Use of Smart Technology Tools for Supporting Public Health Surveillance: From Development of a Mobile Health Platform to Application in Stress Prediction

by

Pedro Elkind Velmovitsky

A thesis

presented to the University of Waterloo

in fulfillment of the thesis requirement

for the degree of

Doctor of Philosophy
in
Public Health Sciences

Waterloo, Ontario, Canada, 2023

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Alex Mariakakis Assistant Professor Department of Computer Science University of Toronto
Supervisor(s)	Plinio Pelegrini Morita Associate Professor School of Public Health Sciences University of Waterloo
Internal Member	Scott Leatherdale Professor and University Research Chair School of Public Health Sciences University of Waterloo
Internal-external Member	Paulo Alencar Adjunct Professor School of Public Health Sciences University of Waterloo

Author’s Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Pedro Elkind Velmovitsky was the sole author for Chapters 1, 2, 3 and 10 which were written under the supervision of Dr. Plinio Pelegrini Morita and were not written for publication.

This thesis consists in part of five manuscripts written for publication . Exceptions to sole authorship of material are as follows:

Research presented in Chapter 4:

Pedro Elkind Velmovitsky was responsible for mapping the surveys to Apple Health Data, reviewing the mobile health research and demographic data as well as writing the manuscript. Merna Kirolos provided support in the collection and review of demographic data as well as in writing the technology adoption sections of the paper. Drs. Plinio Pelegrini Morita, Paulo Alencar, Scott Leatherdale and Donald Cowan supervised the manuscript's writing, editing and revisions. All authors contributed to the article and approved the submitted version.

Velmovitsky, P.E., Kirolos, M., Alencar, P., Leatherdale, S. Cowan, D., & Morita. P.P. (submitted, in revision). Leveraging Mobile Health Technologies for Public Health: A Viewpoint. Submitted to *JMIR Public Health and Surveillance.*

Research presented in Chapter 5:

The primary author Pedro Elkind Velmovitsky was responsible for modelling and developing the mobile health platform. The Ubiquitous Health Technology Lab volunteer Angelo Santiago aided in the design of the app, and intern Mingyang Xu created the database based on the data structures. The developer team at NoClaf Tech helped with development of the API class. The co-authors Drs. Plinio Pelegrini Morita, Paulo Alencar, Scott Leatherdale and Donald Cowan supervised the manuscript's writing, editing and revisions. All authors contributed to the article and approved the submitted version.

Velmovitsky, P.E., Alencar, P., Leatherdale, S. Cowan, D., & Morita. P.P. (submitted, in revision). Modelling and Development of a Mobile Health Data Extraction Platform for Public Health. Submitted to *JMIR mHealth and uHealth.*

Research presented in Chapter 6:

Pedro Elkind Velmovitsky was responsible for designing and conducting the study, including data collection and analyses, and writing the manuscript. Matheus Lotto provided help in the statistical analyses and in the writing of the manuscript. Drs. Plinio Pelegrini Morita, Paulo Alencar, Scott Leatherdale and Donald Cowan supervised the manuscript's writing, editing and revisions. All authors contributed to the article and approved the submitted version.

Velmovitsky, P.E., Lotto, M., Alencar, P., Leatherdale, S. Cowan, D., & Morita. P.P. (Accepted, In Press). Can Heart Rate Variability data from the Apple Watch Electrocardiogram (ECG) Quantify Stress? *Frontiers in Public Health.*

Research presented in Chapter 7:

Pedro Elkind Velmovitsky was responsible for designing and conducting the study, including data collection and analyses, and writing the manuscript. Drs. Plinio Pelegrini Morita, Paulo Alencar, Scott Leatherdale and Donald Cowan supervised the manuscript's writing, editing and revisions. All authors contributed to the article and approved the submitted version.

Velmovitsky, P. E., Alencar, P., Leatherdale, S. T., Cowan, D., & Morita, P. P. (2022). Using apple watch ECG data for heart rate variability monitoring and stress prediction: A pilot study. *Frontiers in Digital Health, 4.*

Research presented in Chapter 8 and 9:

The primary author Pedro Elkind Velmovitsky was responsible for designing and conducting the study, including data collection and analyses, and writing the manuscript. The Ubiquitous Health Technology Lab volunteer Joie Li developed a Python code to compare manually exported Apple Health data with information from the database, and the intern Mingyang Xu developed the REST API service used. Co-authors Drs. Plinio Pelegrini Morita, Paulo Alencar, Scott Leatherdale and Donald Cowan supervised the manuscript's writing, editing and revisions. All authors contributed to the article and approved the submitted version.

Citation for Chapter 8: Velmovitsky, P. E., Alencar, P., Leatherdale, S. T., Cowan, D., & Morita, P. P. (submitted, in revision). Application of a Mobile Health Data Platform for Public Health Surveillance: A Case Study in Stress Monitoring and Prediction. Submitted to *Digital Health.*

Citation for Chapter 9: Velmovitsky, P. E., Alencar, P., Leatherdale, S. T., Cowan, D., & Morita, P. P. (submitted, in revision). Predicting Stress in Daily-Life Routines Using Personal Mobile and Wearable Devices. Submitted to *Internet of Things.*

Abstract

Background Traditional public health data collection methods are typically based on self-reported data and may be subject to limitations such as biases, delays between collection and reporting, costs, and logistics. These may affect the quality of collected information and the ability of public health agencies to monitor and improve the health of populations. An alternative may be the use of personal, off-the-shelf smart devices (e.g., smartphones and smartwatches) as additional data collection tools. These devices can collect passive, continuous, real-time and objective health-related data, mitigating some of the limitations of self-reported information. The novel data types can then be used to further study and predict a condition in a population through advanced analytics. In this context, this thesis' goal is to investigate new ways to support public health through the use of consumer-level smart technologies as complementary survey, monitoring and analyses tools, with a focus on perceived stress. To this end, a mobile health platform (MHP) that collects data from devices connected to Apple Health was developed and tested in a pilot study collecting self-reported and objective stress-related information, and a number of Machine Learning (ML) models were developed based on these data to monitor and predict the stress levels of participants.

Methods The mobile platform was created for iOS using the XCode software, allowing users to self-report their stress levels based on the stress subscale of the Depression, Anxiety and Stress Scale (DASS-21) as well as a single-item LIKERT-based scale. The platform also collects objective data from sensors that integrate with Apple Health, one of the most popular mobile health data repositories. A pilot study with 45 participants was conducted that uses the platform to collects stress self-reports and variables associated with stress from Apple Health, including heart rate, heart rate variability, ECG, sleep, blood pressure, weight, temperature, and steps. To this end, participants were given an iPhone with the platform installed as well as an Apple Watch, Withings Sleep, Withings Thermos, Withings BPM Connect, Withings Body+, and an Empatica E4 (the only device that does not connect to Apple Health but included due to its wide use in research). Participants were instructed to take device measurements and self-report stress levels 6 times per day for 14 days. Several experiments were conducted involving the development of ML models to predict stress based on the data, using Random Forests and

Support Vector Machines. In each experiment, different subsets of the data from the full sample of 45 participants were used. 3 approaches to model development were followed: a) creating generalized models with all data; b) a hybrid approach using 80% of participants to train and 20% to test the model c) creating individualized user-specific models for each participant. In addition, statistical analyses of the data – specifically Spearman correlation and repeated measures ANOVA – were conducted.

Results Statistical analyses did not find significant differences between groups and only weak significant correlations. Among the Machine Learning models, the approach of using generalized models performed well, with f1-macro scores above 60% for several of the samples and features investigated. User-specific models also showed promise, with 82% achieving accuracies higher than 60% (the bottom limit of the state-of-the-art). While the hybrid approach had lower f1-macro scores, suggesting the models could not predict the two classes well, the accuracy of several of these models was in line with the state-of-the-art. Apple Watch sleep features, as well as weight, blood pressure, and temperature, were shown to be important in building the models.

Discussion and Conclusion ML-based models built with data collected from the MHP in real-life conditions were able to predict stress with results often in line with state-of-the-art, showing that smart technology data can be a promising tool to support public health surveillance. In particular, the approaches of creating models for each participant or one generalized model were successful, although more validation is needed in future studies (e.g., with more purposeful sampling) for increased generalizability and validity on the use of these technologies in the real-world. The hybrid approach had good accuracy but lower f1-scores, indicating results could potentially be improved (e.g., possibly with less missing or noisy data, collected in more controlled conditions). For feature selection, important features included sleep data as well as weight, blood pressure and temperature from mobile and wearable devices. In summary, this study indicates that a platform such as the MHP, collecting data from smart technologies, could potentially be a novel tool to complement population-level public health data collection. The predictive stress modelling might be used to monitor stress levels in a population and provide personalized interventions. Although more validation may be needed, this work represents a step in this direction.

Acknowledgments

I would like to acknowledge the financial support from the province of Ontario in the form of an Ontario Trillium Scholarship, the University of Waterloo in the form of several internal scholarships. I would also like to thank MITACS Accelerate Canada, the CSA Group and thinktum in the form of several graduate research scholarships. I

The support and guidance provided by Prof. Plinio P. Morita were essential for the conceptualization and advancement of this thesis. His experience and passion for mobile and ubiquitous health research are not only infectious but also paramount for the progress and completion of this work. Dr. Morita also greatly encouraged me to step out of my comfort zone and participate in several research projects that allowed me to gain more experience in health informatics, develop as an independent researcher and crystallize my future career goals.

I would like to thank Prof. Paulo Alencar and Prof. Scott Leatherdale from my PhD committee for all the ongoing support during every stage of this work. Their guidance and support encouraged me to move forward and better myself at every turn. Thank you for all the time devoted to improving my research. I am also grateful to Prof. Alex Mariakakis, the external examiner, for providing feedback that strengthened the final version of this document.

I would like to thank all the participants that joined in my study and made this a very exciting and fun experience. A special thanks to the volunteers Angelo Santiago who helped to design the app presented in this study, Joie Li who developed a Python code to handle manually exported health data, and Merna Kirolos who helped to review demographic data and write the viewpoint paper; and to the intern Mingyang Xu who created the database and REST API service. To the other volunteers who worked under my supervision at the Ubiquitous Health Technology Lab, whether on tasks related to this thesis or to other lab projects, I would like to extend my sincerest thanks as well.

To my teammates at the Ubiquitous Health Technology Lab, thank you for all the support provided whenever I needed, whether in knowledge or in kind and supportive words. A special thanks to the staff of the School of Public Health Sciences for the support over the years.

Outside my department, I would like to thank Dr. Okey Igboeli, director of the Science and Business program, who, in addition to being a great and supportive research partner, believed in me as a professor and allowed me my first experiences teaching courses in the program.

To Prof. Donald Cowan, my deepest thanks for all the time dedicated to guiding, encouraging and supporting me, as well as providing insight and edits to my manuscripts, through all the years of my doctoral dissertation.

I would also like to thank Prof. Carlos José Pereira de Lucena, Marx Viana, and all my teammates at the Laboratório de Engenharia de Software, Pontifícia Universidade Católica do Rio de Janeiro, as well as the mentors in the Apple Developer Academy PUC-Rio program – Ricardo Venieris, Andrew Costa and Hendi Coelho –, Hillary Srere, Gustavo Robichez, all my colleagues in the program and fellow ambassadors, for helping me pursue a passion in mobile technology development.

Dedication

Obtaining this PhD degree was probably the most challenging thing I ever did. Not only because of the inherent difficulties to completing a doctoral program, but also due to moving to Canada from Brazil. Living independently and alone for the first time, away from family and friends, was both harder and more rewarding that I could ever have imagined.

With this in mind, I would like to dedicate this first and foremost to my parents and sister, Marcio, Fabiana and Vanessa. Their support, and always knowing I could count on them no matter what, encouraged me through this journey. My grandparents, David and Belinha, also believed in me always and were a light for me when I needed it. Thank you all for everything you did and continue to do for me. And I also can't forget to thank Tiffany, Julie, and Brigitte, for all the fun you provided over the years.

To everyone who helped me during my journey in Canada, whether you are in Canada or Brazil, my sincerest thank you. Among the many names that come to mind, I would like to specially thank Vinicius, Matheus, and Raphael, for your support.

Members of the Goodfellas, thank you for always being there with an engaging discussion to distract me when I needed it.

Phillipe, who I can always count on for a good laugh and to discuss life, thank you for being there for me all these years.

Andre and Daniel, thank you for taking my call when I needed it and helping me see the brighter side of things.

To Lucy, thank you for brightening my days.

To Rodrigo, wherever you are, thank you for being a part of my life.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AC: Heart's Acceleration

ACC: Accelerometer Data

AH: Apple Health

AI: Artificial Intelligence

ANS: Autonomic Nervous System

API: Application Programming Interface

ApEn: Approximate Entropy

AR: Autoregressive Spectral Analysis

AUC: Area Under the Curve

AW: Apple Watch

BP: Blood Pressure

BVP: Blood Volume Pulse

BPM: Blood Pressure Monitor

CCHS: Canadian Community Housing Survey

CHMS: Canadian Health Measures Survey

CHSCY: Canadian Health Survey on Children and Youth

CMA/CA: Canadian Census Metropolitan Areas or Census Agglomerations

COVID-19: Coronavirus

CRTC: Canadian Radio-television and Telecommunications Commission

CV: Cross Validation

D: Dataset with All Features

DA: Dataset with only Apple Features

DAO: Data Access Object

DASS-21: Depression, Anxiety and Stress Scale

DAW: Dataset with Apple and Withings Features

DC: Heart's Deceleration

DDSR: Daily-to-daily Self-report

DECG: Dataset with Only ECG Features

DEmpatica: Dataset with Only Empatica Features

DIA: Diastolic Blood Pressure

DT: Decision Tree

DW: Dataset with only Withings Features

ECG: Electrocardiogram

EDA: Electrodermal Activity

EEG: Electroencephalogram

EMA: Ecological Momentary Assessment

FFT: Fast Fourier Transform

GDPR: General Data Protection Regulation

HBSC: Health Behaviours in School-aged Children

HIPAA: Health Insurance Portability and Accountability Act

HR: Heart Rate

HRV: Heart Rate Variability

High Frequency: HF

GSS: General Social Survey

Info.plist: Information Propertly List Files

IoT: Internet of Things

IPAQ: International Physical Activity Questionnaire

IPC: Information and Privacy Commissioner of Ontario

kNN: k-Nearest Neighbours

LDA: Linear Discriminant Analysis

LDKC: Laboratory-to-daily Known Context

LDSR: Laboratory-to-daily self-Report

LF: Low Frequency

LLKC: Laboratory-to-laboratory Known Context

LLSR: Laboratory-to-laboratory Self-report

LOPO: Leave-One-Person Out

LR: Logistic Regression

MAE: Mean Absolute Error

MAP: Mean Arterial Pressure

Max: Maximum

MHP: Mobile Health Platform

Min: Minimum

ML: Machine Learning

MLP: Multi-layer Perceptron

MSE: Multiscale Entropy

NN: Neural Networks

NN50: Number of pairs of NN intervals that are higher than 50ms

PAM: Physical Activity Monitor

PASS: Physical Activity, Sedentary Behaviour and Sleep

PCA: Principal Component Analysis

PII: Personally Identifiable Information

PIPEDA: Protection and Electronic Documents Act

PGHD: Patient-Generated Health Data

PHI: Personal Health Information

PHIPA: Personal Health Information Protection Act

PHIPAA: Personal Health Information Privacy and Access Act

PHIA: Personal Health Information

PNS: Parasympathetic Nervous System

pNN50: NN50 divided by the number of RR intervals

PPG: Photoplethysmography

PSS: Perceived Stress Scale

REB: Research Ethics Board

ROC: Received Operating Characteristic

RF: Random Forest

RMSE: Root Mean Square Error

RMSSD: Square root of the mean sum of the squares of differences between NN intervals in heartbeat

SampEn: Sample Entropy

SD: Standard Deviation

SDNN: Standard deviation of all NN intervals in heartbeat

SES: Socioeconomic Status

SDA: Sleep Dataset with Only Apple Features

SDAW: Sleep Dataset with Withings and Apple Features

SDW: Sleep Dataset with Only Withings Features

SDS: Sleep Dataset with Withings and Apple Only Sleep Features

SMOTE: Synthetic Minority Over-sampling Technique

SNS: Sympathetic Nervous System

SPSS: Statistical Package for Social Sciences

SVM: Support Vector Machine

SYS: Systolic Blood Pressure

Temp: Temperature

U.S.: United States

VLF: Very Low Frequency

UML: Unified Modeling Language

USM: User-Specific Models

WHO: World Health Organization

XML: Extensible Markup Language

Chapter 1 - Introduction

The field of public health has the goal of protecting and improving the health of communities and populations[1]. To accomplish this, public health agencies and programs make use of data to make informed, evidence-based decisions. One of the main sources of data are surveys, typically based on subjective and/or self-reported information [2–6]. While these methods have been widely and successfully used, self-reported data may be subject to significant limitations including social[7–11] and recall bias,[2,3,10–15] loss due to follow-up,[10,11,16] delays between collection and reporting,[17–19] and costs/logistics.[10,18]

Recently, society has experienced a number of technological advancements that make people's lives easier and better. Among these, mobile and wearable devices – such as smartphones, smartwatches, smart rings, wireless scales, among others – have become increasingly popular. Notably, these technologies contain sensing equipment capable of monitoring vital signs, environmental variables, and behavioural metrics, such as: heart rate (HR), sleep, blood pressure (BP), temperature, among others. For example, the Apple Watch Series 4 or higher is equipped with an electrode on the device's digital crown that is able to take a 1-lead, 30 seconds ECG when users place their finger on it [20,21]. Similarly, the same device can measure HR through photoplethysmography (PPG), using green LED lights at the bottom of the device paired with photodiodes to detect blood flow in the wrist [22].

In this manner, mobile, wearable, and Internet of Things (IoT) technologies could potentially be used as additional survey and assessment tools, [23–25] collecting objective data that can mitigate limitations in traditional self-reporting methods, as evidenced by a recent study that used surveys and the Apple Watch data to study heart rate changes in COVID-19 patients.[26]

These technologies are consumer-level, i.e., they can be acquired by regular individuals in a population and not only by research specialists. Further, they have had an amazing adoption rate, with 32 million Canadians owning a smartphone [27] and almost 4 million owning a fitness wearable device [28]. This means that a plethora of health metrics can be collected continuously using these devices, providing rich and useful information for public health agencies. However, little has been done in terms of designing, developing and deploying systems that explore these devices as data collection tools to complement traditional public health surveillance efforts.

The research in this thesis looks to advance this field by presenting the development of a mobile health platform (MHP) targeted at public health. This platform collects and stores data from mobile and wearable devices – specifically from Apple Health (AH), [29,30] a popular repository of health data from sensors, collecting information from smart devices that can be connected to Apple operating systems – and uses it to predict the prevalence of a condition in a population. Specifically, the condition focused on for this work is stress, which was termed the "Health Epidemic of the 21st Century" by the World Health Organization (WHO)[31,32], and as such represent a major public health issue of interest – for instance, the prevalence of chronic stress is increasing worldwide (e.g., over 25% of adults report stress as impairing proper functioning[33]) and can lead to cardiovascular diseases, hypertension, diabetes, among others[34–36]. In addition, stress information is also typically collected through self-report, thus presenting itself as an ideal use case to evaluate a platform that collects objective health data. If stressed individuals can be detected in near real-time by personal devices, interventions can then be applied effortlessly and with precision – for example, the device could prompt the user to open a meditation app for relaxation in case the person is stressed.

The following chapters describe how the ML-based stress prediction models were built, how the mobile health platform was developed and used as a monitoring and collection tool, and lessons learned throughout this process, with the goal of providing public health agents with additional tools and directions for surveillance and protecting population health.

1.1 Structure of the Thesis

This thesis' structure is organized as a series of papers which sequentially present the development of a mobile health platform that extracts Apple Health data for public health, its use in a pilot study that collects real-life stress-related data, and the creation of ML-based models for stress prediction.

Papers were submitted for publication at the time of writing and are presented with amended revisions from their original version following feedback by the thesis committee. Chapters 6 and 7, already published at the time of writing, were included as published. References throughout the thesis were integrated in a single bibliography section and figures/tables numbered in a sequential order. Additional figures and tables included in Supplementary Material for publication are included here in the appendixes, with the text indicating which appendix has the resource.

Each chapter included a foreword which discusses the publication in the context of the thesis, helping to transition from the previous chapter. After the foreword the paper is presented. Finally, in the end of each chapter, an additional section discusses how results of the paper contribute to the thesis as a whole and present additional material that were not included in the papers due to space constraints as applicable. The chapters are as follows:

Chapter 1 contains the introduction, including the structure of the thesis, motivation, and the literature that motivates the papers that follow. Chapter 2 discuss the research goals, including the main research question, while Chapter 3 provides details into objectives, related sub-questions and how each of the chapters in this thesis answers these.

Chapter 4 to 9 correspond to one manuscript each, with individual chapters containing a piece of the full work in this thesis. They include a viewpoint arguing for the use of smart technologies in public health (Chapter 4), a paper detailing the modelling and development of the MHP (Chapter 5), statistical and preliminary Machine Learning analyses (Chapters 6 and 7, respectively) and finally the full analyses and discussions, divided in two papers (Chapters 8 and 9). Chapter 10 is a discussion and conclusion section, detailing the contributions of this thesis, limitations and future directions.

1.2 Motivation

Smart technologies can minimize limitations in traditional data collection efforts. However, they are not currently and consistently being used with an integrated data collection ecosystem in this context. There is a gap in the field of public health surveillance, as agencies are not able to access large volumes of personal, diverse, continuous, and near real-time data to better understand population health. The motivation of this thesis, then, involves the study of new ways to support public health surveillance based on off-the-shelf, consumer-level smart devices that are prevalent and ubiquitous among the population. The overall goal is to understand how these devices can support public health agencies – i.e., entities (generally from governments) that focus on understanding, protecting and improving the health of individuals and communities in a population – through the creation of a mobile health platform (MHP) that will collect and store health data from smart devices, and using data from this platform to study a condition in a population with advanced analytics.

In this manner, the main research question is: *Can smart technology data from AH, efficiently collected and analyzed, improve survey design and support public health surveillance?* To answer this question, the areas that the thesis focuses on are:

- Complementing traditional public health data collection efforts with smart device data, specifically from Apple Health.
- Using a mobile platform to collect Apple Health data as a pilot public health surveillance mobile ecosystem.
- Predict the prevalence of a condition (stress) in a population using advanced analytical methods to handle large and complex data.

The overarching motivation of this thesis, connecting these 3 areas, is to investigate how traditional public health surveillance efforts can be improved through the use of mobile and wearable smart devices. In the following sub-sections, literature related to each component that serves as the basis of the work is presented. Since some of these relate to specific surveys or programming and development tools, grey literature is used when applicable (e.g., to reference specific documentation regarding mobile app development).

1.2.1. Canadian Public Health Surveys

This sub-section describes 3 major Canadian surveys that were investigated for this work. These surveys collect indicators of interest for public health agencies, and as such are used to illustrate how traditional self-reported data could be complemented with the use of smart technologies.

The Canadian Health Measures Survey (CHMS) is a Canadian survey that is voluntary and aims to collect information related to respondents' health, behaviours, diet, among others, with the goal of improving the health of Canadians and better prevent or treat diseases or conditions [6]. The survey is conducted every 2 years; for this thesis, all information from CHMS is pertaining to the most recent completed cycle at the time of writing, Cycle 6, from 2018-2019 [37,38]. The CHMS is comprised of an hour-long interview in the respondents' home (questionnaire topics include, for example, chronic conditions, sleep patterns, physical activity, among others) and a visit to a temporary clinic to collect physical measures such as blood and urine samples and additional self-reported data [37,38]. After the clinician visit, participants are asked to wear an activity monitor for a week [37,38].

The Canadian Community Housing Survey (CCHS) is an annual cross-sectional survey, conducted through electronic questionnaires or in-person/phone interviews, focusing on information related to healthcare services use, health status and conditions, and lifestyle and mental health, to support public health surveillance and research [4]. For this thesis, all information from CCHS pertains to the most recently completed annual component, from 2022.

The Physical Activity, Sleep and Sedentary Behaviour Indicator (PASS), is an annual indicator of the domains of physical activity, sleep, and sedentarism, as indicated by its name, and it is composed of data from several other Canadian surveys including the CCHS and the CHMS [39]. Data in this thesis pertains to the 2017 edition of PASS.

Several of the data collected by these surveys, such as sleep data, physical activity, blood pressure, among others, could potentially be collected through smart devices. As mentioned, however, these devices are not currently being used in this context.

1.2.2 Apple Health, HealthKit Application Programming Interface (API) and Developer Tools

Apple Health is a health data repository that aggregates information from mobile and wearable devices that are designed to be compatible with the platform. This typically includes Apple products, but products from several other manufacturers are designed with Apple Health integration in mind and can be used in conjunction with it, such as devices manufactured by Withings [40]. The repository can be accessed by the user through an iPhone app or a web dashboard.

Apple provides developers with several APIs which can be used to access resources in devices. For example, the MapKit API allows developers to embed Apple Maps in their applications. The HealthKit API, in turn, allows developers to access – with user consent – health data stored in the AH repository. In this manner, it is possible for third-party developers, including researchers, to create apps that collect and store users' health data [29,30]. With the XCode software, developers can create iOS apps [41] using the Swift programming language.

It is important to note that, while this document will focus on AH and related available data at the time of publication, this is not the only mobile health data collection tool or repository available to developers. In addition, several of the devices used in the studies (e.g., devices from Withings, as will be described in later chapters) are compatible with other systems such as Android.

1.2.3 Stress

Stress is the body's normal response to an unexpected situation interpreted as a threat, triggering the body's fight-or-flight response and allowing the individual to deal with the extenuating circumstance. When stress is detected, the sympathetic nervous system (SNS) signals the adrenal glands to release several hormones, such as cortisol, that will help the individual to deal with the situation, causing physiological alterations such as increased heartbeat and glucose levels in the bloodstream with the goal of generating more energy. Once the situation is resolved, the body should ideally return to its normal, pre-stressed state through the parasympathetic nervous system (PNS), re-establishing homeostasis and characterizing an acute stress response [34,42,43]. As a survival mechanism, stress is healthy, helping us to identify and handle threats; however, in daily life, people are exposed to a variety of stressors, not all of them physical in nature (e.g., work or family stressors), which constantly trigger the body's response. This wreaks havoc in several bodily systems, including the gastrointestinal, reproductive, respiratory, immune and cardiovascular [34]. Individuals with high levels of chronic stress are at higher risk for conditions such as hypertension, cardiovascular disease and stroke, among others [34,35].

The prevalence of stress worldwide was high even before the COVID-19 pandemic, and it is increasing. As discussed, the WHO has called it the "Health Epidemic of the 21st Century" [31,32].

Stress results in quantifiable physiological and behavioural changes [35]. For example, the autonomous nervous system (ANS), composed of the SNS and PNS, is directly related to HR, and so this data can help with discriminating stress.

In addition to HR, typically modelled as beats per minute, the heart rate variability (HRV) is also widely used for stress quantification [35,44]. HRV measures the time interval between consecutive beats – the larger the interval, the more resilient or capable to handle stress a person is [44]. Temperature can also be used to discriminate stress [35,45], as well as physical activity [46–48], weight [49,50], blood pressure [35,51] and sleep [46,52,53].

It should be noted that, in addition to these physiological changes, the stress response also has a psychological, subjective component relating to how an individual perceives a threat, i.e., perceived stress. Physiological stress signals do not always correlate to perceived stress, which may vary per individual [54].

1.2.4 Machine Learning

To process increasingly complex, varied, and large data, such as the signals and data types described above, novel methods of advanced analytics are required. One field that deals with the analysis of data, supported by recent computer science and Artificial Intelligence (AI) advancements, is Machine Learning (ML), encompassing "techniques that fit models algorithmically by adapting to patterns in data" [55].

ML models are built on a set of training data and evaluated with a set of test data. There are three types of techniques [55]:

- Supervised learning: involves identifying patterns between variables (known as predictors or features) and measured outcomes (known as classes) to maximize accuracy in predicting these outcomes. The input data consists of examples containing vectors of features and, for each vector, a label indicating what class the vector belongs to.

- Unsupervised learning: involves pattern detection without designation of an outcome of interest.
- Semi-supervised learning: a “combination” of both methods for prediction when a large amount of data is missing.

Stress prediction involves using data labelled as “stress” and “no stress” (for a binary classification problem) to train and test the models, consisting of a supervised learning problem. Can et al. [35] provided a review of stress detection using smartphones/wearables in real-life settings. Widely used methods for stress detection are *random forests (RFs, an ensemble of Decision Trees)* and *Support Vector Machines (SVMs)*[35,56,57]. An additional review conducted for this work and presented in later chapters also found those methods to successful in handling stress-related physiological data, as well as other advantages (such as interpretability for RFs). Further, to provide guidance for researchers, Barro and Amorim [58] evaluated the accuracy of 179 implementations of classifiers on 121 datasets. They found that RFs and SVMs generally achieve the best results. For these reasons these models were chosen for stress prediction in this thesis and so their theoretical underpinnings will be expanded below, first with an explanation of Decision Trees (DTs), which serve the basis for RF, followed by the RF and SVM models.

1.2.4.1 Decision Trees

DTs are hierarchical models in which decision rules are used to perform recursive partitioning on the feature vector space of a problem until a feature subspace with a single class is achieved [57,59–61]. For example, for the feature space in Figure 1, Figure 2 shows a DT that partitions the space forming subspaces. Branch nodes in the tree are parents of two children nodes, while leaf nodes have no children nodes. A decision rule is represented by lines connecting each node and, by following decision rules, DTs achieve a prediction. In Figure 1, by following the rule that X is higher/equal than 1 from Figure 2, the algorithm predicts the class “S”.

To decide the best split values for the feature subspace, several scoring criteria exist. A popular approach is measuring impurity reduction (a 100% purity in a node means the feature subspace contains only one class) by assessing the difference between the impurity in the parent node and the average impurity in the child nodes [59,62,63].

Through these criteria, feature selection is embedded in DTs, requiring fewer data pre-processing [63]. A major advantage of DTs is its interpretability: the output of the method, in addition to the probability of prediction, is the tree itself. Therefore, DTs can help with data interpretation [59,63]. This is particularly important in public health where evidence-based results are fundamental [64].

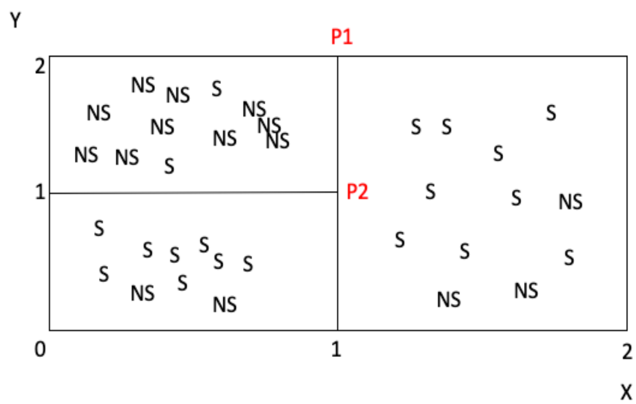


Figure 1: Feature space for features X and Y

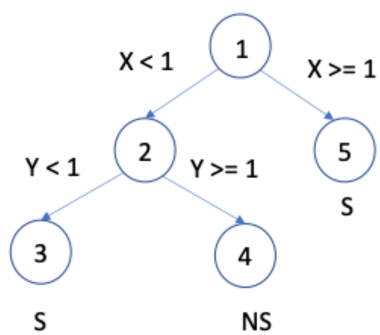


Figure 2: Decision Tree for features X and Y

1.2.4.2 Random Forests

Ensemble methods are a popular ML approach to increase generalization and improve accuracy, in which models are trained on subsamples of the training dataset and their prediction is combined [55,59,61].

RFs are an ensemble method consisting of multiple DTs built on random subsamples (with resampling) of the training dataset, a method known as bagging. The observations left out in the subsamples are called out-of-bag and used for testing the trees. The use of bagging causes DTs to be diverse and capture complex relationships in the data when compared to a single DT [59,61,63,65].

To further increase this diversity, the trees are built on a random subset of features. The prediction of an RF is the average prediction from individual trees using the out-of-bag observations [65].

RFs take advantage of the fact that the format of DTs are largely dependent on the training dataset, and that DTs make few assumptions on the real mapping function between the features and outcomes (unlike a linear regression, for instance, which assumes a linear mapping): the individual DTs built with bagging and random features will have good predictive accuracy while being diverse enough to account for complex structures and variations in the data.[62,63] RFs show high predictive accuracy in several cases [55,66].

By making use of a large number of DTs, RFs are not inherently as interpretable as DTs; while the method provides high accuracy, it does not generate insights into relationships between features[65]. The interpretability of RFs come from a measure of variable importance which assesses the importance of each feature relative to the outcome. Variable importance provides important insights into the relative value of features, including highly correlated ones: if a feature is correlated to a more powerful predictor, random subsampling ensures that there are trees where the powerful predictor is not included [65].

1.2.4.3 Support Vector Machines

To explain SVMs, let’s assume a two-dimensional feature vector plotted as shown in Figure 3 with a line separating two classes. Orange points above the line belong to one class, while blue points below the line belong to a different class. In three-dimensional space, this line becomes a plane and extrapolating this concept to n-dimensions feature spaces it becomes a hyperplane [67]. The hyperplane is selected as the line “in the middle”; in other words, the hyperplane separates the two classes with maximum distance to each of the feature vectors. This maximum-margin hyperplane optimizes an SVMs' accuracy [63,67]: in Figure 4, the dashed and dotted lines are too close to orange/blue points respectively; the solid line is the best boundary. Points farther away from this line have a higher probability of being accurately predicted.

A hyperplane may not be able to divide the data (Figure 5). To handle these "errors" the hyperplane decision boundary must be relaxed, allowing some points to cross it. This relaxation is a parameter called soft margin (parameter C), defined by the user to control how many points are allowed to cross the hyperplane and be misclassified. The soft margin should be flexible to account for variation in the data but not so broad as to allow for many misclassifications [63,67].

SVMs also make use of kernel functions to deal with linearly inseparable data. The kernel function adds dimensions to the data, such as turning a one-dimensional problem into a linearly separable two-dimensional one (Figure 6). The kernel function projects data into higher dimensions to become separable, allowing SVMs to handle linear and non-linear data similarly to DTs/RFs [57,63,67].

For every dataset, a kernel function exists that makes it linearly separable. However, adding dimensions causes the boundaries between classes to fit rigorously to training data, causing overfitting (meaning that the model was too closely fit to the training dataset, capturing noise/error but not the actual relationships in the data; in other words, the model predicts the training data with high accuracy but is not generalizable for new data) [55,59,63]. A kernel function should allow data to be separable without introducing too many dimensions. A major limitation in SVMs is that this function is typically decided by trial-and-error with standard kernels. More rigorous methods to test kernels exist, but they are resource-intensive and do not guarantee that a function not considered in testing will not perform better [67].

Unlike DTs/RFs, SVMs may require robust feature selection and standardization before the learning stage to define the best features to be analyzed and reduce the size of the input data [67]. In terms of interpretability, the hyperplane cannot be visualized for more than 3 features. Therefore, interpretability is more challenging [67].

The next chapter will further discuss the research goals that guided this thesis.

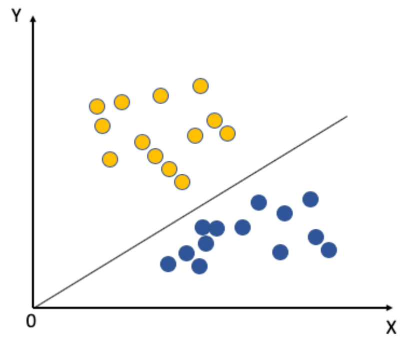


Figure 3: Two-dimension feature space

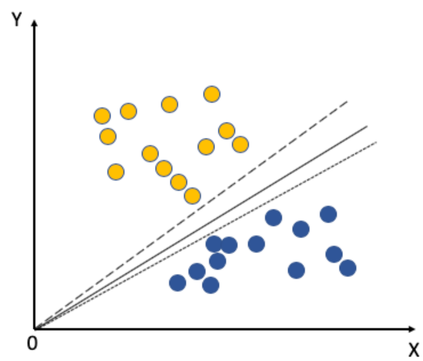


Figure 4: Alternative decision boundaries

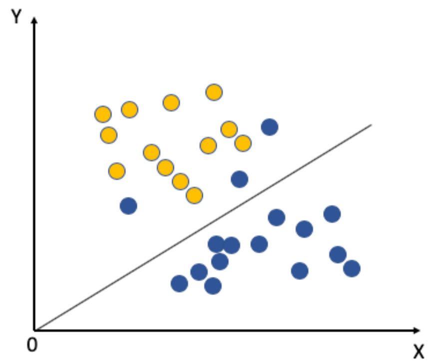


Figure 5: Linearly inseparable two-dimensional feature space

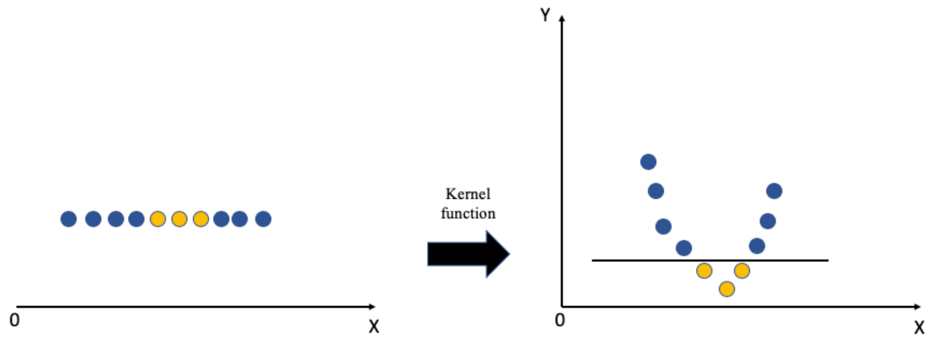


Figure 6: Kernel function (squared values) turning a linearly inseparable 1-dimensional problem into a linearly separable 2-dimensional problem

Chapter 2 – Research Goals

Data collected from smart technologies could complement traditional public health surveillance efforts. This work will focus on investigating the use of these devices for public health and the prevalence of a condition (stress) in the population. Specifically, the development of a mobile system to potentially inform and support next-generation public health surveillance will be presented, as well as variables collected with mobile devices that might complement health surveillance metrics.

