Available: https://www.doi.org/10.5694/mja2.50032

[40] X. Li, J. Dunn, D. Salins, G. Zhou, W. Zhou, S. M. S.-F. Rose, D. Perelman, E. Colbert, R. Runge, S. Rego, R. Sonecha, S. Datta, T. McLaughlin, and M. Snyder, “Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health-Related Information,” *PLoS Biology*, vol. 15, 2017. [Online]. Available: https://doi.org/10.1371/journal.pbio.2001402

[41] S. Guendelman, A. Broderick, H. Mlo, A. Gemmill, and D. Lindeman, “Listening to Communities: Mixed-Method Study of the Engagement of Disadvantaged Mothers and Pregnant Women With Digital Health Technologies,” *J Med Internet Res*, vol. 19, p. e240, 2017. [Online]. Available: https://doi.org/10.2196/jmir.7736

[42] D. Lupton, “Towards critical digital health studies: Reflections on two decades of research in health and the way forward,” *Health:*, vol. 20, 2016. [Online]. Available: https://doi.org/10.1177/1363459315611940

[43] M. Ranney and N. Genes, “Social media and healthcare quality improvement: a nascent field,” *BMJ Quality & Safety*, vol. 25, pp. 389 – 391, 2015. [Online]. Available: https://doi.org/10.1136/bmjqs-2015-004827

[44] M. Loscalzo, K. Clark, and J. Holland, “Successful strategies for implementing biopsychosocial screening,” *Psycho-Oncology*, vol. 20, 2011. [Online]. Available: https://doi.org/10.1002/pon.1930

[45] M. Loscalzo, K. E. Clark, S. Pal, and W. Pirl, “Role of biopsychosocial screening in cancer care,” *Cancer journal*, vol. 19 5, pp. 414–20, 2013. [Online]. Available: https://doi.org/10.1097/PPO.0b013e3182a5bce2

[46] D. Hurley, T. Dusoir, S. McDonough, A. P. Moore, S. Linton, and G. D. Baxter, “Biopsychosocial screening questionnaire for patients with low back pain: preliminary report of utility in physiotherapy practice in Northern Ireland,”

The Clinical journal of pain, vol. 16 3, pp. 214–28, 2000. [Online]. Available: https://doi.org/10.1097/00002508-200009000-00007

[47] L. Kurtinaityte, "E-HEALTH – THE USAGE OF ICT DEVELOPING HEALTH CARE SYSTEM : MULTIPLE-CASE STUDY OF EUROPEAN COUTRIES DENMARK AND LITHUANIA," 2007.

[48] J. Eschler, L. Liu, L. M. Vizer, J. McClure, P. Lozano, W. Pratt, and J. Ralston, "Designing Asynchronous Communication Tools for Optimization of Patient-Clinician Coordination," *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2015, pp. 543–52, 2015.

[49] K. Boydell, M. Hodgins, A. Pignatiello, J. Teshima, H. Edwards, and D. Willis, "Using technology to deliver mental health services to children and youth: a scoping review," *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Academie canadienne de psychiatrie de l'enfant et de l'adolescent*, vol. 23 2, pp. 87–99, 2014.

[50] F. Iorfino, S. P. Cross, T. Davenport, J. Carpenter, E. Scott, S. Shiran, and I. Hickie, "A Digital Platform Designed for Youth Mental Health Services to Deliver Personalized and Measurement-Based Care," *Frontiers in Psychiatry*, vol. 10, 2019. [Online]. Available: https://doi.org/10.3389/fpsyt.2019.00595

[51] F. Keefe, "Behavioral Medicine: A Voyage to the Future," *Annals of Behavioral Medicine*, vol. 41, pp. 141–151, 2011. [Online]. Available: https://doi.org/10.1007/s12160-010-9239-8

[52] K. Ell, H. Oh, and S. Wu, "Integrating Biopsychosocial Intervention Research in a Changing Health Care Landscape," *Research on Social Work Practice*, vol. 26, pp. 28 – 34, 2016. [Online]. Available: https://doi.org/10.1177/1049731515579074

[53] D. Lupton, "Health promotion in the digital era: a critical commentary," *Health promotion international*, vol. 30 1, pp. 174–83, 2015. [Online]. Available: https://doi.org/10.1093/heapro/dau091

[54] S. Ouhbi, A. Idri, and J. L. Fernández-Alemán, "Standards-Based Sustainability Requirements for Healthcare Services in Smart Cities," 2018. [Online]. Available: https://doi.org/10.1007/978-3-319-76669-0_13

[55] M. Gray, 2009. [Online]. Available: https://www.libraryservices.nhs.uk/document_uploads/CKOs/Muir_Gray_CKO_1_Dec_2009.pdf

[56] L. A. Marsch, “Digital health data-driven approaches to understand human behavior,” *Neuropsychopharmacology*, vol. 25, pp. 191–196, 2020. [Online]. Available: https://doi.org/10.1038/s41386-020-0761-5

[57] R. Agrawal and S. Prabakaran, “Big data in digital healthcare: lessons learnt and recommendations for general practice,” *Heredity*, vol. 124, pp. 525 – 534, 2020. [Online]. Available: https://doi.org/10.1038/s41437-020-0303-2

[58] L. Marsch, “Digital health data-driven approaches to understand human behavior,” *Neuropsychopharmacology*, vol. 46, pp. 191 – 196, 2020. [Online]. Available: 10.1038/s41386-020-0761-5

[59] A. Trifan, M. Oliveira, and J. Oliveira, “Passive Sensing of Health Outcomes Through Smartphones: Systematic Review of Current Solutions and Possible Limitations,” *JMIR mHealth and uHealth*, vol. 7, 2019. [Online]. Available: https://doi.org/10.2196/12649

[60] D. Boyd and K. Crawford, “Critical Questions for Big Data,” *Information, Communication & Society*, vol. 15, pp. 662 – 679, 2012. [Online]. Available: https://doi.org/10.1080/1369118X.2012.678878

[61] A. Sharma, R. Harrington, M. McClellan, M. Turakhia, Z. J. Eapen, S. Steinhubl, J. R. Mault, M. D. Majmudar, L. Roessig, K. Chandross, E. Green, B. Patel, A. Hamer, J. Olgin, J. Rumsfeld, M. Roe, and E. Peterson, “Using Digital Health Technology to Better Generate Evidence and Deliver Evidence-Based Care,” *Journal of the American College of Cardiology*, vol. 71 23, pp. 2680–2690, 2018. [Online]. Available: https://doi.org/10.1016/j.jacc.2018.03.523

[62] M. Panth and A. Acharya, “The Unprecedented Role of Computers in Improvement and Transformation of Public Health: An Emerging Priority,” *Indian Journal of Community Medicine : Official Publication of Indian Association of Preventive & Social Medicine*, vol. 40, pp. 8–13, 2015. [Online]. Available: https://doi.org/10.4103/0970-0218.149262

[63] W. Stead and H. S. Lin, “Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions,” 2009.

[64] I. Ioniță and L. Ioniţă, “Applying Data Mining Techniques in Healthcare,” *Studies in Informatics and Control*, vol. 25, 2016. [Online]. Available: https://doi.org/10.24846/v25i3y201612

[65] A. Dhillon and A. Singh, “Machine Learning in Healthcare Data Analysis: A Survey,” *Journal of Biology and Today‘s World*, vol. 8, pp. 1–10, 2019. [Online]. Available: https://doi.org/10.15412/J.JBTW.01070206

[66] S. Vollmer, B. Mateen, G. Bohner, F. Király, R. Ghani, P. Jonsson, S. Cumbers, A. Jonas, K. S. L. McAllister, P. Myles, D. Grainger, M. Birse, R. Branson, K. Moons, G. Collins, J. Ioannidis, C. Holmes, and H. Hemingway, “Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness,” *BMJ*, vol. 368, 2020. [Online]. Available: https://doi.org/10.1136/bmj.l6927

[67] B. Marr, “What is the difference between data mining and machine learning?” [Online]. Available: https://bernardmarr.com/default.asp?contentID=1741

[68] A. Triantafyllidis, Andreas K. Tsanas, “Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature,” *Cancer J*, vol. 21, p. e12286, 2019. [Online]. Available: https://doi.org/10.2196/12286

[69] M. Santhanagopalan, M. Chetty, C. Foale, and B. Klein, “Towards machine learning approach for digital-health intervention program,” *Aust. J. Intell. Inf. Process. Syst.*, vol. 15, pp. 16–24, 2019.

[70] J. Wilkinson, K. Arnold, E. Murray, M. V. Smeden, K. Carr, R. Sippy, M. Kamps, A. Beam, S. Konigorski, C. Lippert, M. Gilthorpe, and P. W. G. Tennant, “Time to reality check the promises of machine learning-powered precision medicine,” *The Lancet. Digital health*, vol. 2 12, pp. e677–e680, 2020. [Online]. Available: https://doi.org/10.1016/S2589-7500(20)30200-4

[71] V. Kaldo, T. Haak, M. Buhrman, S. Alfonsson, H. Larsen, and G. Andersson, “Internet-Based Cognitive Behaviour Therapy for Tinnitus Patients Delivered in a Regular Clinical Setting: Outcome and Analysis of Treatment Dropout,” *Cognitive Behaviour Therapy*, vol. 42, pp. 146–158, 2013. [Online]. Available: https://doi.org/10.1002/jclp.10243

[72] N. Titov, B. Dear, L. Staples, J. Bennett-Levy, B. Klein, R. Rapee, C. Shann, D. Richards, G. Andersson, L. Ritterband, C. Purtell, G. Bezuidenhout, L. Johnston, and O. Nielssen, “Mindspot Clinic: An Accessible, Efficient, and Effective Online Treatment Service for Anxiety and Depression.” *Psychiatric services*, vol. 66 10, pp. 1043–50, 2015. [Online]. Available: https://doi.org/10.1176/appi.ps.201400477

[73] P. Bower, E. Kontopantelis, A. Sutton, T. Kendrick, D. Richards, S. Gilbody, S. Knowles, P. Cuijpers, G. Andersson, H. Christensen, B. Meyer, M. Huibers, F. Smit, A. van Straten, L. Warmerdam, M. Barkham, L. Bilich, K. Lovell, and E. Liu, "Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data," *The BMJ*, vol. 346, 2013. [Online]. Available: https://doi.org/10.1136/bmj.f540

[74] M. Walker, S. Thomas, P. Whitehead, L. A. Condon, R. Fisher, E. Kontou, P. Benford, and C. Cobley, "Biopsychosocial Intervention for Stroke carers (BISC): protocol for a feasibility randomised controlled trial (RCT)," *BMJ Open*, vol. 7, 2017. [Online]. Available: https://doi.org/10.1177/0269215520937039