The objective is to provide public health agencies and workers, health scientists and researchers with potential new tools in their arsenal to gain insights into and improve the health of individuals, communities and populations while capitalizing on advancements in smart technologies, mobile devices and remote sensing. The main research question, as mentioned is: *Can smart technology data from AH, efficiently collected and analyzed, improve survey design and support public health surveillance?*

To answer this question, the 3 major goals of this work are:

(1) Identify data being collected by major Canadian surveys that can be collected with AH, including stress-related variables to inform further objectives.
(2) Develop an MHP that collects objective data from off-the-shelf smart devices supported by AH.
(3) Establish how data from the MHP can be used to study the prevalence of stress in a population by:
    - Developing ML stress prediction models using collected data from the platform.
    - Examining correlations between physiological measures and perceived stress using collected data from off-the-shelf devices.

As discussed in Chapter 1, smart technologies and mobile devices are used by a large part of the population and have embedded sensors that collect a variety of health metrics, many stress related. Traditional surveillance methods have been widely used to gain insights into the health of populations, leading to intervention and protective measures, but they might be subjected to limitations related to self-reported data that might be mitigated through the careful use of objective sensor information from smart technologies. A mapping of variables collected by public health surveys (Goal 1) that could be collected with AH is, therefore, necessary to better understand and illustrate how AH data can improve surveillance.

In addition, while there are a lot of separate studies that use mobile and wearable devices as tools to collect data, as will be seen in subsequent Chapters, there is currently a lack of an ecosystem that integrates mobile and wearable data and use it to support public health initiatives. The pilot MHP presented in this dissertation (Goal 2) was developed to provide insights and directions into how such a system could be developed using consumer-level technologies and products.

Finally, a major part of public health surveillance is the analysis of data to inform decision-making and interventions. With this in mind, the data collected through a pilot study using the MHP is used to predict stress in a population, a major and increasingly prevalent health condition worldwide (Goal 3). ML was chosen as the main analysis method, as it is ideal to deal with large volumes of complex and varied data, while auxiliary statistical analyses were conducted to better understand the data.

The objectives outlined above were distilled into research questions that guided this research. These are outlined in the next Chapter, which also details how the papers described in subsequent chapters framed the research process.

Chapter 3 - Research Questions and Objectives

This chapter provides an overview of the process followed in this research program and the research questions that led to the program, deriving from the goals presented in Chapter 2. In this manner, the goals are further detailed and the questions that frame each objective are presented.

3.1 Apple Health, Major Canadian Surveys and Stress

This goal aims to provide an overview of variables and indicators collected by major Canadian public health surveys that can be collected with AH, as well as an overview of stress, and what stress-related data can be collected with AH.

The specific research questions for this objective are:

1. What data are currently being collected by major Canadian surveys?
2. What data that composes major Canadian surveys can be collected with AH and its associated sensors?
3. What is stress?
4. What data types collected through AH can be used to measure stress?

Question 1 and 2 are answered by the paper in Chapter 4 – specifically in Tables 2-5), while questions 3 and 4 are answered throughout Chapters 6 to 9 which deal with the pilot study that uses AH data to predict stress with ML.

3.2 Development of the MHP

The second objective aims to develop an iOS app that connects to AH and, using Apple's HealthKit API, extracts available stress-related health data with user consent.

The specific research questions for this objective are:

1. What are the necessary requirements and tools for the development of the MHP?
2. How can an MHP be developed that allows easy and flexible access to AH data?

These questions are mainly answered in Chapter 5, which details the elements and tools used to develop the MHP as well as the solution's architecture, modelling and infrastructure. The discussion section in Chapter 8 provides lessons learned from applying the MHP in practice that shed further light into the second question.

3.3 Applicability of the MHP for Stress Prediction

This objective aims to apply ML algorithms to stress-related data from the MHP to create stress prediction models that classify between stressed and non-stressed conditions. This, in turn, will be helpful for public health agencies to determine the prevalence of stress at an individual and a population level. Further, the study will also examine if there are any statistically significant correlations and relationships in the data collected using the MHP.

The specific research questions for this objective are:

1. What is the accuracy of ML stress prediction models based on the MHP data?
2. What features are the most important for these models?
3. Is the data collected with the MHP correlated with perceived stress?

These questions are partly answered in Chapter 7, which shows preliminary ML analyses for ECG features. Chapters 8 and 9 describe the pilot study in detail and the several approaches to develop the models, their accuracy, and feature importance. Chapter 6 presents correlations of the data with perceived stress and additional statistical analyses related to question 3.

Chapter 4 - Advantages and Challenges for Integrating Mobile Health Technologies into Public Health Surveillance

4.1 Foreword

Major challenges in discussing the use of smart technologies – including mobile and wearable devices – for public health surveillance are twofold: first, it is not immediately clear which data from major public health surveys can be complemented by data available in mobile health systems. Second, equity is a major public health principle [68]. However, not every individual has equal opportunity to technology access, use and benefits due to a variety of reasons, from older adults having increased anxiety regarding technology use [69] to lower income individuals not being able to acquire devices [70].

Therefore, public health surveillance efforts intending to integrate smart technologies in their design should be aware of these limitations and plan their study accordingly. For example, collecting older adult data from mobile technologies in a population might not be a recommended approach; on the other hand, as we shall see, data from younger populations should in general be representative of that population.

In addition, the barriers to technology adoption and use are decreasing – for instance, the goal of the Government of Canada is to provide access to high-speed internet to all Canadians by 2030 [71]. Therefore, investigating the use of mobile health systems that integrate data collected from smart technologies into public health now will help us to establish standards, best practices and guidelines regarding the possible effectiveness and use of these methods for surveillance in the future. Further, benefits from the use of mobile health methods may potentially help with policy changes that focus on increased and equitable access to these technologies.

The viewpoint paper presented in this chapter clarifies, quantifies and expand upon the two major challenges described above. A mapping of AH data that could complement 3 major Canadian public health surveys – CCHS, CHMS and PASS detailed in section 1.2.1 – is presented, helping to clarify how mobile and wearable data can support traditional methods of data collection. An overview of studies that use mobile health technology for health research is also provided to showcase the benefits of these methods. In addition, quantitative information on the use of mobile and wearable device users from the 3 largest mobile/wearable companies in Canada are presented, describing the characteristics of these users which can aid with targeted interventions. Finally, major barriers to digital health equity are discussed, including income, age, geographical location and ethnicity.

4.2 Leveraging Mobile Health Technologies for Public Health: A Viewpoint

4.2.1 Abstract

Traditional public health surveillance efforts are generally based on self-reported data. Well-validated, these methods may nevertheless be subjected to limitations such as biases, delays, and costs/logistical challenges. An alternative is the use of smart technologies (e.g., smartphones and smartwatches) to complement self-report indicators. Having embedded sensors that provide zero-effort, passive and continuous monitoring of health variables, these devices generate data that could be leveraged for cases in which the data is related to the same self-report metric of interest. However, some challenges must be considered when discussing the use of mobile health technologies for public health to ensure digital health equity, privacy and best practices. This paper provides an overview of research involving mobile data for public health, including a mapping of variables currently collected by public health surveys that could be complemented with self-report, challenges to technology adoption and considerations on digital health equity – with a specific focus on the Canadian context. Population characteristics from major smart technology brands – a) Apple, b) Fitbit, and c) Samsung – and demographic barriers to the use of technology are provided. We conclude with public health implications and a discussion that public health agencies and researchers should leverage mobile health data while being mindful of current barriers and limitations to device use and access. In this manner, data ecosystems that leverage personal smart devices for public health can be put in place as appropriate, as we move towards a future in which barriers to technology adoption are decreasing.

4.2.2 Introduction

Public health surveillance is the collection, analysis, and dissemination of data to improve population health [72–74]. These data types are the most important source of information to support decision-making and interventions by public health agencies. One of the main sources of data are surveys [2,3]. However, self-reported survey data may have significant limitations related to self-report including social [7–11] and recall biases [2,3,10–15]. These challenges can produce misleading results: for example, Canadian self-reported BMI data were significantly lower than BMI measured directly in a representative sample of adults [75], which can be explained by biases and limitations in self-report [24,75]. Other potential limitations include delays between collection and reporting [17–19], and costs/logistics [10,18].

In this context, an alternative is the use of mobile, wearable, and Internet of Things (IoT) technologies – such as smartphones, smartwatches, and wireless scales – as additional or complementary survey and assessment tools [23–25] which could possibly mitigate some of these challenges, as evidenced by a recent study that used surveys and Apple Watch data to study heart rate (HR) changes in COVID-19 patients [26].

Smart technologies have had an amazing adoption rate, with 32 million Canadians owning a smartphone [27] and almost 4 million Canadians owning a fitness wearable device [28]. Notably, smart technologies have sensors that provide zero-effort monitoring of vital signs, environmental variables, and behavioural metrics, such as heart rate, sleep, and blood pressure (BP), among others [76]. For instance, Apple Health (AH) [29,30], a popular source of health data from sensors, collects information from smart devices that can be connected to Apple operating systems – such as smartwatches, wireless blood pressure cuffs, wireless scales, sleep tracking mats, among others. Sensors manufactured by Apple as well as from different manufacturers can integrate with AH and read and write data to and from it. In this manner, a diverse environment of sensors can be integrated with Apple's data repository.

These data are typically very large and can often be accessed at relatively low costs. Further, the data can be composed of individuals who traditionally may not participate in health studies. Sensor data are also collected continuously, providing richer and more representative objective information which could potentially be used to complement traditional public health self-reporting and reveal new insights into the behaviour of individuals in real-life environments [77].

Velmovitsky et al. [78] provides an example with the Canadian Health Measures Survey (CHMS), a major Canadian public health survey consisting of an interview with the respondent, a visit to a clinic for exams and physical measures, and the use of an activity monitor for a week. While not a traditional surveillance program, the CHMS and similar surveys provide self-reported indicators of interest for public health agencies and so can be used to illustrate the potential of mobile health data to complement traditional self-report. Indeed, several of the CHMS measures, both taken at the clinic and self-reported, could be complemented with data from smart technologies, such as body composition, heart rate, sleep behaviour, and physical activity. In addition to providing additional information, this data could potentially minimize the aforementioned limitations of biases, costs and delays. Further, using data that is passively and continuously collected by personal devices for long periods can provide more accurate and representative data than the weekly fitness tracker [77,78].

However, there are still challenges that need to be overcome if smart, personal devices are to be used for public health, including technological, ethical, and societal challenges. One of the tenants of public health is equity [68]. In the context of smart technologies, *digital health* equity is achieved when individuals have equal opportunity to "benefit from the knowledge and practices related to the development and use of digital technologies to improve health" [79]. Digital health equity can be compromised as not everyone has equal and fair access to technology.

These challenges and limitations must be clearly stated and recognized for public health entities to understand the potential pitfalls of using smart technologies in surveillance. By identifying and being mindful of these, it is possible to plan accordingly and integrate new tools and technologies within public health, moving towards a scenario in which these issues are mitigated and smart technologies could be used to complement data collection.

The goal of this paper is to provide an overview of the potential, as well as limitations, of the use of smart devices for public health, making the case that these tools may improve traditional surveillance methods but careful consideration must be taken before their use. A particular focus on the Canadian context and technology adoption barriers, access and equity will be provided by considering the characteristics of populations that widely use smart devices as well as populations that do not use, or do not have access to, these tools. In addition, a mapping of variables collected in major Canadian surveys that could potentially be gathered with AH is provided to support our view.

4.2.3 Apple Health and Canadian Surveys

Several companies now produce devices capable of capturing data in line with health metrics traditionally found in public health data collection efforts. A major challenge is to identify which data from major surveys can be complemented by data available in mobile health systems. To highlight the existing overlap, we compiled the variables that can be collected with AH (iOS 15.1) and presented them in Table 1 [29,30]. We then compared these to the ones currently captured by 3 major Canadian public health surveys.

Here we present a summary of which AH data could supplement the Canadian Health Measures Survey (CHMS) (Table 2), Canadian Community Housing Survey (CCHS) (Table 3), and Physical Activity, Sedentary Behaviour and Sleep indicators (PASS) (Table 4 and 5 for adults and children respectively) [80],[39], answering the questions of which data are currently being collected by major surveys and that could be complemented with AH. For these analyses, we used the most recently completed survey cycle. Where self-report metrics are composed of many questions, we included examples of these questions and indicated which AH variables could potentially be used to complement the metrics.

As can be seen, several metrics including information on activity, symptoms, sleep, and biological characteristics (e.g., height and weight), among others, can be objectively complemented by AH. These data may also provide more granular and detailed information, complementing traditional public health initiatives based on self-report with more objective data that can be used to gain further insight into the health of populations.

However, it is important to note that – as we will see in the following sections – there are many challenges involved in the use of AH and smart technologies for public health, and that need to be considered and addressed to ensure digital health equity.

4.2.4 Application of Mobile Health Technology in Public Health

Several studies have begun to apply mobile device data for health research. While some of these are not focused on surveillance efforts per se, they highlight how mobile and wearable devices can potentially be used to collect data and gain insights into the health of individuals and study the prevalence of conditions in a population.

An interesting study evaluated the levels of physical activity from players of the popular Pokémon Go mobile app utilizing data from AH and found that the game is associated with short-term physical activity increase, particularly among more sedentary individuals [81]. To collect the data, participants were asked to take screenshots of their AH screen. It is important to note that data can be directly accessed from Apple Health using the HealthKit Application Programming Interface (API), which allows third parties to access – with user consent given in the device – the health data stored in users' AH app, providing powerful tools for researchers to optimize data collection [29,30].

In this context, one of the challenges with mobile health research, as pointed out by Hicks et al. [77], is that researchers typically need multi-disciplinary experience in computer science and health research in order to use these tools to their full potential in their studies. If one wishes to use the HealthKit API, for example, it would be necessary to program a data collection script that uses Apple's programming language, Swift. In other words, researchers looking to use mobile and wearable data for public health need to have knowledge in at least two disparate fields, healthcare (to design proper studies, analyze and interpret the data) and computer science (for data collection with mobile devices)– and having such multi-disciplinary knowledge may be challenging. In case computer science expertise is lacking, public health researchers may be required to find more creative ways to collect the data, as shown in the previous study with the screenshot requested of users. Another path available to collect the data without the need for coding is to export the data directly through the Apple Health app in the Extensible Markup Language (XML) format, although that also may require coding skills to handle the data inside the file.

Also of note, to make it easier for researchers to conduct studies using mobile technology, Apple has introduced the ResearchKit framework enabling the creation of visual consent flows, customizable surveys, and active tasks [30]. An example is the mPower app, developed with ResearchKit, which collects iPhone gyroscope data to better understand Parkinson's disease. Initial results included approximately 10,000 enrolled participants, providing a continuous flow of data from several individuals that consented to their data being used by health researchers around the world [23]. To use ResearchKit, however, research teams should also have knowledge of mobile technology development. This leads to a challenge and opportunity, in that conducting health informatics research needs to involve a multi-disciplinary research team with collaboration between the fields of public health and computer science.

Hicks et al. [77] describes several large-scale observational studies that use commercial mobile and wearable devices, including a study by the authors themselves which used data from over 700,000 activity-tracking app users in 100 countries. This study concluded that inequality in the physical activity levels between different countries is a stronger predictor of obesity than activity levels in the country. The authors point out that novel sources of data from consumer apps allow researchers to gain new insights into the health and behaviors of individuals. This can be enhanced by linking mobile data with other sources, such as administrative datasets. In addition, the approach of using smart technologies, including leveraging data from pre-existing devices, allows the collection of larger observational datasets than were previously thought possible, and could even be used to identify natural experiments in a population.

Further, according to Hicks et al. [77], even if a population is not well-represented in a dataset, it is possible that if the data is large enough there could still potentially be a statistically significant number of participants that follow population distributions and allow for methodologically sound analyses. However, the authors are quick to point out challenges with this approach such as inaccuracy of sensors and missing data. Inequities in technology access may also lead to selection bias as individuals that use the technology or app may not be representative of the general population.

Missing data might be a particular problem for studies dealing with real-life data collection, with a lot of factors outside the researcher's control (e.g., errors in measurement due to movement or caused by individuals forgetting to wear the device or not wearing it correctly) [21,35]. In this case, careful processing of the data must be made, including data imputation algorithms or removing the missing intervals [35].

Velmovitsky et al.[78] discussed the role of Big Data in precision medicine and public health. In particular, an overview of different Big Data types is provided, which include omics, clinical, social (i.e., social media data), patient-generated health data (PGHD) (data from personal smart devices), environmental and demographic data. Among challenges related to mobile health research, difficulty in linking PGHD with clinical data is highlighted as a lot of medical and administrative information may be siloed in providers' systems which are not typically interoperable and cannot be integrated, in addition to security and privacy issues. The authors also suggest areas that could be improved with the use of Big Data, such as disease surveillance. A recent example of the benefits of mobile devices for this field is the previously mentioned observational study that used surveys and Apple Watch data to identify COVID-19 patients [26]. The standard deviation of the interbeat interval of normal sinus beats (SDNN), a heart rate variability (HRV) metric, differed significantly in the 7 days prior to and after COVID-19 diagnosis compared to uninfected periods, suggesting that the Apple Watch could potentially be used as a predictive tool for COVID-19.

Due to its ability to generate large datasets, mobile research can also be used in conjunction with Artificial Intelligence methods, such as Machine Learning (ML), which learns patterns in data to make predictions. Indeed, ML predictive models work best with large datasets, which can be collected through smart technologies. As an example, several studies used ML to forecast COVID-19 incidence, using data from sources such as Google's mobility dataset [82]. Applying ML methods will also require further multi-disciplinary knowledge in computer and data science.

Several efforts are also underway to create ecosystems that allow users to register their devices and continuously donate data for research. For example, the ecobee company, producer of a smart thermostat device, has launched the Donate Your Data program [83,84], which allows thermostat owners to anonymously share their data with researchers. The Ubiquitous Health Technology Lab has conducted studies using this dataset [85–87] and has recently deployed a web platform that allows individuals to access study information and enroll their personal Fitbit and ecobee devices. Once enrolled, data from the devices is collected by the Lab once a day [88].

Recently, Velmovitsky et al. [21] developed a mobile platform that collects Apple Watch ECG data through HealthKit to predict stress levels using ML. By quantifying stress levels, public health agencies could potentially apply interventions such as notifying users or asking if they would like to open a meditation application. Of note, this study gave devices to participants rather than use data already collected from their personal devices, so the dataset used was not particularly large – and so did not represent surveillance, but rather worked as a pilot study to illustrate the benefits of mobile and wearable applications to public health. Preliminary results of the study achieved model accuracies (trained using the entire dataset according to several demographic factors) of around 55-60%, consistent with the low end of state-of-the-art for ML stress prediction models using real-life data. The models had high specificity, accurately identifying when an individual is not stressed but were less successful in predicting when an individual was stressed. The process of collecting HealthKit data in this study and the mobile application are described elsewhere [89,90]. Interestingly, much like the previous work by Hirten et al. [26], this study also found SDNN to be one of the most important features, in this case for predicting stress.

There have also been several studies that compare the accuracy of mobile devices to gold standard measurements. Hart et al. [91] found that the activPal Professional device and the Bouchard Activity Record (a self-report log that assesses time spent sitting, lying, standing and in physical activity) showed moderate to high agreement and correlation for total and concurrent time spent walking and in sedentary behaviour. However, it is not always the case that the results are successful. A study comparing the ActiGraph device with the International Physical Activity Questionnaire (IPAQ) found low to moderate correlations with IPAQ overestimating sitting and vigorous activity, for instance [92]. In fact, a systematic review of wearables found that data may be under- or overestimated in several devices and models [93], and a study found that the Fitbit Flex went against the ActiGraph GT3X+ in reporting steps in free-living conditions (differences increased with the number of steps taken) [94].

Regarding the Apple Watch, the heart's RR intervals measured with the device during relaxation and stress states were shown to have high reliability and agreement with signals obtained from the Polar H7 chest strap [95], suggesting that heart data from

Apple Watch are accurate. There is also limited but promising evidence on the accuracy of Apple Watch sleep data [96]. Lastly, it is important to note that there is growing evidence of inaccuracies in the use of photoplethysmography (PPG) green light signaling in many wearables for individuals with darker skin tones compared to those of lighter skin tones, which may introduce biases in the analyses [97].

It is often challenging to compare the accuracy of mobiles and wearables as studies tend to use different metrics for assessing validity/reliability, making the comparison between devices difficult [93]. In addition, the speed at which new device models are released, or systems updated, causes studies to quickly become obsolete, especially if there are significant differences between the sensors and algorithms used to measure data [24,94]. Differences in models may limit the applicability of mobiles/wearables in population-level studies over time, as it is not possible to guarantee comparability, and that remains a significant issue [24,94]. To date, the literature suggests that several mobile health devices and metrics are in line with gold standard measurements in public health. However, some devices continue to fall below the standard and further development will be required before they can be implemented in public health.

4.2.5 Technology Adoption: Facts and Challenges

As can be seen by the aforementioned studies, although Apple Health is the focus of our review of major Canadian surveys, it is not the only available method for mobile data collection. In this section, we present results relating to the adoption of smart technologies with a focus on the major Canadian companies of wearable devices that also collect health metrics. As of 2022, three companies dominate the Canadian wearable device market: Apple, Samsung, and Fitbit, respectively [98]. A summary is shown in Table 6.

Garmin and Samsung have a similar market share (13%), but we focus on Samsung due to its focus on smart devices and health metrics. In addition we also describe major challenges to technology adoption in the Canadian context. The implications for public health are discussed in the next section.

4.2.5.1 Characteristics of Major Mobile and Wearable Companies

4.2.5.1.1 Apple

Apple has the largest wearable market share in Canada, with 41% of Canadians using an Apple device [98]. Compared to other brands, Apple has the largest share of 18–29-year-old users (35%), and there is no significant difference between the percentage of female and male users. On the other hand, Apple has the lowest share of 50–64-year-old users compared to other brands at approximately 15% [98].

Apple users are typically more educated compared to other companies, with 38% having a bachelor's degree and 19% having a master's or doctoral degree. In addition, 50% of Apple users have a high monthly income. The type of community that Apple users are in also differs from other companies, with approximately 68% living in larger cities. Of note, 67% of Apple users report accessing the internet through their smartwatches compared to other wearable users (53%) [98]. This could be related to Apple users having a higher income which allows them to obtain more devices with Internet access (as further detailed in the subsection on income below). Apple's popularity has grown, with Apple wearable users increasing by 11% in the past 2 years [98].

On a global scale, Apple's geographical segment is primarily in the United States (U.S.) and urban cities [99]. The company's marketing strategy is aimed at consumers with high purchasing power and career focus, such as those in professional executive positions. Further, Apple relies on the loyalty of customers who typically continue to purchase all their electronics from the company. Apple's brand value was approximately $947 billion USD, due in large part to customer loyalty and brand recognition as an exclusive, luxury product [100].

4.2.5.1.2 Fitbit

Fitbit produces the second most used wearable in Canada, with a market share of 38% [101]. In Canada, 61% of Fitbit users are female and Fitbit has the highest share of 50–64-year-old users compared to any other wearables at 32%. On the other hand, Fitbit has the lowest share of 18–29-year-old users compared to other wearables at 20%. 42% of Fitbit users have a high monthly income and 61% of users live in larger cities. A large share of Fitbit users are educated, with 30% having a bachelor's degree and 16% having a master's or doctoral degree. Of note, Fitbit users were found to access the internet less often through their devices (46%) compared to the average wearable user (53%). Further, Fitbit users have a higher percentage of hiking activities in comparison to any other wearable users (26 % and 17%, respectively). They also engage more in aerobic and cardio physical activity compared to other wearable users (23 % and 13%, respectively), suggesting that Fitbit users are in general more interested in fitness and exercising [101].

On the global scale, as of 2021, Fitbit has sold over 127 million wearables worldwide with 111 million registered users [102]. For reference. Apple has the highest share of the wearable device market with 160 million sold globally. Unlike Apple, Fitbit's wearable market share has declined 12% in the past two years [101]. Fitbit was valued at 2.1 billion U.S. dollars when Alphabet Inc purchased the company in 2021 [102].

4.2.5.1.3 Samsung Wearables

Samsung produces the third most used wearable in Canada with a market share of 13% [103]. Compared to other brands, Samsung has the highest share of 30–39-year-old users (30%) and most users are male (57%). In addition, 47% of users have a high monthly income and 21% have a master's or doctoral degree. Several Samsung users (60%) also live in larger cities in Canada [103].

On a global scale, Samsung's geographical segment is primarily in the Asian market sector and urban cities. Globally, Samsung's main users are adults, and their products are marketed toward society in general. Samsung has products that are for users with both low and high purchasing power, expanding the brand's target market [99].

4.2.5.2 Security, Privacy and Data Ownership Issues

Security and privacy issues must be addressed during health data collection, and are particularly important challenges to the collection, storage, and use of data from smart technologies.

In Canada, the Personal Information Protection and Electronic Documents Act (PIPEDA) regulates the collection, use and disclosure of personally identifiable information (PII) for private sector organizations involved in a commercial activity. This includes pharmacies, providers, and laboratories, among others [104]. This federal act applies to all types of PII [105,106]. Several provinces have adopted health sector laws dealing with personal health information (PHI), some of which are deemed substantially similar to PIPEDA and taking precedence in these provinces (Table 7) [105,107,108]. PIPEDA still applies when PHI is transferred provincially/nationally.

PIPEDA is based on ten principles (Table 8) [107,109]. The principle of Safeguards mandates that PII "be protected by security safeguards appropriate to the sensitivity of the information" [106]. Provincial healthcare acts define similar protective measures; for example, PHIPA states that health information custodians must "take steps that are reasonable in the circumstances to ensure that personal health information … is protected against theft, loss and unauthorized use or disclosure…" [109,110]. To inform health custodians, the Information and Privacy Commissioner of Ontario (IPC) listed recommended safeguards (Table 9) [111]. In other words, Canadian privacy laws require that health custodians protect PII by appropriate measures. What constitutes an appropriate measure will depend on the sensitivity of the information and the custodian's circumstances, including type/size of the organization and if the data are shared with third parties [105,111]. Organizations must obtain informed consent for the collection, use and disclosure of PII and state their purposes for data collection (Table 8).

Different countries and regions have different regulations. The Health Insurance Portability and Accountability Act (HIPAA), which applies to subsets of health custodians in the U.S., offers a similar but more comprehensive list of technical, physical and administrative safeguards [112], while the General Data Protection Regulation (GDPR) regulates the handling of PII in the European Union and is considered to be some of the most comprehensive privacy legislation in the world. GDPR and HIPAA guidelines can also help Canadian health custodians to understand their security needs and implement adequate safeguards.

The issue of security and privacy is further complicated when ownership of the data is considered. In other words, is the data owned by the individuals who generated the data, corporations who manufactured the data collection devices, or other stakeholders? While a comprehensive discussion of data governance is outside the scope of this paper, this legal and societal issue still needs to be addressed when discussing data collection with mobile and wearable technologies. Velmovitsky et al.[113] highlights potential trust issues in the data collection process between corporations, third-party solutions, individuals, providers and regulations. In particular, individuals using such technologies need to trust that the corporations (e.g., Apple, Samsung, Fitbit) and research and personal applications (e.g., fitness and research apps) are using the data only for the purposes originally consented to. Regulatory agencies need to make sure regulations (such as the ones discussed above) are being respected by these entities.

Micheli et al.[114] further highlights asymmetries of power regarding technology corporations having large and unrestricted access to data which could result in privacy violations such as the case of a Facebook data leak which enabled Cambridge Analytica to use these data improperly for voter profiling [115]. The authors further highlight governance models proposed in literature, including a) data sharing pools in which data is digitally shared between partners, with contracts stipulating the conditions of use; b) data cooperatives, which are similar to sharing pools but with more involvement of data subjects, which have more control over the data sharing process; c) public data trusts, which involve a public entity accessing citizen and company data; d) and personal data sovereignty, in which data subjects have complete control over their data and sharing permissions. The open issue of data ownership is particularly important in the context of research and public health surveillance, as the gateways offered by companies such as the HealthKit API are controlled by these entities, and as such access to data could potentially be charged in case mobile health data is increasingly used for research.

Researchers who collect, use and disclose PII for non-commercial activities are not typically subject to PIPEDA but must still get approval from appropriate review ethics boards, which typically also require safeguards according to the sensitivity of the data [105,110]. Further, public health agencies are generally not subject to PIPEDA but to federal, provincial and territorial laws dealing with personally identifiable information in their region [116]. For example, the Public Health Agency of Canada is subjected to the federal Privacy Act, which delineates individual privacy rights in relation to the federal government [117].

It should also be noted that applications that allow data sharing between smart technologies typically have their consent mechanisms. For example, the HealthKit API requires that individuals first give consent to each data type for this data collection [29], as shown in Figure A3. In addition to these mechanisms, researchers and public health agencies should still obtain consent for data collection following applicable regulations.

In summary, any third-party entity collecting health data for commercial purposes (e.g., private healthcare providers, mobile app developers) are subject to PIPEDA and must respect the principles to protect personal information. Researchers and public health agencies are subjected to their own ethics boards and privacy regulations, which typically also require obtaining consent for data collection and use.

4.2.5.3 Internet Access by Canadians

As the use of mobile and wearable data in health continues to grow, researchers must acknowledge and address inequalities in technology access. Disparities may lead to selection bias as individuals that use the technology or app may not be representative of the general population.

In 2020, nearly 6% of Canadians did not have internet access at home. 63% felt no need for it, 26% found the service costs too high, and 13% found the equipment costs prohibitive [118]. From 2015 to 2023, Canada's internet users have steadily increased, reaching 36 million. In other words, approximately 94% of Canada's population has access to the internet [119]. In this manner, while internet access remains a barrier for some of the population, most Canadians currently have access to the internet, and this number is projected to increase.

The standards set by the Canadian Radio-television and Telecommunications Commission (CRTC) for internet connectivity are a minimum download speed of 50 Mbps and an upload speed of 10 Mbps. An internet speed of 50 Mbps or more allows Canadians to perform multiple online activities and have various devices connected to the internet at once [118]. Approximately 72% of Canadian households have achieved the CRTC standards for internet connectivity. Regarding mobile data, 80% of Canadians reported having a personal mobile data plan, with only 1.5% reporting a mobile data plan without Internet connection [118]. Without good internet speed and connection, therefore, the use of mobile and wearable devices for data collection are severely limited. The Government of Canada has set a goal of having 98% of Canadians with access to high-speed internet by 2026 and 100% of Canadians by 2030 [71].

In this manner, internet access is a major factor in smart technology adoption. Many barriers limit the use of technology and high-speed internet for Canadians such as household income, age, geographical location, and ethnicity. Understanding these barriers to adoption is important when using mobile health data to address health inequalities, and they will be expanded in the next sections.

4.2.5.4 Income

The most prevalent barrier to internet access for Canadians is low household income. Deloitte's digital equity report found that, of survey participants who did not have a data plan, 68% reported high costs as a barrier [70]. In addition, household income can alter an individual's perception of technology and digital services: people earning over $150,000 CAD annually were likelier to agree (74%) that internet and new technologies had a positive impact on their lives compared to those earning less than $40,000 CAD (49%) [70]. This disparity can manifest itself in differences in the quality of internet service and the range of digital tools individuals have access to.

Indeed, internet speed and household income are highly correlated. Households with lower incomes are more likely to fall below CRTC thresholds compared to households with higher incomes. In fact, most households earning less than $40,000 CAD annually do not meet the CRTC target, which is 19% higher than the national average and 28% higher compared to the highest income category [70]. Furthermore, families with an annual household income of $200,000 CAD or more had access to internet speeds that were approximately 30 Mbps faster than those with an income of less than $20,000 [70]. With an additional 30 Mbps, a household could connect to three more devices, including phones and computers. Without high-speed internet, therefore, digital health equity is severely affected, as individuals may not be able to have internet connection or access to smart technologies.

4.2.5.5 Urban vs Rural Geographical Locations

Due to Canada's vast size and dispersed population, individuals residing in rural and remote geographical locations face additional challenges in accessing high-speed internet. Rural and remote regions encounter distinct challenges concerning internet cost and speed, which can be attributed in large part to Canada's vast size and dispersed population.

Within Canadian Census Metropolitan Areas or Census Agglomerations (CMA/CA), 95% of households had access to a home internet connection. For households residing outside a CMA/CA, this figure drops to 88%. An even greater geographical disparity exists when one considers access to high-speed internet with download speeds of 50 Mbps or more. Only 48% of people living outside CMA/CAs meet the CRTC target compared to 76% of respondents residing within these areas, and 73% have a mobile data plan outside CMA/CAs compared to 81% residing within these areas [118].

In particular, Canadian Indigenous communities are under-represented in the digital landscape: only 39% of First Nation reserves in Canada met the CRTC threshold for high-speed internet [70]. Researchers must consider these geographical disparities in digital equity when implementing and collecting data from devices.

4.2.5.6 Older Adults

Historically, older adults have used less technology than younger populations [120]. In general, older adults typically have higher anxiety when using new technologies, and declining visual, motor, hearing and cognitive impairments can affect technology acceptance [69]. 1 in 3 older adults aged over 75 reported frustration when using unfamiliar technologies [70].

Older adults may also not want to use any additional applications, being limited to call and messaging functions, and decide not to have a device due to cost [69]. In the three brands detailed above, individuals aged 50-64 composed the lowest share for Apple and Samsung (15% and 14%, respectively), with a better representation in Fitbit at 32%. However, a poll conducted during the pandemic revealed that the number of Canadians aged 65 and older who own a smartphone increased (65% in 2020 from 58% in 2019), and 83% of owners use it daily [121]. The pandemic also caused older adults to increase their technology use in general, for example through video calls or using social media to message family and friends [122]. Indeed, in 2020, 72% of Canadians aged over 65 revealed that they now feel confident using technology [122], indicating that, although age could be a barrier to technology adoption, it seems that it is diminishing. This also remains true in other geographical locations, such as in the U.S., where smartphone ownership and social media use among older adults is increasing with 61% of seniors aged over 65 owning a smartphone and 45% using social media, increasing to 83% and 73% for individuals aged 50-64 [120].

Although the technology access gap for older adults is becoming smaller, barriers still remain that need to be addressed to ensure equitable access, including individual (e.g., physical aging, sensory impairments, cognitive limitations) and technological barriers [123]. Indeed, one of the most frequent obstacles to accessing technology among older adults is physical aging, particularly hearing and vision impairments. Decreases in motor control, such as tremors in the hands, also difficult the use of devices, especially with small screens. Lack of experience with technology, perception of their own proficiency in using devices, and a general aversion to technology may also difficult adoption among other adults [123]. Further, technological functional barriers, such as small screen and text sizes, as well as complex functionalities that are intuitive and/or assume the user has prior experience with the technology, also negatively impact adoption. The limited availability of technology devices suited or adapted to older adults also poses a challenge. Finally, the cost associated with purchasing electronic devices and data is also a significant factor limiting technology adoption; while this is true for most populations, it particularly affects older adults if they rely on a restricted or fixed income, such as government pensions [123].

4.2.5.7 Ethnicity

An individual's ethno-cultural background can affect technology adoption. For example, individuals of Middle Eastern, North African, and South Asian descent are more likely to view cost as a significant barrier to accessing digital technologies compared to both the national average and individuals of European descent [70].

Further, racially motivated discrimination, cyberbullying, and harassment are prevalent in online spaces. Individuals of Indigenous, Middle Eastern, Asian, or African descent are likelier (60%) to have experienced online bullying or discrimination compared to individuals of Caucasian or European descent (25%) [70].

On the other hand, it is also important to note that Canadians of Indigenous, Middle Eastern, Asian, or African descent utilize the internet as a means of connecting with others who share their ethnocultural background and finding individuals who can relate to their experiences. In Canada, 80% of Indigenous individuals utilized the internet to maintain regular connections with members of their community, which is significantly higher than the national average of 50% [70]. Nevertheless, digital inequity remains a significant challenge for Indigenous communities across Canada, relating to historical failures in recognizing Indigenous rights, which have contributed to longstanding and wide-ranging socio-economic disparities between Indigenous and non-Indigenous populations.