[75] H. Mason and C. Wiggins, "A Taxonomy of Data Science," 2010. [Online]. Available: https://doi.org/10.1155/2018/2964816

[76] B. Lehman, D. David, and J. A. Gruber, "Rethinking the biopsychosocial model of health: Understanding health as a dynamic system," *Social and Personality Psychology Compass*, vol. 11, 2017. [Online]. Available: https://doi.org/10.1111/SPC3.12328

[77] R. Gatchel, Y. Peng, M. Peters, P. Fuchs, and D. Turk, "The biopsychosocial approach to chronic pain: scientific advances and future directions," *Psychological bulletin*, vol. 133 4, pp. 581–624, 2007. [Online]. Available: https://doi.org/10.1037/0033-2909.133.4.581

[78] A. Tripathi, A. Das, and S. K. Kar, "Biopsychosocial Model in Contemporary Psychiatry: Current Validity and Future Prospects," *Indian Journal of Psychological Medicine*, vol. 41, pp. 582 – 585, 2019. [Online]. Available: https://doi.org/10.4103/IJPSYM.IJPSYM_314_19

[79] E. Babalola, P. Noel, and R. White, "The biopsychosocial approach and global mental health: Synergies and opportunities," *Indian Journal of Social Psychiatry*, vol. 33, pp. 291 – 296, 2017. [Online]. Available: https://doi.org/10.4103/ijsp.ijsp_13_17

[80] R. C. Smith, "Making the biopsychosocial model more scientific-its general and specific models," *Social science & medicine*, p. 113568, 2020. [Online]. Available: https://doi.org/10.1016/j.socscimed.2020.113568

[81] N. B. Anderson, “Levels of analysis in health science. A framework for integrating sociobehavioral and biomedical research,” *Ann N Y Acad Sci*, vol. 840, pp. 563–76, 1998. [Online]. Available: https://doi.org/10.1111/j.1749-6632.1998.tb09595.x

[82] R. M. Kaplan, “Behavior as the central outcome in health care,” *American Psychologist*, vol. 45, pp. 1211–1220, 1990. [Online]. Available: https://doi.org/10.1037/0003-066X.45.11.1211

[83] G. Schwartz, “Testing the biopsychosocial model: The ultimate challenge facing behavioral medicine?” *Journal of Consulting and Clinical Psychology*, vol. 50, no. 6, pp. 1040–1053, dec 1982. [Online]. Available: https://doi.org/10.1037/0022-006X.50.6.1040

[84] J. D. Matarazzo, “Behavioral health and behavioral medicine: frontiers for a new health psychology,” *Am Psychol*, vol. 35, pp. 807–17, 1980. [Online]. Available: https://doi.org/10.1037//0003-066x.35.9.807

[85] J. Suls and A. Rothman, “Evolution of the Biopsychosocial Model: Prospects and Challenges for Health Psychology,” *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, vol. 23, pp. 119–25, 2004. [Online]. Available: https://doi.org/10.1037/0278-6133.23.2.119

[86] H. Kusnanto, D. Agustian, and D. Hilmanto, “Biopsychosocial model of illnesses in primary care: A hermeneutic literature review,” *Journal of family medicine and primary care*, vol. 7, pp. 497–500, 2018. [Online]. Available: https://doi.org/10.4103/jfmpc.jfmpc_145_17

[87] A. J. Wijma, C. P. van Wilgen, M. Meeus, and J. Nijs, “Clinical biopsychosocial physiotherapy assessment of patients with chronic pain: The first step in pain neuroscience education,” *Physiother Theory Pract*, vol. 32, pp. 368–84, 2016. [Online]. Available: https://doi.org/10.1080/09593985.2016.1194651

[88] B. Andersen, “Biobehavioral outcomes following psychological interventions for cancer patients,” *Journal of consulting and clinical psychology*, vol. 70 3, pp. 590–610, 2002. [Online]. Available: https://doi.org/10.1037//0022-006X.70.3.590

[89] M. Antoni, D. Cruess, S. Cruess, S. Lutgendorf, M. Kumar, G. Ironson, N. Klimas, M. Fletcher, and N. Schneiderman, “Cognitive-behavioral stress management intervention effects on anxiety, 24-hr urinary norepinephrine output, and T-cytotoxic/suppressor cells over time among symptomatic HIV-infected gay

men," *Journal of Consulting and Clinical Psychology*, vol. 68, pp. 31–45, 2000. [Online]. Available: https://doi.org/10.1037/0022-006X.68.1.31

[90] J. Blumenthal, A. Sherwood, E. Gullette, A. Georgiades, and D. Tweedy, "Biobehavioral approaches to the treatment of essential hypertension," *Journal of consulting and clinical psychology*, vol. 70 3, pp. 569–89, 2002. [Online]. Available: https://doi.org/10.1037/0022-006X.70.3.569

[91] B. Tong, C. Sung, and J. Sánchez, "Using the biopsychosocial model to predict sense of community for persons with serious mental illness," *Journal of Mental Health*, pp. 1–9, 2019. [Online]. Available: https://doi.org/10.1080/09638237.2019.1581330

[92] T. Smith and J. M. Ruiz, "Psychosocial influences on the development and course of coronary heart disease: current status and implications for research and practice," *Journal of consulting and clinical psychology*, vol. 70 3, pp. 548–68, 2002. [Online]. Available: https://doi.org/10.1037//0022-006x.70.3.548

[93] R. Niaura and D. Abrams, "Smoking cessation: progress, priorities, and prospectus," *Journal of consulting and clinical psychology*, vol. 70 3, pp. 494–509, 2002. [Online]. Available: https://doi.org/10.1037//0022-006x.70.3.494

[94] Q. Government, "Tailored Support - Recovery your way," 2019. [Online]. Available: https://www.worksafe.qld.gov.au/rehabilitation-and-return-to-work/recovering-from-injury-or-illness/working-together-to-support-recovery/tailored-support-recovery-your-way

[95] H. Kamal, R. Tiwari, J. Behera, and B. Hasan, "Personality Variables and Marital Satisfaction:A Systematic Review," *Indian Journal of Health and Wellbeing*, vol. 9, pp. 534–541, 2018.

[96] F. Zaheri, M. Dolatian, M. Shariati, M. Simbar, A. Ebadi, and S. B. H. Azghadi, "Effective Factors in Marital Satisfaction in Perspective of Iranian Women and Men: A systematic review," *Electronic Physician*, vol. 8, pp. 3369 – 3377, 2016. [Online]. Available: https://doi.org/10.19082/3369

[97] A. P. Adigeb and A. P. Mbua, "The Influence of Psychosocial Factors on Marital Satisfaction Among Public Servants in Cross River State," *Global Journal of Human-Social Science Research*, vol. 15, 2015.

[98] T. F. Robles, R. Slatcher, J. M. Trombello, and M. M. McGinn, "Marital quality and health: a meta-analytic review." *Psychological bulletin*, vol. 140 1, pp. 140–187, 2014.

[99] W. Boyce and B. Ellis, "Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity." *Development and psychopathology*, vol. 17 2, pp. 271–301, 2005.

[100] M. N. Silverman and P. Deuster, "Biological mechanisms underlying the role of physical fitness in health and resilience," *Interface Focus*, vol. 4, 2014.

[101] K. Hara, J. Bjørngaard, H. Jacobsen, P. Borchgrevink, R. Johnsen, T. Stiles, S. Brage, and A. Woodhouse, "Biopsychosocial predictors and trajectories of work participation after transdiagnostic occupational rehabilitation of participants with mental and somatic disorders: a cohort study," *BMC Public Health*, vol. 18, 2018. [Online]. Available: 10.1186/s12889-018-5803-0

[102] T. Farre, Albert Rapley, "The New Old (and Old New) Medical Model: Four Decades Navigating the Biomedical and Psychosocial Understandings of Health and Illness," *Healthcare (Basel, Switzerland)*, vol. 5, p. 88, 2017. [Online]. Available: https://doi.org/10.3390/healthcare5040088

[103] R. C. Smith, A. H. Fortin, F. Dwamena, and R. M. Frankel, "An evidence-based patient-centered method makes the biopsychosocial model scientific," *Patient education and counseling*, vol. 91, pp. 265–270, 2013. [Online]. Available: https://doi.org/10.1016/j.pec.2012.12.010

[104] O. Freudenreich, N. Kontos, and J. Querques, "The Muddles of Medicine: A Practical, Clinical Addendum to the Biopsychosocial Model," *Psychosomatics*, vol. 51, pp. 365–369, 2010. [Online]. Available: https://doi.org/10.1016/S0033-3182(10)70717-9

[105] D. Wade and P. Halligan, "The biopsychosocial model of illness: A model whose time has come," *Clinical Rehabilitation*, 2017. [Online]. Available: https://doi.org/10.1177/0269215517709890

[106] N. Karunamuni, I. Imayama, and D. Goonetilleke, "Pathways to well-being: Untangling the causal relationships among biopsychosocial variables," *Soc Sci Med*, p. 112846, 2020. [Online]. Available: https://doi.org/10.1016/j.socscimed.2020.112846

[107] S. Hewitt, R. Sephton, and G. Yeowell, "The Effectiveness of Digital Health Interventions in the Management of Musculoskeletal Conditions: Systematic Literature review," *Journal of Medical Internet Research*, vol. 22, 2020. [Online]. Available: https://doi.org/10.2196/15617

[108] A. Ahmadvand, R. Gatchel, J. Brownstein, and L. Nissen, "The Biopsychosocial-Digital Approach to Health and Disease: Call for a Paradigm Expansion," *Journal of medical Internet research*, vol. 20, pp. e189–e189, 2018. [Online]. Available: https://doi.org/10.2196/jmir.9732

[109] A. Thiel, G. Sudeck, H. Gropper, F. M. Maturana, T. Schubert, D. Srismith, M. Widmann, S. Behrens, P. Martus, B. Munz, K. Giel, S. Zipfel, and A. M. Nieß, "The iReAct study – A biopsychosocial analysis of the individual response to physical activity," *Contemporary Clinical Trials Communications*, vol. 17, p. 100508, 2020. [Online]. Available: https://doi.org/10.1016/j.conctc.2019.100508

[110] C. C. Coelhoso, P. R. Tobo, S. S. Lacerda, A. H. Lima, C. R. C. Barrichello, E. Amaro Jr, and E. H. Kozasa, "A New Mental Health Mobile App for Well-Being and Stress Reduction in Working Women: Randomized Controlled Trial," *J Med Internet Res*, vol. 21, p. e14269, 2019. [Online]. Available: https://doi.org/10.2196/14269