4.2.6 Implications for Public Health

The sections above illustrate how data from smart technologies could possibly be used to support health sciences and public health efforts, complementing self-report metrics with objective data and leveraging personal devices to provide continuous, passive data collection from large populations.

Indeed, studies focusing on mobile datasets are already underway. While traditional data collection methods, typically focused on self-report, have years of use and validation – as evidenced by the major surveys in use – they might be complemented by objective sensor data collected passively through smart technologies that are widely adopted by Canadians and worldwide. This could allow researchers and public health specialists to access a larger volume of continuous, real-world, and real-time data for decision support and to gain new insights into the health of individuals and populations. These stakeholders should still respect applicable security and privacy regulations, mandates from review ethics boards, and obtain user consent.

Given that, a system that allows users to share their data with public health organizations – such as a larger version of the mobile platform suggested by Velmovitsky et al. [21] and by the Ubiquitous Health Technology Lab [88] – might be beneficial in supporting health efforts, research and interventions.

However, scientists conducting studies based on mobile health population data must be aware of the barriers and challenges identified above and take them into account when designing their studies and collecting data. For example, these methods may not be appropriate for certain ethnicities, older populations or individuals that possess lower income or are located in geographically distant areas. It is possible that the Apple Health repository, which was the focus of this paper, may not be the best approach for data collection depending on the population and study design, and data from other health repositories should be considered (e.g., for physical activity studies, it is possible Fitbit might be a better choice). Digital health equity concerns must be addressed to ensure all populations benefit from the use of smart devices, and in case populations without equal access to mobile technologies or internet are part of the study, special care must be taken to avoid digital exclusion.

In addition, to mitigate some of these challenges, scientists can consider the characteristics of the population that uses each device. For example, if lower-income populations are the focus of a study, it would make more sense to leverage Samsung personal devices than Apple ones, as Apple products target individuals with higher purchasing power. On the same token, Fitbit devices can better target individuals with a prior interest in physical activity. Careful consideration must also be taken to ensure the devices have prior evidence suggesting the collected data has good agreement and correlation with gold standard measurements.

Another important factor to note is that a lot of these barriers are already recognized by the Government of Canada and other stakeholders, and efforts are in place to mitigate or eliminate them over the next years. As mentioned, the Canadian government has a goal of enabling all Canadians to access high-speed internet by 2030 [124], and older adults are becoming increasingly comfortable with technology. In a few years, it is likely that some of the barriers may not be present anymore. In addition, new technologies – such as the implementation of 5G – can greatly reduce some of the challenges, for example by increasing the number of devices that can be connected to a single point as well as the speed of data collection and transfer [125].

If researchers and public health organizations develop methods and guidelines for collecting and using personal health data now – with careful considerations on current issues regarding adoption, access, privacy, ownership and equity – they will be more prepared to use this information in the future if and when those barriers are greatly diminished. For example, new standards and best practices can be created on how to obtain, process, secure and store mobile health data; how to deal with different device models; how to obtain consent in studies using mobile data; or special considerations for certain populations. In addition, more studies using mobile and wearable data might generate more evidence to decision-makers on whether these devices could improve health of populations.

The lack of interoperability between devices, which adds additional complexities, must also be considered: if a public health agency develops a system that extracts data from Fitbit devices, for instance, the same data pipeline will not work for Samsung or Apple products. Different programming languages, APIs, and protocols need to be used. This may affect potential studies as having larger and more robust datasets from a larger population would lead to more representative and quality data. The issue of interoperability must be carefully considered when designing and creating population-wide data collection systems and should also be integrated into the development of standards and best practices.

In conclusion, the sooner population-wide data collection and surveillance systems using mobile technology are in place, the sooner specialists can take advantage of these data. On the same token that contact tracing apps had to be developed quickly during the COVID-19 pandemic as there was not a wide system available and in place for managing disease spread before it in most countries, by being proactive, anticipating the need and investigating these systems in parallel to the process of eliminating barriers to device and internet access, health scientists will be better prepared to deal with the challenges of tomorrow while taking advantage of the opportunities that the future will bring.

Table 1: Variables Collected by AH, per Group

Group	**Variable Name**
Activity	Flights Climbed

Activity	Steps
Activity	Walking + Running Distance
Activity	Active Energy
Activity	Exercise Minutes
Activity	Resting Energy
Activity	Stand Hour
Activity	Cardio Fitness(VO2 max)
Activity	Workouts
Activity	Cycling Distance
Activity	Downhill Snow Sports Distance
Activity	NikeFuel
Activity	Pushes
Activity	Swimming Distance
Activity	Swimming Strokes
Activity	Wheelchair Distance
Activity	Stand Minutes
Activity	Move Minutes
Mindfulness	Mindful Minutes
Nutrition	Biotin
Nutrition	Caffeine
Nutrition	Calcium
Nutrition	Carbohydrates
Nutrition	Chloride
Nutrition	Chromium
Nutrition	Copper
Nutrition	Dietary Cholesterol
Nutrition	Dietary Energy
Nutrition	Dietary Sugar
Nutrition	Fiber

Nutrition	Folate
Nutrition	Iodine
Nutrition	Iron
Nutrition	Magnesium
Nutrition	Manganese
Nutrition	Molybdenum
Nutrition	Monounsaturated fat
Nutrition	Niacin
Nutrition	Pantothenic Acid
Nutrition	Phosphorus
Nutrition	Polyunsaturated Fat
Nutrition	Potassium
Nutrition	Protein
Nutrition	Riboflavin
Nutrition	Saturated Fat
Nutrition	Selenium
Nutrition	Sodium
Nutrition	Thiamin
Nutrition	Total Fat
Nutrition	Vitamin A
Nutrition	Vitamin B12
Nutrition	Vitamin B6
Nutrition	Vitamin C
Nutrition	Vitamin D
Nutrition	Vitamin E
Nutrition	Vitamin K
Nutrition	Water
Nutrition	Zinc
Sleep	In Bed

Sleep	Asleep
Body Measurements	Body Fat Percentage
Body Measurements	Body Mass Index
Body Measurements	Height
Body Measurements	Weight
Body Measurements	Lean Body Mass
Body Measurements	Waist Circumference
Body Measurements	Basal Body Temperature
Body Measurements	Body Temperature
Body Measurements	Electrodermal Activity
Heart	Heart Rate
Heart	Resting Heart Rate
Heart	Walking Heart Rate Average
Heart	Cardio Fitness (VO2 Max)
Heart	Cardio Fitness Notifications
Heart	Peripheral Perfusion Index
Heart	High Heart Rate Notifications
Heart	Low Heart Rate Notifications
Heart	Irregular Rythym Notifications
Heart	Heart Rate Variability (HRV)
Heart	Blood Pressure
Heart	Electrocardiogram (ECG)
Symptoms	Abdominal Cramps
Symptoms	Acne
Symptoms	Appetite Changes
Symptoms	Bladder Incontinence
Symptoms	Bloating
Symptoms	Body and Muscle Ache
Symptoms	Breast Pain

Symptoms	Chest Tightness or Pain
Symptoms	Chills
Symptoms	Congestion
Symptoms	Constipation
Symptoms	Coughing
Symptoms	Diarrhea
Symptoms	Dizziness
Symptoms	Dry Skin
Symptoms	Fainting
Symptoms	Fatigue
Symptoms	Fever
Symptoms	Headache
Symptoms	Heartburn
Symptoms	Hot Flashes
Symptoms	Loss of Smell
Symptoms	Loss of Taste
Symptoms	Lower Back Pain
Symptoms	Memory Lapse
Symptoms	Mood Changes
Symptoms	Nausea
Symptoms	Night Sweats
Symptoms	Pelvic Pain
Symptoms	Rapid, Pounding or Fluttering Heartbeat
Symptoms	Runny Nose
Symptoms	Shortness of Breath
Symptoms	Skipped Hearbeat
Symptoms	Sleep Changes
Symptoms	Sore Throat
Symptoms	Vaginal Dryness

Symptoms	Vomiting
Symptoms	Wheezing
Vitals	Heart Rate
Vitals	Blood Pressure
Vitals	Body Temperature
Vitals	Respiratory Rate
Vitals	Blood Glucose
Vitals	Menstruation
Vitals	Blood Oxygen
Cycle Tracking	Abdominal Cramps
Cycle Tracking	Acne
Cycle Tracking	Appetite Changes
Cycle Tracking	Basal Body Temperature
Cycle Tracking	Bladder Incontinence
Cycle Tracking	Bloating
Cycle Tracking	Breast Pain
Cycle Tracking	Cervical Mucus Quality
Cycle Tracking	Constipation
Cycle Tracking	Contraceptives
Cycle Tracking	Diarrhea
Cycle Tracking	Dry Skin
Cycle Tracking	Fatigue
Cycle Tracking	Hair Loss
Cycle Tracking	Headache
Cycle Tracking	Hot Flashes
Cycle Tracking	Lactation
Cycle Tracking	Lower Back Pain
Cycle Tracking	Memory Lapse
Cycle Tracking	Menstruation

Cycle Tracking	Mood Changes
Cycle Tracking	Nausea
Cycle Tracking	Night Sweats
Cycle Tracking	Ovulation Test Result
Cycle Tracking	Pelvic Pain
Cycle Tracking	Pregnancy
Cycle Tracking	Pregnancy Test Result
Cycle Tracking	Progesterone Test Result
Cycle Tracking	Sexual Activity
Cycle Tracking	Sleep Changes
Cycle Tracking	Spotting
Cycle Tracking	Vaginal Dryness
Hearing	Headphone Audio Levels
Hearing	Audiogram
Hearing	Environmental Sound Levels
Hearing	Noise Notifications
Hearing	Headphone Notifications
Respiratory	Cardio Fitness(VO2 Max)
Respiratory	Forced Expiratory Volume, 1 sec
Respiratory	Forced Vital Capacity
Respiratory	Inhaler Usage
Respiratory	Oxygen Saturation
Respiratory	Peak Expiratory Flow Rate
Respiratory	Respiratory Rate
Respiratory	Six-Minute Walk
Mobility	Double Support Time
Mobility	Step Length
Mobility	Walking Speed
Mobility	Walking Asymmetry

Mobility	Walking Steadiness
Mobility	Stair Speed: Up
Mobility	Stair Speed: Down
Mobility	Cardio Fitness (VO2 Max)
Mobility	Six-Minute Walk
Mobility	Walking Steadiness Notifications
Other	Alcohol Consumption
Other	Blood Alcohol Content
Other	Blood Glucose
Other	Handwashing
Other	Inhaler Usage
Other	Insulin Delivery
Other	Number of Times Fallen
Other	Sexual Activity
Other	Toothbrushing
Other	UV Index
Immunizations	COVID-19 Vaccine Records
Health Details	Name
Health Details	Date of Birth
Health Details	Sexual Activity
Health Details	Blood Type
Health Details	Fitzpatrick Skin Type
Health Details	Wheelchair Distance
Health Details	Medication That Affects Heart Rate

Table 2: CHMS Measures related to Possible AH Variables

Type	**Description/Question Example**	**AH**	**Possible AH Variables**
Clinic Component			

Height	Standing height measured during clinic visit	Yes	Height
Weight	Weight measured during clinic visit	Yes	Weight
Neck Circumference	Neck circumference measured during clinic visit	No	N/A
Waist Circumference	Waist circumference measured during clinic visit	Yes	Waist circumference
Resting blood pressure	Device applied in the clinic to measure resting blood pressure	Yes	Blood pressure
Heart Rate	Device applied in the clinic to measure heart rate	Yes	Heart Rate, Resting Heart Rate, Walking Heart Rate Average, Heart Rate Variability
Vision Assessment	Consists of several tests in the clinic: visual acuity, intraocular pressure, visual field, and retinal photography	No	N/A
Cardiovascular fitness	Measured using the Canadian Aerobic Fitness test, in which individuals go up and down the steps for several minutes to measure "the efficiency of lungs and heart in delivering oxygen to the exercising muscles as well as the efficiency of these exercising muscles in using the oxygen".	Yes	Flights Climbed, Steps, Walking + Running Distance, Active Energy (kcal), Exercise Minutes, Resting Energy, VO_2 Max, Workouts, Oxygen Saturation, Forced Expiatory Volume 1 sec, Forced Vital Capacity, Peak Expiratory Flow Rate, Respiratory Rate

Grip Strength	Measured in the clinic with a device called dynamometer that is squeezed as hard as the individual can.	No	N/A
Sit and Reach	Measured in the clinic, individuals sit on a mat and lean forward at the hips	No	N/A
Bone Mineral Content	X-ray at the clinic	No	N/A
Vertical jumps	Measured in the clinic with two tests: multiple two-legged hoping test, and vertical jump test	No	N/A
Level of Physical Activity	Measures the intensity, time, duration and frequency of the activity with a physical activity monitor wore for seven days following the clinic visit	Yes	Steps, Walking + Running Distance, Active Energy (kcal), Exercise Minutes, Resting Energy, Stand Hour, VO_2 Max, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Weelchair Distance, Stand Minutes
Blood samples	Blood samples from respondent at clinic	No	N/A

Urine samples	Urine samples from respondent provided at home	No	N/A
Household Questionnaire			
General health (GEN)	Using a scale of 0 to 10, where 0 means "Very dissatisfied" and 10 means "Very satisfied", how do you feel about your life as a whole right now?	No	N/A
Health Utility Index (HUI)	Are you usually free of pain or discomfort? How would you describe your usual ability to remember things? How often do you use a wheelchair? Are you able to walk at all?	Yes, for some questions	Abdominal Cramps, Body and Muscle Ache, Breast Pain, Chest Tightness or Pain, Headache, Lower Back Pain, Pelvic Pain, Sore Throat, Wheelchair Use, Audiogram, Environmental Sound Levels, Headphone Audio Levels, Noise Notifications, Headphone Notifications, Walking Speed, Step Length, Six-Minute Walk, Stair Speed: Up, Stair Speed: Down, Six-Minute Walk, Cardio Fitness, Walking Steadiness,

			Walking Asymmetry, Memory Lapse
Chronic conditions (CCC)	Have you had any asthma symptoms or asthma attacks in the past 12 months? Do you have high blood pressure? Do you currently take insulin for your diabetes? In the past month, did you take pills to control your blood sugar?	Yes, for some questions	Inhaler Usage, Blood Pressure, Insulin Delivery, Blood Glucose
Vision (VIS)	Have you ever had glaucoma?	No	N/A
Sleep Apnea	Without the use of sleeping aids, how often do you usually have trouble going to sleep or staying asleep? Using a scale from 0 to 10, where 0 means "no sleepiness" and 10 means "extremely sleepy", how would you assess your sleepiness during a typical day?	Yes, to some questions	Sleep Time In Bed, Sleep Time Asleep, Sleep Changes
Pregnancy (PRS)	Are you pregnant?	Yes	Pregnancy, Pregnancy Test, Progesterone Test
Menopause (MEN)	Have you had a menstrual period in the last 12 months?	Yes, to some questions	Mensturation

Fracture History (FRH)	Have you fallen in the past 12 months?	Yes, to some questions	Number of Times Fallen
Fracture Details (FRD)	Which bone(s) did you break or fracture (on that occasion)?	No	N/A
Medication Use (MEU)	Have you taken or used any other prescription medications in the past month?	No	N/A
Steroids and Osteoporosis Medications (SOM)	Have you ever used steroids administered by inhalation, for example, Flovent, Pulmicort or Vanceril? Do not include nasal sprays.	No	N/A
Height and Weight (HWT)	How tall are you without shoes on?	Yes	Height, Weight
Meat Consumption (MFC)	Now I'd like to ask about the use of omega-3 enriched eggs in the eggs and egg dishes you just reported.	No	N/A
Milk and Dairy Product Consumption (MDC)	hat kind of enriched milk substitutes do you usually drink or use on cereal?	No	N/A
Grain, Fruit and Vegetable Consumption (GFV)	Now, a few questions about grains, fruits and vegetables. Remember, think about all the foods you eat, both meals and snacks, at home and away from home.	No	N/A

Dietary Fat Consumption (DFC)	Remember, think about all the foods you eat, both meals and snacks, at home and away from home.	No	N/A
Water and Soft Drink Consumption (WSD)	How much water, in cups, do you usually drink at home?	No	N/A
Salt Consumption (SLT)	What type of salt is usually used?	No	N/A
Physical activities - Adults (PAA)	In the last seven days, how much time in total did you spend doing vigorous activities that caused you to be out of breath?	Yes	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Weelchair Distance, Stand Minutes, Move Minutes
Physical activities for youth (PAY)	You have reported a total of ^DV_PAYTOTAL minutes of physical activity. Of these activities, were there any of vigorous intensity, meaning they caused you to be out of breath?	Yes, for most questions excluding questions which include location (e.g., In the last seven days, did you use active ways like walking or	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness,

		cycling to get to places such as [school, the bus stop, the shopping centre, work/school, the bus stop, the shopping centre/the bus stop, the shopping centre, work/the bus stop, the shopping centre] or to visit friends?)	Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Weelchair Distance, Stand Minutes, Move Minutes
Physical Activity of Children (CPA)	Over a typical or usual week, on how many days are you physically active for a total of at least 60 minutes per day?	Yes, for most questions excluding questions which include location (e.g., About how many hours a week do you usually take part in physical activity that makes you out of breath or warmer than usual: in your class time at school?).	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Weelchair Distance, Stand Minutes, Move Minutes
Time Spent Outdoors (TSD)	During a weekday, did you go to school (including kindergarten)?	No	N/A
Sedentary Activities (SAC)	In the last seven days, how much of your free time did you spend: reading books, magazines or newspapers, including in	No	N/A

	electronic formats? Include time spent reading as part of your homework, but do not include time spent reading at work, during class time, while travelling in a vehicle or while exercising.		
Neighbourhood Environment (NBE)	What is the main type of housing in your neighbourhood?	No	N/A
Smoking (SMK)	In your lifetime, have you smoked a total of 100 or more cigarettes (about 4 packs)?	No	N/A
Electronic Cigarette (ELC)	Have you ever tried an electronic cigarette, also known as an e-cigarette?	No	N/A
Exposure to Second-Hand Smoke (ETS)	Is smoking allowed inside this home?	No	N/A
Exposure to Second-Hand Vapor (ETV)	Overall, in the past month, how often were you exposed to second-hand vapour inside this home?	No	N/A
Alcohol Use (ALC)	During the past 12 months, that is have you had a drink of beer, wine, liquor or any other alcoholic beverage?	No	N/A
Illicit Drug Use (IDU)	Have you ever used or tried marijuana, cannabis or hashish?	No	N/A

Sexual Behaviour (SXB)	In the past 12 months, have you had sexual intercourse?	Yes, to some questions	Sexual Activity
Birth Control (BCL)	In total, over your lifetime, how many years did you use birth control pills?	Yes	Contraceptives
Maternal Breastfeeding (MBF)	Have you ever given birth?	Yes, to some questions	Pregnancy
Breastfeeding (BRF)	Did you breastfeed your baby?	No	N/A
Pregnancy Information (PRG)	Did [you/she] smoke during [your/her] pregnancy?	No	N/A
Birth Information (BIR)	How much did you weigh at birth?	No	N/A
Breastfeeding Information (BRI)	For how long did [you/she] breastfeed?	No	N/A
Labour market activity minimum - LMAM	Last week, did you work at a job or business? (regardless of the number of hours)	No	N/A
Labour market activity Sublock Labour force status - LMA2	Last week, did you have a job to start at a definite date in the future?	No	N/A
Labour market activity Sublock Class of worker - LMA3	Were you an employee or self-employed?	No	N/A
Industry (LMA4)	What was the name of your business?	No	N/A

Labour market activity Sublock Occupation - LMA5	What was your work or occupation?	No	N/A
Labour Market Hours of Work (LMH)	On average, how many hours do you usually work per week?	No	N/A
Immigration Block (IMG)	In what country were you born?	No	N/A
Aboriginal minimum - AMB	Are you an Aboriginal person, that is, First Nations, Métis or Inuk/Inuit? First Nations includes Status and Non-Status Indians.	No	N/A
Population Group (PG)	You may belong to one or more racial or cultural groups on the following list.	No	N/A
Language Extended (LAE)	Of English or French, which language(s) do you speak well enough to conduct a conversation?	No	N/A
Education Minimum Block with Concept (EDM)	What type of educational institution [are you attending/did you attend]?	No	N/A
Education Sublock School attendance "currently" - ESC1	Are you currently attending school, college, CEGEP or university?	No	N/A

Total Personal Income (TPI)	Can you estimate in which of the following groups your personal income falls?	No	N/A
Total Household Income (THI)	Can you estimate in which of the following groups your household income falls?	No	N/A
Administration Information (ADM)	Was this interview conducted on the telephone or in person?	No	N/A
Activity Monitor	Use of activity monitor for a week	Yes	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Wheelchair Distance, Stand Minutes, Move Minutes

Table 3: CCHS Measures related to Possible AH Variables

Type	Question Example	AH	Possible AH Variables
Respondent Availability	May I speak to [First name of household contact] [Last	Yes	The owner of the phone is the respondent

	name of household contact]?		
Proxy Respondent (PRX)	Is there someone [Minimum age of person providing proxy interview] or older who could provide us with some information on behalf [First name of household contact]?	No	N/A
Verification (VER2)	Are you [First name of specific respondent] [Last name of specific respondent]? What is your date of birth?	Yes	Date of Birth, respondent is the owner of the phone
Date of birth (AGE)	What is [your] date of birth?	Yes	Date of Birth
Sex and Gender (GDRA)	What was [your] sex at birth?	Yes, for some questions	Biological Sex
Relationship with confirmation (RWC)	What is the relationship…of: [Name of specific respondent] ([Age of specific respondent]) to: [Name of secondary respondent] ([Age of secondary respondent])?	No	N/A
Main activity (MAC)	In the past 12 months, was your main activity working at a job or business?	No	N/A

Main activity (MA)	During the past 12 months, what was your main activity?	No	N/A
Main Activity (EDC)	Are you currently attending a school, college, CEGEP or university?	No	N/A
General health (GEN)	In general, how is your health? Thinking about the amount of stress in your life, how would you describe most of your days?	No	N/A
Life satisfaction measures (LSM)	How do you feel aboutyour life as a whole right now?	No	N/A
Pregnancy	Are you pregnant?	Yes	Pregnancy, Pregnancy Test Result, Progesterone Test Result
Height and weight (HWT)	How tall are you without shoes on?	Yes	Height, Weight
Weight perception (WTP)	Do you consider yourself overweight, underweight or just about right?	No	N/A
COVID-19 (COV2)	In the last 3 months, have you experienced any of the symptoms that led you to believe that you had COVID-19, such as fever, headache, sore throat, runny	Yes, for some questions	Headache, Fever, Runny Nose, Sore Throat, Shortness of Breath, Wheezing, Immunizations

	nose, difficulty breathing or tiredness? Have you been vaccinated against COVID-19?		
Vaccination passeport COVID-19 (PVC)	Some public health authorities are considering establishing a COVID-19 vaccination passport or have already done so. Is such a passport a motivation for you to get vaccinated?	No	N/A
COVID-19 (COV3)	If an additional dose of the COVID-19 vaccine is offered to stimulate your immune system or to fight against variants, how likely is it that you would get it?	No	N/A
Chronic conditions (CCC)	Do you currently take insulin for your diabetes? Do you have high blood pressure? Do you currently take insulin for your diabetes? In the past month, did you take pills to control your blood sugar? Do you have heart disease?	Yes, for some questions	Inhaler Usage, Blood Pressure, Insulin Delivery, Blood Glucose, Medication That Affects Heart, Dietary Cholesterol, Mood Changes, High/Low Heart Rate Notifications, Irregular Rhythm Notifications

Chronic conditions (CC1)	Do you have an anxiety disorder? Do you have Alzheimer's disease or any other dementia?	Yes, for some questions	Mood Changes, Memory Lapse, Fatigue
Abilities (WDM)	Do you have difficulty doing any of these activities? Difficulty hearing, even if using a hearing aid Do you have difficulty doing any of these activities? Difficulty walking or climbing steps	Yes, for some questions	Audiogram, Environmental Sound Levels, Headphone Audio Levels, Noise Notifications, Headphone Notifications, Walking Speed, Step Length, Six-Minute Walk, Stair Speed: Up, Stair Speed: Down, Six-Minute Walk, Cardio Fitness, Walking Steadiness, Walking Asymmetry
Injuries (INJ)	In the past 12 months, did you have any of the following injuries? A head injury or concussion	No	N/A
Oral health (OHM)	In general, how would you rate the health of your mouth	No	N/A
Oral health (OHM3)	How often do you usually see a dental professional?	No	N/A

Changes made to improve health (CIH)	In the past 12 months, did you do anything to improve your health? (For example, lost weight, quit smoking, increased exercise.)	Yes, for some questions	Weight, Steps, Walking + Running Distance, Workouts, Cycling Distance
Eating Habits (EAH)	In the past 30 days, how many times did you eat food from a restaurant? In the past 30 days, how many times did you eat the following fruits and vegetables?	No	N/A
Physical activities - adults 18 years and older (PAA)	In the past 7 days, did you do sports, fitness or recreational physical activities? In the past 7 days, on which days did you do these other activities that made you sweat at least a little and breathe harder?	Yes, for the majority of questions excluding questions which include location (e.g., In the past 7 days, did you do any other physical activities while at work, in or around your home or while volunteering?)	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Weelchair Distance, Stand Minutes, Move Minutes
Physical activities for youth (PAY)	In the past 7 days, did you do any other physical activities?	Yes, for most questions excluding questions which include location (e.g., In the	Steps, Walking + Running Distance, Active Energy,

	You have reported a total of [total hours of active transportation + total hours of recreational physical activities + total hours of other physical activity] hours of physical activity. Of these activities, were there any of vigorous intensity, meaning they caused you to be out of breath?	past 7 days, did you do sports, fitness, or recreational physical activities while at [school or day camp, including during physical education classes, during your breaks and any other time you played indoors or outdoors/school, including during physical education classes, during your breaks and any other time you played indoors or outdoors/day camp, including any time you played indoors or outdoors]?)	Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Wheelchair Distance, Stand Minutes, Move Minutes
Use of protective equipment (UPE)	In the past 12 months, have you participated in any of these activities? 1: Bicycling 2: In-line skating or rollerblading 3: Downhill skiing 4: Snowboarding 5: Skateboarding 6: Playing ice hockey	Yes, to questions identifying activities (however AH does not have information on protective equipment use)	Workouts, Cycling Distance, Downhill Snow Sports Distance, Swimming Distance
Sedentary behaviours (SBE)	On a school or work day, how much of your free time did you spend watching television or a screen on any	No (while in theory sedentary behaviour could be measured from smartphones, the questions in CCHS ask how much free time respondents	N/A

	electronic device while sitting or lying down?	spent watching television or a screen, which is not something AH can capture)	
Sleep (SLP)	How long do you usually spend sleeping each night? How often do you have trouble going to sleep or staying asleep?	Yes, for some questions	Sleep Time In Bed, Sleep Time Asleep, Sleep Changes
Current smoking status (CSS)	Have you ever smoked a whole cigarette?	No	N/A
Smoking – past use (SPU)	Have you ever smoked cigarettes daily?	No	N/A
Electronic cigarettes and vaping (ECV)	Have you ever tried an e-cigarette or vaping device?	No	N/A
Electronic cigarettes and vaping 2 (ECV2)	During the past 30 days, on how many days did you vape the following products? An e-liquid with nicotine	No	N/A
Alcohol use (ALC)	Have you ever had a drink in your lifetime?	No	N/A
Medication use — pain relievers (PRM)	In the past 12 months, have you taken any codeine products?	No (AH may not be capable of tracking medication intake but it does allow for pain symptom self-reporting)	N/A
Cannabis use (CAN)	Have you ever used or tried cannabis?	No	N/A

Maternal experiences (MEX)	Are you taking a vitamin supplement containing folic acid? Have you given birth in the past 5 years?	Yes, for some questions	Folate, Iron, Weight, Pregnancy, Pregnancy Test, Progesterone Test, Lactation
Smoking during maternal experience (MXS)	In the 3 months before your pregnancy with [your last child], or before you realized you were pregnant, did you smoke cigarettes?	No	N/A
Alcohol use during maternal experience (MXA)	In the 3 months before your pregnancy with [your last child], or before you realized you were pregnant, did you drink any alcohol?	No	N/A
Flu shots (FLU)	In the past 12 months, have you had a seasonal flu vaccine?	No	N/A
Regular health care provider (RHC)	Which of the following health care providers do you regularly consult with?	No	N/A
Labour market activities (LMAM)	Last week, did you work at a job or business?	No	N/A
Labour market activities (LMAM3)	Were you an employee or self-employed?	No	N/A
Labour market activities (LMAM4)	What kind of business, industry or service was this?	No	N/A

Labour market activities (LMAM5)	What kind of work were you doing?	No	N/A
Labour market activities (LMAM6)	[Excluding overtime, on average, how many paid hours do you usually work per week?/On average, how many hours do you usually work per week?]	No	N/A
Labour market activities (LBF)	Did you have more than one job or business last week?	No	N/A
Telework (LM)	In the past 30 days, in which of these locations did you work the most hours?	No	N/A
Place of birth, immigration and citizenship (IM)	Where were you born?	No	N/A
Indigenous identity (ABM)	Are you First Nations, Métis or Inuk (Inuit)?	No	N/A
Population group (PG)	Are you? [List of Population Groups]	No	N/A
Language	Can you speak English or French well enough to conduct a conversation?	No	N/A
Sexual orientation (SOR)	What is your sexual orientation?	No	N/A
Home care services (HMC)	In the past 12 months, what type of home care services have been received?	No	N/A

Insurance coverage (INL)	Do you have insurance that covers all or part of the cost of your long term care, including home care?	No	N/A
Insurance coverage (INP)	Do you have insurance that covers all or part of the cost of your prescription medications?	No	N/A
Prescriptions cost (PCN)	In the past 12 months, did [you] do any of the following because of the cost of [your] prescriptions?	No	N/A
Food security (FSC)	The food that you [and other household members] bought just didn't last, and there wasn't any money to get more.	No	N/A
Administration information (ADMC)	For which province or territory is your health number?	No	N/A
Total household income (INC)	What is your best estimate of total household income received by all household members, from all sources, before taxes and deductions, during the year ending December 31, [Past year]?	No	N/A

Table 4: PASS (Adult) Measures related to Possible AH Variables

Type	Description	Source	AH	Possible AH Variables

Physical activity guideline adherence	Percentage of adults who meet physical activity guidelines by accumulating at least 150 minutes of moderate-to-vigorous physical activity each week, in bouts of 10 minutes or more	CHMS (2016-2017)	Yes	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Max, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Wheelchair Distance, Stand Minutes, Move Minutes
Total moderate-to-vigorous physical activity amount	Average number of minutes per day adults are engaged in moderate-to-vigorous physical activity	CHMS (2016-2017)	Yes	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Max, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Wheelchair Distance, Stand Minutes, Move Minutes
Occupational physical activity and active chores amount	Average number of hours per week adults report doing physical activities while at work, in or around their home or while volunteering	CCHS (2018)	No	N/A
Leisure time physical activity amount	Average number of hours per week adults report doing sports, fitness or recreational physical activities, organized or non-organized, that lasted a	CCHS (2018)	Yes	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Max, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming

	minimum of 10 continuous minutes			Strokes, Wheelchair Distance, Stand Minutes, Move Minutes
Sports participation amount	Percentage (%) of population who reported regularly participating in any sports during the past 12 months	General Social Survey (GSS) (2016)	Yes	Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes
Active travel amount	Percentage (%) of adults who report walking or cycling to work or school/Average number of hours per week adults report using active ways like walking or cycling to get to places	CCHS (2018)	No	N/A
Intention level	Percentage (%) of adults who, when thinking about the next six months, intend to be physically active	Physical Activity Monitor (PAM) (2014-2015)	No	N/A
Enjoyment level	Percentage (%) of adults who report that physical activity is generally pleasant	PAM (2014-2015)	No	N/A

Confidence level	Percentage (%) of adults who report they are confident that they could regularly do a total of 30 minutes or more of moderate physical activity three or four times a week	PAM (2014-2015)	No	N/A
Physical health status	Percentage (%) of adults who report their health is "very good" or "excellent"	CCHS (2019)	No	N/A
Mental health status	Percentage (%) of adults who report their mental health is "very good" or "excellent"	CCHS (2019)	No	N/A
Presence of parks and recreation facilities	Percentage (%) of adults that "somewhat agree" or "strongly agree" that their neighbourhood has several free or low cost recreation facilities, such as parks, walking trails, bike paths, recreation centers, playgrounds, public swimming pools, etc.	CCHS RR (2011)	No	N/A
Presence of active transport infrastructure	Percentage (%) of adults who report their community has	CCHS RR (2011)	No	N/A

	infrastructure that supports walking or biking (well-maintained sidewalks or designated bike areas for biking)			
Shower access at work	Percentage (%) of adults who report having access to showers or change rooms at or near work	CCHS (2007-2008)	No	N/A
Total sedentary time amount	Average number of hours per day adults spend sedentary, excluding sleep time	CHMS (2016-2017)	Yes	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Max, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Wheelchair Distance, Stand Minutes, Move Minutes
Recreational screen time amount	Average number of hours per day adults report watching television, DVDs, or videos or spending time on a computer, tablet, or other hand-held electronic device e.g. watching videos, playing computer/video games, emailing or surfing the Interne	CHMS (2014-2015)	No	N/A

Nighttime sleep amount	Average number of hours adults report sleeping in a 24-hour period	CHMS (2014-2015)	Yes	Sleep Time In Bed, Sleep Time Asleep
Sleep quality — sleep continuity	Percentage (%) of adults who report having trouble going to sleep or staying asleep "most of the time" or "all of the time"	CHMS (2014-2015)	Yes	Sleep Time In Bed, Sleep Time Asleep , Sleep Changes

Table 5: PASS (Children and Youth) Measures related to Possible AH Variables

Type	**Description**	**Source**	**AH**	**Possible AH Variables**
Physical activity guideline adherence	Percentage (%) of children and youth who meet physical activity recommendations by accumulating at least 60 minutes of moderate-to-vigorous physical activity per day	CHMS (2016-2017)	Yes	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Max, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Wheelchair Distance, Stand Minutes, Move Minutes
Total moderate-to-vigorous physical activity amount	Average number of minutes per day children and youth are engaged in moderate-to-vigorous physical activity	CHMS (2016-2017)	Yes	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Max, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming

				Strokes, Wheelchair Distance, Stand Minutes, Move Minutes
24-hour movement	Percentage (%) of children and youth who meet the Canadian 24-Hour Movement Guidelines for Children and Youth	CHMS (2014-2015)	Yes	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Max, Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Wheelchair Distance, Stand Minutes, Move Minutes
School physical activity amount	Average number of hours per week youth in Grades 6 to 10 report taking part in physical activity that makes them out of breath or warmer than usual during class time at school/Average number of hours per week that parents report their children spend doing physical activity during class time at school	Health Behaviours in School-aged Children (HBSC) (2014) and CHMS (2018-2019)	No	N/A
Sports participation amount (leisure time)	Percentage (%) of Canadian parents who report that their children participated in	Canadian Health Survey on Children and Youth	Yes (although in this case it wouldn't be	Workouts, Cycling Distance, Downhill Snow Sports Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes

	sports in the last 12 months	(CHSCY) (2019)	reported by the parents)	
Active play amount (leisure time)	Percentage (%) of children who accumulate 3 hours or less per week of active play (unstructured physical activity) outside of school	CHMS (2016-2017)	No	N/A
Active travel amount	Percentage (%) of youth who report walking or cycling to work or school/Average number of hours per week youth report using active ways like walking or cycling to get to placesschool/Average number of hours per week adults report using active ways like walking or cycling to get to places	CCHS (2018)	No	N/A
Physical health status	Percentage (%) of youth who report their health is "very good" or "excellent"/Percentage (%) of parents who	CCHS (2019), CHMS (2018-2019)	No	N/A

	report the health of their child is "very good" or "excellent"			
Mental health status	Percentage (%) of youth who report their mental health is "very good" or "excellent"	CCHS (2019)	No	N/A
Level of parental support	Percentage (%) of Canadian parents who report "often" or "very often" playing active games with their children in the past year	PAM (2014-2015)	No	N/A
Level of peer support	Percentage (%) of youth in Grades 9 and 10 who report that most of their friends "often" participate in organized sports activities with others	HBSC (2014)	No	N/A
Level of community safety	Percentage (%) of Canadian parents who identify safety concerns as a barrier to their children's physical activity	CHSCY (2019)	No	N/A
Presence of parks and	Percentage (%) of youth who "somewhat agree" or "strongly	CCHS RR (2011)	No	

recreation facilities	agree" that their neighbourhood has several free or low cost recreation facilities, such as parks, walking trails, bike paths, recreation centers, playgrounds, public swimming pools, etc.			
Supportive policies at school	Percentage (%) of schools that have a committee that overseas policies and practices concerning physical activity (e.g. health action team)	HBSC — Admin (2014)	No	
Sedentary behaviour recommendation adherence	Percentage (%) of children and youth who report meeting sedentary behaviour recommendations by spending 2 hours or less per day watching television or using a computer during leisure-time	CHMS (2018-2019)	No	
Amount of sedentary time	Average number of hours per day children and youth spend sedentary, excluding sleep time	CHMS (2016-2017)	Yes	Steps, Walking + Running Distance, Active Energy, Exercise Minutes, Resting Energy, Stand Hour, Cardio Fitness, Max, Workouts, Cycling Distance, Downhill Snow Sports

				Distance, NikeFuel, Pushes, Swimming Distance, Swimming Strokes, Wheelchair Distance, Stand Minutes, Move Minutes
Recreational screen time amount	Average number of hours per day youth report watching television, DVDs, or videos or spending time on a computer, tablet, or other hand-held electronic device e.g. watching videos, playing computer/video games, emailing or surfing the Internet	CHMS (2018-2019)	No	N/A
Time spent outdoors	Average number of hours per day children spend outside	CHMS (2014-2015)	No	N/A
Sleep recommendation adherence	Percentage (%) of children and youth who report meeting sleep recommendations by obtaining adequate sleep: 9-11 hours per night for ages 5 to 13 years and 8-10 hours per night for ages 14 to 17 years	CHMS (2014-2015)	Yes	Sleep Time In Bed, Sleep Time Asleep

Amount of sleep in 24-hour period	Average number of hours children and youth report sleeping in a 24-hour period	CHMS (2014-2015)	Yes	In Bed, Asleep
Sleep quality — sleep continuity	Percentage (%) of children and youth who report having trouble going to sleep or staying asleep "most of the time" or "all of the time"	CHMS (2014-2015)	Yes	In Bed, Asleep, Sleep Changes
Electronic media in the bedroom	Percentage (%) of children and youth who use electronic devices in the bedroom before falling asleep	CHSCY (2019)	No(although this information can be obtained from devices, it is not from AH)	N/A

Table 6: Characteristics of Apple, Fitbit and Samsung Users

Device	% Market Share	% BSc and Above	% Larger Cities	% Gender	% High Income	Predominant Age (18-64)
Apple	41%	57%	68%	Equal share of male and female users	50% High Income	18-29 years (35%)
Fitbit	38%	46%	61%	61% Female	42% High Income	50-64 (32%)
Samsung	13%	50%	60%	57% Male	47% High Income	18-29 (34%)

Table 7: Canadian provincial health laws deemed substantially similar to PIPEDA

Province	Provincial privacy laws
Ontario	Personal Health Information Protection Act (PHIPA)
New Brunswick	Personal Health Information Privacy and Access Act (PHIPAA)
Newfoundland and Labrador	Personal Health Information Act (PHIA)
Nova Scotia	Personal Health Information Act (PHIA)

Table 8: 10 principles of PIPEDA

Principle	Description
Accountability	PII is the responsibility of the organization that controls it. The organization must designate individuals that are accountable for compliance with PIPEDA's principles.
Identifying Purposes	The purposes for the collection of PII must be stated by the organization before or during collection.
Consent	Consent of individuals whose data is being collected, used or disclosed is required.
Limiting Collection	Collection of PII will be limited to what is necessary to fulfill the purposes outlined by the organization. The information must be collected legally.
Limiting Use, Disclosure, and Retention	Use and disclosure of PII will be limited to the purposes for which it was collected, except when required by law or if consent from the individual is sought again. The information will be retained until the purposes are fulfilled.
Accuracy	PII will be as accurate, up-to-date and complete as required to complete the

	purposes outlined for its collection.
Safeguards	Appropriate safeguards to the sensitivity of the information will be employed by the organization to protect PII.
Openness	Policies and practices concerning the management of PII will be made available.
Individual Access	If it is requested by an individual, organizations must inform individuals of the existence of PII and related uses. An individual can contest the accuracy of the information and have it changed.
Challenging Compliance	Individuals can challenge compliance with the above principles by an organization.

Table 9: Examples of physical, administrative and technical safeguards recommended by IPC

Physical	Locked cabinets
	Restricted office access
	Alarm systems
Administrative	Security clearance
	Confidentiality agreements
	Training
	Regular monitoring of compliance with security policies
Technical	Login and password
	Encryption
	Firewalls
	Malware detection software

4.3 Discussion

The information and point-of-view presented here serves to inform and support the rest of the work in this thesis. Mobile health technologies could increase the access and quality of care provided to individuals and populations. Despite many barriers still remaining, a lot of the challenges and limitations presented in the paper are becoming increasingly reduced, paving the way for a future with digital health equity in which consumer-level mobile and wearable technologies could safely monitor and deliver care become pervasive and ubiquitous. This thesis represents a tentative step towards this direction. In the next chapter, the development and infrastructure of the MHP is presented.

Chapter 5 - Development of a Mobile Health Platform for Individual and Population-Level Surveillance

5.1 Foreword

The next step is the development of the mobile health platform (MHP), an iOS app that uses Apple's HealthKit API to collect Apple Health data and which is used as a tool for the pilot study presented in subsequent chapters. The MHP forms the crux of a potential mobile health data ecosystem to be used in public health surveillance initiatives.

This paper describes the tools used for the creation of the MHP, which are provided by Apple to developers. Brief introductions are given on the HealthKit API, XCode software, Swift programming language and SwiftUI declarative syntax. While these tools are specific to iOS development, other systems have similar methods. The interface of the MHP is presented, as well as the flow of the app.

The multi-modal data that were collected by the MHP were included due to the relationship of extracted features with stress: HR/HRV [35,44], temperature [35,44], physical activity [46–48], weight [49,50], blood pressure [35,51], and sleep [46,52,53]. The devices and data collection modes will be discussed in subsequent chapters.

In terms of infrastructure, the Unified Modelling Language (UML) diagram to create the system is also included. Detailed explanations on the MHP's architecture and classes that were programmed for the creation of the app are included with justifications on why several modelling and development choices were made.

It should be noted that, although AH was the chosen health repository in this study, there are many other personal devices from different manufacturers that collect health data. The previous chapter discussed Fitbit and Samsung in addition to Apple, for instance, and each device has its own characteristics and properties. While this chapter and the following sections discuss data from AH, this technology is a tool for the studies; rather, our focus is on personal mobile devices.

Finally, the database structure is also described, which details how the health data types that were collected in the stress pilot study (described in subsequent chapters) were stored, which additional information related to these were collected and how some of the data types differ in terms of data structures from one another.

5.2 Modelling and Development of a Mobile Health Data Extraction Platform for Public Health

5.2.1 Abstract

Background: The goal of public health is to improve the health of populations. To accomplish this, data is needed to support decision-making. These data are typically obtained through self-report methods that may be subject to biases, costs and delays. In this context, objective data from sensors embedded in personal smart devices present themselves as potential alternative measurement tools.

Objective: The goal of this work is to describe the modelling and development a mobile health platform (MHP) that collects health data from the Apple Health repository, to be used as a tool for real-world data collection.

Methods: The MHP was developed using Apple's XCode software and the Swift programming language based on a real-world pilot study protocol involving stress-related health data from Apple Health.

Results: The MHP interface, modelling and architecture is detailed, including the UML diagram, app flow, and health data type structures. Future directions for new versions of the MHP are also presented.

Conclusions: A similar system to the MHP could be used for public health surveillance, complementing traditional self-report metrics when applicable. Researchers and public health specialists must respect applicable security and privacy regulations when collecting digital health data.

5.2.2 Introduction

The field of public health focuses on understanding and improving the health of populations [126]. To support decision-making, data on individual and population health is needed [72–74]. Data collection efforts are often based on self-report [78],[80] which may be subject to limitations including several biases, delays between data collection and reporting, high costs and logistical issues [2,3,10–15,17,18].

In this context, smart devices could be used as additional monitoring tools. Personal, off-the-shelf, consumer-level devices often come equipped with sensors that are able to collect a variety of health-related data. These data, in turn, could possibly be used to complement traditional public health data collection initiatives. For example, the Canadian Community Housing Survey [4] collects self-reported metrics on physical activity, sleep, and blood pressure, among others that could be complemented with objective sensor data. As an example, the Apple Watch collects data on activity and sleep that could be used in conjunction with subjective, self-reported information on these metrics [89,90]. Given that data from these devices is also generally collected in a passive manner, in real-life environments, and for long periods (i.e., for as long as users wears their devices), they are likely to collect more representative information of users' behaviours and lifestyles.

With this in mind, a pilot study is being planned that uses data from mobile and wearable devices to predict stress in a population through the use of Machine Learning models. To collect data for this study, a Mobile Health Platform (MHP) will be developed. This platform will be an iOS app that extracts data from Apple Health [89] [127], a popular health data repository available in iOS systems and that stores data from Apple devices as well as devices from other manufacturers that possess integration with the system.

The goal of this paper is to describe the modelling and development of the MHP used in the study as a data collection tool. While a prototype, systems similar to the MHP could be leveraged by researchers and public health specialists to collect large volumes of continuous and near real-time data from personal devices with arguably less constraints than traditional data collection efforts. Further, while the MHP is developed for iOS, a similar platform and architecture could be implemented for other systems. We briefly describe the study protocol and present how this app was developed, the platform's architecture and its database structure. The discussion section details future directions for the platform and how such a platform could be implemented in real-world scenarios as an "add-on" to self-reporting applications.

5.2.3 Methods

5.2.3.1 Study Protocol

This section briefly describes the study protocol, with the goal of providing an increased understanding of the modelling choice made for creating the MHP (described in the next sub-section.

In the proposed study, 45 participants will be given an iPhone with the MHP installed and use the following mobile and wearable devices for a period of 2 weeks to collect health data:

- Apple Watch Series 6 with watchOS 8.3, collecting steps, heart rate, heart rate variability, electrocardiogram and sleep data.
- Withings Sleep, collecting sleep and heart rate data.
- Withings Blood Pressure Monitor (BPM) Connect, collecting heart rate and blood pressure data.
- Withings Thermos, collecting temperature data.
- Withings Body+, collecting weight data.
- Empatica E4, collecting heart rate and heart rate variability data.

All these devices, with the exception of the Empatica E4, possess integration with Apple Health, i.e., data collected by these devices will be stored in Apple Health. The Empatica E4 is a research device and as such will be used to complement results with data exported manually.

Participants will be asked to collect data from these devices 6 times a day while also self-reporting their stress levels by answering 8 questions. These questions are based on the 7 questions composing the stress scale of the Depression, Anxiety and Stress Scale (DASS-21) [128] and an additional question asking how users are feeling right now from “Feeling Great” to “Stressed Out” [129]. More details on the questions and answer options can be found in Table B1 in Appendix B.

With this protocol in mind, the MHP should be able to collect the aforementioned health data from Apple Health, differentiating the devices used for data collection (since different devices may collect the same type of data). In addition, the MHP should allow users to self-report their stress levels several times a day.

5.2.3.2 Mobile Health Platform

Apple provides developers with several tools to develop apps for its systems (e.g., iOS, macOS). The tools that are used in this study to develop the MHP are:

- **HealthKit** The HealthKit Application Programming Interface (API) allows developers to access user data from the Apple Health repository (after user consent is obtained through the API).

- **XCode and Swift** In order to develop apps for Apple systems, developers can use XCode, Apple's native program for creating iOS apps [41]. The programming language supported by XCode and used in this work is Swift [130].

- **SwiftUI** To create app interfaces in a more intuitive manner, Apple recently introduced the SwiftUI tool, providing a declarative syntax in which developers can implement the desired interface with little code by explicitly stating and ordering the interface elements and functions [131].

Figure A1 in Appendix A shows the interface of XCode with a standard example from Apple [41]. XCode, Swift, Swift UI and the HealthKit API were used in developing the MHP. As mentioned, the goal of the prototype is to collect stress-related data from Apple Health in order to investigate the development of ML stress prediction models in real-life settings. With this in mind, the prototype will collect the following data from Apple Health through the HealthKit API: a) weight, b) steps, c) heart rate, d) heart rate variability, e) blood pressure, f) sleep, g) electrocardiogram (ECG), h) temperature, f) workout. The prototype is described in the next section.

5.2.4 Results

Figure 7 presents the MHP interface and flow. The top-left screen is shown while the app collects new data that was uploaded to Apple Health since the previous time the MHP was opened. Once that is finished, the user is taken to the top-right screen, titled

"Apple Health Variables", which shows a list of collected variables (while the MHP is enabled to collect workout data, although that information was not used for the pilot study as it required the participants to specifically do the activity).

On the tab at the bottom of the screen, the middle icon leads the user to the bottom-left screen, titled "Stress Questionnaires", which shows the completed stress questionnaires for the day and how many are remaining. Once the "Fill Questionnaire" button is clicked, users are taken to the bottom-right screen, where they can answer the 8 stress-related questions (assigned a random order to appear each time).

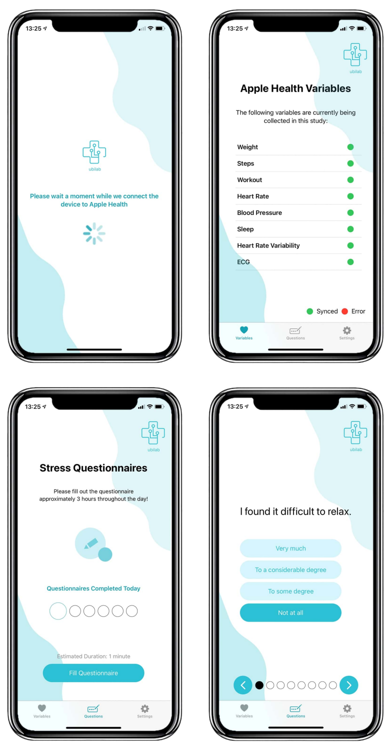


Figure 7: Interface of the MHP

Figure 8 presents the Unified Modelling Language (UML) diagram for the MHP. For the model classes, we represent them with the standard UML notation. For SwiftUI views in the *View* container, we include the names of the views. All classes in the *Model* container follow the Singleton design pattern (represented here by the green icon), meaning there is only one instance of these classes throughout the app's lifecycle [132].

When the user is entering the app for the first time, the login view is presented. The user can also sign up (Figure A2 in Appendix A). For the pilot study, the researchers will sign up for the participants through an email containing their participant ID. Looking at Figure 8, the *LoginManager* class is responsible for managing the login functions, including adding new users to the database and checking for existing users. If the user is logging in for the first time, the *HealthKitView* will request authorization to access the data (Figure A3 in Appendix A) through the *HKManager* class (in the study, this screen does not appear to participants as they give consent using a form following review ethic board protocols and are signed in by researchers before the study begins).

Once authorization is given (or retrieved, if it was given at a previous point in time), the *HKManager* calls the *HKDAO* class. This class was named with the Data Access Object (DAO) suffix as it encapsulates all access to a data source, although it does not directly follow the DAO design pattern. In this case, the *HKDAO* class handles all access to HealthKit data through queries. These queries are hardcoded into the code, as each of the variables that the user consents to being retrieved from Apple Health must be hardcoded together with their specific permission requests.

HealthKit [133] possesses many different types of queries, such as *HKStatisticsQuery*, which performs statistical calculations over a set of samples [134]; *HKSourceQuery*, which returns a list of sources for the data type [135]; or *HKActivitySummaryQuery*, which reads summaries of activities [136]. For the MHP, we make use of the *HKAnchoredObjectQuery* – which returns new data points since the previous query [137]– and the *HKObserverQuery,* which monitors Apple Health and provides updates when there are any changes, such as when new data is added [138]. Specifically in the MHP implementation, the observer query implements the anchored query, meaning that when there is an observed change in Apple Health pertaining to new data, an anchored query is triggered to collect these new data. This can be seen in the methods of the *HKDAO* class, with both observer and anchored queries for each data type.

Once new Apple Health data is retrieved, the *UserDAO* class is responsible for formatting and sending it to the database. The data is encoded into a struct defined in the code called *HKData,* which stores an id, value, dates, device, and any additional information through the several update functions for each data type, and is then uploaded to the database through the *API* class using the *uploadData* method.

Of note, in *UserDAO,* the *updateStress* method uploads the results of the stress questionnaire as an array of integer tuples. This is because the values of the questionnaire are represented in the following format: *(Int: Int)*, meaning a question number and a stress rating. For example, (2,3) means that for question 2 the user rated a level of 3 as their stress state at the moment. In summary, the *uploadData* function is responsible for storing Apple Health data in the database while *updateStress* does the same for answers to the stress questionnaires.

The observer queries are set up to enable the background delivery of data. However, each data type possesses different minimum periods for background data delivery, which are defined by Apple and cannot be changed [139]. In other words, each data type can only be collected by the queries at specific intervals. In addition, background delivery is scheduled automatically according to the usage of the app. If the user constantly opens the app, background data collection will occur constantly; however, the opposite is also true. In addition, there is a risk that the iPhone's operating system will stop the background queries. For this reason, in addition to setting up the background data delivery, new queries are activated and override previous ones every time the app is opened, and the login phase completed to avoid missing new Apple Health data.

As mentioned, while the data retrieval from Apple Health and upload to the database occurs (the process typically lasts a few seconds, depending on data size and Internet connection), the user is shown the top-left screen of Figure 7. When the process completes, the app directs the user to the top-right screen. There, the user can see all variables collected and if the process was successful (represented by a green dot on the side of the variables' names) as well as go to the *QuestionnaireView* through the tab options. This will take the users to the bottom-left screen of Figure 7 and, clicking on the "Fill Questionnaire" button directs them to the bottom-right screen to answer stress self-reports. Once the questionnaire is completed, the answers are saved in the *User* class and sent to the database through the *UserDAO* and *API* classes as discussed. Finally, users can go to the *ProfileView*, shown in Figure A4 in Appendix A, which thanks participants for their participation, displays their study participation ID and provides an e-mail in case of any questions.

In addition to the flow described above, two additional classes, *NotificationManager* and *NetworkMonitor*, are activated when the app starts (in the *AppDelegate.swift* class which provides methods to be trigerred when the app is opened). *NotificationManager* configures notifications every 3 hours starting at 9am reminding users of the data collection process. Users are instructed to start data collection when they wake up and not to necessarily follow the notification timing; rather, the notifications are meant to serve as reminders of the ongoing study and data collection to participants. The *NetworkMonitor* class activates when a drop in connectivity is detected, blocking the app's functions and displaying the screen in Figure A5 to avoid any errors due to inexistent connectivity.

Figure A6 in Appendix A shows the database structure. Device and user information are stored as well as health data. The information in the Devices table was uploaded prior to the data collection in the database, since devices were distributed to participants.

Each data type typically has an *id*, *value* (e.g., in Kg for weight, in beats per minute for heart rate, and in Celsius for temperature, among others), *start* and *end dates* for the data collection, *device id*, and any additional information provided by Apple Health. Some data types have different characteristics, such as sleep, which has both a decimal value and a category representing if the user was asleep or awake. ECG is different from most data types as it has additional attributes (e.g., the algorithm version). Also, the actual ECG data is stored in the *voltage_measurement* fields as a dictionary with timestamps and the voltage measured for that timestamp. This dictionary contains 15360 measurements, which are assembled into the ECG signal used for analyses.

5.2.5 Discussion

The MHP presented in the previous section, created using Apple's mobile development tools, is able to capture health-related data from the Apple Health repository

– which includes data from Apple devices, such as the Apple Watch, and data from other devices that integrate with Apple systems, such as devices manufactured by Withings – while also allowing users to self-report on their stress levels. This prototype version will be used as a data collection tool for a study collecting participant data in real-world settings with the goal of predicting stress levels of individuals.

While the prototype MHP was developed to be used in this study extracting data from devices distributed to participants, if successful, a similar system could be deployed by health researchers and public health agencies to extract data from personal devices already in use by the population. In Canada, where this study takes place, 32 million individuals own a smartphone [27] and almost 4 million possess a fitness wearable device [28]. These devices passively and continuously collect data such as physical activity, sleep, and heart rate which could be used in research to generate new insights into the health of individuals and populations while mitigating limitations of traditional self-report data collection efforts. Collecting objective data (or even self-report, as is the case of the stress questionnaires) constantly over shorter periods can reduce biases, and it is much simpler and cost-effective to deploy an app in the App Store, for example, than handling the logistics and costs of deploying full-scale surveys [78].

Leveraging personal devices also relieves the burden on users to collect data. Rather than a "heavy" data collection process such as the one presented in the study protocol, a long-term system could extract more data points collected over longer periods. Since most devices collect data passively, this could also minimize follow-up losses compared to other studies using self-report methods.

Improvements can also be made to the prototype MHP for future versions. First, a backup feature should be implemented in case data cannot be sent from the device to the database (e.g., due to low connectivity or malfunction in the server). In case of error, data should be saved in the device and sent to the database again at a later point in time. In addition, other data types supported by Apple Health could be hardcoded into the source code, allowing the system to collect more health-related data that is of interest to researchers and public health officials.

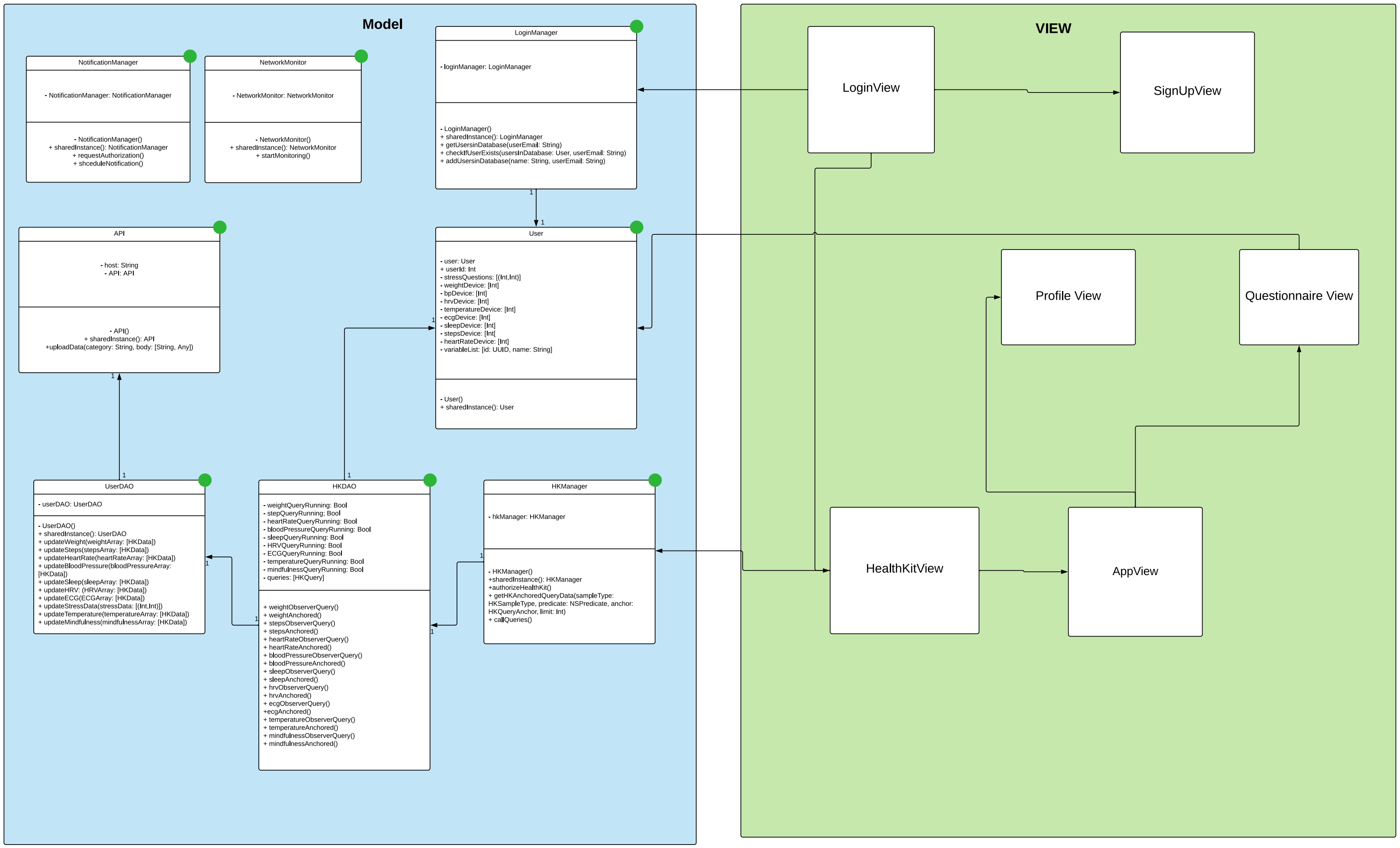


Figure 8: UML Class Diagram for the MHP

Additional features could also be introduced to encourage app use, such as allowing individuals to track the progression of their health metrics. In this manner, the app will become both a public health surveillance tool and a health management system, beneficial for individuals and public health researchers. Future studies should also explore the development of similar applications for other systems, such as Android, to understand their commonalities and differences to Apple systems.

It should be noted that, in addition to the technical requirements described here, there are legal and ethical issues that should be addressed when collecting personally identifiable information and digital data. While a full discussion of these is outside the scope of this paper, researchers and public health agencies must obtain user consent for data collection and use according to regulations – regardless of whether users already consented to their data being collected through the HealthKit API process. In Canada, these regulations may include, for example, the Personal Information Protection and Electronic Documents Act (PIPEDA), which regulates the use of personally identifiable information in commercial activities; the Privacy Act, which regulates the use of data by federal public health agencies and other entities; and regulations from review ethics boards for research projects. Researchers and public health officials looking to leverage mobile health data in their studies should be mindful of and respect applicable legislation for their specific region and case.

5.2.6 Conclusion

This paper presented the motivation, modelling and development of an iOS-based MHP that collects, and stores Apple Health data related to stress while also enabling stress self-report. This MHP will be used as a data collection tool in a future study to investigate its effectiveness for public health monitoring systems that leverages personal mobile and wearable devices.

Such a system could complement traditional self-report efforts and possibly enable the collection of larger and more representative data in a more simple and cost-effective way. Future versions of the MHP can include improvements such as data backups and features that allow individuals to track changes in their health metrics to encourage use. Further, researchers and public health officials should be mindful of applicable security and privacy legislations and their requirements when collecting personally identifiable digital health information.

5.3 Discussion

As the tool used to collect data for the pilot study, integrating information from several devices that can be used in conjunction with Apple's operating system, the MHP forms the basis of this thesis and of a potential public health surveillance system implemented using the tools and data described in the work.

In this chapter, a detailed description of the MHP, its architecture, modelling and storage choices were presented. Further information on the practical application of the MHP in the pilot study is described in the following Chapters. In particular, Chapter 8 describes lessons learned and future directions from the practical application of the MHP in the pilot study.

Chapter 6 – Pilot Study and Preliminary Statistical Methods with ECG Data for Stress Quantification

6.1 Foreword

With the MHP developed and included in iPhone devices for study purposes (the platform was developed for the study and not currently available for download by the public at large), the next of the research process involved data collection to evaluate the platform with a pilot study. Among the many data types collected, ECG data from the Apple Watch presents itself as an interesting case. The Apple Watch ECG collects a 30-second 1-lead ECG (as opposed to the standard 12-lead ECG which places electrodes in the user's body) when users place their finger on the digital crown of the watch [127]. This is a very simple, quick, and non-invasive data collection method. Further, HRV data can be derived from ECG and are commonly used in the identification of stress [140].

Therefore, before presenting the full study protocol and ML analyses, this chapter focuses solely on ECG data and traditional analysis methods, such as Spearman correlation and ANOVA, to study how HRV data derived from the Apple Watch ECG can be used to quantify stress. This paper uses a subset of the data, considering only healthy participants, to conduct the analysis. The goal was to provide preliminary information on the pilot study and data processing, apply more traditional health analysis methods to the data, and increase knowledge on the usefulness of the Apple Watch ECG for stress detection. To the best of our knowledge, this paper and others presented in this thesis are some of the first works to investigate the use of the Apple Watch ECG in this context.

6.2 Can Heart Rate Variability data from the Apple Watch Electrocardiogram (ECG) Quantify Stress?

6.2.1 Abstract

Chronic stress has become an epidemic with negative health risks including cardiovascular disease, hypertension, and diabetes. Traditional methods of stress measurement and monitoring typically relies on self-reporting. However, wearable smart technologies offer a novel strategy to continuously and non-invasively collect objective health data in the real-world. A novel electrocardiogram (ECG) feature has recently been introduced to the Apple Watch device. Interestingly, ECG data can be used to derive Heart Rate Variability (HRV) features

commonly used in the identification of stress, suggesting that the Apple Watch ECG app could potentially be utilized as a simple, cost-effective, and minimally invasive tool to monitor individual stress levels. Here we collected ECG data using the Apple Watch from 36 health participants during their daily routines. Heart rate variability (HRV) features from the ECG were extracted and analyzed against self-reported stress questionnaires based on the DASS-21 questionnaire and a single-item LIKERT-type scale. Repeated measures ANOVA tests did not find any statistical significance. Spearman correlation found very weak correlations ($p<0.05$) between several HRV features and each questionnaire. The results indicate that the Apple Watch ECG cannot be used for quantifying stress with traditional statistical methods, although future directions of research (e.g., use of additional parameters and Machine Learning) could potentially improve stress quantification with the device.

6.2.2 Introduction

According to the WHO, stress is the "Health Epidemic of the 21st Century" [141]. Over a quarter of U.S. adults report such high levels of daily stress that they are not able to function properly [33]. Stress, as a survival mechanism, is normal and healthy: stress allows the body to generate more energy to deal with a potential threat. The stress response is modulated by the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS). The SNS is responsible for triggering a response to unexpected threats to generate energy and resources for the body – the fight-or-flight response – by signalling adrenal glands to release adrenalin and cortisol, which lead to several physiological changes including increased heart rate, blood pressure, and respiration [34,42]. Once the acute stressors are removed, the PNS functions to relax the body, returning it to its normal state [34,42].

Despite the necessity of a stress response to survival, chronic exposure to stressors can lead to severe health consequences including cardiovascular diseases, hypertension, obesity, and diabetes [34–36]. Chronic stress is an increasingly observed condition worldwide. High levels of daily stress are reported by 38% of U.S. adults aged 40-49 years and 33% of adults aged 50-59 years [142]. In Canada, daily stress was highest amongst individuals between 35-49 years (27.8%) followed by individuals aged 50-64 years (22%) and 18-34 years (21.9%) [31]. Individuals over 65 years reported the lowest levels of stress [31]. Chronic stress is estimated to cost over USD 300 billion annually in associated healthcare expenses, reduced job performance, and absenteeism

[141,143]. Workplace stress is connected with 120,000 premature deaths annually [144]. The COVID-19 pandemic has amplified this crisis: a recent survey by the American Psychological Association discovered that approximately 80% of respondents identify the pandemic as a major source of stress in their life and almost 70% reported increased levels of stress owing to COVID-19 [32].

The identification of stress and the application of interventions should be a public health priority. Research data on stress is typically collected through self-reporting surveys, which may have limitations such as low response rates, recall and social bias, cost and delays [21]. Smart technologies, such as mobile and wearable devices, have recently been identified as useful tools to measure health parameters. Several of these technologies have embedded sensors that collect objective health data such as sleep, blood pressure, and heart rate [89,90]. In particular, an electrocardiogram (ECG) feature for detecting atrial fibrillation has been introduced to the Apple Watch device [20,89,90]. Unlike the standard 12-lead ECGs, which use electrodes connected to the body, the Apple Watch ECG collects a 30-second 1-lead ECG when users place their finger on an electrode located in the digital crown of the device [127]. Interestingly, ECG data can be used to derive Heart Rate Variability (HRV) features which are commonly used in the identification of stress [140]. This suggests that the Apple Watch ECG app could potentially identify and monitor individual stress. Apple Watch applications could use this information to provide instant user feedback and interventions, such as suggesting the use of meditation apps [145]. Furthermore, the use of a wearable data collection device would improve stress research data by eliminating recall biases and increasing population sample sizes. However, compared to longer measurements, there is not a large amount of evidence suggesting that ultra-short HRV measurements are reliable [146].

The goal of this paper was to explore the associations between HRV data collected from the Apple Watch ECG app with perceived stress levels in a real-life study. To the best of our knowledge, this is the first paper that provides statistical analyses of data derived from the Apple Watch ECG for stress detection, studying the reliability of these short-term measurements, and it is a continuation of previous work that uses a set of the same data, from 40 participants, to create Machine Learning (ML) stress prediction models [21]. ECG data from the Apple Watch ECG app was collected from 36 participants in a real-world setting over 2 weeks. We were able to identify significant, albeit weak, correlations between several HRV features and self-reported stress states, as well as significant differences between groups. Results from this study support the continued development of wearable ECG sensors as tools to measure stress.

The paper is organized as follows: section 2 described related work, including previous studies that used different sets of the same data for creating Machine Learning models; section 3 describes the methods, while section 4 presents the results and section 5 discusses our findings. Finally, section 6 presents the conclusions.

6.2.3 Related Work

This paper is an extension of previous work performed by the authors that uses data from 40 participants, to derive HRV features from the Apple Watch ECG data and use that data to create machine learning (ML) models for stress prediction – specifically using Random Forest and Support Vector Machines [21]. The models, trained on subsets of the data according to age, gender, income, profession, and health status, found a weighted f1-score lying approximately between 55-65%, which is in line with the state-of-the-art for stress prediction using ML, although towards the low end. The models possessed high specificity – i.e., in general they were capable of successfully predicting when an individual is not stressed – but were less successful when predicting the stressed state. Notably, feature importance of the Random Forest models was calculated to determine, for each model, what features were most important in determining the prediction results. Although they vary per model, in general the heart's acceleration (AC) and deceleration (DC) capacity were some of the most important features, present in most of the models. Another noteworthy feature is the standard deviation of interbeat intervals (SDNN). A more detailed explanation of HRV features and the feature extraction process is provided in the methods section.

Data from the same study, this time from 27 participants, was used by Benchekroun et al. [147], although in this case the HRV data was derived from the Empatica E4 device rather than from the Apple Watch ECG. The Empatica E4 device collects data continually as opposed to cross-sectionally, providing larger datasets. Random Forests trained on this data in an area under the receiver operating characteristic (ROC) curve (ROC AUC) of 0.79 and a macro f1-score of 75%. Further, a cross dataset analysis was performed in which models were trained on a laboratory dataset and tested on the Empatica E4 data, achieving a ROC AUC of 65% and a f1-macro score of 62%.

MCcraty et al. [148] performed repeated measures ANOVA analysis on HRV metrics of 24 patients with panic disorder and healthy control, finding differences in features such as the SDNN index, Total Power of VLF, Normalized LF/HF ratio, among others. Hong et al. [149] conducted repeated ANOVA analyses for participants, finding changes in HF and RMSDD.

Seipäjärvi et al. [150] studies stress and HRV in a laboratory setting among participants in different age groups and health status, finding that with the application of stressors differences in HRV could be observed. Föhr et al. [48] investigated the association between physical activity, HRV and subjective stress measured with the perceived stress scale (PSS), finding significant changes between physical activity and HRV with stress. However, using ecological momentary assessments, Martinez et al. [151] found a significant but small relationship between HRV and stress, where only a small amount of variance was explained by models. The author's concluded that HRV might be a good proxy for stress in controlled settings with specific stressors applied, but not in real-life. Silva et al. [152] conducted Spearman correlation analysis between the perceived stress scale (PSS-14) and 5-minute HRV variables at rest, and found weak to moderate correlation for the low frequency (LF) band. A similar Spearman correlation analysis was done in this study between HRV features and stress.

The Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology provide widely used guidelines for the analyses of HRV data and were of great help in guiding this research [153]. The authors in Acharya et al. provided an extensive review of HRV metrics [140], while several papers explored the feasibility and characteristics of analyzing HRV data. For example, Benchekroun et al. [154] discussed the impact of missing data on several HRV-related metrics and the best interpolation techniques to handle this situation.

It is important to note that there is limited research on the reliability of ultra-short-term HRV measurements (less than 5 minutes) when compared to long-term methods. Baek et al. studies ultra-short-term measurements to define recommended minimum intervals for each of these metrics to be valid [146]. In general, each metric has a different recommended interval, varying from seconds to minutes. Shaffer et al. [155] conducted a review of ultra-short-term heart rate variability norms, finding that most studies did not use criterion validity to study if the procedures produce comparable results with validates measurement procedures, applying other metrics (e.g., Pearson correlation) which may be insufficient to provide evidence of comparable methods. Studies that did use more appropriate metrics (such as Baek et al. [146] mentioned previously) typically found that different metrics will depend on different intervals. Munoz et al. [156], for example, found that a minimum of 10s was required for RMSSD and 30s for SDNN. The authors also found that ultra-short-term measurements are extremely sensitive to artifacts. For example, a single false heartbeat can alter the HRV metrics, and so special care must be taken when analyzing the data. In short, while ultra-short-term recordings such as the ones used in this study have potential due to its increased accessibility and ease-of-use, there is a lack of robust evidence base to assert that these recordings can be used as proxies for longer recordings. In this study, as will be described, the Kubios Premium Sofware was used to process the data to mitigate issues with noise or artificats.

In addition, while Apple Watch ECG data was shown to be successful in detecting atrial fibrillation [157] there is also a lack of a robust evidence base on how the HRV data derived from the Apple Watch ECG compares to gold standards. A study by Saghir et al. [158] found good results, showing that the agreement between the Apple Watch ECG and a standard 12-lead ECGs to be moderate to strong in health adults. In other words, there is promising but limited evidence both on ultra-short-term recordings and on how Apple Watch ECG data compares to more traditional, longer-term measurement methods.

It should also be noted that, while on this work we are specifically focusing on HRV derived from ECG – HRV being an essential parameter in stress quantification – other metrics, such as electrodermal activity (EDA), can also be considered for analyses [159].