[111] A. T. Reyes, "A Mindfulness Mobile App for Traumatized covid-19 Healthcare Workers and Recovered Patients: A response to "the use of digital Applications and COVID-19," *Community Mental Health Journal*, vol. 56, pp. 1204–1205, 2020. [Online]. Available: https://doi.org/10.1007/s10597-020-00690-9

[112] M. Wilson, J. Doyle, and G. McTaggart, "Designing Digital Mindfulness-based Interventions for Older Informal Carers," 2018. [Online]. Available: https://doi.org/10.14236/ewic/HCI2018.52

[113] B. Meskó, Z. Drobni, É. Bényei, B. Gergely, and Z. Győrffy, "Digital health is a cultural transformation of traditional healthcare," *mHealth*, vol. 3, p. 38, 2017. [Online]. Available: https://doi.org/10.21037/mhealth.2017.08.07

[114] P. V. Klasnja and W. Pratt, "Healthcare in the pocket: Mapping the space of mobile-phone health interventions," *Journal of biomedical informatics*, vol. 45 1, pp. 184–98, 2012. [Online]. Available: https://doi.org/10.1016/j.jbi.2011.08.017

[115] A. Papa, M. Mital, P. Pisano, and M. Giudice, "E-health and wellbeing monitoring using smart healthcare devices: An empirical investigation," *Technological Forecasting and Social Change*, vol. 153, p. 119226, 2020. [Online]. Available: https://doi.org/10.1016/J.TECHFORE.2018.02.018

[116] A. Kazi, S. A. Qazi, N. Ahsan, S. Khawaja, F. Sameen, M. Saqib, M. A. K. Mughal, Z. Wajidali, S. Ali, R. M. Ahmed, H. Kalimuddin, Y. Rauf, F. Mahmood, S. Zafar, T. A. Abbasi, K.-U.-R. Khoumbati, M. Abbasi, and L. K. Stergioulas, "Current Challenges of Digital Health Interventions in Pakistan: Mixed Methods Analysis," *Journal of Medical Internet Research*, vol. 22, 2020. [Online]. Available: https://doi.org/10.2196/21691

[117] F. Fatehi, M. Samadbeik, and A. Kazemi, "What is Digital Health? Review of Definitions," *Studies in health technology and informatics*, vol. 275, pp. 67–71, 2020. [Online]. Available: https://doi.org/10.3233/SHTI200696

[118] T. Kowatsch, L. Otto, S. Harperink, A. Cotti, and H. Schlieter, "A design and evaluation framework for digital health interventions," *IT - Information Technology*, vol. 61, pp. 253–263, 2019. [Online]. Available: https://doi.org/10.1515/itit-2019-0019

[119] M. Orton, S. Agarwal, P. Muhoza, L. Vasudevan, and A. Vu, "Strengthening Delivery of Health Services Using Digital Devices," *Global Health: Science and Practice*, vol. 6, pp. S61 – S71, 2018. [Online]. Available: https://doi.org/10.9745/GHSP-D-18-00229

[120] G. Alkhaldi, K. Modrow, F. Hamilton, K. Pal, J. Ross, and E. Murray, "Promoting Engagement With a Digital Health Intervention (HeLP-Diabetes) Using Email and Text Message Prompts: Mixed-Methods Study," *Interactive journal of medical research*, 2017. [Online]. Available: https://doi.org/10.2196/ijmr.6952

[121] E. Murray, E. Hekler, G. Andersson, L. Collins, A. Doherty, C. Hollis, D. Rivera, R. West, and J. Wyatt, "Evaluating Digital Health Interventions: Key Questions and Approaches," *Am J Prev Med*, 2016. [Online]. Available: https://doi.org/10.1016/j.amepre.2016.06.008

[122] J. Schneider, P. Malinowski, P. Watson, and L. P., "The role of mindfulness in physical activity: a systematic review," *Obesity Reviews*, 2019. [Online]. Available: https://doi.org/10.1111/obr.12795

[123] C. Parsons, K. Jensen, A. Roepstorff, L. Fjorback, and C. Linehan, "Designing technology tools to support engagement in mindfulness-based interventions: an analysis of teacher and student experiences," *Digital Health*, vol. 5, 2019. [Online]. Available: https://doi.org/10.1177/2055207619868550

[124] M. Wilson, J. Doyle, and G. McTaggart, "Designing Digital Mindfulness-based Interventions for Older Informal Carers," 2018. [Online]. Available: https://doi.org/10.14236/ewic/hci2018.52

[125] R. Mitchell and K. Pallaris, "Encouraging New Encounters: Digital Design Patterns to Support Social Wellbeing," 2018.

[126] S. Kamper, A. Apeldoorn, A. Chiarotto, R. Smeets, R. Ostelo, J. Guzmán, and M. V. van Tulder, "Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis," *The BMJ*, vol. 350, 2015. [Online]. Available: https://doi.org/10.1136/bmj.h444

[127] K. M. Carpenter, S. Stoner, J. Mundt, and B. L. Stoelb, "An Online Self-help CBT Intervention for Chronic Lower Back Pain," *The Clinical Journal of Pain*, vol. 28, p. 14–22, 2012. [Online]. Available: https://doi.org/10.1097/AJP.0b013e31822363db

[128] H. Chhabra, S. K. Sharma, and S. Verma, "Smartphone app in self-management of chronic low back pain: a randomized controlled trial," *European Spine Journal*, vol. 27, pp. 2862–2874, 2018. [Online]. Available: https://doi.org/10.1007/s00586-018-5788-5

[129] A. B. Irvine, H. Russell, M. Manocchia, D. E. Mino, T. C. Glassen, R. Morgan, J. Gau, A. J. Birney, and D. Ary, "Mobile-Web App to Self-Manage Low Back Pain: Randomized Controlled Trial," *Journal of Medical Internet Research*, vol. 17, 2015. [Online]. Available: https://doi.org/10.2196/jmir.3130

[130] D. Falzon, M. Raviglione, E. Bel, C. Gratziou, D. Bettcher, and G. Migliori, "The role of eHealth and mHealth in tuberculosis and tobacco control: a WHO/ERS consultation," *European Respiratory Journal*, vol. 46, pp. 307 – 311, 2015. [Online]. Available: https://doi.org/10.1183/09031936.00043315

[131] M. Loeckx, R. Rabinovich, H. Demeyer, Z. Louvaris, R. Tanner, N. Rubio, A. Frei, C. D. Jong, E. Gimeno-Santos, F. M. Rodrigues, S. Buttery, N. Hopkinson, G. Büsching, A. Strassmann, I. Serra, I. Vogiatzis, J. Garcia-Aymerich, M. I.

Polkey, and T. Troosters, “Smartphone-Based Physical Activity Telecoaching in Chronic Obstructive Pulmonary Disease: Mixed-Methods Study on Patient Experiences and Lessons for Implementation,” *JMIR mHealth and uHealth*, vol. 6, 2018. [Online]. Available: https://doi.org/10.2196/mhealth.9774

[132] C. Y. Hui, R. Walton, B. McKinstry, T. Jackson, R. Parker, and H. Pinnock, “The use of mobile applications to support self-management for people with asthma: a systematic review of controlled studies to identify features associated with clinical effectiveness and adherence,” *Journal of the American Medical Informatics Association*, vol. 24, p. 619–632, 2017. [Online]. Available: https://doi.org/10.1093/jamia/ocw143

[133] V. Poberezhets, H. Pinnock, I. Vogiatzis, and V. Mishlanov, “Implementation of digital health interventions in respiratory medicine: a call to action by the European Respiratory Society m-Health/e-Health Group,” *ERJ Open Research*, vol. 6, 2020. [Online]. Available: https://doi.org/10.1183/23120541.00281-2019

[134] C. Guo, H. Ashrafian, S. Ghafur, G. Fontana, C. Gardner, and M. Prime, “Challenges for the evaluation of digital health solutions—A call for innovative evidence generation approaches,” *NPJ Digital Medicine*, vol. 3, 2020. [Online]. Available: https://doi.org/10.1038/s41746-020-00314-2

[135] L. Morrison, “Theory-based strategies for enhancing the impact and usage of digital health behaviour change interventions: A review,” *Digital health*, vol. 1, 2015. [Online]. Available: https://doi.org/10.1177/2055207615595335

[136] S. Garrido, C. Millington, D. Cheers, K. Boydell, E. Schubert, T. Meade, and Q. Nguyen, “What Works and What Doesn’t Work? A Systematic Review of Digital Mental Health Interventions for Depression and Anxiety in Young People,” *Frontiers in Psychiatry*, vol. 10, 2019. [Online]. Available: https://doi.org/10.3389/fpsyt.2019.00759

[137] S. Agarwal, A. Labrique, A. Lefevre, G. Mehl, T. Tamrat, L. Vasudevan, and K. Zeller, “Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment,” 2016.

[138] WHO, “WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening,” vol. 61, p. 124, 2019. [Online]. Available: https://www.who.int/reproductivehealth/publications/digital-interventions-health-system-strengthening/en/

[139] D. Arigo, D. E. Jake-Schoffman, K. Wolin, E. Beckjord, E. B. Hekler, and S. L. Pagoto, "The history and future of digital health in the field of behavioral medicine," *Journal of behavioral medicine*, vol. 42, pp. 67–83, 2019. [Online]. Available: https://doi.org/10.1007/s10865-018-9966-z

[140] C. Scharff and R. Verma, "Scrum to support mobile application development projects in a just-in-time learning context," in *CHASE '10*, 2010. [Online]. Available: https://doi.org/10.1145/1833310.1833315

[141] D. Dhingra and A. Dabas, "Global Strategy on Digital Health," *Indian Pediatrics*, vol. 57, pp. 356–358, 2020. [Online]. Available: https://doi.org/10.1007/s13312-020-1789-7

[142] Q. Pham, J. Shaw, P. P. Morita, E. Seto, J. N. Stinson, and J. A. Cafazzo, "The Service of Research Analytics to Optimize Digital Health Evidence Generation: Multilevel Case Study," *J Med Internet Res*, vol. 21, p. e14849, 2019. [Online]. Available: https://doi.org/10.2196/14849

[143] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," *Health Aff (Millwood)*, vol. 33, pp. 1123–31, 2014. [Online]. Available: https://doi.org/10.1377/hlthaff.2014.0041

[144] P. Amirian, F. Loggerenberg, and T. Lang, "Data Science and Analytics," 2017. [Online]. Available: https://doi.org/10.1007/978-3-319-62990-2_2

[145] V. Jain and J. M. Chaterjee, *Machine Learning with Health Care Perspective*. Cham: Springer, 2020.

[146] Q. Pham, G. Graham, C. Carrion, P. P. Morita, E. Seto, J. N. Stinson, and J. A. Cafazzo, "A Library of Analytic Indicators to Evaluate Effective Engagement with Consumer mHealth Apps for Chronic Conditions: Scoping Review," *JMIR Mhealth Uhealth*, vol. 7, p. e11941, 2019. [Online]. Available: https://doi.org/10.2196/11941

[147] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining," *Journal of Data Warehousing*, vol. 5, no. 4, 2000. [Online]. Available: https://www.bibsonomy.org/bibtex/28efa10127187a460bc800ed4a055d028/kdubiq

[148] R. Wirth, “CRISP-DM: Towards a standard process model for data mining,” in *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 2000, pp. 29–39.

[149] A. Dåderman and S. Rosander, “Evaluating Frameworks for Implementing Machine Learning in Signal Processing : A Comparative Study of CRISP-DM, SEMMA and KDD,” 2018.

[150] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. R. Shearer, and R. Wirth, “CRISP-DM 1.0: Step-by-step data mining guide,” 2000.

[151] Z. Bosnjak, O. Grljevic, and S. Bosnjak, “CRISP-DM as a framework for discovering knowledge in small and medium sized enterprises’ data,” in *2009 5th International Symposium on Applied Computational Intelligence and Informatics*, 2009, pp. 509–514. [Online]. Available: https://doi.org/10.1109/SACI.2009.5136302

[152] C. Ponsard, M. Touzani, and A. Majchrowski, “Combining Process Guidance and industrial Feedback for Successfully Deploying Big Data Projects,” *Open Journal of Big Data (OJBD)*, vol. 3, pp. 26–41, 2017. [Online]. Available: https://doi.org/10.13140/RG.2.2.23182.41285

[153] S. Institute, U. Sas, and E. M. Software, “Data Mining and the Case for Sampling - A SAS Institute Best practices paper Solving Business problems Using SAS® Enterprise Miner™ Software.”

[154] H. J. G. Palacios, R. A. Toledo, G. A. H. Pantoja, and Ã. Navarro, “A comparative between crisp-dm and semma through the construction of a modis repository for studies of land use and cover change,” *Advances in Science, Technology and Engineering Systems Journal*, vol. 2, pp. 598–604, 2017.

[155] A. A. do Nascimento, Givanildo Santana de Oliveira, “An Agile Knowledge Discovery in Databases Software Process,” vol. 7696, pp. 56–64, 2012. [Online]. Available: https://doi.org/10.1007/978-3-642-34679-8_6

[156] P. Franková, M. Drahošová, and P. Balco, “Agile Project Management Approach and its Use in Big Data Management,” *Procedia Computer Science*, vol. 83, pp. 576–583, 2013. [Online]. Available: https://doi.org/10.1016/j.procs.2016.04.272

[157] F. Foroughi and P. Luksch, “Data Science Methodology for Cybersecurity Projects,” *ArXiv*, 2018. [Online]. Available: https://doi.org/10.5121/CSIT.2018.80401

[158] C. Byrne, *Development workflows for Data Scientists*. O'Reilly, 2017.

[159] T. Dineva, Kristina Atanasova, "OSEMN process for working over data acquired by IOT devices mounted in beehives," vol. 7, 2018.

[160] S. Guhr, J.-H. Martenson, H. Laser, J. Gless, D. Amendt, B. Schantze, S. Gerbel, and B. Jandric, "Data Science as a Service - Prototyping for an Enterprise Self-Service Platform for Reproducible Research," 2019.

[161] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent advances and applications of machine learning in solid-state materials science," *npj Computational Materials*, vol. 5, no. 1, p. 83, 2019. [Online]. Available: https://doi.org/10.1038/s41524-019-0221-0

[162] F. Fabris, J. P. Magalhães, and A. Freitas, "A review of supervised machine learning applied to ageing research," *Biogerontology*, vol. 18, pp. 171 – 188, 2017. [Online]. Available: https://doi.org/10.1007/s10522-017-9683-y

[163] J. Dougherty, R. Kohavi, and M. Sahami, "Supervised and Unsupervised Discretization of Continuous Features," in *Proceedings of the Twelfth International Conference on International Conference on Machine Learning*, ser. ICML'95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995, p. 194–202. [Online]. Available: https://doi.org/10.1016/b978-1-55860-377-6.50032-3

[164] J. E. Dobson, "Can An Algorithm Be Disturbed?: Machine Learning, Intrinsic Criticism, and the Digital Humanities," *College Literature*, vol. 42, pp. 543–564, 2015. [Online]. Available: https://doi.org/10.1353/LIT.2015.0037

[165] N. S. Downing, A. Cloninger, A. Venkatesh, A. Hsieh, E. Drye, R. Coifman, and H. Krumholz, "Describing the performance of U.S. hospitals by applying big data analytics," *PLoS ONE*, vol. 12, 2017. [Online]. Available: https://doi.org/10.1371/journal.pone.0179603

[166] Q. C. Hoang, Xuan Dau Nguyen, "Botnet Detection Based On Machine Learning Techniques Using DNS Query Data," *Future Internet*, vol. 10, no. 5, p. 43, 2018. [Online]. Available: https://doi.org/10.3390/fi10050043

[167] U. C. Kothari and M. Momayez, "Machine Learning: A Novel Approach to Predicting Slope Instabilities," *International Journal of Geophysics*, vol. 2018, p. 4861254, 2018. [Online]. Available: https://doi.org/10.1155/2018/4861254

[168] E. Alpaydin, *Introduction to Machine Learning*, 3rd ed., 2014.

[169] T. Hofmann, "Unsupervised Learning by Probabilistic Latent Semantic Analysis," *Machine Learning*, vol. 42, no. 1, pp. 177–196, 2001. [Online]. Available: https://doi.org/10.1023/A:1007617005950

[170] G. Marshall, "Statistical methods in the atmospheric sciences," *Meteorological Applications*, vol. 14, no. 2, pp. 205–205, 2007. [Online]. Available: https://doi.org/10.1002/met.16

[171] S. Mani, Y. Chen, X. Li, L. Arlinghaus, A. B. Chakravarthy, V. Abramson, S. R. Bhave, M. A. Levy, H. Xu, and T. E. Yankeelov, "Machine learning for predicting the response of breast cancer to neoadjuvant chemotherapys," *Journal of the American Medical Informatics Association : JAMIA*, vol. 20, no. 4, pp. 688–695, 2013. [Online]. Available: https://doi.org/10.1136/amiajnl-2012-001332

[172] C. Xia, L. Fu, Z. Liu, H. Liu, L. Chen, and Y. Liu, "Aquatic Toxic Analysis by Monitoring Fish Behavior Using Computer Vision: A Recent Progress," *Journal of Toxicology*, 2018. [Online]. Available: https://doi.org/10.1155/2018/2591924

[173] X. Yan, Y. Bai, S. Fang, and J. an Luo, "A kernel-free quadratic surface support vector machine for semi-supervised learning," *Journal of the Operational Research Society*, vol. 67, pp. 1001–1011, 2016. [Online]. Available: https://doi.org/10.1057/jors.2015.89

[174] G. Huang, S. Song, J. N. D. Gupta, and C. Wu, "Semi-Supervised and Unsupervised Extreme Learning Machines," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2405–2417, 2014. [Online]. Available: https://doi.org/10.1109/TCYB.2014.2307349

[175] R. Rastogi and P. Saigal, "Tree-Based Localized Fuzzy Twin Support Vector Clustering with Square Loss Function," *Applied Intelligence*, vol. 47, no. 1, p. 96–113, Jul. 2017. [Online]. Available: https://doi.org/10.1007/s10489-016-0886-8

[176] K. Torshin, Ivan Rudakov, "On the theoretical basis of metric analysis of poorly formalized problems of recognition and classification," *Pattern Recognition and Image Analysis*, vol. 25, pp. 577–587, 2015. [Online]. Available: https://doi.org/10.1134/S1054661815040252

[177] A. Muscoloni, J. M. Thomas, S. Ciucci, G. Bianconi, and C. V. Cannistraci, "Machine learning meets complex networks via coalescent embedding in the hyperbolic space," *Nature communications*, vol. 8, no. 1, p. 1615, 2017. [Online]. Available: https://doi.org/10.1038/s41467-017-01825-5

[178] A. Gosavi, "Reinforcement Learning: A Tutorial Survey and Recent Advances," *INFORMS Journal on Computing*, vol. 21, no. 2, pp. 178–192, 2009. [Online]. Available: https://doi.org/10.1287/ijoc.1080.0305

[179] N. Justesen, P. Bontrager, J. Togelius, and S. Risi, "Deep Learning for Video Game Playing," *IEEE Transactions on Games*, vol. 12, pp. 1–20, 2020. [Online]. Available: https://doi.org/10.1109/TG.2019.2896986

[180] E. Perot, M. Jaritz, M. Toromanoff, and R. de Charette, "End-to-End Driving in a Realistic Racing Game with Deep Reinforcement Learning," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 474–475, 2017. [Online]. Available: https://doi.org/10.1109/CVPRW.2017.64

[181] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034, 2015. [Online]. Available: https://doi.org/10.1109/ICCV.2015.123

[182] Y.-t. Liu, Shuai-shi Tian, "Facial Expression Recognition Method Based on Gabor Wavelet Features and Fractional Power Polynomial Kernel PCA," pp. 144–151, 2010. [Online]. Available: https://doi.org/10.1007/978-3-642-13318-3_19

[183] M. Bojarski, D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to End Learning for Self-Driving Cars," *ArXiv*, 2016. [Online]. Available: https://arxiv.org/abs/1604.07316

[184] W. M. Guzella, Thiago Caminhas, "A review of machine learning approaches to spam filtering," *Expert Systems with Applications*, vol. 36, pp. 10 206–10 222, 2001. [Online]. Available: https://doi.org/10.1016/j.eswa.2009.02.037

[185] P. Chan and S. Stolfo, "Toward Scalable Learning with Non-Uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection," in *KDD*, 1998.

[186] G. Kesavaraj and S. Sukumaran, "A study on classification techniques in data mining," *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pp. 1–7, 2013. [Online]. Available: https://doi.org/10.1109/ICCCNT.2013.6726842

[187] Z. Zia, S. K. Tipu, and M. I. Khan, "Research on Association Rule Mining," 2012.

[188] I. Witten and E. Frank, "Data mining: practical machine learning tools and techniques with Java implementations," in *SGMD*, 2002.

[189] S. Chakrabarti, E. Cox, E. Frank, R. H. Güting, J. Han, X. Jiang, M. Kamber, S. Lightstone, T. P. Nadeau, R. Neapolitan, D. Pyle, M. Refaat, M. Schneider, T. Teorey, and I. Witten, "Data Mining - Know It All," 2008.