6.2.4 Methods

6.2.4.1 Participant Recruitment

Healthy participants (n = 36) were recruited from the University of Waterloo as well as through Facebook Ads and Kijiji (a Canadian website that allows users to advertise products and services). Participants had to live close to the Kitchener-Waterloo region in Ontario for devices to be delivered in person. Participants were offered CAD 100.00 for two weeks of data collection. This study was approved by the University Waterloo Research Ethics Board (REB [43612]). Data collection took place between December 2021 and December 2022. Table 1 shows the characteristics of the study participants. Participants were aged 18 years or older. For the analyses described in this paper, we considered only healthy participants, i.e., who did not drink or smoke, did not have any chronic conditions or take prescription medications.

6.2.4.2 Data Collection

This study followed the Ecological Momentary Assessment (EMA) methodology to obtain self-reports closer to the event to approximate real-life scenarios [160]. Participants were given an iPhone 7 with iOS 15.0 and an Apple Watch Series 6 with watchOS 8.3 for 2 weeks. The Apple Watch contained the ECG app, and a Mobile Health Platform (MHP) was installed on the iPhone. The MHP was used to collect health data, including ECG recordings, from the iPhone's Apple Health app data repository [21,89,90,147].

Users were instructed to perform an ECG measurement on the Apple Watch ECG app 6 times during the day in approximately three-hour intervals followed by the stress questionnaire (below) on the iPhone. Figure 9 shows the study protocol (times are included for reference purposes; participants were asked to collect data as soon as they woke up).

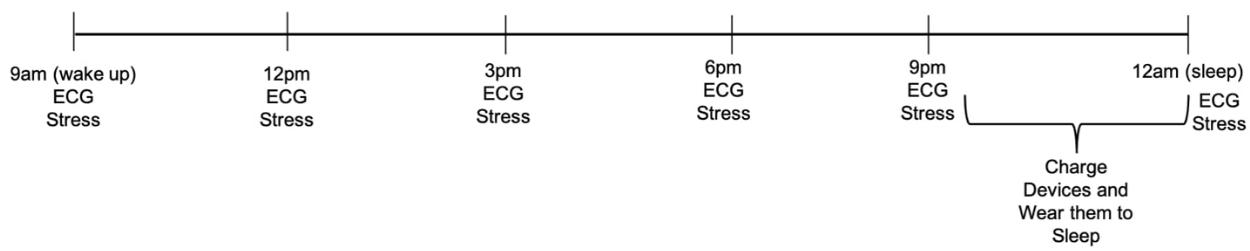


Figure 9: Study Protocol

The app installed in the iPhone, termed the Mobile Health Platform (MHP), can collect health data saved on the iPhone's health data repository, the Apple Health app, including the ECG recordings. The MHP collected this data, which were then saved in our database using the JSON format (for each ECG reading there are 15360 voltage measurements and associated timestamps in milliseconds, forming the 30-second ECG). The MHP also contains a tab with the stress questionnaires to be completed, which will be described next. Figure 10 shows the interface of the MHP, including the additional variables collected in the study.

We noticed that several participants had difficulty managing the study protocol with their daily life responsibilities. Therefore, we asked participants to use the devices for additional days to compensate as applicable.

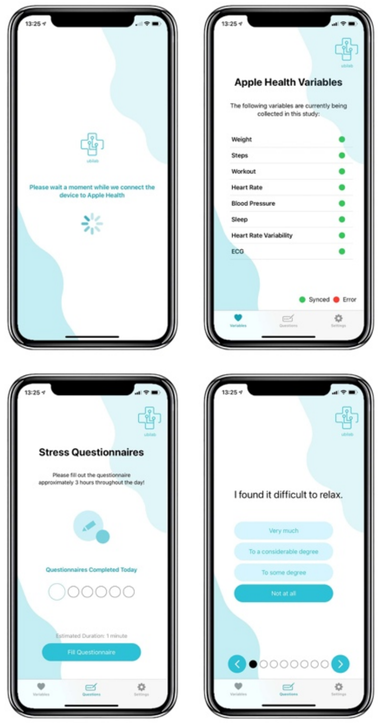


Figure 10: MHP Interface

Of note, this study is part of a larger cross-sectional study that investigates the use of smart technologies for stress detection. As part of this larger study, in addition to the Apple Watch and iPhone, participants were also given additional devices capable of collecting other data, such as the Withings Blood Pressure Monitor and the Empatica E4. Since this is not the focus of the paper we will not describe the use of these devices further, but more information on these expanded protocols is provided in [21,89,90,147].

6.2.4.3 Stress Questionnaires

As there are a limited number of validated stress questionnaires for the EMA with a validation period relevant to this study, we used the stress subscale of the Depression, Anxiety, and Stress Scale (DASS-21) for our stress questionnaire. While the DASS-21 is usually applied over a week, there is promising evidence of using DASS-21 with EMA [128]. In addition, Wang et al. [129] used a single-item measure that, while lacking validation in the literature, was used successfully for stress prediction and is moderately correlated with robust stress questionnaires.

The following questionnaire on a LIKERT-type scale was used for our study. Questions 1-7 are related to the DASS-21 and question 8 comprises the single-item measure used by Wang et al.

1. I found it hard to wind down;
2. I felt that I was using a lot of nervous energy;
3. I found myself getting agitated;
4. I found it difficult to relax;
5. I tended to over-react to situations;
6. I was intolerant of anything that kept me from getting on with what I was doing;
7. I felt that I was rather touchy;
8. Right now, I am…

Questions 1-7 have the options: "Not at all", "To some degree", "To a considerable degree", and "Very much", while Question 8 has "Stressed Out", "Definitely stressed", "A little stressed", "Feeling good", and "Feeling great". The questions were displayed to the user in a random order each time the questionnaire was filled in the MHP, and compose the perceived stress, i.e., the degree to which a stressfull situation affects an individual, is measured.

In addition to self-reporting stress throughout the day, participants were asked to self-report their stress levels at the beginning of the study with the single-item measure (results shown in Table 1).

6.2.4.4 Data Pre-Processing

To obtain the HRV features from the ECG readings, we made use of Kubios Premium 3.5.0, a widely used software that analyzes and extracts features from several heart-related signals [140,161]. The JSON ECG data was exported into a CSV format and each voltage measurement was sorted by timestamp. The CSV file was imported into Kubios.

Kubios automatic beat correction feature was used and any samples that contained more than 5% of corrected beats were removed. In addition, any ECG sample classified as Poor Recording or Inconclusive by the ECG app was also removed from the analysis [127]. Frequency features were calculated using both the Fast Fourier Transform (FFT) and Autoregressive Spectral Analysis (AR). A list of the features generated by Kubios based on the 30-second ECG signal is presented in Table 11 [140,161].

The scores of the DASS-21 questions summed together were multiplied by 2. If the score was bigger than 14, the sample was classified as "stress" according to DASS-21 guidelines [162]. For the single-item measure, the sample was classified as "stress" if the score was bigger than 2, as that would represent the user being at least "a little stressed." If the DASS-21 score or the single-item score were classified as "stress," the measurement was classified as the "stress" state.

6.2.4.5 Statistical Analysis

Statistical analyses were performed through the Statistical Package for Social Sciences (v. 28.0; SPSS, Chicago, IL, USA). Using baseline stress scores from the Single-Item measure at the beginning of the study, repeated measures ANOVA analyses were conducted followed by Tukey's Post-Hoc test in case of statistically significant features. In addition, Spearman´s non-parametric correlation test was applied to detect the correlation between each ECG variable with the quantitative DASS-21 and Single Item questionnaire scores. For all analyses, $P<.05$ was considered statistically significant. While correlations were performed for every feature, to limit the potential of biases ANOVA analyses were conducted with a subset of the features (Table 3) as seen in other works [148,149]. In addition, for the analyses, we considered 13 days of data for each participant (the minimum days of all participants in the study).

6.2.5 Results

To determine whether HRV data collected from an Apple Watch ECG was associated with perceived stress level, we recruited 36 healthy participants to participate in a real-life study. Using the Apple Watch ECG app and an iPhone app developed for this study, users were instructed to collect ECG readings and complete a stress questionnaire 6 times during the day in approximately three-hour intervals for 2 weeks, as well as fill an initial survey about perceived stress levels prior to data collection. Table 2 lists the HRV features captured by the Apple Watch ECG. Questionnaires comprised 8 questions based on the DASS-21 [162] and the measure used by Wang et al [129] as mentioned in the previous section.

Participants were predominantly female (64%) (Table 1). 61% were employed and 36% were students. Participants were mostly South Asian, White, or Latin American (31%, 25%, and 19% respectively), and reported low to medium income (44% and 36%, respectively The average of days a participant had in the study was 17.1 (± 2.5), and an average of 59 (± 16.0) ECG recordings. Participants were also asked to self-report their stress levels at the beginning of the study with the single-item measure (results shown in Table 10).

Table 10: Study Population Characteristics

Participants (N = 36)	*Frequency*	*Percentage*
Age		
18-24	12	33
25-34	10	28
35-44	10	28
45-64	3	8
Above 65	1	3
Gender		
Male	13	36
Female	23	64
SES		
Low (0-$30,000)	16	44
Medium ($30,000– $100,000)	13	36
High (Above $100,000)	4	12
Do not wish to disclose	3	8
Profession		
Full-time	17	47
Part-time	3	8
Student	13	36
Self-employed/Other	2	6
Retired	1	3
Ethnicity		
Black or African American	3	8
Black and Southeast Asian	1	3
Chinese	4	11
Indian	1	3
Latin American	7	19
South Asian	11	31
White	9	25
Self-Reported Stress Level, Beginning of Study		
1 (Great)	0	0
2 (Good)	8	22

3 (A little stressed)	15	42
4 (Definitely Stressed)	11	31
5 (Stressed Out)	2	6

Table 11: Kubios HRV Features derived from Apple Watch ECG

Name	Description
ECG_Mean HR	Mean of heart rate from ECG(ms)
ECG_ SD HR	Standard deviation of instantaneous heart rate from ECG (1/min)
ECG_Min HR	Minimum instantaneous heart rate calculated using 5 beat moving average from ECG(1/min)
ECG_Max HR	Maximum instantaneous heart rate calculated using 5 beat moving average from ECG (1/min)
HRV-1	Heart rate variability collected as SDNN with the Apple Watch
ECG_PNS Index	Parasympathetic nervous system activity compared to normal resting values
ECG_SNS Index	Sympathetic nervous system activity compared to normal resting values
ECG_Stress Index	Square root of Baevsky's stress index
ECG_Mean RR	Mean of R-R intervals (ms)
ECG_SDNN	Standard deviation of R-R intervals (ms)
ECG_RMSSD	Square root of the mean squared differences between successive RR intervals f(ms)
ECG_DC	Heart rate deceleration capacity (ms)
ECG_DCMod	Modified DC computer as a two-point difference (ms)
ECG_AC	Heart rate acceleration capacity (ms)
ECG_ACMod	Modified AC computer as a two-point difference (ms)
ECG_FFT LF	Fast Fourier Transform Low Frequency band components (Hz)
ECG_FFT HF	Fast Fourier Transform High Frequency band components (Hz)
ECG_AR LF	Autoregressive Low Frequency band components (Hz)
ECG_AR HF	Autoregressive High Frequency band components (Hz)
ECG_FFT Absolute Power LF	Fast Fourier Transform Absolute Power of Low Frequency band components (ms2)
ECG_FFT Absolute Power HF	Fast Fourier Transform Absolute Power of High Frequency band components (ms2)
ECG_AR Absolute Power LF	Autoregressive Absolute Power of Low Frequency band components (ms2)
ECG_AR Absolute Power HF	Autoregressive Absolute Power of High Frequency band components (ms2)
ECG_FFT Relative Power LF	Fast Fourier Transform Relative Power of Low Frequency band components (%)
ECG_FFT Relative Power HF	Fast Fourier Transform Relative Power of High Frequency band components (%)
ECG_AR Relative Power LF	Autoregressive Relative Power of Low Frequency band components (%)

ECG_AR Relative Power HF	Autoregressive Relative Power of High Frequency band components (%)
ECG_FFT Normalized Power LF	Fast Fourier Transform Normalized Power of Low Frequency band components (n.u)
ECG_FFT Normalized Power HF	Fast Fourier Transform Normalized Power of High Frequency band components (n.u)
ECG_FFT Total Power	Fast Fourier Transform Total Power (ms2)
ECG_FFT LF/HF	Fast Fourier Transform ratio between low and high frequency
ECG_AR Normalized Power LF	Autoregressive Normalized Power of Low Frequency band components (n.u)
ECG_AR Normalized Power HF	Autoregressive Normalized Power of High Frequency band components (n.u)
ECG_AR Total Power	Autoregressive Total Power (ms2)
ECG_AR LF/HF	Autoregressive ratio between low and high frequency
ECG_SD1	The standard deviation perpendicular to the line-of-identity in Poincaré plot (ms)
ECG_SD2	The standard deviation along the line-of-identity in Poincaré plot (ms)
ECG_SD2/SD1	Ratio between SD2 and SD1 (ms)

As described in the previous section, using the questionnaire score, measurements were designated as self-perceived "stress" if (a) the DASS-21 questions were classified as "stress" according to a DASS-21 greater than 14; or (b) the single-item measure was classified as "stress" if the score was greater than 2. Measurements that did not meet this cut-off were designated as "no stress".

Repeated measures ANOVA test was performed to compare differences recorded by the Apple Watch ECG and self-perceived stress. No statistical significance was revealed (Table 12). To determine which ECG variables correlated with stress, we applied a Spearman´s non-parametric correlation analysis between HRV features and self-perceived stress, divided by each of the stress scores (DASS-21 and Single-Item measure). Spearman correlation coefficients (r) and p-values were calculated and shown in Table 13.

Regarding DASS-21, several features were shown to have a weak correlation including: *SNS Index, Stress Index, SDNN, SD HR, Min HR, RMSSD, NN50, pNN50, RR Tri Index, TINN, DC, DC mod, AC, AC mod, FFT Absolute Power VLF, FFT Absolute Power LF, FFT Absolute Power HF, FFT Absolute Power VLF log, FFT Absolute Power LF log, FFT Absolute Power HF log, AR Absolute Power VLF, AR Absolute Power LF, AR Absolute Power HF, AR Absolute*

Power VLF log, AR Absolute Power LF log, AR Absolute Power HF log, FFT Total Power, AR Total Power, SD1, SD2.

The Single-Item measure significant correlations were: *Stress Index, SDNN, SD HR, TINN, DC, AC, FFT Absolute Power VLF, FFT Absolute Power LF, FFT Absolute Power VLF log, FFT Absolute Power LF log, AR Absolute Power LF, AR Absolute Power LF log, FFT Total Power, SD2, SD1/SD2.*

Table 12: Repeated measures ANOVA for HRV parameters with baseline self-perceived stress (p<0.05)

Parameter	Source	Sum of Squares	Mean Square	F	P-value
AC	Days	1293	108	0.656	0.794
	Days x Self-Perceived Stress	1332	111	0.675	0.775
DC	Days	2119	177	0.733	0.719
	Days x Self-Perceived Stress	1242	103	0.430	0.952
RMSSD	Days	1390	116	0.443	0.945
	Days x Self-perceived stress	2293	191	0.731	0.721
SDNN	Days	678	56.5	0.372	0.973
	Days x Self-perceived stress	667	55.6	0.366	0.975
Stress Index	Days	150	12.5	0.856	0.592
	Days x Self-perceived stress	158	13.2	0.901	0.546
FFT Absolute Power LF	Days	8.52e+6	710175	0.672	0.779
	Days x Self-perceived stress	1.70e+7	1.41e+6	1.337	0.195
FFT Absolute Power HF	Days	8.85e+6	737302	0.600	0.843
	Days x Self-perceived stress	1.44e+7	1.20e+6	0.978	0.469
AR Absolute Power LF	Days	3.24e+7	2.70e+6	0.787	0.664
	Days x Self-perceived stress	5.12e+7	4.27e+6	1.245	0.250
AR Absolute Power HF	Days	1.08e+9	8.96e+7	0.956	0.491

	Days x Self-perceived stress	1.43e+9	1.19e+8	1.269	0.234
FFT Relative Power LF	Days	1591	132.6	0.895	0.552
	Days x Self-perceived stress	979	81.6	0.551	0.881
FFT Relative Power HF	Days	2048	171	1.031	0.419
	Days x Self-perceived stress	1812	151	0.912	0.535
AR Relative Power LF	Days	2136	178	1.091	0.366
	Days x Self-perceived stress	1431	119	0.731	0.721
AR Relative Power HF	Days	2258	188	1.056	0.396
	Days x Self-perceived stress	1829	152	0.856	0.593

Table 13: Correlation coefficients (r) and p value for Spearman´s non-parametric correlation analysis.

Variables	**DASS-21**		**Single item**	
	r	*P*	r	*P*
ECG PSN Index	0.039	0.070	-0.010	0.653
ECG SNS Index	-0.075	0.001*	-0.026	0.227
ECG Stress Index	-0.105	0.001*	-0.046	0.036*
ECG Mean RR	0.033	0.131	0.014	0.528
ECG SDNN	0.109	0.001*	0.044	0.041*
ECG Mean HR	-0.033	0.131	-0.014	0.528
ECG SD HR	0.102	0.001*	0.048	0.027*
ECG Min HR	-0.050	0.021*	-0.022	0.303
ECG Max HR	-0.014	0.522	-0.013	0.554
ECG RMSSD	0.077	0.001*	0.001	0.957
ECG NN50	0.064	0.003*	-0.003	0.892
ECG pNN50	0.069	0.001*	-0.002	0.925
ECG RR Tri Index	0.091	0.001*	0.034	0.120
ECG TINN	0.110	0.001*	0.045	0.038*
ECG DC	0.099	0.001*	0.058	0.008*
ECG Dcmod	0.076	0.001*	0.008	0.721
ECG AC	-0.105	0.001*	-0.073	0.001*
ECG ACmod	-0.075	0.001*	-0.014	0.528
ECG FFT Absolute Power VLF	0.081	0.001*	0.051	0.020*
ECG FFT Absolute Power LF	0.102	0.001*	0.047	0.030*

ECG FFT Absolute Power HF	0.099	0.001*	0.033	0.127
ECG FFT Absolute Power VFL log	0.081	0.001*	0.051	0.020*
ECG FFT Absolute Power LF log	0.102	0.001*	0.047	0.030*
ECG FFT Absolute Power HF log	0.099	0.001*	0.033	0.127
ECG AR Absolute Power VLF	0.088	0.001*	0.032	0.134
ECG AR Absolute Power LF	0.099	0.001*	0.044	0.041*
ECG AR Absolute Power HF	0.085	0.001*	0.016	0.472
ECG AR Absolute Power VLF log	0.088	0.001*	0.032	0.134
ECG AR Absolute Power LF log	0.099	0.001*	0.044	0.041*
ECG AR Absolute Power HF log	0.085	0.001*	0.016	0.472
ECG FFT Relative Power VLF	-0.028	0.191	-0.003	0.903
ECG FFT Relative Power LF	-0.009	0.685	0.010	0.658
ECG FFT Relative Power HF	0.025	0.247	-0.006	0.782
ECG AR Relative Power VLF	-0.019	0.371	-0.009	0.684
ECG AR Relative Power LF	-0.009	0.676	0.032	0.135
ECG AR Relative Power HF	0.022	0.304	-0.020	0.350
ECG FFT Normalized Powers LF	-0.023	0.297	0.007	0.763
ECG FFT Normalized Powers HF	0.023	0.291	-0.006	0.788
ECG FFT Total Powers	0.110	0.001*	0.053	0.014*
ECG FFT LFHF	-0.023	0.292	0.006	0.773
ECG AR Normalized Powers LF	-0.019	0.393	0.025	0.254
ECG AR Normalized Powers HF	0.019	0.379	-0.024	0.267
ECG AR Total Power	0.099	0.001*	0.034	0.118
ECG AR LFHF	-0.019	0.385	0.024	0.262
ECG SD1	0.077	0.001*	0.001	0.953
ECG SD2	0.115	0.001*	0.057	0.008*
ECG SD1SD2	0.021	0.330	0.081	0.001*

6.2.6 Discussion

Overall, some HRV features captured by the Apple Watch weakly correlate to the stress questionnaires. Repeated measures ANOVA test and Tukey's Post-Hoc test indicated that Apple Watch ECG features in the current study design cannot statistically differentiate between stress states in a real-world setting. Therefore, the answer of "can Heart Rate Variability data from the Apple Watch ECG Quantify Stress" with the use of the statistical methods investigated in this work seems to be no.

Regarding Spearman correlations, while several features in the domains (time domain, frequency domain, non-linear) were shown to have a significant correlation with the DASS-21 and single-item measure, all were weak. Nevertheless, interesting points can be made by comparing the differences between the two questionnaires.

In general, the significant correlations between HRV features and the single-item measure are a subset of the ones from DASS-21. One of the main differences in the correlations between the DASS-21 and the single-item measure is that the latter does not seem to be significantly correlated to the absolute power high frequency components (FFT Absolute Power HF and AR Absolute Power HF). In this way, the use of both questionnaires for the study seem to complement each other in capturing differing dimensions of self-perceived stress, although it should be noted that the weak correlations may limit the validity of these results.

Interestingly, Silva et al [152], also found weak to moderate correlations using the Spearman test while comparing HRV metrics with stress from the PSS-14 questionnaire but failed to find any significant correlations except for the LF band. Given that participants' measurements were taken at rest and the PSS-14 stress scores were in the mid to low range, it is possible that physiological changes owing to stress affected the correlation values in our current work. Indeed, several factors may have affected the quality of the data. First, being a "real-life" experiment, data may be subjected to noise and errors in measurements. For example, respondents may forget to take measurements throughout the day, take the measurements incorrectly, or be influenced by the Hawthorne Effect in which respondents change their behaviour because they are being monitored. On the same token, elements such as sweat, or movement may affect the measurement. These factors may have influenced the results, leading to potentially inaccurate data. Future work should explore data collection of ECG in controlled conditions, potentially with an intervention (e.g., applying stressors in a lab) to evaluate the robustness of this data. This recommendation is also in line with Martinez et al. conclusions that HRV may be best represented in controlled environments with specific stressors [151]. While this would diminish the validity of ECG data to be used in real-life scenarios to identify stress, it could provide further clues as to how the relationship between these variables work and new directions of research. Further, future work on this dataset can consider the distribution of the data per day and HRV diurnal fluctuations, which could provide more significant and illuminating results.

In addition, a convenience sample was used in this pilot study, and as can be seen by Table 10, there is a predominance of females and participants with low to medium SES which may affect the external validity of the results. Finally, since we used the EMA methodology, we decided to combine both the DASS-21 and the single-item measure for stress classification, which can potentially affect how individuals report stress and may lead to some of the contradictory findings in terms of group differences presented here. On that note, this study focused on perceived stress, i.e., the degree to which a situation perceived as stressful affects individuals. In this context, subjective ratings of stress may be affected by each participant's internalized definition of stress, which in turn may influence responses [54]. Nevertheless, the fact that several significant – albeit weak – correlations were found are encouraging and additional, more controlled, and stratified experiments should be conducted to confirm and clarify these relationships between the HRV features from the Apple Watch ECG and self-perceived stress.

As described in the Related Work section, there is promising but limited evidence on the reliability of ultra-short-term measurements and the Apple Watch ECG when compared to traditional measurement methods and data. It is possible that inaccuracies in the Apple Watch ECG led to a lack of statistical differences between stress states in this study. In addition to controlled experiments, future research could also consider using different methods of ultra-short-term data collection to verify the results. Given that weak correlations were found, the use of additional parameters in addition to simply the Apple Watch ECG might also help with quantifying stress. Indeed, several physiological and behavioural variables have been widely used in stress research. This could include brain activity measured through electroencephalogram (EEG), electrodermal activity (EDA), speech, mobile phone usage, among others [35]. Physical activity [46–48] and sleep [46,52,53] could also be potentially used to discriminate stress and can also be collected passively with the Apple Watch sensors – if ECG and other Apple Watch data were successfully used in conjunction to differentiate between stressed states, potential solutions could focus simply on the Apple Watch for stress quantification, which would be of great value in studying the prevalence of these conditions and providing feedback to users. Finally, the use of Machine Learning for prediction, as previously mentioned, has shown promising results [21], and further studies also using other parameters could help improve prediction accuracy and realize the potential of the Apple Watch for stress studies.

6.2.7 Conclusions

The use of an Apple Watch ECG to quantify individual stress was piloted in a real-world scenario. Significant but weak correlations were found between several HRV features and measures of self-perceived stress. This study highlights the potential usefulness of the Apple

Watch ECG as a minimally invasive tool for stress monitoring, quantification, and intervention, although more robust evidence is needed to establish the relationships between the data and its relevancy.

6.3 Additional Data and Discussion

The results from this paper suggest that traditional statistical methods cannot detect strong correlations or differences for stressed/non-stressed states using the Apple Watch ECG data. This remains true for other data types collected in the pilot study. While this chapter focused solely on HRV data derived from the Apple Watch ECG, given that correlation analyses found significant albeit weak values, I have included an additional table (Table 14) showing the correlation results from selected other data types. This was not included with the paper, as it was focused solely on ECG and additional data types were not yet processed at the time of submission. The features for Table 14 were selected based on their importance for models developed in Chapters 8 and 9. Similarly to the ECG, although there are statistically significant relationships, results do not show strong correlations or even moderate correlations. A comprehensive list and description of these data types is presented in Table B2 on Appendix B. It should also be noted that, as mentioned, participants found the protocol burdensome, with some requiring additional days for study completion. For this specific subset of participants, the average days in the study was 16.69 (± 2.4), and 13 days were used for the ANOVA analyses. The subset of data had 60% of data labelled as no stress (1612 data points) and 40% as stress (1117 data points).

Table 14: Correlation coefficients (*r*) and p value for Spearman´s non-parametric correlation analysis (p<0.05)

Variables	DASS-21		Single item	
	r	*P*	r	*P*
Apple Watch Mean HR - Interval	-0.068	<0.001*	-0.008	0.704
Apple Watch Max HR - Interval	0.034	0.098	0.033	0.107
Apple Watch Min HR - Interval	-0.123	<0.001*	-0.014	0.497
Withings Total Time Asleep	-0.021	0.358	-0.094	<0.001*
Apple Watch Total Time Asleep	-0.015	0.483	-0.009	0.673
Apple Watch Number of Wake-ups	-0.111	<0.001*	-0.117	<0.001*
Apple Watch Consolidated Time Awake During Sleep	-0.037	0.079	-0.045	0.034*
Apple Watch Total Time in Bed	0.009	0.668	0.068	0.001*

Apple Watch % of Time Asleep While in Bed	0.007	0.731	-0.073	<0.001*
Withings Number of Wake-ups	-0.049	0.032*	-0.087	<0.001*
Withings Consolidated Time Awake During Sleep	-0.004	0.852	-0.045	0.046*
Withings Total Time in Bed	-0.012	0.595	-0.073	0.001*
Withings % of Time Asleep While in Bed	-0.030	0.192	-0.047	0.038
Time Spent in Light Stage	-0.081	<0.001*	-0.114	<0.001*
Time Spent in Deep Stage	-0.031	0.165	-0.080	<0.001*
Time Spent in REM Stage	0.131	<0.001*	0.078	<0.001*
Apple Watch Mean Steps	-0.042	0.028*	-0.057	0.002*
Apple Watch Max Steps	-0.030	0.115	-0.023	0.222
Apple Watch Min Steps	0.011	0.563	0.037	0.053
Short Term Mean HR	-0.046	0.015*	-0.045	0.019*
Short Term Max HR	-0.013	0.491	-0.017	0.366
Short Term Min HR	-0.072	<0.001*	-0.062	0.001*
Temperature	0.028	0.152	0.058	0.004*
Weight	-0.011	0.635	-0.054	0.017*
HRV-1	0.087	<0.001*	0.034	0.107
Systolic Blood Pressure	0.073	<0.001*	0.018	0.347
Empatica AC	-0.158	<0.001*	-0.098	0.001*
Empatica DC	0.172	<0.001*	0.095	0.002*
Empatica LF HF	-0.056	0.063	0.018	0.580
Empatica RMSSD	0.115	<0.001*	0.033	0.271
Empatica SDNN	0.144	<0.001*	0.075	0.013*
Empatica Stress Index	-0.149	<0.001*	-0.078	0.010*

Chapter 7 - Preliminary Machine Learning Analyses for Stress Prediction with Apple Watch ECG Data

7.1 Foreword

While correlation and repeated measures ANOVA analyses did not provide good results, the ECG data was also used in the creation of RF and SVM models. A subset of participants was used for this analysis, as the full study was not yet completed at this time. This subset included healthy and non-healthy participants – the word “healthy” here, and in the rest of the work, being shorthand for individuals that did not report any chronic disease or illness, use of prescription drugs, alcohol consumption or smoking. Further, this preliminary analysis did not consider any advanced data pre-processing techniques. The goal was to study the behaviour of RF and SVM with preliminary data, gaining a better understanding of how ML models can be used to predict stress. While other methods could have potentially be used to train the models, RF was chosen due to its interpretability, while both RF and SVMs are able to handle more complex data types.

Similarly to Chapter 6, the focus of this paper was on the Apple Watch ECG data due to its potential in being a quick and non-invasive tool to monitor stress in real-time. This chapter presents more details on preliminary feature selection done on the data as well as how the ML models were trained and tested, including different stratifications of participants according to traits such as age, gender, income, profession, and health status. A discussion on feature importance for the RF models is also presented.

7.2 Using Apple Watch ECG Data for Heart Rate Variability Monitoring and Stress Prediction: A Pilot Study

7.2.1 Abstract

Stress is an increasingly prevalent mental health condition that can have serious effects on human health. The development of stress prediction tools would greatly benefit public health by allowing policy initiatives and early stress-reducing interventions. The advent of mobile health technologies including smartphones and smartwatches has made it possible to collect objective, real-time, and continuous health data. We sought to pilot the collection of heart rate variability data from the Apple Watch electrocardiograph (ECG) sensor and apply machine learning techniques to develop a stress prediction tool. Random Forest (RF) and Support Vector

Machines (SVM) were used to model stress based on ECG measurements and stress questionnaire data collected from 33 study participants. Data were stratified into socio-demographic classes to further explore our prediction model. Overall, the RF model performed slightly better than SVM, with results having an accuracy within the low end of state-of-the-art. Our models showed specificity in their capacity to assess "no stress" states but were less successful at capturing "stress" states. Overall, the results presented here suggest that, with further development and refinement, Apple Watch ECG sensor data could be used to develop a stress prediction tool. A wearable device capable of continuous, real-time stress monitoring would enable individuals to respond early to changes in their mental health. Furthermore, large-scale data collection from such devices would inform public health initiatives and policies.

7.2.2 Introduction

Stress is an often overlooked determinant of health. High stress levels are linked to severe health problems such as depression, obesity, and cardiovascular diseases [34]. Unfortunately, 1 in 5 Canadian citizens report experiencing high levels of stress daily [163]. Increased awareness of mental health has emphasized the need for more timely stress monitoring and early intervention, and the collection of population-wide stress data could support public health initiatives and interventions.

Self-reporting continues to be the gold standard for monitoring stress. These methods face challenges and limitations such as social and recall bias [10,11], loss due to follow-up [10,11], delays between collection and reporting [18], and costs/logistics [10,18]. However, the link between stress and multiple biomarkers has revealed opportunities to develop technologies to quantify stress. One such feature is heart rate variability (HRV) which is now routinely quantified through an electrocardiograph (ECG). ECGs have been widely used for stress prediction, and are typically performed at a healthcare facility which limits their accessibility. The development of rapid point-of-care or self-monitoring devices would improve patient outcomes, providing invaluable information for public health agencies and real-time interventions (e.g., guided meditations) [89,90].

Digital technologies, including smartphones and wearable smartwatches, are pervasive in our lives. In 2020, the number of Apple Watch users worldwide was estimated at 100 million [164]. In line with the modern health trend toward patient self-care, these technologies now include sensors designed to continuously collect health data with minimal user effort [165]. Collected health parameters include steps, heart rate, blood pressure, and sleep. These technologies now generate massive quantities of objective data. Further, the datasets obtained from this novel, real-life data can be used to create prediction models using Machine Learning (ML), allowing public health agencies to better understand and study the prevalence of a condition in a population.

Apple Health, a popular source of digital health data, has recently introduced an ECG sensor to their Apple Watch device [89] [127]. The sensor, which is similar to a 1-lead ECG, collects 30 seconds of data through an electrode placed on the device's digital crown [20]. According to Apple, studies have shown good agreement in classifying the rhythm of the Apple Watch ECG compared to standard 12-lead ECGs, and in a clinical trial of 600 participants the ECG sensor had 99.6% specificity when classifying synus rhythm and 98.3% sensitivity for atrial fibrillation [127].

ECG data collected from this wearable device could potentially be employed to predict stress: users would simply take a non-invasive 30-second ECG and get instant feedback on their stress levels. It is currently unclear whether the brief 30-second ECG reading will be sufficient for stress prediction.

The goal of this work was to pilot the use of Apple Watch ECG data for stress prediction. This analysis is part of the development of a Mobile Health Platform (MHP), which collects Apple Health data from several mobile and wearable devices [89,90]. We collected ECG and stress questionnaire data from 36 participants over 2 weeks with the platform. We applied the machine learning models Random Forests (RFs) and Support Vector Machines (SVMs), as these models have been successfully used in stress prediction literature [35]. To the best of our knowledge, this is the first work that utilizes Apple Watch ECG for stress prediction. We found that the models performed at the low end of the state-of-the-art stress prediction technology. We were able to identify several HRV features, as well as socio-demographic classes which impacted the accuracy of the model. The results suggest that, with further development, Apple Watch ECG sensors could be employed for mobile, real-time stress prediction.

7.2.2.1 Related Work

The authors of Can et al. [35] provide a survey of stress prediction in real-life scenarios with mobile health technologies. As can be seen in this survey, and supported by stress prediction literature, successful methods for stress detection are Random Forests (RFs) and Support Vector Machines (SVMs), which were selected for this study.

Examples of studies that use these methods include Hovsepian et al. [166], which trains an SVM using ECG and respiration data in both laboratory and real-life settings. The model outputs the probability that a user is stressed with an accuracy of 90% in the laboratory and 72% in real-life. Muaremi et al. [167] collected ECG, respiration, galvanic skin response, sleep data and posture of sleeping individuals, achieving good accuracy with SVMs (73%) and RF (71%). Gjoreski et al. [168] use laboratory data to build RFs that predict stress with an accuracy of 83%; then, the RF model is used as an output to train an SVM that achieves 76% accuracy on real-life data. Can et al. [169] used heart rate variability and electrodermal activity data for real-life stress prediction, achieving 68% accuracy with SVM and 66% with RF. Based on these considerations and review results [35], the state-of-the-art accuracy for stress detection in real-life settings lies approximately between 60% and 80%.

Regarding HRV analyses, the Task Force of The European Society of Cardiology and the North American Society of Pacing and Electrophysiology provides guidelines on the measurement and analyses of HRV data [153], which were of great help for this work (more details are described in the sections below). Further, Acharya et al. provide a review of HRV metrics and their meaning [140], while Benchekroun et al. analyze the impact of missing data on several metrics and studied different interpolation techniques to handle missing data [154]. Baek et al. analyzed several of these metrics on ultra-short term measurements and defines the minimum time interval for each of these metrics to be valid when compared to standard measurements, and found that each metric is different with some requiring only a few seconds of data while others require several minutes [146]. The same work also showed that HRV can vary according to factors such as age [146].

7.2.3 Materials and Methods

7.2.3.1 Data Collection

Participants were given an iPhone 7 with iOS 15.0 and an Apple Watch Series 6 containing an installed Apple Watch ECG app (WatchOS 8.3) for two weeks. Following the Ecological Momentary Assessment (EMA) methodology [160], which enables self-reporting to approximate real-life scenarios, users were asked to perform an ECG reading using the app. EMAs are further described in section 2.3. They were instructed to collect data 6 times during the day in approximately three-hour intervals. Before the ECG collection, participants were asked to complete a stress questionnaire on the iPhone using the MHP. Figure 11 shows the study protocol (the times are just reference; participants were asked to start measurements at wake-up).

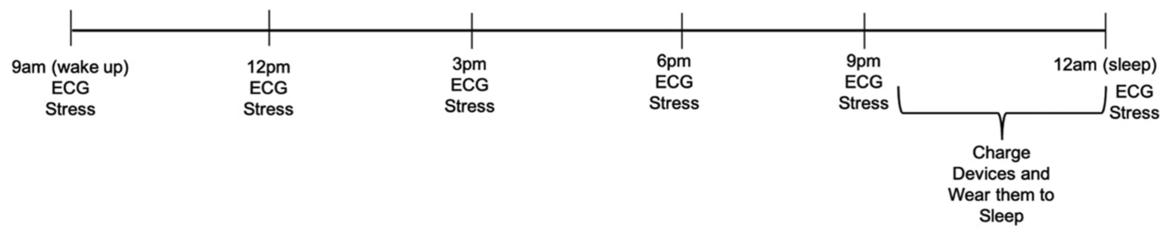


Figure 11: Study Protocol

7.2.3.2 Apple Watch ECG application

WatchOS 8.3 is an application capable of recording ECG measurements via an Apple Watch version 4 or higher. Briefly, ECG measurements requires users to open the ECG app and place their finger on the digital crown of the device and remain still for 30 seconds [170]. The instructions distributed to the users can be found in Appendix C. ECG readings were automatically stored in Apple's HealthKit API. We extracted the API data through the MHP and saved it in JSON format on our database.