[190] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health Information Science and Systems*, vol. 2, 2014. [Online]. Available: https://doi.org/10.1186/2047-2501-2-3

[191] D. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," *Health affairs*, vol. 33 7, pp. 1123–31, 2014. [Online]. Available: https://doi.org/10.1377/hlthaff.2014.0041

[192] E. A. Mohammed, B. Far, and C. Naugler, "Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends," *BioData Mining*, vol. 7, pp. 22 – 22, 2014. [Online]. Available: https://doi.org/10.1186/1756-0381-7-22

[193] Y. Wang and N. Hajli, "Exploring the path to big data analytics success in healthcare," *Journal of Business Research*, vol. 70, pp. 287–299, 2017. [Online]. Available: https://doi.org/10.1016/J.JBUSRES.2016.08.002

[194] E. Commission, "Communication on Enabling the Digital Transformation of Health and Care in the Digital Single Market; Empowering Citizens and Building a Healthier Society," 2020. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/

[195] K. Kolasa and G. Koziński, "How to Value Digital Health Interventions? A Systematic Literature Review," *International Journal of Environmental Research and Public Health*, vol. 17, 2020. [Online]. Available: https://doi.org/10.3390/ijerph17062119

[196] E. E. Schadt, "Molecular networks as sensors and drivers of common human diseases," *Nature*, vol. 461, no. 7261, pp. 218–23, 2010. [Online]. Available: https://doi.org/10.1038/nature08454

[197] M. C. McClellan, J. King, "Genetic heterogeneity in human disease," *Cell*, vol. 141, no. 2, pp. 210–7, 2010. [Online]. Available: https://doi.org/10.1016/j.cell.2010.03.032

[198] T. Hastie, R. Tibshirani, and J. Friedman, *Linear Methods for Regression*. Springer New York, 2009. [Online]. Available: https://doi.org/10.1007/978-0-387-84858-7_3

[199] E. Rietveld, H. C. C. de Jonge, J. Polder, Y. Vergouwe, H. Veeze, H. Moll, and E. Steyerberg, "Anticipated Costs of Hospitalization for Respiratory Syncytial Virus Infection in Young Children at Risk," *The Pediatric Infectious Disease Journal*, vol. 23, pp. 523–529, 2004. [Online]. Available: https://doi.org/10.1097/01.inf.0000129690.35341.8d

[200] M. A. Cucciare and W. O'donohue, "Predicting future healthcare costs: how well does risk-adjustment work?" *Journal of health organization and management*, vol. 20 2-3, pp. 150–62, 2006. [Online]. Available: https://doi.org/10.1108/14777260610661547

[201] P. Krijnen, B. C. Jaarsveld, M. Hunink, and J. Habbema, "The effect of treatment on health-related quality of life in patients with hypertension and renal artery stenosis," *Journal of Human Hypertension*, vol. 19, pp. 467–470, 2005. [Online]. Available: https://doi.org/10.1038/sj.jhh.1001847

[202] T. Hastie and R. Tibshirani, "Generalized Additive Models," 2014. [Online]. Available: https://doi.org/10.1002/9781118445112.stat03141

[203] C. Reddy and Y. Li, "A review of clinical prediction models," in *Healthcare Data Analytics*, 2015. [Online]. Available: https://doi.org/10.1201/b18588-13

[204] D. Pregibon, "Logistic Regression Diagnostics," *Annals of Statistics*, vol. 9, pp. 705–724, 1981. [Online]. Available: https://doi.org/10.1214/aos/1176345513

[205] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of biomedical informatics*, vol. 35 5-6, pp. 352–59, 2002. [Online]. Available: https://doi.org/10.1016/S1532-0464(03)00034-0

[206] D. Wingerchuk, V. Lennon, S. Pittock, C. Lucchinetti, and B. Weinshenker, "Revised diagnostic criteria for neuromyelitis optica," *Neurology*, vol. 66, pp. 1485–1489, 2006. [Online]. Available: https://doi.org/10.1212/01.wnl.0000216139.44259.74

[207] I. Kononenko, "Inductive and Bayesian learning in medical diagnosis," *Appl. Artif. Intell.*, vol. 7, pp. 317–337, 1993. [Online]. Available: https://doi.org/10.1080/08839519308949993

[208] M. Pepe, "The Statistical Evaluation of Medical Tests for Classification and Prediction," *Technometrics*, vol. 47, pp. 245 – 245, 2003. [Online]. Available: https://doi.org/10.1002/sim.2185

[209] M. J. Aspinall, "Use of a Decision Tree to Improve Accuracy of Diagnosis," *Nursing Research*, vol. 28, p. 182–185, 1979. [Online]. Available: https://doi.org/10.1097/00006199-197905000-00014

[210] J. Quinlan, "C4.5: Programs for Machine Learning," 1992.

[211] G. V. Kass, "An Exploratory Technique for Investigating Large Quantities of Categorical Data," *Journal of The Royal Statistical Society Series C-applied Statistics*, vol. 29, pp. 119–127, 1980. [Online]. Available: https://doi.org/10.2307/2986296

[212] L. Breiman, J. Friedman, R. Olshen, and C. J. Stone, "Classification and Regression Trees," 1983. [Online]. Available: https://doi.org/10.2307/2530946

[213] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123–140, 1996. [Online]. Available: https://doi.org/10.1023/A:1018054314350

[214] S. Gambhir, C. Hoh, M. Phelps, I. Madár, and J. Maddahi, "Decision tree sensitivity analysis for cost-effectiveness of FDG-PET in the staging and management of non-small-cell lung carcinoma," *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 37 9, pp. 1428–36, 1996. [Online]. Available: https://doi.org/10.1007/BF03000105

[215] W. Long, J. L. Griffith, H. Selker, and R. D'Agostino, "A comparison of logistic regression to decision-tree induction in a medical domain," *Computers and biomedical research, an international journal*, vol. 26 1, pp. 74–97, 1993. [Online]. Available: https://doi.org/10.1006/cbmr.1993.1005

[216] F. Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain,” *Psychological review*, vol. 65 6, pp. 386–408, 1958. [Online]. Available: https://doi.org/10.1037/h0042519

[217] L. Bottaci, P. Drew, J. Hartley, M. B. Hadfield, and J. Monson, “Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions,” *The Lancet*, vol. 350, pp. 469–472, 1997. [Online]. Available: https://doi.org/10.1016/S0140-6736(96)11196-X

[218] N. Ganesan, K. Venkatesh, M. A. Rama, and A. Palani, “Application of Neural Networks in Diagnosing Cancer Disease using Demographic Data,” *International Journal of Computer Applications*, vol. 1, pp. 81–97, 2010. [Online]. Available: https://doi.org/10.5120/476-783

[219] P. Lisboa and A. Taktak, “The use of artificial neural networks in decision support in cancer: A systematic review,” *Neural networks : the official journal of the International Neural Network Society*, vol. 19 4, pp. 408–15, 2006. [Online]. Available: https://doi.org/10.1016/j.neunet.2005.10.007

[220] P. Lisboa, “A review of evidence of health benefit from artificial neural networks in medical intervention,” *Neural networks : the official journal of the International Neural Network Society*, vol. 15 1, pp. 11–39, 2002. [Online]. Available: https://doi.org/0.1016/S0893-6080(01)00111-3

[221] M. Mazurowski, P. Habas, J. Zurada, J. Lo, J. Baker, and G. Tourassi, “Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance,” *Neural networks : the official journal of the International Neural Network Society*, vol. 21 2-3, pp. 427–36, 2008. [Online]. Available: https://doi.org/10.1016/j.neunet.2007.12.031

[222] X. Zhang, Lingsong Lin, “Some considerations of classification for high dimension low-sample size data,” *Statistical Methods in Medical Research*, vol. 2, no. 5, pp. 537–550, 2013. [Online]. Available: https://doi.org/10.1177/0962280211428387

[223] S. Bishop, M. Lau, S. Shapiro, L. Carlson, N. Anderson, J. Carmody, Z. Segal, S. Abbey, M. Speca, D. M. Velting, and G. Devins, “Mindfulness : A proposed operational definition,” *Clinical Psychology-science and Practice*, vol. 11, pp. 230–241, 2006. [Online]. Available: https://doi.org/10.1093/clipsy.bph077

[224] S. Shapiro, L. Carlson, J. Astin, and B. Freedman, "Mechanisms of mindfulness." *Journal of clinical psychology*, vol. 62 3, pp. 373–8, 2006. [Online]. Available: https://doi.org/10.1002/jclp.20237

[225] P. Malinowski, "Neural mechanisms of attentional control in mindfulness meditation," *Frontiers in Neuroscience*, vol. 7, 2013. [Online]. Available: https://doi.org/10.3389/fnins.2013.00008

[226] K.-Z. J., *Wherever you go, there you are*. Piatkus, 2004.

[227] K. L. Gratz and L. Roemer, "Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale," *Journal of Psychopathology and Behavioral Assessment*, vol. 26, pp. 41–54, 2004. [Online]. Available: https://doi.org/10.1023/B:JOBA.0000007455.08539.94

[228] L. Hallion, S. A. Steinman, D. Tolin, and G. Diefenbach, "Psychometric Properties of the Difficulties in Emotion Regulation Scale (DERS) and Its Short Forms in Adults With Emotional Disorders," *Frontiers in Psychology*, vol. 9, 2018. [Online]. Available: https://doi.org/10.3389/fpsyg.2018.00539

[229] S. H. Lovibond and P. Lovibond, "Depression Anxiety and Stress Scales (DASS)," 2013.

[230] A. Coker, O. Coker, and D. Sanni, "Psychometric properties of the 21-item Depression Anxiety Stress Scale (DASS-21)," *African Research Review*, vol. 12, pp. 135–142, 2018. [Online]. Available: https://doi.org/10.4314/afrrev.v12i2.13

[231] S. Lamers, G. Westerhof, E. Bohlmeijer, P. T. ten Klooster, and C. Keyes, "Evaluating the psychometric properties of the Mental Health Continuum-Short Form (MHC-SF)," *Journal of clinical psychology*, vol. 67 1, pp. 99–110, 2011. [Online]. Available: https://doi.org/10.1002/jclp.20741

[232] C. Keyes, "The mental health continuum: from languishing to flourishing in life," *Journal of health and social behavior*, vol. 43 2, pp. 207–22, 2002. [Online]. Available: https://doi.org/10.2307/3090197

[233] M. Scheier, C. Carver, and M. Bridges, "Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test," *Journal of personality and social psychology*, vol. 67 6, pp. 1063–78, 1994. [Online]. Available: https://doi.org/10.1037//0022-3514.67.6.1063

[234] J. Gustems-Carnicer, C. Calderón, and M. Santacana, "Psychometric properties of the Life Orientation Test (LOT-R) and its relationship with psychological well-being and academic progress in college students," *Revista Latinoamericana De Psicologia*, vol. 49, pp. 19–27, 2017. [Online]. Available: https://doi.org/10.1016/j.rlp.2016.05.001

[235] R. Kessler, G. Andrews, L. Colpe, E. Hiripi, D. Mroczek, S. Normand, E. Walters, and A. Zaslavsky, "Short screening scales to monitor population prevalences and trends in non-specific psychological distress," *Psychological medicine*, vol. 32 6, pp. 959–76, 2002. [Online]. Available: https://doi.org/10.1017/s0033291702006074

[236] F. T. Reshvanloo, H. Kareshki, M. Amani, S. Esfandyari, and M. Torkamani, "Psychometric Properties of the Kessler psychological distress scale (K6) based on classical test theory and Item-response theory," *Razi Journal of Medical Sciences*, vol. 26, pp. 20–33, 2020.

[237] C. J. Caspersen, K. Powell, and G. M. Christenson, "Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research," *Public health reports*, vol. 100 2, pp. 126–31, 1985.

[238] C. Garber, B. Blissmer, M. Deschenes, B. Franklin, M. Lamonte, I. M. Lee, D. Nieman, and D. Swain, "Quantity and Quality of Exercise for Developing and Maintaining Cardiorespiratory, Musculoskeletal, and Neuromotor Fitness in Apparently Healthy Adults: Guidance for Prescribing Exercise," *Medicine and science in sports and exercise*, vol. 43, pp. 1334–59, 2011. [Online]. Available: https://doi.org/10.1249/MSS.0b013e318213fefb

[239] D. Warburton, P. Katzmarzyk, R. Rhodes, and R. Shephard, "Evidence-informed physical activity guidelines for Canadian adults," *Canadian journal of public health = Revue canadienne de sante publique*, vol. 98 Suppl 2, pp. S16–68, 2007. [Online]. Available: https://doi.org/10.1139/H07-123

[240] M. Hillsdon, C. Foster, and M. Thorogood, "Interventions for promoting physical activity," *The Cochrane database of systematic reviews*, vol. 1, p. CD003180, 2005. [Online]. Available: https://doi.org/10.1002/14651858

[241] F. Guerrera, S. Renaud, F. Tabbó, and P. Filosso, "How to design a randomized clinical trial: tips and tricks for conduct a successful study in thoracic disease domain," *Journal of thoracic disease*, vol. 9 8, pp. 2692–2696, 2017. [Online]. Available: https://doi.org/10.21037/jtd.2017.06.147

[242] TechAmerica, “Demystifying Big Data: A Practical Guide to Transforming the Business of Government,” TechAmerica Reports, Report, 2012.

[243] G. Kapil, A. Agrawal, and P. R. Khan, “A study of big data characteristics,” 2016. [Online]. Available: https://doi.org/10.1109/CESYS.2016.7889917

[244] S. T. Mueller, “The Psychology Experiment Building Language (Version 0.13),” 2013. [Online]. Available: http://pebl.sourceforge.net

[245] P. Natarajan, J. Frenzel, and D. Smaltz, “Demystifying Big Data and Machine Learning for Healthcare,” 2017.

[246] L. Blonde, K. Khunti, S. B. Harris, C. Meizinger, and N. S. Skolnik, “Interpretation and Impact of Real-World Clinical Data for the Practicing Clinician,” *Advances in therapy*, vol. 35, pp. 1763–1774, 2018. [Online]. Available: https://doi.org/10.1007/s12325-018-0805-y

[247] T. BMønsted, “Achieving veracity: A study of the development and use of an information system for data analysis in preventive healthcare,” *Health Informatics Journal*, vol. 25, pp. 491–495, 2019. [Online]. Available: https://doi.org/10.1177/1460458218796665

[248] V. Lukoianova, Tatiana Rubin, “Veracity Roadmap: Is Big Data Objective, Truthful and Credible?” *ASIST 2013: 76th Annual Meeting of the Association for Information Science and Technology: Beyond the Cloud: Rethinking Information Boundaries*, vol. 24, 2013. [Online]. Available: https://doi.org/10.7152/acro.v24i1.14671

[249] M. García Lozano, J. Brynielsson, U. Franke, M. Rosell, E. Tjörnhammar, S. Varga, and V. Vlassov, “Veracity assessment of online data,” *Decision Support Systems*, vol. 129, p. 113132, 2020. [Online]. Available: https://doi.org/10.1016/j.dss.2019.113132

[250] V. Gupta, M. Sengupta, J. Prakash, and B. C. Tripathy, “Production of Recombinant Pharmaceutical Proteins,” *Basic and Applied Aspects of Biotechnology*, pp. 77 – 101, 2016. [Online]. Available: https://doi.org/10.1007/978-981-10-0875-7_4

[251] R. Matousek, P. Dobkin, and J. Pruessner, “Cortisol as a marker for improvement in mindfulness-based stress reduction,” 2010. [Online]. Available: https://doi.org/10.1016/j.ctcp.2009.06.004

[252] H. Selye, *The Stress of Life.* New York: McGraw-Hill, 1956.

[253] M. U. Ahmed, S. Begum, and M. S. Islam, "Heart Rate and Inter-Beat Interval Computation to Diagnose Stress Using ECG Sensor Signal," Mälardalen University,Västerås, Sweden, Report 1929, 2010.

[254] J. B. Priest, P. N. Roberson, and S. B. Woods, "In Our Lives and Under Our Skin An Investigation of Specific Psycho-Biological Mediators Linking Family Relationships and Health Using the Biobehavioural Family Model," *Family Process*, 2018. [Online]. Available: https://doi.org/10.1111/famp.12357

[255] S. Nikolopoulos, A. Alexandridi, S. Nikolakeas, and G. Manis, "Experimental Analysis of Heart Rate Variability of Long-Recording Electrocardiograms in Normal Subjects and Patients with Coronary Artery Disease and Normal Left Ventricular Function," *Journal of Biomedical Informatics*, vol. 36, no. 3, pp. 202–217, 2003. [Online]. Available: https://doi.org/10.1016/J.JBI.2003.09.001

[256] D. Rebergen, S. Nagaraj, E. Rosenthal, M. Bianchi, M. van Putten, and M. Westover, "ADARRI: A Novel Method to Detect Spurious R-peaks in the Electrocardiogram for Heart Rate Variability Analysis in the Intensive Care Unit," *Journal of clinical monitoring and computing*, vol. 32, no. 1, pp. 53–61, 2018. [Online]. Available: https://doi.org/10.1007/s10877-017-9999-9

[257] D. Williams, C. Cash, C. Rankin, A. Bernardi, J. Koenig, and J. Thayer, "Resting Heart Rate Variability Predicts Self-Reported Difficulties in Emotion Regulation: A Focus on Different Facets of Emotion Regulation," *Frontiers in Psychology*, vol. 6, 2018. [Online]. Available: https://doi.org/10.3389/fpsyg.2015.00261

[258] F. B.-C. Epstein, S. Anthony, L, and M. Ronald, "The Biopsychosocial Model 25 Years Later: Principles, Practice, and Scientific Inquiry," *Annals Of Family Medicine*, vol. 2, no. 6, pp. 576–582, 2004. [Online]. Available: https://doi.org/10.1370/afm.245

[259] T. Hao, J. Rogers, H.-Y. Chang, M. Ball, K. Walter, S. Sun, C.-H. Chen, and X. Zhu, "Towards Precision Stress Management: Design and Evaluation of a Practical Wearable Sensing System for Monitoring Everyday Stress," 2017. [Online]. Available: https://doi.org/10.2196/iproc.8441

[260] R. R. Kroll, J. Boyd, and D. M. Maslove, “Accuracy of a Wrist-Worn Wearable Device for Monitoring Heart Rates in Hospital Inpatients: A Prospective Observational Study,” 2016. [Online]. Available: https://doi.org/10.2196/jmir.6025

[261] A. Sano, S. M. Taylor, A. W. McHill, L. K. Barger, E. Klerman, and R. Picard, “Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones: Observational Study,” 2018. [Online]. Available: https://doi.org/10.2196/jmir.9410

[262] J. E. Richards, “The Statistical Analysis of Heart Rate A Review,” *Psychophysiology*, vol. 17, no. 2, 1980. [Online]. Available: https://doi.org/0.1111/J.1469-8986.1980.TB00129.X

[263] Forbes, “Wearable Tech Market To Be Worth $34 Billion by 2020,” 2018. [Online]. Available: https://www.forbes.com/sites/paullamkin/2016/02/17/wearable-tech-market-to-be-worth-34-billion-by-2020

[264] E. Sharashova, “Decline in Resting Heart Rate, its Association with Other Variables, and its Role in Cardiovascular Disease,” Thesis, 2016.

[265] S. J. Kang, E. Kim, and K. J. Ko, “Effects of Aerobic Exercise on the Resting Heart Rate, Physical Fitness, and Arterial Stiffness of Female Patients with Metabolic Syndrome,” *The Journal of Physical Therapy Science*, vol. 28, no. 6, pp. 1764–1768, 2016.