7.2.3.3 Stress Questionnaires

There are a limited number of validated stress questionnaires for EMA-style data collection. To mitigate this issue, we used the stress subscale of the Depression, Anxiety, and Stress Scale (DASS-21) as there is promising evidence of using DASS-21 with EMA (Questions 1-7) [128]. This was combined with a single-item measure (Question 8) used successfully for stress measurement with a moderate correlation to robust stress questionnaires [129].

The following 8 questions, on a LIKERT-type scale [128], were designated as the Stress Questionnaire for participants:

1. I felt that I was using a lot of nervous energy;
2. I found it hard to wind down;
3. I found myself getting agitated;
4. I found it difficult to relax;

5. I tended to over-react to situations;
6. I was intolerant of anything that kept me from getting on with what I was doing;
7. I felt that I was rather touchy;
8. Right now, I am…

Questions 1-7 have the options: “Not at all”, “To some degree”, “To a considerable degree”, and “Very much”, while Question 8 has “Stressed Out”, “Definitely stressed”, “A little stressed”, “Feeling good”, and “Feeling great”. The questions were displayed to the user in a random order each time the questionnaire is filled.

7.2.3.4 Mobile Health Platform

As discussed, we developed a mobile health platform (MHP) using Apple’s software for creating iOS apps, (XCode, version 12.5.1). The MHP acted as a user interface: automatically collecting data from Apple Health (via HealthKit) and allowing users to complete the stress questionnaire [29]. More details are provided in the Results section.

7.2.3.5 Study Population

Participants were recruited from the University of Waterloo (students) and online advertisements (workers; Facebook Ads and Kijiji). Participants were local to the Kitchener-Waterloo region in Ontario, Canada. Participants were initially only included if they were healthy. This requirement was subsequently relaxed to allow ‘unhealthy’ participants (chronic disease or illness, prescription drug use, or frequent use of alcohol or drugs). Participants were offered CAD 100.00 for two weeks of data collection. Additional data collection beyond two weeks was requested from some participants who had missed measurements (less than 6 measurements per day). This study was approved by University Waterloo Research Ethics Board (REB [43612]). Participant consent for data collection was obtained before device distribution. Data from 40 participants were collected. After applying the data cleaning and pre-processing described below, 7 participants had less than 50% of data points available. Therefore, these participants were excluded, and the subsequent analysis was done on 33 participants. Table 15 shows the characteristics of the study participants. Of note, 27% of participants were male and 73% were female; 24% were aged 18-24, 30% were aged 25-34 and the same proportion was found for participants aged 35-44, 12% were aged 45-64, and only 3% (1 participant) was aged over 65. The average BMI was 27.2 (±6.70), and participants had an average of 65.1 (±11.80) valid ECG recordings.

Table 15: Study Population Characteristics

Participants (N = 33)	*Frequency*	*Percentage*
Age		
18-24	8	24
25-34	10	30
35-44	10	30
45-64	4	12
Above 65	1	3
Gender		
Male	9	27
Female	24	73
SES		
Low (0-$30,000)	15	45
Medium ($30,000– $100,000)	15	45
High (Above $100,000)	2	6
Do not wish to disclose	1	3
Profession		
Full-time	14	42
Part-time	3	9
Student	13	39
Self-employed/Other	2	6
Retired	1	3
Ethnicity		
Black or African American	2	6
Chinese	4	12
Indian	1	3
Latin American	8	24
South Asian	6	18
White	12	36
Health Status		
Healthy	26	79
Chronic Disease or Illness, Prescription Drug Use,	7	21

Smoking or Alcohol

7.2.3.6 Data Pre-Processing and Analysis

We exported the ECG data from HealthKit into a CSV format and sorted each ECG voltage measurement by timestamps. We removed any ECG measurement that was classified as Poor Recording or Inconclusive by the ECG app [127]. The CSV file was imported into Kubios Premium 3.5.0 to determine heart rate variability (HRV) signals [161,171].

In order to apply signal filtering, we used the Kubios automatic beat detection feature as well as automatic noise detection, which excludes all segments marked as noise – the default Medium setting was used for noise segments. Kubios also has an automatic artefact correction method which was used for this analysis, and any samples containing more than 5% of corrected beats was removed. A list of the features generated by Kubios is presented in Table 16 [140,161]. Kubios automatically calculates a list of features for HRV analysis [140,161]. However, some features could not be calculated by the software with the 30 seconds measurements, and so these features were not used. The full list of Kubios features used for the analyses are mentioned in Table 16.

Table 16: HRV Features

Time-Domain Features	
Name	Description
PNS Index	Parasympathetic nervous system activity compared to normal resting values
SNS Index	Sympathetic nervous system activity compared to normal resting values
Stress Index	Square root of Baevsky's stress index
Mean RR	Mean of R-R intervals
SDNN	Standard deviation of R-R intervals
Mean HR	Mean of heart rate
STD HR	Standard deviation of instantaneous heart rate
Min HR	Minimum instantaneous heart rate calculated using 5 beat moving average
Max HR	Maximum instantaneous heart rate calculated using 5 beat moving average
RMSSD	Square root of the mean squared differences between successive RR intervals
DC	Heart rate deceleration capacity
DCMod	Modified DC computer as a two-point difference
AC	Heart rate acceleration capacity
ACMod	Modified AC computer as a two-point difference

Frequency-Domain Features	
FFT LF	Fast Fourier Transform Low Frequency band components
FFT HF	Fast Fourier Transform High Frequency band components
AR LF	Autoregressive Low Frequency band components
AR HF	Autoregressive High Frequency band components
FFT Absolute Power LF	Fast Fourier Transform Absolute Power of Low Frequency band components
FFT Absolute Power HF	Fast Fourier Transform Absolute Power of High Frequency band components
AR Absolute Power LF	Autoregressive Absolute Power of Low Frequency band components
AR Absolute Power HF	Autoregressive Absolute Power of High Frequency band components
FFT Relative Power LF	Fast Fourier Transform Relative Power of Low Frequency band components
FFT Relative Power HF	Fast Fourier Transform Relative Power of High Frequency band components
AR Relative Power LF	Autoregressive Relative Power of Low Frequency band components
AR Relative Power HF	Autoregressive Relative Power of High Frequency band components
FFT Normalized Power LF	Fast Fourier Transform Normalized Power of Low Frequency band components
FFT Normalized Power HF	Fast Fourier Transform Normalized Power of High Frequency band components
FFT Total Power	Fast Fourier Transform Total Power
FFT LF/HF	Fast Fourier Transform ratio between low and high frequency
AR Normalized Power LF	Autoregressive Normalized Power of Low Frequency band components
AR Normalized Power HF	Autoregressive Normalized Power of High Frequency band components
AR Total Power	Autoregressive Total Power
AR LF/HF	Autoregressive ratio between low and high frequency
Non-Linear Features	
SD1	The standard deviation perpendicular to the line-of-identity in Poincaré plot
SD2	The standard deviation along the line-of-identity in Poincaré plot
SD2/SD1	Ratio between SD2 and SD1

In addition, several features were excluded following recommendations made by the Task Force of The European Society of Cardiology and the North American Society of Pacing and Electrophysiology [153]: we removed pNN50 and NN50 as they are highly correlated with the RMSSD, and the RMSSD was preferred. The TINN, HRV Tri Index, VLF, and log measurements were removed as they were indicated for longer time periods than that measured here. Finally, features that were highly correlated were identified using the Pearson correlation method ($r = 0.95$) and removed.

Participant stress states for each measurement were determined based on the results of the stress questionnaires. Measurements were categorized as “stress” or “no stress” based on the following criteria. The scores of the DASS-21 questions (Questions 1-7) were summed together and multiplied by 2; if the score was greater than 14, the sample was classified as “stress” [162]. For the single-item measure (Question 8), the sample was classified as “stress” if the score was greater than 2. To integrate data from two separate questionnaires, if either the DASS-21 score or the single-item score was classified as “stress”, the sample was classified as “stress”.

We divided the dataset into 70% for training and validation and 30% for testing. We used 10-fold cross-validation for training the RFs and SVMs, which were developed using sci-kit learn. These models were chosen as they are widely and successfully used in stress prediction literature [35]. The “GridSearchCV” function was used to tune the model parameters and find the best ones. The data were normalized using sci-kit learn’s “StandardScaler” function for optimization.

The models were trained to the entire dataset as well as the subset of healthy participants. Given the relationship between HRV measures and demographics [146,172–174], we trained models based on age (18-24 years, 25-34 years, 35-44 years, and 45-65 years), gender (male, female), income (< $30,000 CAD, > $30,000 CAD), and profession (student, worker). For each model, we calculated feature importance for the RF model using the mean decrease in impurity (a 100% purity in a node means the decision tree’s node contains only one class, and by assessing the difference between the impurity in the parent and child nodes we can calculate the best split in the tree and use it as a proxy for feature importance). For categories that had only one participant we did not perform the model analyses.

7.2.4 Results

We sought to pilot the use of machine learning with Apple Watch ECG data as a step towards developing a wearable device for stress prediction. We recruited students and staff from the University of Waterloo (Ontario, Canada) to participate in a two-week study. Participants were given an iPhone 7 and an Apple Watch Series 6. It is important to note that this study is part of a larger wearable study involving other devices such as wireless blood pressure cuffs; for this study, we focus specifically on the iPhone and Apple Watch and on ECG measurements alone. The details of the other study are described elsewhere [89,90].

Users were asked to collect ECG measures using the Apple Watch ECG app six times during the day at approximately three-hour intervals. Before the acquisition of each ECG, participants were asked to complete a stress questionnaire on the MHP developed for the study, which also updated new ECG measurements to our database. The MHP app interface is depicted in Figure 12.

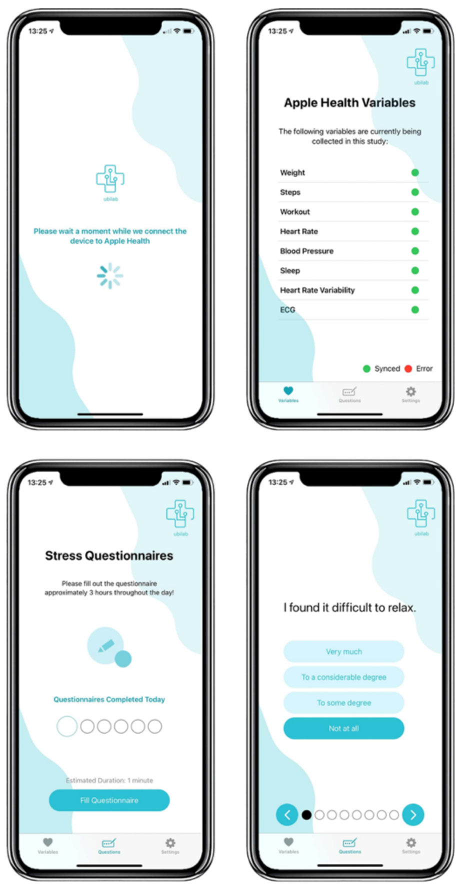


Figure 12: MHP Interface

As there are few validated stress questionnaires, we made use of the stress portion of the Depression, Anxiety, and Stress Scale (DASS-21) in conjunction with a single-item measure that has been used successfully in previous stress prediction studies [129]. In total, we acquired 2421 ECG/survey measures from 33 participants after data cleaning and pre-processing. Readings were classified as “stress” or “no stress” based on the answers to the questionnaire [162]. We applied the machine learning models Random Forests (RFs) and Support Vector Machines (SVMs) to train the model.

Table 17 shows a summary of the results for each trained model, described in more detail below.

Table 17: Metrics for Each Trained Model

	Random Forest				**Support Vector Machine**			
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	***Support***
Complete Dataset	*No Stress*	0.58	0.66	0.61	0.58	0.58	0.58	359
	Stress	0.52	0.43	0.47	0.50	0.51	0.51	306
	Accuracy	-	-	0.55	-	-	0.54	665
	Weighted Average	0.55	0.55	0.55	0.54	0.54	0.54	665
Healthy Subjects	*No Stress*	0.59	0.73	0.65	0.57	0.57	0.57	283
	Stress	0.50	0.34	0.40	0.45	0.45	0.45	221
	Accuracy	-	-	0.56	-	-	0.52	504
	Weighted Average	0.55	0.56	0.54	0.52	0.52	0.52	504
Subjects Aged 18-24	*No Stress*	0.64	0.88	0.74	0.63	0.75	0.69	95
	Stress	0.50	0.19	0.27	0.43	0.31	0.36	59
	Accuracy	-	-	0.62	-	-	0.58	154
	Weighted Average	0.58	0.62	0.56	0.56	0.58	0.56	154
Subjects Aged 25-34	*No Stress*	0.39	0.16	0.23	0.42	0.14	0.22	69
	Stress	0.67	0.87	0.76	0.67	0.89	0.77	133
	Accuracy	-	-	0.63	-	-	0.64	202
	Weighted Average	0.57	0.63	0.58	0.58	0.64	0.58	202
Subjects Aged 35-44	*No Stress*	0.69	0.91	0.78	0.68	0.67	0.68	116
	Stress	0.61	0.26	0.37	0.43	0.45	0.44	65
	Accuracy	-	-	0.67	-	-	0.59	181
	Weighted Average	0.66	0.67	0.63	0.59	0.59	0.53	181
Subjects Aged 45-64	*No Stress*	0.70	0.84	0.76	0.68	0.86	0.76	50
	Stress	0.53	0.33	0.41	0.50	0.26	0.34	27
	Accuracy	-	-	0.66	-	-	0.65	77
	Weighted Average	0.64	0.66	0.64	0.62	0.65	0.61	77
	No Stress	0.64	0.68	0.66	0.57	0.56	0.57	95

Male Participants	*Stress*	0.62	0.58	0.60	0.52	0.53	0.52	85
	Accuracy	-	-	0.63	-	-	0.55	180
	Weighted Average	0.63	0.63	0.63	0.55	0.55	0.55	180
Female Participants	*No Stress*	0.59	0.55	0.57	0.59	0.67	0.63	250
	Stress	0.52	0.56	0.53	0.55	0.47	0.50	214
	Accuracy	-	-	0.55	-	-	0.58	464
	Weighted Average	0.56	0.55	0.55	0.57	0.58	0.57	464
Low SES Participants	*No Stress*	0.62	0.74	0.68	0.64	0.63	0.64	180
	Stress	0.45	0.32	0.37	0.45	0.45	0.45	118
	Accuracy	-	-	0.57	-	-	0.56	298
	Weighted Average	0.55	0.57	0.56	0.56	0.56	0.56	298
Medium and High SES Participants	*No Stress*	0.53	0.42	0.47	0.55	0.48	0.51	161
	Stress	0.53	0.64	0.58	0.55	0.62	0.58	167
	Accuracy	-	-	0.53	-	-	0.55	328
	Weighted Average	0.53	0.53	0.52	0.55	0.55	0.55	328
Students	*No Stress*	0.57	0.51	0.54	0.60	0.54	0.56	134
	Stress	0.54	0.60	0.57	0.56	0.62	0.59	128
	Accuracy	-	-	0.55	-	-	0.58	262
	Weighted Average	0.56	0.55	0.55	0.58	0.58	0.58	262
Workers	*No Stress*	0.55	0.69	0.61	0.56	0.69	0.62	200
	Stress	0.49	0.35	0.41	0.52	0.39	0.45	170
	Accuracy	-	-	0.53	-	-	0.55	370
	Weighted Average	0.52	0.53	0.52	0.54	0.55	0.54	370

7.2.4.1 Stress Prediction Models Using Total Dataset and Subset with Healthy Subjects

The RF and SVM models were trained against the complete data set. The complete dataset was fairly balanced, with the “stress” class representing 46% of all test examples (306 out of 665 in the test dataset). Due to class imbalances, we reported the F1-score weighted. The best accuracy was achieved by the RF model with 55% compared to 54% for the SVM model (Table 3). Weighted averages were similar to the accuracy. Recall and precision were higher for the “no stress” class when compared to the “stress” class, with the SVM having a higher recall for the “stress” class than the RF. Results indicated that, when using ECG measurements from a wearable device in a real-life setting, both the RF and SVM machine learning models approached the lower end of state-of-the-art accuracy levels for predicting stress levels.

As there are multiple heart rate variability (HRV) parameters determined by the ECG test, we sought to identify the most important features for the RF algorithm. The top 10 features were determined using the mean decrease in impurity. Figure 13 shows that the top feature was the ECG heart rate deceleration capacity (DC) (Table 16).

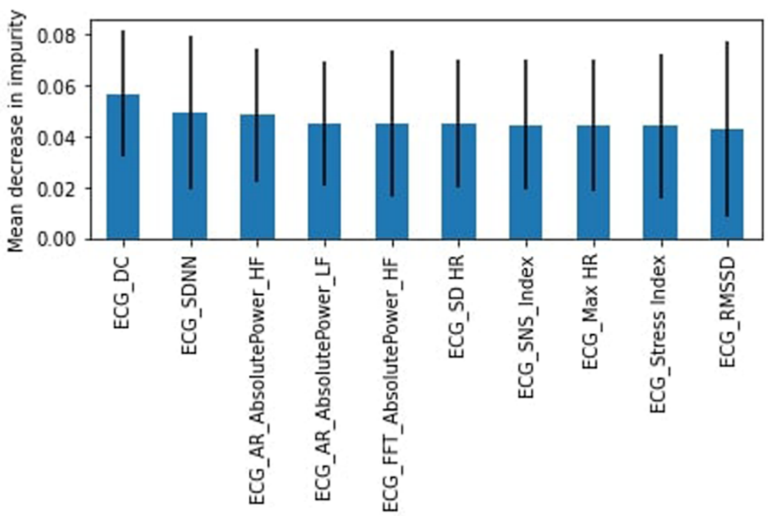


Figure 13: RF feature importance, complete dataset

Originally, all participants involved in this study were healthy; however, due to difficulty in finding study subjects, we relaxed the criteria to allow participants that were not healthy (chronic disease or illness, prescription drug use, or frequent use of alcohol or drugs). Again we found the RF model outperformed the SVM model. The healthy subset achieved a slightly lower weighted average for the “stress” class of 54% for the RF model with a recall of 34% (45% for SVM) and precision of 50%. DC was again identified as the most important feature (Figure 14) followed by the heart’s acceleration capacity (AC).

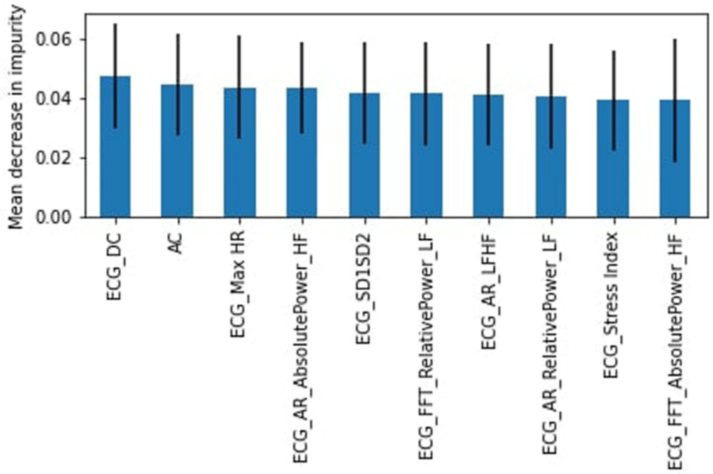


Figure 14: RF feature importance, healthy subjects

7.2.4.2 Impact of Age on Stress Prediction Models

Given that age influences HRV [146,153], we trained the RF and SVM models based on age to see if we could improve the weighted average.

For the 18-24 years group, the "stress" class represented 38% of the data. The RF model outperformed the SVM model with an accuracy of 62% and an F1-score weighted of 56%. The recall and precision in the "stress" class were 19% and 50% respectively. Low Frequency Absolute Power calculated with FFT was identified as the most important feature (Figure 15).

In the 25-34 years group, the "stress" class is the majority, representing 66% of the dataset. Here the SVM model slightly outperformed the RF model with an accuracy of 64% compared to 63%. The F1-score weighted was 58% for both models. In the "stress" category, the recall (87%) and precision (67%) were high but with a corresponding loss of recall (16%) and precision (39%) in the "no stress" class. The most important feature was the standard deviation of intervals, SDNN (Figure 16).

In the 35-44 years group, the "stress" class was the minority, representing 36% of the dataset. The RF had a higher accuracy of 67% (F1-score weighted of 63%) compared to 59% for the SVM model (F1-score weighted of 53). The SVM had higher recall than the RF for the "stress" class (45% to 26%), but lower precision (43% to 61%). The AC was the most important feature (Figure 17).

Finally, for the 45-64 years group, the "stress" class comprised 54% of the dataset. We found that the RF model performed better than the SVM with an accuracy of 66% and an F1-

score weighted of 64%. The “stress” class had a low recall of 33% and a 53% precision. AC was the most important feature as well (Figure 18).

To determine which features were most commonly identified as important across all age groups, we determined the frequency which with features appeared in the top 10. Figure 19 shows that the deviation of the instantaneous heart rate (SD HR), heart acceleration capacity (AC) and AR Low Frequency Absolute Power were the most important features across all age-related models.

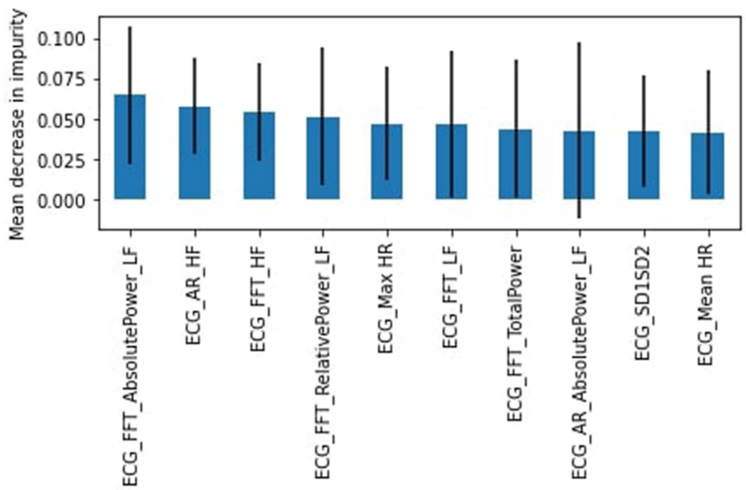


Figure 15: RF feature importance, subjects aged 18–24

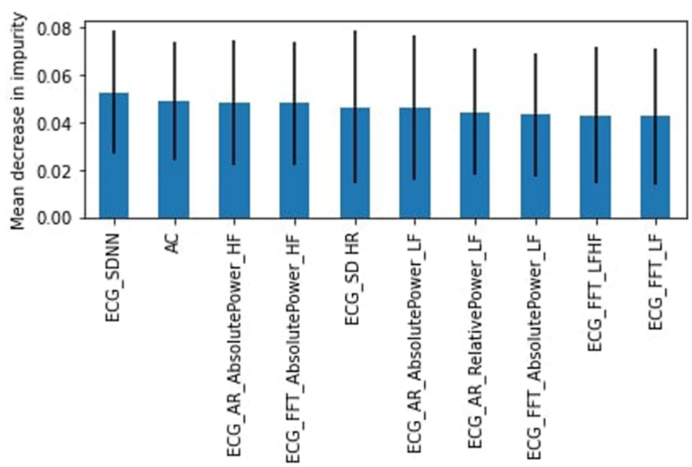


Figure 16: RF feature importance, subjects aged 25–34.

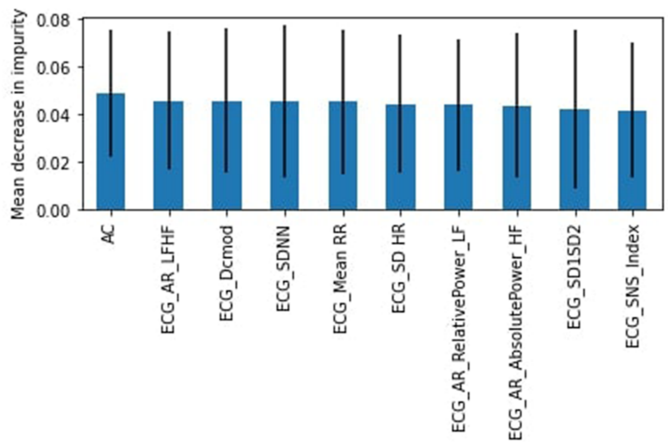


Figure 17: RF feature importance, subjects aged 35–44

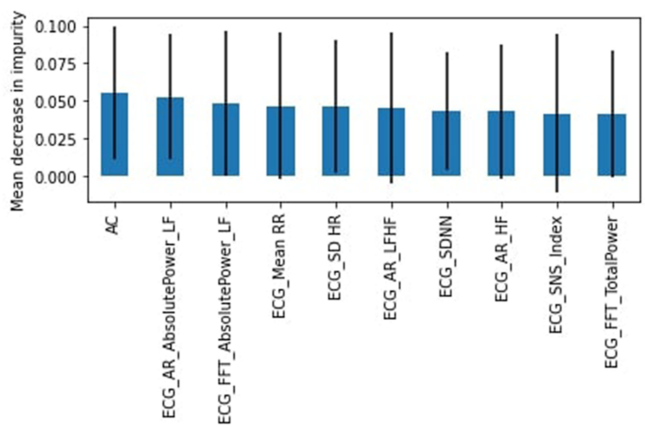


Figure 18: RF feature importance, subjects aged 45–64

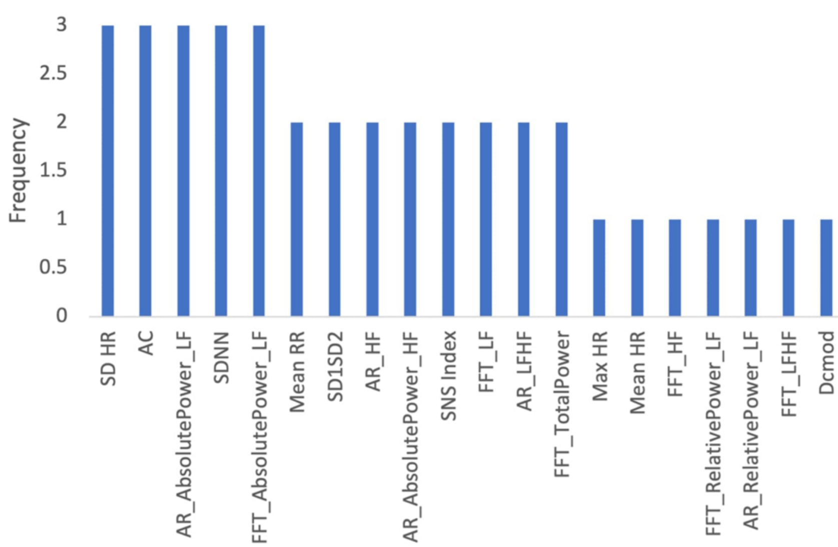


Figure 19: Frequency of features in age related models

7.2.4.3 Impact of Gender on Stress Prediction Models

Evidence suggests that gender has an impact on HRV [172]. To determine whether our stress prediction would improve if we accounted for gender, we trained the RF and SVM learning models for males and females. The "stress" class represented slightly more than 45% of the datasets.

The RF model performed better for the male participants with an accuracy and F1-score weighted of 63%. The precision was 62% and recall was 58% for the "stress" class. The PNS Index was the most important feature (Figure 20).

In contrast, the SVM model performed better for female participants with an accuracy of 58% and weighted average of 57%. The "stress" class had a precision of 55% and a recall of 47%. SDNN was found to be the most important feature (Figure 21).

Figure 22 shows the frequency with which each feature appeared as the 10 most important features across both gender-related models. DC, PNS Index, and FFT High Frequency Absolute Power were identified as the most frequently important HRV features.

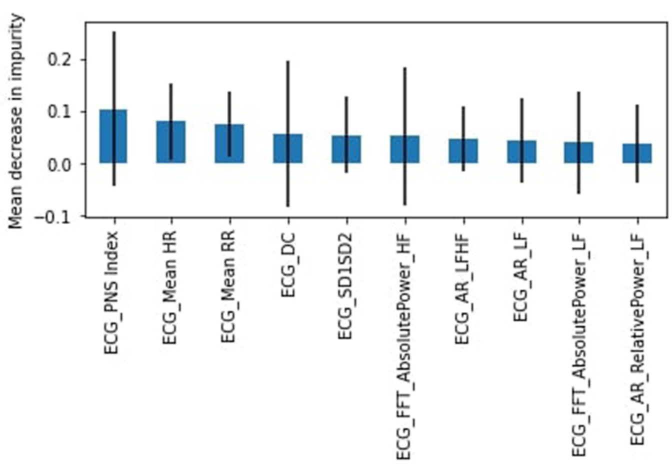


Figure 20: RF feature importance, male participants

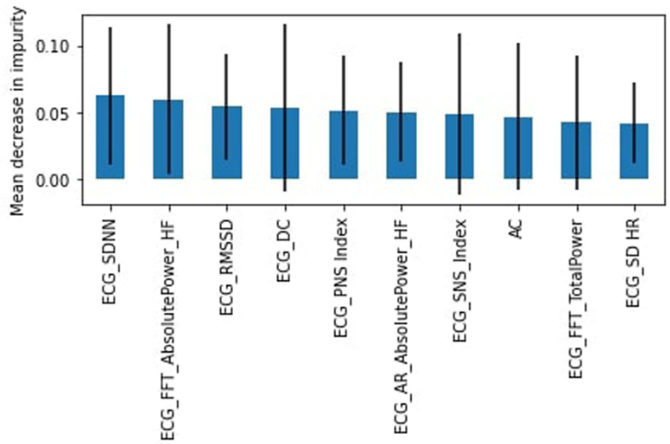


Figure 21: RF feature importance, female participants

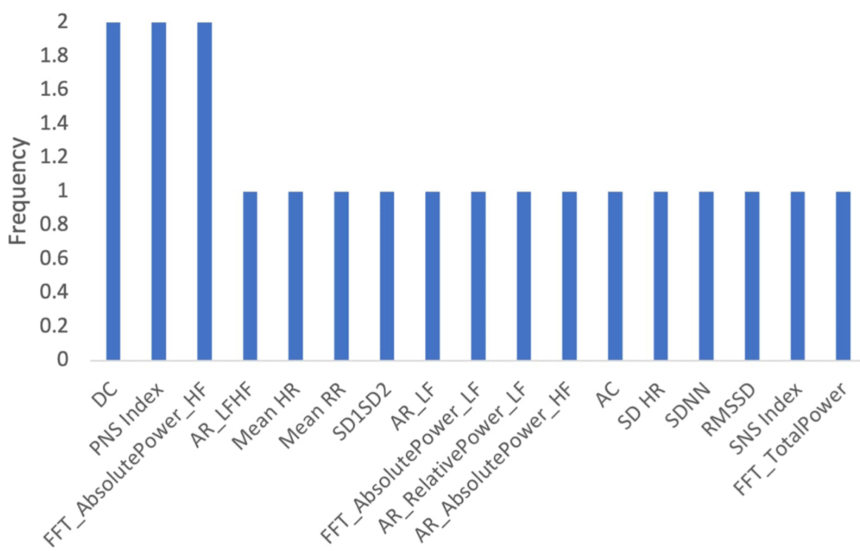


Figure 22: Frequency of Features in gender-related models

7.2.4.4 Impact of Socioeconomic Status on Stress Prediction Model

As there are large socioeconomic disparities in cardiovascular disease and HRV, we sought to train our machine learning models based on socioeconomic status (SES) [173] . Participants were considered to be in the low SES category if their net income was < 30,000 CAD based on an approximation from the Canadian tax cut-off for low-income populations [175]. Those with incomes above this threshold were considered medium-to-high SES.

For the low SES group, the “stress” class comprises 40% of the dataset. The SVM model performed better with an accuracy and F1-score weighted of 56%, recall of 45%, and precision of 45% for the “stress” class. The most important feature was DC, the heart deceleration capacity (Figure 23).

For the medium and high SES participants, the "stress" class represented 51% of the dataset. The SVM model performed slightly better than the RF model with accuracy and F1-score weighted of 55%. For the "stress" class, the recall was 62% (slightly higher for RF at 64%) and precision was 55%. DC, the heart deceleration capacity, was again identified as the most important feature (Figure 24).

Figure 25 shows the frequency of features that appeared as the 10 most important features across both income-related models. The most frequently identified features were DC, SDNN, FFT Absolute Power HF, and AC.

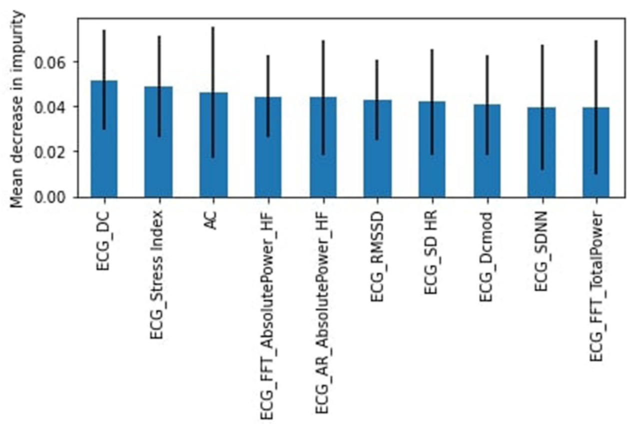


Figure 23: RF feature importance, Low SES participants

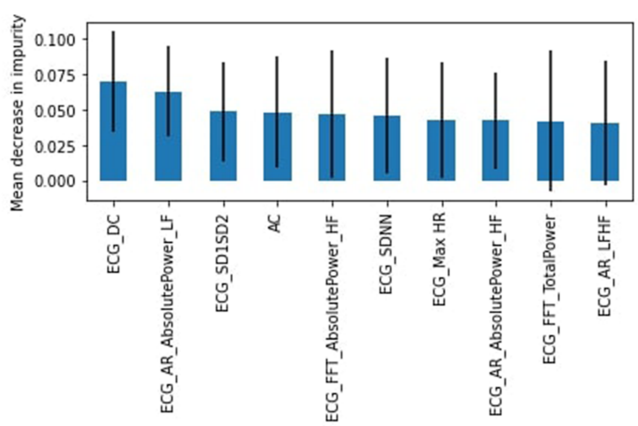


Figure 24: RF feature importance, medium and high SES participants

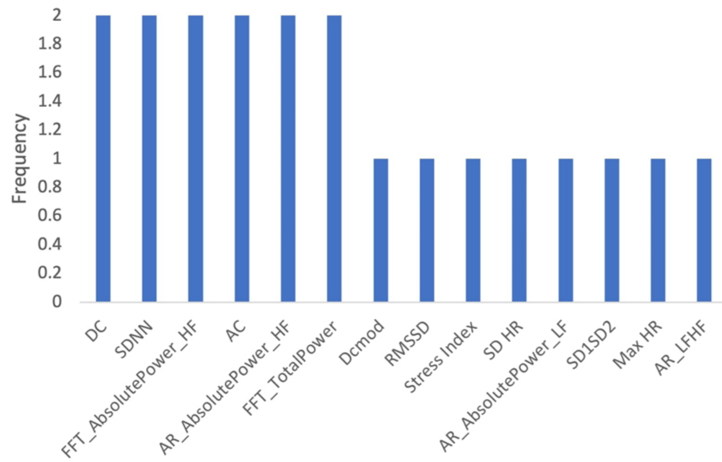


Figure 25: Frequency of features in income-related models

7.2.4.5 Impact of Profession on Stress Prediction Model

Occupational stress is associated with cardiovascular disease and HRV [172]. As such, we trained our stress prediction models based on participant occupation. Participants were categorized as workers (full-time, part-time, self-employed, or other) and students. We did not train a model for the retired participant as only one participant was in that category.

The SVM had better accuracy (58%) and F1-score weighted (58%) when models were trained for students. The "stress" class represented 49% of the dataset with a recall of 62% and precision of 56%. The AR High Frequency Absolute Power was the most important feature (Figure 26).

When we trained the model for workers, the "stress" class represented 46% of the dataset. The SVM model slightly outperformed the RF in accuracy (55% compared to 53%) and F1-score weighted (54% to 52%). The SVM had a better recall (39%) and precision (52%) for the "stress" class. Figure 27 shows the 10 most important features, with the AC as the most important feature.

We determined which features appeared most frequently as the top 10 most important features across both datasets (Figure 28). The most important features were DC, SDNN, AC, and SD HR.

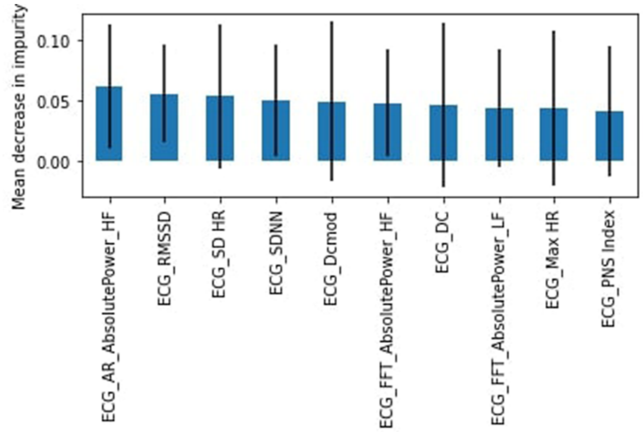


Figure 26: RF feature importance, students

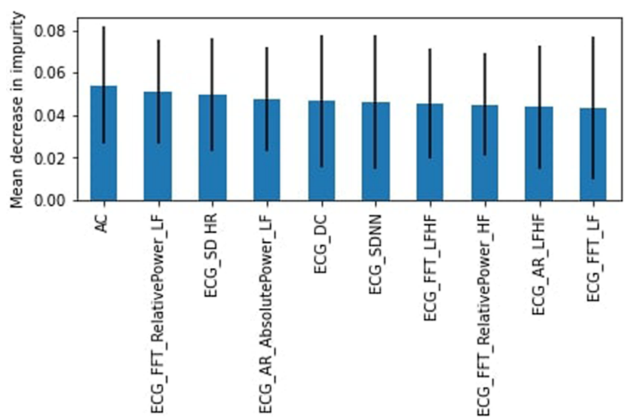


Figure 27: RF feature importance, workers

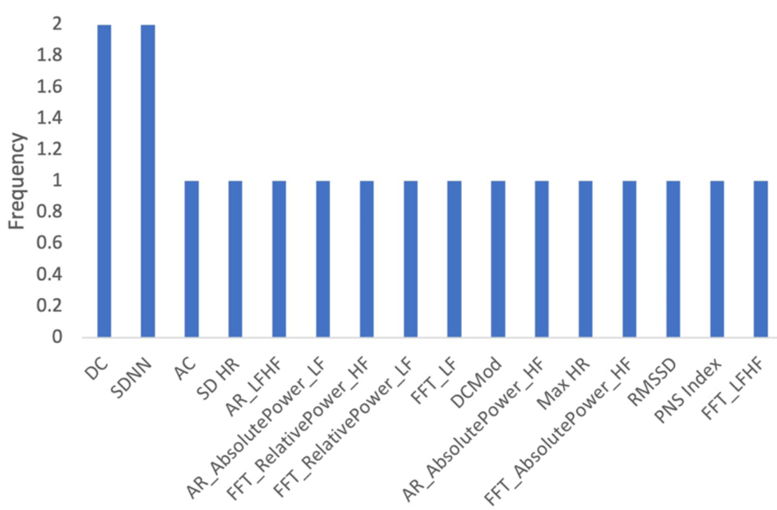


Figure 28: Frequency of features in profession-related models

7.2.4.6 Priority HRV Features for Stress Prediction Models

Several HRV features collected during the ECG measurements were identified as important across the models trained for the entire dataset, age, gender, socioeconomic status, and profession. We determined the frequency with which each feature appeared as the "10 most important features" across all 10 models described above (Figure 19). The top features identified were SDNN, AC, and DC.

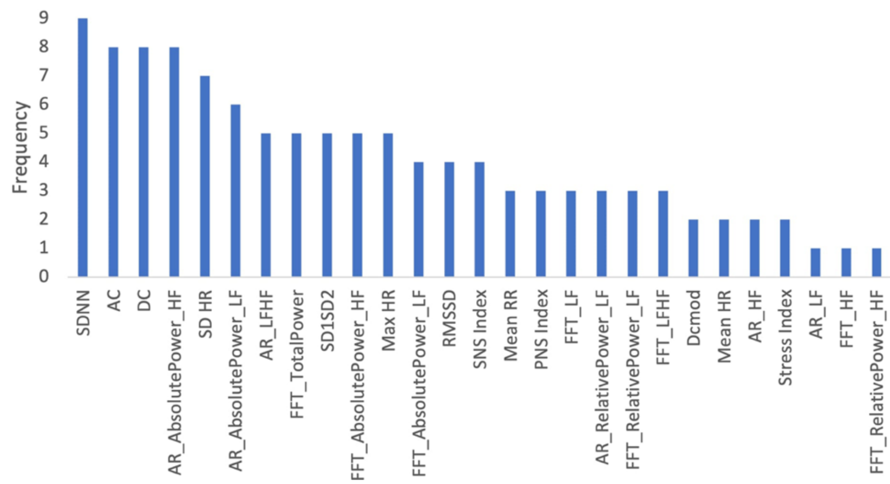


Figure 29: Frequency of features in all models

7.2.5 Discussion

Here we piloted the use of an Apple Watch ECG sensor to predict participant stress levels. Overall, both models performed similarly in different circumstances, achieving F1-weighted scores ranging from 52% to 64%. The state-of-the-art accuracy for stress detection in real-life settings lies approximately between 60% and 80% [35]. In general, the "stress" models had a high level of precision but lower recall. The "no stress" models performed generally well with a recall typically above 60%. Considering the ultra-short duration of the ECG measurements performed here compared to the standard, as well as the nature of real-life measurements, the results presented were quite promising.

Divisions by gender, profession or income were found to be good proxies for the prediction models, although more data seems to be needed for improvement. In the majority of cases, the models performed better for the "no stress" class compared to the "stress" class. As the fraction of data falling into the "no stress" class was often greater, the performance discrepancy may be related to class imbalance. Future studies should explore over- and under-sampling techniques to improve the models. Overall, while the models have high specificity, predicting "no stress" states relatively well, they currently lack the predictive power to accurately predict the "stress" states. Future work should focus on frequency-domain metrics and implement novel approaches for data analyses. Additional stress-related variables could also be integrated into the analyses, as well as exploring training and testing datasets based on subjects rather than randomly.

The heart acceleration (AC) and deceleration capacity (DC) were some of the most valuable HRV features included in the model, being present in most, if not all, of the 10 most important features in all models described. This is interesting as AC and DC are relatively new indicators in HRV analyses and lack research with a focus on stress; these results, then, can indicate new avenues of research focusing on these metrics for stress prediction [176]. The SDNN, one of the most widely used metrics for time-domain HRV, was also present in most models. Frequency-domain features were commonly identified as important as well. This was consistent with the Task Force recommendations; frequency-domain metrics are better at capturing variations in HRV than time-domain metrics for short measurement periods of time.

Still regarding feature importance, it is important to note the wide error bands for most of the calculated mean decrease in impurity, which points to the fact that the different trees in the random forest models are varied to take into account all complexities in the data. Most features possess similar wide error bands, and that the features described above are repeated throughout the models, suggesting that they are the most important ones and should be evaluated carefully. One limitation of this study was skewed population representation: participants were primarily white females. As such, there may have been insufficient data to accurately train the models for other representative groups. As well, due to limited participant numbers, it was challenging to stratify characteristics. For example, socioeconomic status and profession were only stratified into two categories which may be insufficient to capture demographic features. Here we applied the use of RF and SVM to train the models [35], however, other methods may perform better including Deep Learning approaches. Future work could apply Deep Learning methods using the raw signals from each participant's ECG measurements as time series data.

To the best of our knowledge, this is the first study to use Apple Watch ECG data to predict the stress levels of individuals. The results are currently in the low-end of state-of-the-art; as mentioned above, stratifying participants can improve accuracy, and larger studies that allow further stratification of the cohort might achieve even better results. In addition, data was collected in real-life conditions which can potentially introduce noise in the data. On the same token, stress self-report was used as the ground truth for a given moment in time, which might not always reflect physiological parameters. Since the results were promising with these factors potentially introducing noise in the data, and given the novelty of the data type, conducting further studies in a controlled setting, such as applying stressors in a lab environment, could give us additional insights into the relationship between Apple Watch ECG data and stress. In addition, since the Apple Watch can also collect additional data such as sleep and physical activity, it should also be interesting to use ECG data with other stress-related variables, as they can complement the data and increase the models' predictive power.

7.2.6 Conclusion

This study presented an analysis of Apple Watch ECG data from 33 participants. To the best of our knowledge, this is the first study to use Apple Watch ECG data to predict stress levels of individuals. RF and SVM models were developed for the task, with the models performing similarly.

Further, the results are in line with the start-of-the-art for stress prediction, although at the low-end. This is very promising considering the ultra-short-term and real-life nature, as well as the novelty of, the Apple Watch ECG data. However, while the current models have high specificity, predicting "no stress" states relatively well, it lacks the predictive power to accurately predict the "stress" states as of yet. Future work should focus on the AC, DC, SDNN as well as frequency-domain metrics and implement novel approaches for data analyses, such as Deep Learning, as well as integrating additional stress-related variables into the analyses.

Overall, the results from the pilot study validate the continued development of wearable ECG technology and suggest that, with further refinement, models can likely achieve stress prediction with state-of-the-art quality. In that way, we can develop near real-time, non-intrusive stress detection, monitoring, and intervention applications using a technology that is already widely popular and accepted by the population, leading to better health outcomes.

7.3 Discussion

When considering the accuracy metrics, the results in several of the stratifications were close to or at the bottom of the state-of-the-art (around 60%) for stress prediction. However, a lot of results were close to 50%, highlighting the need for further validation.

As we shall see, further data pre-processing, including missing data imputation, and more robust inclusion/exclusion criteria for features with missing components will improve the performance of these models (f1-macro scores will be the focus of later chapters as they consider both *stress* and *no stress* classes as having equal importance). However, features derived from the ECG were not as important as others in future chapters. In this manner, coupled with the results from Chapter 6, using HRV data from the Apple Watch ECG by itself does not seem to be a good approach for stress quantification, although there is the potential for improvement with further research, potentially in more controlled environments. The next chapter describes the pilot study in full, including all data types collected and devices used.

Chapter 8 - Applying MHP in Public Health Surveillance: Stress Prediction and Lessons Learned

8.1 Foreword

Following the preliminary analyses and models shown in Chapters 6 and 7, this chapter discusses the pilot study in its entirety. This includes an overview of all data types collected, features derived from these data, and mobile and wearable devices used to collect it. It also uses data from all participants rather than a subset as was done in Chapters 6 and 7. Some subsets from Chapter 7 are eliminated (e.g., people aged above 45) due to the low number of participants in these stratifications, which can affect the results.

The ML models presented in this chapter were trained in a similar manner to the ones presented in Chapter 7, i.e., using data from all participants and dividing the dataset into training and testing sets, with 10-fold cross-validation. However, the choice was made to use an 80-20 split for training and testing rather than 70-30 as used in Chapter 7. Since the dataset collected was not particularly large – although larger than most studies in literature, as we will see – especially when stratifying participants, an 80-20 split enabled more data for the model learning process.

Apple Watch ECG data is also included in this paper, and separate models using only these data are included. The data was subject to more data pre-processing than the dataset used in Chapter 7, namely missing data imputation using k-nearest neighbours, which led to better results.

Further, separating the entire dataset into training and testing without accounting for individual participants is not the only approach for model training with stress prediction that is used in the literature, and might have disadvantages such as the models benefiting from seeing the data from participants in the test set. Chapter 9 will deal with alternative methods of creating the models.

In addition to presenting the development and results of ML models with the entire dataset, this paper also discusses lessons learned from applying the current version of the MHP in practice for data collection in the pilot study, including limitations of the prototype and challenges encountered with several devices selected for the study. Future directions for new versions of the MHP and other similar surveillance systems are also presented. Finally, implications for public health from these lessons are discussed.

8.2 Application of a Mobile Health Data Platform for Public Health Surveillance: A Case Study in Stress Monitoring and Prediction

8.2.1 Abstract

Background: Public health surveillance involves the collection, analysis and dissemination of data to improve population health. The main sources of data for public health decision-making are surveys, typically comprised of self-report which may be subject to biases, costs and delays. To complement subjective data, objective measures from sensors could potentially be used. Specifically, advancements in personal mobile and wearable technologies enable the collection of real-time and continuous health data.

Objective: In this context, the goal of this work is to apply a mobile health platform (MHP) that extracts health data from the Apple Health repository to collect data in daily-life scenarios and use it for the prediction of stress, a major public health issue.

Methods: A pilot study was conducted with 45 participants over 2 weeks, using the MHP to collect stress-related data from Apple Health and perceived stress self-reports. Apple, Withings and Empatica devices were distributed to participants and collected a wide range of data, including heart rate, sleep, blood pressure, temperature, and weight. These were used to train random forests and support vector machines. The SMOTE technique was used to handle imbalanced datasets.

Results: Accuracy and f1-macro scores were in line with state-of-the-art models for stress prediction above 60% for the majority of analyses and samples analyzed. Apple Watch sleep features were particularly good predictors, with most models with these data achieving results around 70%.

Conclusions: A system such as the MHP might be used for public health data collection, potentially complementing traditional self-reporting methods when possible. The data collected with the system showed promise for monitoring and predicting stress in a population.

8.2.2 Introduction

The goal of public health is to improve and protect the health of communities and populations [126]. In order to understand the characteristics of a population and where treatments and interventions are more effective, public health agencies typically conduct surveillance efforts to collect and analyze data [72–74]. These efforts are traditionally focused on self-report, such as surveys and questionnaires. For example, the Canadian Health Measures Survey [78] and the

Canadian Community Health Survey [80] are major surveys that collect data on the characteristics, behaviour and health of Canadians. However, subjective and self-reported data may be subject to limitations such as biases, delays, costs and logistics [2,3,10–15,17,18].

New advancements in sensing and remote monitoring technologies allow the ubiquitous and effortless monitoring of objective health data with the use of smart devices and Internet of Things (IoT) solutions [177]. For example, smartphones can typically collect movement data; smart thermostats are able to collect temperature and movements around the house; and smartwatches collect a range of variables from heart rate to steps and sleep [89,90,177]. These technologies might potentially be used in complement to traditional data collection techniques, collecting objective data that can mitigate challenges associated with self-report – as evidenced by a number of recent studies that use mobile and wearable technologies to gain new insights into the health of individuals [21,26,77,81,85,147,178]. Further, given the personal nature of these devices, it could be possible to leverage data that is being passively collected in real-life environments for long periods. For instance, smartwatches typically collect heart rate and steps data from individuals wearing them throughout the day, without any action required on the user's part. This could provide new and large sources of continuous, real-world data collected with relatively low effort that might allow scientists to conduct novel health research [78].

Indeed, many efforts are being put into place to create platforms that allow individuals to share their data for research. For example, the ecobee smart thermostat company has a program called Donate Your Data [83,84], which enables the anonymous sharing of device information with researchers [25,86,87]. The Ubiquitous Health Technology Lab at the University of Waterloo has developed a web platform that enables the enrolment of personal Fitbit and ecobee devices for research: data from the devices is continuously collected once a day once enrolment is complete [88]. The Digital Epidemiology and Population Health Laboratory (DEPtH Lab) at Western University have developed the Smart Platform, which allows researchers to engage with personal devices of patients [179].

However, these mobile health datasets also bring new set of challenges. Larger datasets, generated at faster speeds than previous data collection efforts, with a variety of formats, structures, and even different data collection periods, require new methods of processing and analyzing data. Therefore, to handle this Big Data, Machine Learning (ML) techniques have been shown to be useful tools in analysis, discovery and prediction [55].

The goal of this study is to contribute to the informatization of public health data collection efforts, first by introducing the Mobile Health Platform (MHP), an iOS app that leverages data from personal devices that are stored in Apple Health (AH) [89 127], a popular health data repository. We describe how this app was applied in a pilot study with 45 participants, collecting a plethora of health data (e.g., sleep, steps, heart rate, blood pressure, among others). In turn, the collected data is used to predict stress states through the creation of ML models, specifically random forests (RFs) and support vector machines (SVM). Related work on stress prediction and the rationale behind stress as a use case are presented in the next subsection.

8.2.2.1 Related Work – Stress and Machine Learning

Stress is a major public health issue, with the World Health Organization calling it the "health epidemic of the 21st century" [141], and its prevalence is increasing. While stress is a normal response to an unexpected situation – generating energy and enabling the individual to deal with a threatening circumstance – the body should, ideally, return to its normal state once the situation is resolved [35,180]. Long-term, constant exposure to stressors can increase the risk for hypertension, cardiovascular diseases, and stroke, among others [34,35]. It is estimated that stress places a burden of over $300 billion USD annually on health costs and job performance [141,143], leading to 120,000 preventable deaths when coupled with a lack of health insurance [144]. The pandemic has also greatly affected the stress levels of individuals: according to a recent survey from the American Psychological Association, almost 70% of respondents experienced increased levels of stress due to COVID-19 [32].

Stress is typically collected for public health initiatives and in real-world environments through self-report [4,35,166]. In this way, it is an ideal use case for the MHP, which collects objective data that can be used for stress prediction. Indeed, many studies have sought to use ML coupled with mobile and wearable technologies to predict stress. For example, a study used daily self-report for 4 months coupled with variables such as physical activity and heart rate variability (HRV) to predict stress in 35 participants [46]. Logistic Regression was used in a generalized model – using data from all participants and a leave-one-person-out (LOPO) validation method – with 53% accuracy, and an individualized model – using data from each participant using a leave-one-day-out validation procedure – with 61% accuracy. Jin et al. [45] use data from 6 participants obtained with the Empatica E4 device such as blood volume pulse and electrodermal activity for 4 weeks, applying RFs and SVMs on a generalized model that uses 10-fold cross-validation to train and tune the model and a 10% validation set for testing, achieving an Area Under the Curve of 87.3% (RF) and 82.1% (SVM). Can et al. [169] collects heart rate (HR) and electrodermal activity data from 14 participants, both in the lab and in real life during one week, training several combinations of models (e.g., models developed using the real-life data, laboratory data, or combination of both). Several ML algorithms are used on a generalized dataset. In particular, a 68% accuracy is achieved with SVMs and 52% accuracy with RF using 10-fold cross-validation and 80- 20 train-test datasets for data collected in real-life.

In addition to daily self-report, many studies use stressors applied in a laboratory or controlled environment [42,166,181]. Akmandor and Jha [181] use ECG, respiration, blood pressure, and other variables from 32 participants to develop generalized – i.e., combining data from all participants into one dataset – and individualized models using SVMs and k-nearest neighbors (kNN), dividing the datasets into train, test and validation sets to validate the models without cross-validation. They achieved an accuracy of 89.2% (kNN) and 83.1% (SVM) for the generalized models and 94.5% (kNN) and 86.7% (SVM) for individualized ones. Liao et al. [182] use neural networks to develop generalized models based on attention and meditation (as opposed to stress and non-stress states) in the laboratory with EEG data from 7 participants, achieving f1-scores of 60% for the attention state but of only 1% for the meditation state.

As can be seen by the examples above, the state-of-the-art accuracy for stress prediction seems to lie between 60%-80%, decreasing for studies using real-life data. Further, there are many different ways to predict stress. Indeed, the works above vary widely according to the number of participants, period of data collection, algorithms used to develop the models, and metrics and methods applied to validate the models, among other factors. In this work, based on the results of previous studies and a survey by Can et al. [35], we elected to use RFs and SVMs, as they were successfully used in a variety of studies to predict stress. These models will be used to predict stress based on data collected from the MHP to test its efficacy in monitoring and predicting this condition in a population.

8.2.3 Methods

8.2.3.1 Recruitment and Study Protocol

We recruited participants from the University of Waterloo as well as through Facebook groups and Ads. Kijiji, a Canadian website that allows users to advertise products and services, was also used. Inclusion criteria consisted of participants aged 18 years and older, and initially involved participants that did not have any chronic condition, take any medication or consume alcohol/smoke frequently. The latter criteria were later relaxed due to a difficulty in finding participants, and this was accounted for in the analyses as will be expanded in the following subsections. Since devices were delivered to participants in-person, they were required to be located near the Kitchener-Waterloo region in Ontario.

45 participants were recruited for the study and were offered CAD 100.00 for two weeks of data collection. Participants were given the following devices (per manufacturer):

- **Apple**: iPhone 8 with iOS 15.0 and Apple Watch Series 6 with watchOS 8.3.

- **Withings**: Withings Sleep, Withings Blood Pressure Monitor (BPM) Connect, Withings Thermos and Withings Body+

- **Empatica**: Empatica E4 wristband

Of this list, the Empatica E4 is the only one that is not considered a personal, consumer-level device, and was included due to its extensive use in stress prediction literature. During our experiments, we conducted analyses excluding Empatica data to test if this research-grade device had a significant impact on the results.

Table 18 describes the variables collected in each device. In Appendix C, the User Manual shared with each participant providing instructions on how devices should be installed and used is included. A 1-hour video call was also scheduled with each participant to go over the manual, make sure the devices were installed and working properly, and answer any questions about the protocol.

This study followed the Ecological Momentary Assessment (EMA) methodology, which strives to obtain self-reports closer to events in daily life to approximate real-world scenarios and obtain accurate data [160]. Therefore, users were instructed to collect data 6 times during the day (starting at wake-up and finishing at sleep), in approximately three-hour intervals according to their daily routine. This included taking Weight, Blood Pressure, Heart Rate Variability, ECG and Temperature measurements and filling out the stress self-report forms. Apple Watch and iPhone Steps, Apple Watch HR, and Empatica E4 data were collected continuously without patient involvement. The protocol is shown in Figure 30, with the times shown being illustrative.

Table 18: Variables Collected and Devices Used in Study

Variable List	Devices Distributed to Participants						
Name	**iPhone (with MHP Installed)**	**Apple Watch**	**Empatica E4**	**Withings Sleep**	**Withings Blood Pressure Monitor Connect**	**Withings Thermos**	**Withings Body +**
Weight (Kg)							
Steps							
Heart Rate, HR (Bpm)							
Blood Pressure, BP (mmHg)							
Sleep							
Heart Rate Variability, HRV							
Electrocardiogram, ECG (mV)							
Temperature (Celsius)							
Stress Self-Report							

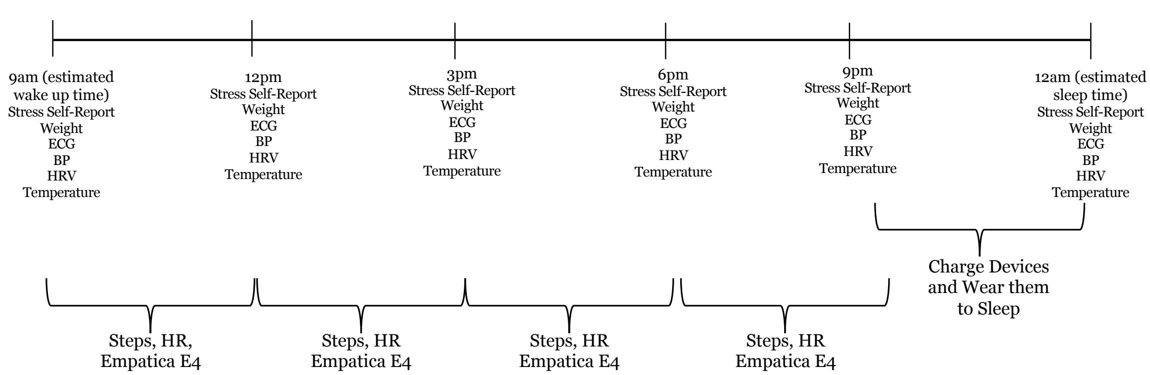


Figure 30: Study Protocol

The iPhone contained the prototype MHP. Participants were instructed on how to use the platform to complete stress self-reports (see User Manual in Appendix C for detailed instructions). The MHP uses the HealthKit Application Programming Interface (API) provided by Apple to extract Apple Health data [29].

Figure 31 shows the interface of the MHP, extracting Apple Health data automatically and allowing users to self-report their stress levels. The MHP is used as the data collection tool for the study: users open the app to answer the stress questionnaires (at which point all new Apple Health measurements are synced with the research database) and proceed to take the measurements, following the instructions in the User Manual in Appendix C.

Notably, the HRV data on the Apple Watch is collected throughout the day based on user behavior, but to trigger collection for our study, we used the Breathe app, an Apple Watch mindfulness application that asks users to breathe in and out for several minutes (more information in the User Manual) [95]. To avoid affecting stress levels, we asked users to do this as the last step in the data collection protocol.

In addition, while the Empatica E4 collects data continuously when active, we noticed that it constantly disconnected from the iPhone through its Bluetooth connection. Therefore, users were asked to constantly check if the Empatica was still active and, if not, to establish the connection again (see User Manual). However, as shall be described in further sections, this resulted in a lot of missing data from this device. Several participants encountered difficulties managing the study protocol with their daily life routines. In these cases, we asked participants to use the devices for additional days. This study was approved by the University Waterloo Research Ethics Board (REB [43612]). Data collection occurred between December 2021 and December 2022.

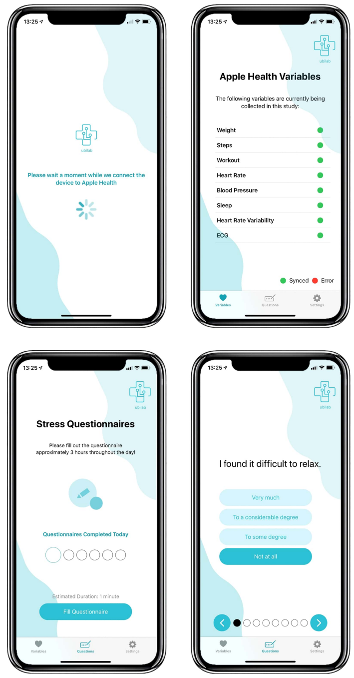


Figure 31: MHP Interface

8.2.3.2 Stress Self-Report

One of the challenges encountered when designing the study was a lack of validated stress questionnaires for EMA, as most stress questionnaires have a validated period of days or weeks. To mitigate this issue, we made use of the stress subscale of the Depression, Anxiety, and Stress Scale (DASS-21), comprised of seven questions related to stress. While the DASS-21 is usually applied over a week, there is promising evidence of using this questionnaire with EMA [128]. In addition, Wang et al. [129] used a single-item measure that, while lacking validation in the literature, was successfully applied for stress quantification and is moderately correlated with robust stress questionnaires [129].

In our study, we used both questionnaires comprising the 8 questions that are asked to participants. Questions 1-7 are related to the DASS-21, and question 8 comprises the single-item measure used by Wang et al [129]. The stress-related questions and respective questionnaires are shown in Table B1 in Appendix B.

Following DASS-21 guidelines, the score is multiplied by 2 and, if it is above 14, we consider users as stressed in that moment [162]. For the single-item measure, if the user answers "A little stressed" or above, we consider users having stress. For each questionnaire filled, if the DASS-21 or the single-item measure (or both) show stress, that data collection period is marked as stress. In other words, if in the moment of data collection the DASS-21 score is marked as "stress" and the single-item measure is marked as "no stress", or vice-versa, that data point will be labelled as "stress". The questions are displayed to the user in a random order each time the questionnaire is filled. Figure A25 in Appendix A shows an example of the dataset.

8.2.3.3 Data Collection and Pre-Processing

For all variables and features, we evaluated if the participant had any data points missing. In the case of a missing data point, if that was an isolated event, i.e., if less than two data points were consecutively missed, we used the average between the next and previous data points for the participant. In case more than two data points were missing, i.e., two or more consecutive data points were missing, we used the k-nearest neighbors algorithm to estimate the value based on the proximity of features that are not missing. This was done using SciKit Learn's KNN Imputer method with number of neighboring samples set to 5 [183]. In addition, features were included if they had at least 30% of data for the specific user. This number was used to balance the amount of data used in the analyses while empirically assessing the KNNImputter behaviour and model performance for missing data.

Next, we describe the processing of variables extracted from each device in Table 18. An exhaustive list and description of all features used in the study are shown in Table B2 in Appendix B.

8.2.3.4 Steps

For steps data, only Apple Watch information was used. Since data was collected through the HealthKit API differentiating between the Apple Watch and iPhone, it was not possible to integrate data from these two devices without avoiding duplication of information. From the Apple Watch steps data, we extracted the mean, maximum and minimum number of steps for the time interval between the start and end dates of the data point.

Unlike other data types, for the steps data we did not use averages or KNNImputer for missing data, as it was possible the user simply did not walk during the time period in question. Figure A26 in Appendix A shows example data points in the dataset for steps.

8.2.3.5 Heart Rate

For HR data, we also focused on Apple Watch as the device collects data throughout the day over infrequent periods. The BPM Connect device only calculates HR when the participant is using the equipment, and Withings Sleep only collects data during the night. We calculated the mean, maximum and minimum heart rate for the time interval, measured as beats per minute.

In addition, we noticed that during the data collection protocol, when the user activates the Breathe or ECG apps, the device typically shortens the HR data collection period to milliseconds. Therefore, we also extracted a Short-Term HR feature in which we only consider the millisecond data close to the time the user filled out the stress self-report, rather than the entire HR data from the 3-hour time interval between the start and end dates of the data point. For the Short-Term features, we also extracted the mean, maximum and minimum. Finally, we used the data collected from the Apple Watch and for any other devices (e.g., Withings Sleep) during the night, i.e., after the last data point collected on day *t-1* and before the first data point on day *t*. Figure A27 shows a snapshot of the dataset with the Apple Watch HR features.

We also considered Empatica HR data. More specifically, we used the *BVP.csv* file supplied by Empatica, which provides raw data, to extract several HR and HRV features. This file was processed using the Kubios HRV Premium 3.5.0 software [161]. More details on Kubios and HRV will be described in the next sub-section. For HR, Kubios provided the mean, maximum, minimum, and standard deviation of HR for the Empatica HR data during the time interval.

Finally, ECG data from the Apple Watch ECG app was also processed using Kubios. The HR output of ECG is similar to the Empatica, with the mean, maximum, minimum, and standard deviation of HR for the ECG data during the time interval

8.2.3.6 Heart Rate Variability/ECG

In AH, HRV is measured as the standard deviation of beat-to-beat measurements (SDNN). The Apple Watch measures SDNN using photoplethysmography (PPG), a technique in which a green LED light is used to detect the amount of blood flowing in the wrist [22], irregularly throughout the day. In addition to the passive collection of this metric with the Apple Watch, we also asked users to leave the Breathe app open for 5 minutes as the final step of the data collection process to trigger the HRV data collection, as previously mentioned. To process features, we calculated both the SDNN from the Breathe app and the SDNN collected throughout the day.

Despite not being able to control when the Apple Watch would collect the metric, we noticed that there was typically at least one SDNN data point collected per time interval. On the other hand, many users failed to activate the Breathe app during the data collection process, leading to missing data. Given that, in the real-life deployment of a system such as the MHP it is more feasible to depend on passively collected data, we decided to use the SDNN metric collected by the Apple Watch throughout the day rather than using the Breathe app as a trigger. This feature was named HRV-1 (see Table B2 in Appendix B).

ECG data, composed of timestamps and voltage measurements which generate a 30-second measurement on the Apple Watch ECG app (see User Manual for more details), was processed using Kubios into HRV data. In terms of program parameters, Kubios automatic beat correction algorithm was used, and the automatic noise detection was set to medium.

It should be noted that there is limited evidence on the use of ultra-short HRV measurements [146,184] such as the ones provided by the Apple Watch ECG app. To the best of our knowledge, this is one of the first works to use the Apple Watch ECG data in stress prediction. Following recommendations of the Task Force of The European Society of Cardiology and the North American Society of Pacing and Electrophysiology [153], we removed ECG several features, as follows:

- When it comes to time-domain measures, the RMSSD is highly correlated with the pNN50 and the NN50, and the RMSSD is preferred. Therefore, we removed pNN50 and NN50.
- TINN, HRV Tri Index, VLF and log measurements were removed as they seem to be more indicated for longer periods, and the ECG measurement is 30 seconds long.

Finally, Empatica E4 data was also processed into HRV data using the Kubios software, as shown in Table B2. To capture physiological states during the time of data collection, we used 10-minute intervals close to the data collection point. Ideally, the intervals started 5 minutes before the time of the stress self-report and continued for 5 minutes after, but due to the amount of noise in the data as well as missing data due to connectivity issues, this was not always possible. Therefore, data was processed as close to the time of data collection as possible. In case there were not 10 full minutes of quality data close to data collection time, we used a cut-off of 5 minutes, i.e., at least 5 minutes of data were required to be included in the study and processed in Kubios. We removed the same features as above with the exception of very low-frequency and log components which may capture relevant information for longer measurements.

Figure A28 shows the Apple Watch HRV-1 feature in the dataset; Figure A29 shows ECG HRV features; and Figure A30 shows Empatica HRV features.

8.2.3.7 Weight, Blood Pressure, and Temperature

Data on weight, blood pressure and temperature were included in this study. For blood pressure, in addition to systolic and diastolic pressure, the mean arterial pressure was used as a feature. While true MAP can only be obtained with intrusive devices, it is possible to estimate it with the following formula: (sys + 3* dys)/3 [181,185].

Due to the size and weight of the Withings Body+ smart scale, participants found it difficult to bring this device with them during their daily routine. Therefore, they were instructed to only take weight measurements when they were at home and had access to the scale (which also reflects how such measurements would be taken in the real-world). For this reason, we included weight data in this study despite large gaps, using the KNNImputter to fill these. Figure A31 shows snapshots of weight, blood pressure and temperature features.

8.2.3.8 Sleep

As mentioned in the HR subsection, we calculated mean, maximum and minimum HR during the night (between the last data point collected from the previous day and the first from the current day).

In addition, we calculated sleep features from the Apple Watch and the Withings Sleep device. From both devices, we calculated the following sleep features: Total Time Asleep, Number of Wake-Ups, Time Awake During Sleep, Total Time in Bed, and Percentage of Time Asleep While in Bed. Withings Sleep also provided additional information on the time the participant spent in Light, Deep and REM stages, respectively (of note, the Apple Watch recently introduced an update on sleep monitoring that also collects data on sleep stages, but that was not available at the time of the study) [186].

To calculate Apple Watch sleep durations, we made use of both the Apple Watch and iPhone. The Apple Watch calculates times spent asleep. To calculate time in bed, Apple systems make use of the iPhone's sleep calendar feature. Users were asked to include an estimate of their sleep schedules on the iPhone (see User Manual). Those values are updated based on when participants are using the phone and were used to estimate when the user went to sleep.

Because the relationship between sleep and mental health is potentially bidirectional [53,187], we also created features offsetting the day for 2 additional days and 2 days before. For example, if a feature was collected at time *t,* the *t-2* equivalent would place this value two days before, *t-1* at one day before, *t+1* at 1 day after, and *t+2* at 2 days after. In order to not greatly increase our feature set, initial RF models were used to calculate the feature importance (based on mean decrease in impurity) of different offset days, with 2 days before and 2 days after being extremely prevalent among the most important features. Therefore, the majority of sleep features included were from *t-2, t+2,* and *t.* The features from other day offsets that were included, based on the tests, were *t+1* Apple Watch Mean HR, *t+1* Apple Watch Max HR, *t+1* Apple Watch Min HR, *t+1 Withings Total Time Asleep, t+1* Apple Watch Number of Wake-Ups*, t+1* Apple Watch Time Awake During Sleep *, t-1* Apple Watch Mean HR, *t-1* Apple Watch Max HR, *t-1* Apple Watch Min HR, *t-1 Withings Total Time Asleep, t-1 Apple Watch Total Time In Bed.*

Finally, because sleep data is collected at a different frequency than each of the EMA data collection – EMA is collected approximately 6 times a day while sleep data is collected once per day –, sleep data were included for the entire day (e.g., every measurement of day *t* will have the same sleep features) to maximize the amount of granular stress information collected. Figure A32 shows sleep features for *t.*

8.2.3.9 Feature Selection and Normalization

In addition to the removal of features mentioned in the previous section, before every experiment, highly correlated features (with a Pearson correlation coefficient higher than 0.95) were removed. While RF is not affected by the difference in units, we normalized the data for input in the SVM models using *SciKit Learn's* Standard Scaler method [188].

8.2.3.10 Analyses/Experiments

A number of experiments with different subsets of the data were conducted to allow us to better understand the predictive power of each device/manufactures (Empatica, Apple, Withings) for stress prediction. Further, we conducted the analyses with and without the sleep data due to its different data collection periods. We conducted the following experiments excluding sleep:

- Dataset with all features, D (n = 22)
- Dataset with only ECG features, DECG (n = 42)
- Dataset with only Apple features, DA (n = 42)
- Dataset with only Withings Features, DW (n = 44)
- Dataset with Apple and Withings Features, DAW (n = 41)
- Dataset with Only Empatica Features, DEmpatica(n = 27).

Adding sleep features to the datasets, we conducted the following experiments:

- Sleep Dataset with only Apple features, SDA (n = 34)
- Sleep Dataset with only Withings Features, SDW (n = 34)
- Sleep Dataset with Withings and Apple Features, SDAW (n = 27)
- Sleep Dataset with Withings and Apple Only Sleep Features, SDS (n = 27)

DECG and DA contain the same participants, with different features. In case a participant possessed a feature with less than 30% of the data, this participant was removed from datasets using the feature. For example, if a user possesses less than 30% of the Empatica features, they were not included in the D or DEmpatica sets, and similarly for other features in each dataset. For the datasets including sleep features, we did not include Empatica data as this would result in very small datasets. Table B3 shows the participant characteristics for each dataset, and each separate sampling is further discussed in the Results section.

As mentioned in the Related Work section, there are many different ways to train, test and validate the models in the literature. For these analyses, given that in real-world deployment public health agencies would collect a large amount of data from populations, we elected to train generalized models, meaning using data from all participants. Randomly, 80% of the dataset is used for training/validation and hyper-parameter tuning with 10-fold cross-validation, while 20% is used for testing.