[266] C. R. Sotomayor and C. M. Gallagher, “The Team Based Biopsychosocial Model: Having a Clinical Ethicist as a Facilitator and a Bridge Between Teams,” *HEC Forum*, vol. 31, pp. 75–83, 2019. [Online]. Available: https://doi.org/10.1007/s10730-018-9358-3

[267] V. D. Giovinazzo and M. Novarese, “The meaning of happiness: attention and time perception,” *Mind & Society*, vol. 15, pp. 207–218, 2016. [Online]. Available: https://doi.org/10.1007/S11299-015-0180-1

[268] S. Bezdjian, L. Baker, D. I. Lozano, and A. Raine, “Assessing inattention and impulsivity in children during the Go/NoGo task,” *The British journal of developmental psychology*, vol. 27 Pt 2, pp. 365–83, 2009. [Online]. Available: https://doi.org/10.1348/026151008X314919

[269] M. Moniz, S. N. d. Jesus, E. Gonçalves, J. Viseu, A. Pacheco, and S. Moreira, "Portuguese Version of Simple Go/No-Go Task: Influence of Age in Attention and Response Inhibition Reaction Time," *Psychology*, vol. 07, no. 02, pp. 254–257, 2016. [Online]. Available: https://doi.org/10.4236/psych.2016.72027

[270] R. Whelan, "Effective analysis of reaction time data," *Psychological Record*, vol. 58, no. 3, pp. 475–482, 2008. [Online]. Available: https://doi.org/10.1007/BF03395630

[271] Osborne, "Notes on the use of data transformations," *Practical Assessment, Research and Evaluation*, vol. 8, no. 6, 2002. [Online]. Available: https://doi.org/10.7275/4VNG-5608

[272] S. Lo and S. Andrews, "To transform or not to transform: using generalized linear mixed models to analyse reaction time data," *Front Psychol*, vol. 6, p. 1171, 2015. [Online]. Available: https://doi.org/10.3389/fpsyg.2015.01171

[273] A. Heathcote, S. J. Popiel, and D. J. K. Mewhort, "Analysis of Response-Time Distributions - an Example Using the Stroop Task," *Psychological Bulletin*, vol. 109, no. 2, pp. 340–347, 1991. [Online]. Available: https://doi.org/0.1037/0033-2909.109.2.340

[274] R. H. Hohle, "Inferred components of reaction times as functions of forepaid duration," *Journal of Experimental Psychology*, no. 69, pp. 382–386, 1965. [Online]. Available: https://doi.org/10.1037/h0021740

[275] D. A. Balota and D. H. Spieler, "Word frequency, repetition, and lexicality effects in word recognition tasks: beyond measures of central tendency," *J Exp Psychol Gen*, vol. 128, no. 1, pp. 32–55, 1999. [Online]. Available: https://doi.org/10.1037//0096-3445.128.1.32

[276] M. P. Henriquez-Henriquez, P. Billeke, H. Henriquez, F. J. Zamorano, F. Rothhammer, and F. Aboitiz, "Intra-Individual Response Variability Assessed by Ex-Gaussian Analysis may be a New Endophenotype for Attention-Deficit/Hyperactivity Disorder," *Front Psychiatry*, vol. 5, p. 197, 2014. [Online]. Available: https://doi.org/0.3389/fpsyt.2014.00197

[277] Y. Lacouture and D. Cousineau, "How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times," *Tutorials in Quantitative Methods for Psychology*, vol. 4, no. 1, pp. 35–45, 2008. [Online]. Available: https://doi.org/10.20982/tqmp.04.1.p035

[278] S. Sternberg, "RTs and the Ex-Gaussian Distribution Page 1 Reaction Times and the Ex-Gaussian Distribution: When is it Appropriate?"

[279] E. M. Palmer, T. S. Horowitz, A. Torralba, and J. M. Wolfe, "What are the shapes of response time distributions in visual search?" *J Exp Psychol Hum Percept Perform*, vol. 37, no. 1, pp. 58–71, 2011. [Online]. Available: https://doi.org/10.1037/a0020747

[280] C. V. Dolan, H. L. van der Maas, and P. C. Molenaar, "A framework for ML estimation of parameters of (mixtures of) common reaction time distributions given optional truncation or censoring," *Behav Res Methods Instrum Comput*, vol. 34, no. 3, pp. 304–23, 2002. [Online]. Available: https://doi.org/10.3758/BF03195458

[281] D. S. Wilks, "Maximum-Likelihood-Estimation for the Gamma-Distribution Using Data Containing Zeros," *Journal of Climate*, vol. 3, no. 12, pp. 1495–1501, 1990. [Online]. Available: https://doi.org/10.1175/1520-0442(1990)003<1495:MLEFTG>2.0.CO;2

[282] H. C. S. Thom, "A Note on the Gamma Distribution," *Monthly Weather Review*, vol. 86, no. 4, pp. 117–122, 1958. [Online]. Available: https://doi.org/10.1175/1520-0493(1958)086<0117:anotgd>2.0.co;2

[283] G. J. Husak, J. Michaelsen, and C. Funk, "Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications," *International Journal of Climatology*, vol. 27, no. 7, pp. 935–944, 2007. [Online]. Available: https://doi.org/10.1002/joc.1441

[284] V. Ricci, "Fitting Distributions with R," 2005.

[285] P. T. Schickedanz and G. F. Krause, "A Test for the Scale Parameters of Two Gamma Distributions using the Generalized Likelihood Ratio," *Journal of Applied Meteorology*, vol. 9, no. 1, pp. 13–16, 1970. [Online]. Available: https://doi.org/10.1175/1520-0450(1970)009<0013:ATFTSP>2.0.CO;2

[286] K. Krishnamoorthy, M. Lee, and W. Xiao, "Likelihood ratio tests for comparing several gamma distributions," *Environmetrics*, vol. 26, no. 8, pp. 571–583, 2015. [Online]. Available: https://doi.org/10.1002/env.2357

[287] L. Degroote, A. DeSmet, I. De Bourdeaudhuij, D. Van Dyck, and G. Crombez, "Content validity and methodological considerations in ecological momentary

assessment studies on physical activity and sedentary behaviour: a systematic review," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 17, p. 35, 2012. [Online]. Available: https://doi.org/10.1186/s12966-020-00932-9

[288] A. A. Stone and S. Shiffman, "Ecological Momentary Assessment (EMA) in Behavioral Medicine," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 16, pp. 199–202, 1994. [Online]. Available: https://doi.org/10.1093/abm/16.3.199

[289] S. Shiffman, A. A. Stone, and M. R. Hufford, "Ecological momentary assessment," *Annual review of clinical psychology*, vol. 4, pp. 1–32, 2008. [Online]. Available: https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

[290] S. Shiffman and A. A. Stone, "Capturing momentary, self-report data: A proposal for reporting guidelines," *Annals of Behavioral Medicine*, vol. 24, pp. 236–243, 2002. [Online]. Available: https://doi.org/10.1207/S15324796ABM2403_09

[291] G. F. Dunton, "Ecological Momentary Assessment in Physical Activity Research," *Exerc Sport Sci Rev*, vol. 45, pp. 48–54, 2017. [Online]. Available: https://doi.org/10.1249/jes.0000000000000092

[292] T. B. Wray, J. E. Merrill, and P. M. Monti, "Using Ecological Momentary Assessment (EMA) to Assess Situation-Level Predictors of Alcohol Use and Alcohol-Related Consequences," *Alcohol Res*, vol. 36, pp. 19–27, 2014.

[293] N. Schüz, J. Cianchi, S. Shiffman, and S. G. Ferguson, "Novel Technologies to Study Smoking Behavior: Current Developments in Ecological Momentary Assessment," *Current Addiction Reports*, vol. 2, pp. 8–14, 2015. [Online]. Available: https://doi.org/10.1007/s40429-015-0039-x

[294] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *Journal of health and social behavior*, vol. 24 4, pp. 385–96, 1983. [Online]. Available: https://doi.org/10.2307/2136404

[295] I. Pratama, A. Permanasari, I. Ardiyanto, and R. Indrayani, "A review of missing values handling methods on time-series data," 2016. [Online]. Available: https://doi.org/10.1109/ICITSI.2016.7858189

[296] P. D. Allison, "Missing Data," 2001. [Online]. Available: https://doi.org/10.4135/9780857020994.n4

[297] Y. Dong and C. Peng, "Principled missing data methods for researchers," *SpringerPlus*, vol. 2, 2013. [Online]. Available: https://doi.org/10.1186/2193-1801-2-222

[298] S. JL., "Multiple imputation: a primer," 1999. [Online]. Available: https://doi.org/10.1177/096228029900800102

[299] D. A. Bennett, "How can I deal with missing data in my study?" 2001. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/11688629/

[300] L. S. F. Barbara G. Tabachnick, "Using Multivariate Statistics," 2013.

[301] H. Toutenburg, "Rubin, D.B.: Multiple imputation for nonresponse in surveys," 1990. [Online]. Available: https://doi.org/10.1007/BF02924688

[302] Y. He, "Missing data analysis using multiple imputation: getting to the heart of the matter," 2010. [Online]. Available: https://doi.org/10.1161/circoutcomes.109.875658

[303] T. Benichou, B. Pereira, M. Mermillod, I. Tauveron, D. Pfabigan, S. Maqdasy, and F. Dutheil, "Heart rate variability in type 2 diabetes mellitus/: A systematic review and meta-analysis," 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0195166

[304] Ö. Çokluk and M. Kayri, "The Effects of Methods of Imputation for Missing Values on the Validity and Reliability of Scales," *Kuram Ve Uygulamada Egitim Bilimleri*, vol. 11, pp. 303–309, 2011.

[305] M. Santhanagopalan, M. Chetty, C. Foale, S. Aryal, and B. Klein, "Modeling neurocognitive reaction time with gamma distribution," in *Proceedings of the Australasian Computer Science Week Multiconference*, ser. ACSW '18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: https://doi.org/10.1145/3167918.3167941

[306] C. Batini and M. Scannapieco, "Data and Information Quality - Dimensions, Principles and Techniques," in *Data-Centric Systems and Applications*, 2016.

[307] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, vol. 4, 2018. [Online]. Available: https://doi.org/10.1016/j.heliyon.2018.e00938

[308] K. Das and R. N. Behera, "A Survey on Machine Learning: Concept, Algorithms and Applications," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 5, 2017.

[309] R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Solano, and O. Rendon, "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," *Journal of Internet Services and Applications*, vol. 9, pp. 1–99, 2018. [Online]. Available: https://doi.org/10.1186/s13174-018-0087-2

[310] F. Ogwueleka, S. Misra, R. Colomo-Palacios, and L. Fernández, "Neural Network and Classification Approach in Identifying Customer Behavior in the Banking Sector: A Case Study of an International Bank," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 25, pp. 28–42, 2015. [Online]. Available: https://doi.org/10.1002/hfm.20398

[311] Z. C. Lipton, "A Critical Review of Recurrent Neural Networks for Sequence Learning," *ArXiv*, 2015. [Online]. Available: https://arxiv.org/abs/1506.00019

[312] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," *Neural computation*, vol. 9, 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[313] A. Yadav, C. K. Jha, and A. Sharan, "Optimizing LSTM for time series prediction in Indian stock market," *Procedia Computer Science*, vol. 167, pp. 2091–2100, 2020. [Online]. Available: https://doi.org/10.1016/j.procs.2020.03.257

[314] X. Ma and E. Hovy, "End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF," vol. 1, pp. 1064–1074, 2016. [Online]. Available: https://doi.org/10.18653/v1/P16-1101

[315] A. Søgaard and Y. Goldberg, "Deep multi-task learning with low level tasks supervised at lower layers," pp. 231–235, 2016. [Online]. Available: https://doi.org/10.18653/v1/P16-2038

[316] C. C. Aggarwal, *Data classification : algorithms and applications*. Chapman & Hall/CRC, 2015. [Online]. Available: https://doi.org/10.5555/2685756

[317] Z. Xing, J. Pei, and E. Keogh, "A Brief Survey on Sequence Classification," *SIGKDD Explor. Newsl.*, vol. 12, no. 1, p. 40–48, Nov. 2010. [Online]. Available: https://doi.org/10.1145/1882471.1882478

[318] L. Wei and E. Keogh, “Semi-supervised time series classification,” in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’06. New York, NY, USA: Association for Computing Machinery, 2006, p. 748–753. [Online]. Available: https://doi.org/10.1145/1150402.1150498

[319] F. Sebastiani, “Machine Learning in Automated Text Categorization,” *ACM Comput. Surv.*, vol. 34, no. 1, p. 1–47, Mar. 2002. [Online]. Available: https://doi.org/10.1145/505282.505283

[320] V. Tan, Pang-Ning Kumar, “Discovery of Web Robot Sessions Based on their Navigational Patterns,” *Data Mining and Knowledge Discovery*, vol. 6, no. 1, 2002. [Online]. Available: https://doi.org/10.1023/A:1013228602957

[321] O. Duskin and D. G. Feitelson, “Distinguishing Humans from Robots in Web Search Logs: Preliminary Results Using Query Rates and Intervals.” [Online]. Available: https://doi.org/10.1145/1507509.1507512

[322] X. Liu, P. Zhang, and D. Zeng, “Sequence Matching for Suspicious Activity Detection in Anti-Money Laundering.” Berlin, Heidelberg: Springer-Verlag, 2008, p. 50–61. [Online]. Available: https://doi.org/10.1007/978-3-540-69304-8_6

[323] M. Deshpande and G. Karypis, “Evaluation of Techniques for Classifying Biological Sequences,” ser. PAKDD ’02. Berlin, Heidelberg: Springer-Verlag, 2002, p. 417–431. [Online]. Available: https://doi.org/10.1007/3-540-47887-6_41

[324] R. Rizzo, A. Fiannaca, M. L. Rosa, and A. Urso, “A Deep Learning Approach to DNA Sequence Classification,” in *CIBB*, 2015. [Online]. Available: https://doi.org/10.1007/978-3-319-44332-4_10

[325] E. N. Yolacan, “Learning from sequential data for anomaly detection,” 2014. [Online]. Available: https://doi.org/10.1.1.682.7818

[326] T. Lane and C. Brodley, “Temporal sequence learning and data reduction for anomaly detection,” *ACM Trans. Inf. Syst. Secur.*, vol. 2, pp. 295–331, 1999. [Online]. Available: https://doi.org/10.1145/322510.322526

[327] A. Nanopoulos, R. Alcock, and Y. Manolopoulos, “Feature-based classification of time-series data,” 2001. [Online]. Available: https://doi.org/10.5555/766914.766918

[328] E. J. Keogh and S. Kasetty, “On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration,” *Data Mining and Knowledge Discovery*, vol. 7, pp. 349–371, 2004. [Online]. Available: https://doi.org/10.1145/775047.775062

[329] M. Kadous and C. Sammut, “Classification of Multivariate Time Series and Structured Data Using Constructive Induction,” *Machine Learning*, vol. 58, pp. 179–216, 2005. [Online]. Available: https://doi.org/10.1007/s10994-005-5826-5

[330] C. Li, L. Khan, and B. Prabhakaran, “Real-time classification of variable length multi-attribute motions,” *Machine Learning*, vol. 10, pp. 163–183, 2006. [Online]. Available: https://doi.org/10.1007/s10115-005-0223-8

[331] S. Norgaard, R. Saeedi, and A. Gebremedhin, “Multi-Sensor Time-Series Classification for Activity Tracking Under Variable Length,” *IEEE Sensors Journal*, vol. 20, pp. 2701–2709, 2020. [Online]. Available: https://doi.org/10.1109/JSEN.2019.2953938

[332] C. Yang, Z. Chen, and C. Yang, “Sensor Classification Using Convolutional Neural Network by Encoding Multivariate Time Series as Two-Dimensional Colored Images,” *Sensors (Basel, Switzerland)*, vol. 20, 2020. [Online]. Available: https://doi.org/10.3390/s20010168

[333] N. Reimers and I. Gurevych, “Optimal Hyperparameters for Deep LSTM/-Networks for Sequence Labeling Tasks,” *ArXiv*, 2017. [Online]. Available: https://arxiv.org/abs/1707.06799

[334] F. Hutter, H. Hoos, and K. Leyton-Brown, “An efficient approach for assessing hyperparameter importance,” vol. 2, pp. 1130–1144, 2014. [Online]. Available: https://doi.org/10.5555/3044805.3044891

[335] B. Nakisa, M. N. Rastgoo, A. Rakotonirainy, F. Maire, and V. Chandran, “Long Short Term Memory Hyperparameter Optimization for a Neural Network Based Emotion Recognition Framework,” *IEEE Access*, vol. PP, pp. 1–1, 2018. [Online]. Available: https://doi.org/10.1109/ACCESS.2018.2868361

[336] J. Snoek, H. Larochelle, and R. Adams, “Practical Bayesian Optimization of Machine Learning Algorithms,” *Advances in Neural Information Processing Systems*, vol. 4, 2012. [Online]. Available: https://doi.org/10.1155/2018/7310496

[337] Z. Xing, J. Pei, and E. J. Keogh, "A brief survey on sequence classification," *SIGKDD Explor.*, vol. 12, pp. 40–48, 2010. [Online]. Available: https://doi.org/10.1145/1882471.1882478

[338] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," *J. Mach. Learn. Res.*, vol. 13, no. null, p. 281–305, Feb. 2012. [Online]. Available: https://doi.org/10.5555/2188385.2188395

[339] P. Sarkar, P. Janardhan, and P. Roy, "Prediction of sea surface temperatures using deep learning neural networks," 2020. [Online]. Available: https://doi.org/10.1007/s42452-020-03239-3

[340] K. Vijayaprabakaran and K. Sathiyamurthy, "Towards activation function search for long short-term model network: A differential evolution based approach," *Journal of King Saud University - Computer and Information Sciences*, 2020. [Online]. Available: https://doi.org/10.1016/j.jksuci.2020.04.015

[341] A. Farzad, H. Mashayekhi, and H. Hassanpour, "A comparative performance analysis of different activation functions inLSTM networks for classification," *Neural Computing and Applications*, vol. 31, 2019. [Online]. Available: https://doi.org/10.1007/s00521-017-3210-6

[342] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay," *ArXiv*, 2018. [Online]. Available: https://arxiv.org/abs/1803.09820

[343] L. Nordgren, E. Monell, A. Birgegård, J. Bjureberg, and H. Hesser, "Factor Structure of the Difficulties in Emotion Regulation Scale in Treatment Seeking Adults with Eating Disorders," *Journal of Psychopathology and Behavioral Assessment*, vol. 42, pp. 111–126, 2020. [Online]. Available: https://doi.org/10.1007/s10862-019-09765-8

[344] K. L. Gratz, M. Z. Rosenthal, M. Tull, C. Lejuez, and J. Gunderson, "An experimental investigation of emotion dysregulation in borderline personality disorder." *Personality Disorders: Theory, Research, and Treatment*, pp. 18–26, 2009. [Online]. Available: https://doi.org/10.1037/1949-2715.S.1.18

[345] E. Sloan, K. Hall, R. Moulding, S. Bryce, H. Mildred, and P. Staiger, "Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: A systematic review."

Clinical psychology review, vol. 57, pp. 141–163, 2017. [Online]. Available: https://doi.org/10.1016/j.cpr.2017.09.002

[346] L. Tull, Matthew T. Roemer, "Emotion Regulation Difficulties Associated with the Experience of Uncued Panic Attacks: Evidence of Experiential Avoidance, Emotional Nonacceptance, and Decreased Emotional Clarity," *Behavior Therapy*, vol. 38, pp. 378–391, 2007. [Online]. Available: https://doi.org/10.1016/j.beth.2006.10.006

[347] L. Villalta, P. Smith, N. Hickin, and A. Stringaris, "Emotion regulation difficulties in traumatized youth: a meta-analysis and conceptual review," *European Child & Adolescent Psychiatry*, vol. 27, pp. 527–544, 2018. [Online]. Available: https://doi.org/0.1007/s00787-018-1105-4

[348] E. Visted, J. Vøllestad, M. B. Nielsen, and E. Schanche, "Emotion Regulation in Current and Remitted Depression: A Systematic Review and Meta-Analysis," *Front Psychol*, vol. 9, p. 756, 2018. [Online]. Available: https://doi.org/10.3389/fpsyg.2018.00756

[349] E. Thayer, J. F. Sternberg, "Beyond heart rate variability: vagal regulation of allostatic systems," *Ann N Y Acad Sci*, vol. 1088, pp. 361–72, 2006. [Online]. Available: https://doi.org/10.1196/annals.1366.014

[350] J. Blechert, M. Lajtman, T. Michael, J. Margraf, and F. H. Wilhelm, "Identifying anxiety states using broad sampling and advanced processing of peripheral physiological information," *Biomed Sci Instrum*, vol. 42, pp. 136–41, 2006.

[351] U. Hess, A. Kappas, G. J. McHugo, J. T. Lanzetta, and R. E. Kleck, "The facilitative effect of facial expression on the self-generation of emotion," *International Journal of Psychophysiology*, vol. 12, pp. 251–265, 1992. [Online]. Available: https://doi.org/10.1016/0167-8760(92)90064-I

[352] Y. Zhang, R. Fu, L. Sun, Y. Gong, and D. Tang, "How Does Exercise Improve Implicit Emotion Regulation Ability: Preliminary Evidence of Mind-Body Exercise Intervention Combined with Aerobic Jogging and Mindfulness-Based Yoga," *Frontiers in Psychology*, vol. 10, 2019. [Online]. Available: https://doi.org/10.3389/fpsyg.2019.01888

[353] C. Racine, K. Kalra, M. Ceide, N. J. Williams, F. Zizi, M. V. Mendlowicz, and G. Jean-Louis, "Sleep Duration, Insomnia Symptoms, and Emotion Regulation

among Black Women," *J Sleep Disord Ther*, vol. 2, 2013. [Online]. Available: https://doi.org/10.4172/2167-0277.1000122

[354] J. Kim, D. Marcusson-Clavertz, K. Yoshiuchi, and J. Smyth, "Potential benefits of integrating ecological momentary assessment data into mHealth care systems," *Biopsychosocial Medicine*, vol. 13, 2019. [Online]. Available: https://doi.org/10.1186/s13030-019-0160-5.

[355] F. Abramovich and M. Pensky, "Classification with many classes: Challenges and pluses," *J. Multivar. Anal.*, vol. 174, 2019. [Online]. Available: https://doi.org/10.1016/j.jmva.2019.104536