In addition, since data was collected in real-life environments, many users had a predominance of one class over another (e.g., with a lot of data points classified as *no stress* compared to *stress*, or vice-versa). For this reason, we conducted the analyses with imbalanced classes as well as with balanced classes using the SMOTE (Synthetic Minority Over-sampling Technique) method on *SciKit Learn,* which upsamples the minority class [147,189]. The technique is applied only to the training sets in the examples described above before cross-validation to make sure the model is tested on real data [147]. In other words, only the train sets are balanced.

Finally, due to the relationship between stress measures and factors such as sex, age, income, work, and health, we trained the following models [21]:

(1) Total: comprised of data from all participants in each subset.
(2) Age: models in the age range of 18-24, 25-34 and 35-44 were trained. We decided not to train models in the 45-64 range due to the scarcity of participants in this interval. By the same token, we did not train a model for participants aged above 65 as only one participant was in that category.
(3) Sex: we trained models for male and female participants. We did not train a model for the participant that self-identified as gender fluid as only one participant was in that category.
(4) Income: we trained models for participants belonging to low socioeconomic status (SES), comprising participants that earn less than CAD 30,000, and participants belonging to middle and high SES. The CAD 30,000 cut-off point was based on an approximation of the Canadian tax cut-off for low-income populations [175].
(5) Profession: we trained models for workers (full-time, part-time, and participants that are self-employed or classified as other) and students. We did not train a model for the retired participant as only one participant was in that category.
(6) Healthy: we trained a model removing participants that reported chronic diseases, illnesses, frequent alcohol or drug use, or prescription drug use.

For each of these divisions, we trained the model with binary classification (*stress* vs *no stress*), reporting accuracy, f1-weighted and f1-macro score in tables B4 and B5 from Appendix B. Finally, while the SVM model performs many transformations to fit the data, making it harder to obtain information on the importance of features, we calculated feature importance for the RF model using the mean decrease in impurity [21]. A 100% purity in a node means the decision tree's node contains only one class, and by assessing the change in impurity between parent and child nodes we can calculate the best split in the tree and use it as a proxy for feature importance.

Figure 32 and Figure 33 show the process of obtaining the different datasets and training the models.

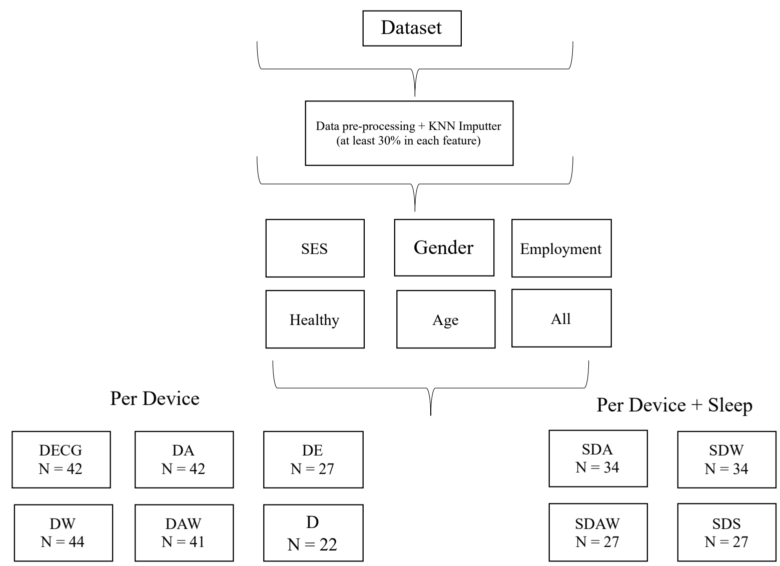


Figure 32: Division into Datasets Per Device and Per Device + Sleep

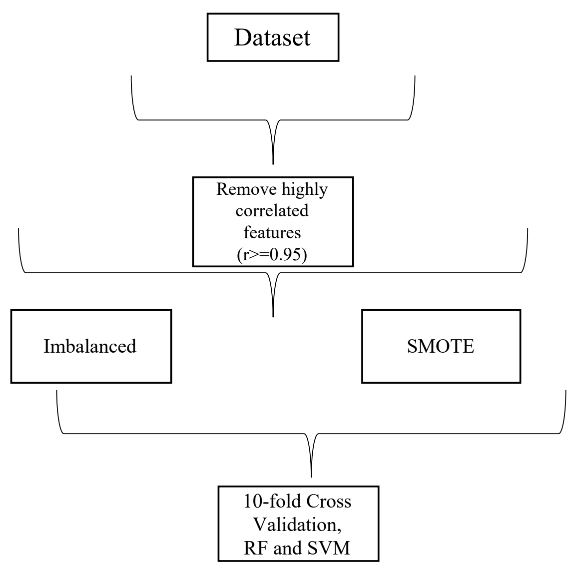


Figure 33: Training Generalized Models

8.2.4 Results

8.2.4.1 Population Data Characteristics

Table 2 details the full sample (n=45) used in the study. Most participants are aged 44 years or younger (87%), female (67%), with low (44%) or medium (40%) income, and workers (62%). In terms of ethnicity, a majority identified as white (33%), South Asian (24%) or Latin American (22%). Most participants (80%) did not have chronic diseases or illnesses, use prescription drugs or frequently consumed alcohol/smoke. The average of days a participant had in the study was 17.1 (± 2.5), and participants had an average of 78.91 (± 11.0) data points. Data quality did not visibly differ between participants based on number of days required for data collection.

Table 19: Participant Characteristics

Participants (N = 45)	*Frequency*	*Percentage*
Age		
18-24	13	29
25-34	14	31
35-44	12	27
45-64	5	11
Above 65	1	2
Sex/Gender		
Male	14	31
Female	30	67
Gender Fluid	1	2
SES		
Low (0-$30,000)	20	44
Medium ($30,000– $100,000)	18	40
High (Above $100,000)	4	9
Do not wish to disclose	3	7

Profession		
Full-time	21	47
Part-time	5	11
Student	16	36
Self-employed/Other	2	4
Retired	1	2
Ethnicity		
Black and Southeast Asian	1	2
Black or African American	3	7
Chinese	4	9
Indian	1	2
Latin American	10	22
South Asian	11	24
White	15	33
Health Status		
Healthy	36	80
Chronic Disease or Illness, Prescription Drug Use, Smoking or Alcohol	9	20

For the 45 participants, 43% of total answers were classified as stress (1539), while the remaining 57% (2012) were labelled as no stress. This proportion is maintained, approximately, for each of the subsamples used in the analyses (Table B3 in Appendix B).

8.2.4.1 Models

This section discusses the results of each of our models. The focus of this section is on the f1-macro metric, as accuracy may not reflect imbalance in classes (if one class is predicted well but another is not, the accuracy may still be high), so the f1-score which calculates the harmonic mean between precision and recall is preferred. Further, the macro average treats both classes as being of equal importance.

Table 20 shows the f1-score for the experiments. More detail on other metrics, including specifics for each class, accuracy, f1-weighted, precision and recall can be found in Table B4 (without sleep) and B5 (with sleep). Feature importance in each dataset is presented in Tables B10 to B19 in Appendix B for each dataset.

8.2.4.1.1 Generalized

8.2.4.1.1.1 Without Sleep Data

In this section, we discuss the results of the models for each of the datasets. As can be seen by Table 20, which shows the results for each dataset, most contain results above 60% when considering all participants in the specific samples. In particular, D, DAW, and DW results for all participants are above 65% for RF, indicating that Withings features seem to be good predictors and that RF generally works better than SVMs, which generally has lower results. Datasets containing only Apple features, and in particular only ECG, perform worse compared to others, although several results are still above 60%.

Looking only at the healthy participants when compared to all, in most cases the f1-macro score varies slightly (e.g., dropping from 63% to 60% for RF in DA or improving from 66% to 68% for RF in D). In general, stratifying participants by gender and employment improves results, although that is also not always the case – especially with female participants. When stratifying according to income, the divisions with low-income participants typically perform worse, while divisions containing participants with medium to high-income show improvement, often with an f1-macro average above 70%. Finally, stratifying by age seems to worsen results in most cases.

SMOTE results show mild improvement over results without it in most cases, especially for RF. However, many examples indicate worsening results – especially for SVM – or do not demonstrate any improvement.

In terms of feature importance for the RF model, tables B10 to B15 in Appendix B show the top 10 most important features and importance value – calculated as the mean decrease in impurity – for each dataset and stratification. Looking into each of these, we investigated the top 10 features that repeat across strata. For example, for Gender, the only top 10 feature in the D dataset that appears in both Male and Female stratification is Weight, as can be seen in Figure A7 in Appendix A with a frequency of 2 (meaning the feature appeared in the 2 gender-related models). We did the same for Income (Low/Medium and High), Employment (Workers and Students), Age (18-24, 25-34, and 35-44) and finally, between datasets with all participants and datasets with only healthy participants. Figures A7-A15 show this frequency for the stratifications in each dataset. DW was not included in these analyses as it only has 5 features.

When considering the importance of all features in D, a mix of features from different sources can be seen, including Apple data from the ECG (ECG_DC, ECG_AR_AbsolutePower_HF, ECG_AR_AbsolutePower_LF, ECG_Stress Index), Apple data from HRV (HRV-1) Withings blood pressure data (MAP, dia), and Empatica data (Empatica_AR_RelativePower_LF, Empatica_AR_LFHF). This mix of modalities is maintained throughout other stratifications in D, although with different features for different strata (such as sys and Weight in Male stratification). Interestingly, ECG_DC is present in several analyses and repeated in stratifications such as All/Healthy, Income, Employment, and Age. This feature is also prominent in other datasets that it is present.

Withings features are very prominent in most datasets that mix features from different devices, ranking high among the most important features (e.g., MAP or weight are ranked among the top 2 most important features in several stratifications in D and DAW), and Withings features repeat among stratifications in these datasets. Also, of note, the User feature appears frequently in DECG, DA, DAW, and DEmpatica.

8.2.4.1.1.2 With Sleep Data

In general, adding sleep data to the dataset improves results, especially with the RF algorithm, with most results above 65% and an f1-macro score of 70% being commonplace. There were few cases in which the metrics worsened – the latter is mainly seen in SDAW on stratifications such as gender and income. In particular, male participants from SDAW showed a great decrease from f1-macro scores around 70% to results in the low fifties. This might be due to a decrease in the number of male participants from DAW to SDAW (29% to 22%, respectively), which may not have given the model enough data to be trained accurately. For several sleep datasets, upsampling classes using SMOTE resulted in slight improvements, although a worsening result can be seen in some cases. Since sleep data is repeated over a day, the SMOTE method possibly did not accurately synthesize minority class samples.

Finally, while the dataset using only sleep features, SDS, typically showed results on par with the other models, it also produced the worst results among the sleep datasets, notably on male participants (f1-macro score below 50%) and on participants aged 35-44 (53% f1-macro score with SVM-SMOTE), suggesting these features are more robust when used in conjunction with others.

When looking at feature importance, sleep features typically dominate the datasets in Tables B16 to B19. Interestingly, in datasets that mix Apple and Withings features (SDAW, SDS), there is still a predominance of Apple Watch sleep features, although several Withings non-related features (e.g., MAP, Weight, sys, temp) are prevalent among the top 10 most important features in each stratification. When looking at SDW, containing only Withings features, there is also a prevalence of non-sleep related features, and these features are generally repeated among stratifications.

The Apple Watch Consolidated Time Awake During Sleep feature, and its offsets for *t+1* and *t+2*, also repeat among stratifications in SDA, SDS and SDAW.

Table 20: Macro-F1 Score Results for Generalized Models

Dataset with all features (D)				
	RF	**SVM**	**RF-SMOTE**	**SVM-SMOTE**
All	0.66	0.65	0.68	0.65
Gender - Male	0.68	0.7	0.63	0.64
Gender - Female	0.68	0.61	0.66	0.65
Employment - Student	0.61	0.62	0.6	0.62
Employment - Worker	0.62	0.61	0.63	0.58
Income - Low	0.62	0.63	0.63	0.59
Income - Medium High	0.73	0.7	0.71	0.7
Age – 18-24	0.51	0.39	0.52	0.47
Age – 25-34	0.56	0.62	0.6	0.44
Age – 35-44	0.5	0.55	0.54	0.62
Healthy	0.68	0.68	0.67	0.62

Dataset with only ECG features (DECG)				
All	0.62	0.53	0.61	0.54
Gender - Male	0.61	0.6	0.62	0.59
Gender - Female	0.62	0.59	0.62	0.56
Employment - Student	0.64	0.6	0.62	0.6
Employment - Worker	0.6	0.56	0.59	0.47
Income - Low	0.53	0.55	0.53	0.46
Income - Medium High	0.67	0.61	0.67	0.58
Age – 18-24	0.61	0.58	0.57	0.51
Age – 25-34	0.59	0.59	0.59	0.51
Age – 35-44	0.58	0.56	0.61	0.49
Healthy	0.58	0.55	0.62	0.57
Dataset with only Apple Features (DA)				
All	0.63	0.58	0.63	0.53
Gender - Male	0.63	0.61	0.64	0.58
Gender - Female	0.59	0.53	0.59	0.58
Employment - Student	0.65	0.54	0.63	0.58
Employment - Worker	0.63	0.57	0.61	0.54
Income - Low	0.54	0.52	0.58	0.55
Income - Medium High	0.67	0.6	0.66	0.59
Age – 18-24	0.54	0.56	0.53	0.55
Age – 25-34	0.6	0.57	0.63	0.56
Age – 35-44	0.59	0.61	0.65	0.52
Healthy	0.6	0.56	0.62	0.57
Dataset with Apple and Withings Features (DAW)				
All	0.67	0.6	0.69	0.61
Gender - Male	0.73	0.7	0.71	0.7
Gender - Female	0.67	0.63	0.66	0.58
Employment - Student	0.65	0.64	0.62	0.61
Employment - Worker	0.66	0.63	0.67	0.64
Income - Low	0.53	0.58	0.59	0.56

Income - Medium High	0.72	0.65	0.71	0.67
Age – 18-24	0.6	0.59	0.63	0.6
Age – 25-34	0.69	0.63	0.66	0.58
Age – 35-44	0.61	0.69	0.67	0.57
Healthy	0.62	0.63	0.6	0.6
Dataset with only Withings Features (DW)				
All	0.69	0.62	0.66	0.64
Gender - Male	0.65	0.68	0.63	0.63
Gender - Female	0.63	0.57	0.64	0.62
Employment - Student	0.65	0.66	0.64	0.6
Employment - Worker	0.66	0.64	0.66	0.63
Income - Low	0.58	0.62	0.55	0.57
Income - Medium High	0.73	0.65	0.69	0.69
Age – 18-24	0.51	0.55	0.53	0.55
Age – 25-34	0.6	0.57	0.63	0.56
Age – 35-44	0.62	0.64	0.65	0.56
Healthy	0.62	0.58	0.61	0.58
Dataset with only Empatica Features				
All	0.64	0.6	0.65	0.61
Gender - Male	0.64	0.7	0.67	0.59
Gender - Female	0.68	0.66	0.67	0.66
Employment - Student	0.67	0.66	0.68	0.63
Employment - Worker	0.65	0.67	0.65	0.65
Income - Low	0.56	0.57	0.57	0.56
Income - Medium High	0.65	0.66	0.67	0.66
Age – 18-24	0.48	0.53	0.58	0.47
Age – 25-34	0.68	0.67	0.67	0.62
Age – 35-44	0.54	0.68	0.62	0.61
Healthy	0.61	0.61	0.63	0.6
Sleep Dataset with only Apple Features (SDA)				
All	0.7	0.65	0.73	0.66

Gender - Male	0.63	0.63	0.65	0.64
Gender - Female	0.71	0.66	0.71	0.64
Employment - Student	0.69	0.67	0.69	0.67
Employment - Worker	0.66	0.65	0.68	0.61
Income - Low	0.7	0.61	0.71	0.66
Income - Medium High	0.75	0.72	0.72	0.67
Age – 18-24	0.7	0.65	0.71	0.64
Age – 25-34	0.68	0.61	0.69	0.64
Age – 35-44	0.68	0.67	0.68	0.61
Healthy	0.72	0.67	0.71	0.64
Sleep Dataset with Apple and Withings Features (SDAW)				
All	0.73	0.69	0.72	0.71
Gender - Male	0.53	0.53	0.53	0.45
Gender - Female	0.7	0.67	0.71	0.66
Employment - Student	0.74	0.7	0.74	0.74
Employment - Worker	0.67	0.65	0.7	0.63
Income - Low	0.74	0.7	0.74	0.71
Income - Medium High	0.67	0.69	0.67	0.68
Age – 18-24	0.75	0.74	0.73	0.75
Age – 25-34	0.71	0.65	0.69	0.68
Age – 35-44	0.63	0.68	0.65	0.53
Healthy	0.69	0.67	0.7	0.7
Sleep Dataset with Withings Features (SDW)				
All	0.69	0.58	0.67	0.68
Gender - Male	0.79	0.81	0.8	0.8
Gender - Female	0.68	0.66	0.67	0.65
Employment - Student	0.71	0.73	0.73	0.73
Employment - Worker	0.67	0.65	0.68	0.64
Income - Low	0.67	0.68	0.69	0.68
Income - Medium High	0.71	0.73	0.71	0.73
Age – 18-24	0.67	0.72	0.65	0.67

Age – 25-34	0.73	0.75	0.7	0.74
Age – 35-44	0.72	0.73	0.73	0.68
Healthy	0.72	0.7	0.7	0.72
Sleep Dataset with Withings and Apple Only Sleep Features (SDS)				
All	0.7	0.69	0.7	0.71
Gender - Male	0.45	0.45	0.47	0.49
Gender - Female	0.71	0.71	0.71	0.71
Employment - Student	0.73	0.74	0.71	0.74
Employment - Worker	0.66	0.65	0.63	0.62
Income - Low	0.69	0.71	0.73	0.73
Income - Medium High	0.66	0.66	0.63	0.65
Age – 18-24	0.65	0.6	0.65	0.61
Age – 25-34	0.67	0.67	0.68	0.69
Age – 35-44	0.6	0.6	0.64	0.53
Healthy	0.73	0.73	0.73	0.74

8.2.5 Discussion

8.2.5.1 Stress Models

While the use of different metrics for model evaluation in literature, as well as the different strategies for data collection and model training, difficult comparisons between works, the results of the Generalized models achieved good to great accuracy between 60% and 70% (Tables B4 and B5), which is in line with the state-of-the-art – particularly for studies that use real-life data (see Table B6 in Appendix B with studies labelled DDSR, meaning they were trained with daily life self-report labels). The promising f1-macro score values shown in Table 20 indicate the models are able to predict the two classes. This indicates that the MHP provided accurate and representative data and that a similar backend system might potentially be deployed by public health agencies for data collection and monitoring of a condition in a population, such as stress. This is especially promising considering the models were built on data collected from personal, consumer-level, off-the-shelf devices rather than using data from research-grade equipment.

In terms of which features to collect, Apple Watch sleep features are very prominent among the sleep datasets, usually ranking higher than other features. Other features such as Temperature, Weight, and Blood Pressure also appear as good predictors when looking at feature importance (all Withings related). Datasets that integrated sleep features achieved the best results, with an f1-macro score and accuracy typically above 70%. In addition, feature importance results suggest that offsets of the data should be considered. Specifically, *t + 2* and *t - 2* demonstrated good results, although offsetting for a day also provided important features in specific cases. Since this study, Apple has updated its sleep data collection to include additional features such as sleep stages [190], which could improve performance even more – especially as time Spent in REM was one of the few Withings sleep features that appeared repeatedly among stratifications.

In terms of non-sleep data, datasets that contained Withings data (D, DW, DAW) typically performed better than others. Coupled with the prevalence of Withings non-sleep features among the important features in the sleep datasets, this suggests that using these devices to collect temperature, weight and blood pressure would be an interesting avenue of research to follow.

Encouragingly, Empatica E4 data did not seem to greatly affect the models, as they usually performed well without this data, specially sleep-related models. Given this, public health agencies could potentially leverage data from personal, consumer-level devices, rather than having to resort to medical-grade wearables such as the Empatica E4.

When looking into Apple Watch ECG data alone, the metrics worsen compared to other datasets, although they are generally above 60%. This is an improvement over previous work by the authors using the ECG dataset without any missing data imputation and with only a subset of the data [21]. Much like in the previous work, ECG_DC and, to a lesser extent, ECG_AC, featured prominently in ECG-related models. In the present case, ECG_DC also features in datasets containing other data modalities in addition to ECG. Empatica_DC was also a prominent feature when looking at the Empatica data and among some stratifications. Therefore, the heart's deceleration (DC) and even the heart's acceleration (AC) seem to be valuable HRV metrics for the models, being constantly present among the top 10 most important features. AC and DC are relatively new indicators in HRV studies, and it would be interesting to conduct further research into stress using AC and DC to establish if they can be robustly used to differentiate stress states.

When considering the f1-scores for each class in Tables B4 and B5, generally, the "no stress" class seems to outperform the "stress" class, especially for the non-sleep datasets (Table B4). This suggests that the models typically have higher specificity than sensitivity.

8.2.5.2 Performances for Different Samples in Datasets

Given the different samples use for each dataset, it is worthwhile to look at how the datasets differ regarding these samples and how this may affect results (Table B3 in Appendix B). D contains fewer participants in the 18-24, 25-34 and 35-44 ranges, which may account for this dataset generally having poor results in these stratifications when compared to others. However, D has good results on the male/female and low/medium and high-income stratifications despite possessing half the number of men, women and participants in low to middle income (with high-income participants removed entirely). Therefore, it is likely that the division of fewer participants into 3 distinct age categories did not provide enough information to train the model, and future research collapsing the age category into fewer divisions (e.g., young vs. old) as opposed to the intervals presented here could lead to better results.

D also has lower results than other datasets for profession, likely due to the removal of several workers from the dataset. Interestingly, while D has a 7% increase in the proportion of unhealthy participants, it also has the highest non-sleep related metric for the healthy stratification. With a decrease from 36 healthy participants in the full sample to 16 in D, a higher homogenization of participants likely led to better metrics. In particular, of the healthy participants in D, 63% were female, 50% were students, and 56% were low income. This suggests that creating models for more than one stratification (e.g., healthy women) may lead to better results. On the other hand, it will reduce the dataset even further, which may affect the model's effectiveness. Future work, with more purposeful sampling could lead to further insights into this avenue of research.

DEmpatica is another dataset that had a large number of participants excluded, and similar observations to D can be made: results in the age stratifications were poor, with other stratifications performing better. Stratifying by profession fared better in this dataset, with a larger number of workers and students. The number of healthy participants also increased, leading to lower metrics – likely due to more heterogeneous participants. When looking at demographics, DEmpatica has a higher proportion of healthy females (65%), a slight decrease in healthy low-income participants (55%), and a higher decrease in students (40%), suggesting that future sampling of healthy participants by profession could lead to additional insights.

DECG/DA contain the same sample of participants, with very similar proportions to the entire sample. These datasets did not generally perform as well as other models containing Withings features, likely due to lower predictive power in non-sleep related Apple features. Indeed, DAW – with 4 fewer participants than the total sample of 45 while still maintaining similar proportions – performed better, reflecting the higher feature importance of Withings features. DW has very similar proportions to DAW with only 1 participant removed from the total sample and, when comparing this dataset to DAW, we can see that using only Withings features led to good results, sometimes better than using Apple and Withings features combined.

In general, regardless of how the samples varied – and indeed the sleep datasets had more users removed due to missing data –, sleep datasets performed better than datasets without sleep features, highlighting their importance. SDA had major changes regarding income (increase from 44% to 50% in low-income and 9% to 0% in high-income participants). Interestingly, these reductions led to slightly better results in general when compared to other datasets, likely due to increased homogenization of participants. Most low-income participants in this dataset were female (71%) students (77%) aged 18-24 (53%) while most medium-income participants were female (64%) workers (94%), aged 25-34 (50%).

Despite having fewer participants, SDAW maintained good results, indicating the importance of both sleep-related features and non-sleep related Withings features. This is reflected also in SDS, which uses the same sample but with worsening results due to the removal of non-sleep Withings features. Both SDAW and SDS had a 9% decrease in male participants; given the already low prevalence of men, this may explain the poor results for these stratifications. Finally, as we saw in the previous subsection, Withings features are particularly important for SDW.

8.2.5.3 Limitations, Lessons Learned and Implications for Public Health

In this section, we describe empirical lessons derived from deploying the MHP prototype in real-life and developing the ML models and discuss implications for the potential use of a similar system in public health. In the process, we also discuss study limitations as well as mitigation strategies where applicable.

First, a limitation of this study is that, due to convenience sampling, most participants were female, typically white and young. Looking at sex/gender in particular, most datasets contain approximately 30% men and 70% women, and their model results metrics are similar. As mentioned, on SDW and SDS, the male stratification performs poorly; in these datasets, the proportion of male participants is closer to 20%, which could indicate there is a lack of data from these participants to accurately train the models. The same may have happened for other stratifications, such as income and age, which performed poorly in several cases. As mentioned, future studies with more purposeful sampling could lead to better stratifications and further insights into how models will perform for different traits.

The RF model in general performed better than the SVM model. The SMOTE method was used in this work with mixed results to handle class imbalances, and careful consideration must be taken to generate synthetic data for public health decision-making. However, if a system similar to the MHP is deployed in the real-world, it would potentially capture much larger datasets from each user – for example, by collecting data points for longer periods –, generating datasets with more examples in each class that would balance the collected dataset, this mitigating this issue. To mitigate this, it would be interesting to investigate the use of reinforcement learning, in which the model learns from mistakes through a reward and punishment approach [191]. For example, the mobile application could display the prediction result to the user and ask if this prediction is correct. This feedback will provide rewards or punishment to the model, improving future predictions. The collection of context information (e.g., what the user is doing at the time of data collection, or which stressors affected them) could potentially improve results in all approaches[168].

In terms of feature importance, Apple Watch sleep features as well as Withings non-sleep related features (such as temperature, weight and blood pressure) were shown to be important for models and should be prioritized. On that note, a limitation of the data collection method was the amount of missing data. 10 participants had over 70% missing Apple Watch sleep data, and an additional 10 had missing Withings sleep data. While the missing Apple Watch sleep data was likely due to participants not wearing the device while sleeping, the Withings Sleep device is placed below the mattress, and so should be available at all times when connected to a plug. Discussing the device with participants during the video call and looking at the data, it seems this error is more likely due to limitations with the device.

Indeed, several challenges occurred with the Withings Sleep. First, the device needs to connect to a Wi-Fi network. One participant could not use the device as their network was part of a university, and a certificate was needed to connect. Since the certificate could not be downloaded onto Withings Sleep, the device did not work for this individual. In addition, the integration between Withings Sleep and Apple Health did not work consistently. The MHP collects data from the Apple Health app, and the Withings Sleep device is synced to the Withings proprietary Health Mate app (see User Manual for more details on this app). Health Mate, then, integrates with Apple Health after the data-sharing option is selected in the app. Despite this, in many situations Health Mate did not share sleep data with Apple Health, requiring the data sharing option on Health Mate to be turned on and off by the researcher or participants. There were also many situations where researchers did not observe sleep data being synced from the Withings Sleep device and, after asking participants to reset the iPhone, data collection proceeded as normal. One Withings Sleep device also broke during the study. Finally, although the Withings Sleep device was reset to factory mode before being provided to a participant, there were many cases during the video call where participants were required to reset it to this mode again. Given these issues, and that Withings Sleep data was not shown to be among the most important predictive features, Apple Watch sleep data can be prioritized in further deployments.

The Empatica E4 device also demonstrated issues during data collection. To collect real-life data, the device had to be used in Bluetooth mode, which required a connection with the E4 Realtime app on the iPhone (see User Manual in Appendix C). However, if the device was out of range, no alert was given when the device disconnected and the data collection stream interrupted. To mitigate this issue, participants were asked to monitor the E4 Realtime app and reconnect the device as needed. In addition, we also experienced technical issues with the Empatica devices, as during the 1-year data collection period, 3 out of 4 available devices broke and had to be fixed before collection could proceed. Movement also introduced a lot of noise in the data, making a lot of it unusable. For this reason, half of participants (28 out of 45) did not have at least 30% of the necessary Empatica data to use in the analyses. As discussed, given that Empatica features were not among the most important features, this device does not need to be included in future work to complement data from personal devices.

The prototype version of the MHP did not have any backup features in case data sharing between the device and the research database did not work (for example, due to low connectivity). To mitigate this issue, manually exported data from Apple Health was compared to the MHP data to make sure missing information due to issues in data sharing is considered in the analysis. Future versions of the MHP should contain a failsafe in case the connection does not work. In particular for the current iOS system, the Information Property List Files (*info.plist*) in the device might be an interesting solution: *Info.plist* is a structured text file, available for edits in Apple's iOS development software XCode [41], that contains information describing the app's configuration, and data can be stored in it using a dictionary format [192]. Storing information that could not be shared with the database on this file, retrieving it and sending it again would be a potential backup solution that would not require any online storage.

The prototype MHP contained data types that were hardcoded into the source code. In case researchers need to collect additional data types, the code for requesting additional authorizations in Apple Health and for additional HealthKit queries needs to be developed. One interesting update to future versions would be to allow researchers to customize which data they want to collect. By hardcoding queries for Apple Health data types and allowing users to activate them through the app's interface, the MHP would be flexible in enabling public health scientists to collect different types of data depending on their need. On the same token, the devices used in the study were hardcoded into the app prior to data collection, but a real-world deployment could enrol new devices belonging to each user by obtaining device information from the data sources in Apple Health obtained through HealthKit.

Of note, the MHP queries also enabled data collection in the background, i.e., while the app is not terminated but also not being currently used. However, if the app is not regularly used, the background queries will not be triggered constantly and may be terminated by the iPhone's operating system. Our solution to this was creating new queries every time the app is opened and terminating the app every time it is placed in the background to make sure the queries are triggered. Ideally, users would constantly utilize the system, triggering background data collection. While out of the scope of this work, to encourage people to use the MHP, there are many possible strategies, such as gamification [193] or developing an interface that allows users to monitor and manage their health in addition to collecting the data [194]. Finally, public health agencies can use such a system not only for monitoring but also for intervention. For example, if a user's stress levels are high, the MHP could trigger a meditation app (potentially the Breathe app on the Apple Watch for Apple systems). This feedback and intervention loop can reach a lot of people in real-time and fulfill the mission of public health – improving the quality of life of populations. While a real-time loop would not allow for the use of features collected "in the future", i.e., sleep features from *t+2*, they could still make use of the rest of the features that were shown to be good predictors. In addition, since the public health issue in question is chronic stress, real-time intervention may not necessarily be required – rather, long-term, constant stress patterns can be evaluated, and feedback provided based on them.

In terms of study design, the data collection protocols required several devices and apps to be used during participants' daily life, which can be quite intrusive and demanding. In particular, many had difficulties leaving the Breathe app open for 5 minutes to collect HRV data without making any movements, which would stop data collection. For this reason, we decided not to use the 5-minute HRV data, instead using HRV collected randomly throughout the day (HRV-1). Interestingly, HRV-1 is an effective feature, appearing among the top 10 most important features in many stratifications in datasets containing Apple data.

In the case of real-world deployment, however, it cannot be expected that users will possess all devices described in the study or that they will collect data constantly. To mitigate this issue, as we discussed, the Apple Watch was one of the most important device used in this study, particularly its sleep data, which produced better models. Weight, collected through the Withings Body+, was also shown to be an important feature – even though weight data was only collected when participants were at home – and it is very common for individuals to have scales as a regular household item. Even if they are not wireless devices, manual input of weight into the phone can still be done. The same can be said for thermometers, and to a lesser extent blood pressure cuffs (all collected with Withings devices). Therefore, it should not be hard to obtain data for the most important features. Further, to reduce the burden on users, it might be more useful to ask for fewer data points over a longer period (e.g., once in the morning and once at night for several months) rather than for several data collected throughout the day. As discussed, this could help with obtaining datasets with more examples of each prediction class. In this way, public health researchers might be able to obtain large datasets to build models and study individual health while placing less burdens on users.

While Apple Health was the chosen tool in this study to aggregate sensor data, there are other systems that could potentially be used. All Withings devices mentioned in this study – and, as we have seen, that had features identified as important predictors – are also compatible with

Android devices, for example. On the same token, the findings should translate to other equipment that collect data similarly to the devices used here, although that is not a gurantee and future work could focus on using other mobile health repositories for increased generalizability.

Finally, while the results are promising, since a given participant could have had data in both training and test set and that the *KNNImputter* method was applied to the entire dataset before split into train and test, future work should focus on additional strategies for model development and evaluation such as leave-one-person-out cross validation to increase the generalizability of results.

8.2.6 Conclusions

In this study, we developed an MHP that collects data from Apple Health – which in turn integrates data from mobile and wearable devices – that could potentially be used in public health data collection efforts. To test its efficacy, we predicted the stress states of individuals using ML models of RF and SVM based on Apple Health data gathered throughout participants' daily routines and collected using a mobile health platform.

Additional future work should implement improvements on the MHP such as backup in case of an error in data sharing, customization of data collection and encouragement for people to use the platform (e.g., gamification, health management features). In addition, further validation of the models in more controlled environments (such as in a lab, where stressors can be applied and controlled to generate balanced labeled data) would allow more robust evidence of their efficacy. Purposeful sampling will also allow the generation of more robust models for different stratifications.

The development of the MHP and stress prediction models indicates that mobile systems have the potential to be successfully used for health data collection. RF models perform well, and sleep data from the Apple Watch as well as Withings features, such as weight and temperature, are important predictors. This work suggests that health data from smart technologies might be used to monitor data for public health surveillance, and Apple and Withings devices could be used to study and predict conditions in a population such as stress. The platform presented here represents a tentative step towards a future in which smart technologies could potentially be used in conjunction with self-report data collection methods to enable new insights into the health of populations.

8.3 Discussion

The ML models presented in this paper were successful in predicting stressed states, especially the RF model, with metrics in line with the state-of-the-art (especially when considering models developed using real-world data). These results are especially promising when considering that models were built using data from mobile and wearable personal devices, rather than using research devices as most stress studies (Table B6). Sleep features from the Apple Watch were shown to be important to create models, as well as weight, temperature and BP data from Withings devices. This suggests that a system such as the MHP might be a valuable tool for quantifying and predicting stress in a population.

However, several limitations and challenges for public health implementation were discussed, including the need to collect enough examples of each class in the data, missing data, and technical problems with devices. A discussion of how such a system could work in a real-world deployment, and additional features on the MHP such as a backup failsafe, were also presented.

The next chapter discusses other approaches to train and test the ML models. This included using 80% of *participants* to train and 20% to test the model, as opposed to 80% of *all data*. This is more challenging as the model is tested on new, previously unseen data. Another approach is the creation of individualized, user-specific models that use data from a participant to create a model specific to that participant.

Chapter 9 - Stress and Additional Approaches for Prediction with Mobile and Wearable Devices

9.1 Foreword

While the results presented in Chapter 8 were promising, the strategy of following a train-test split and cross-validation on the entire dataset is not the only one used in the stress prediction literature. Other approaches include using leave-one-person-out cross-validation [166], or creating individualized models for each participant [195].

This chapter investigates these approaches for stress prediction when using the data collected from the MHP. An adaption is made to leave-one-person-out cross-validation in which 80% of participants are used to train the model and 20% to test the model in order to avoid biased models due to imbalanced classes in the testing datasets. The goal of this "hybrid" approach is to allow more examples of each class to be in the test set while providing the model with new data in each iteration to test. Since this approach was shown to have high variance in results, i.e., the choice of which participants will be in the training set and which will be used for testing affected the metrics, 50 train-test loops were made, and their average calculated for evaluation.

While the previous chapter focused on the public health aspect of the thesis, this chapter focuses more on stress, including more background information on the prevalence of the condition and how it affects the body, as well as more detailed related work on stress prediction using ML.

9.2 Predicting Stress in Daily-Life Routines Using Personal Mobile and Wearable Devices

9.2.1 Abstract

Background: Stress is an important modifiable health issue. With new advancements in sensing and remote monitoring tools, consumer-level devices (e.g., smartphones, smartwatches) have embedded sensors that monitor health-related variables such as heart rate, steps and sleep. These variables could potentially be used to monitor and predict stress in individuals, leveraging their personal devices.

Objective: Investigate whether health data from mobile and wearable devices can be successfully used to predict stress.

Methods: A pilot study was conducted with 45 participants for 2 weeks. Participants received an iPhone and several mobile and wearable devices (e.g., Apple Watch, Withings Blood Pressure Monitor). The iPhone had an app installed that collected device data from the Apple Health repository and allowed participants to self-report their stress levels. Random Forests and Support Vector Machines were used to predict stress states based on these data, utilizing two approaches: using data from all participants, and creating individual user-specific models for each participant.
Results: Models using data from all participants had f1-macro scores of around 50% for different analyses and stratifications. User-specific individualized models typically performed with f1-macro scores above 65% when accounting for class imbalances using the SMOTE technique. Sleep features were shown to be important for model development.
Conclusions: Data from personal-level mobile and wearable devices show promise in predicting stress using individual data, especially the approach of creating user-specific models. Given that the development of models based on real-world data is challenging, the approach of using generalized models could be improved with robust data collection in more controlled environments. Regardless of the approach used, special care must be taken to gather quality data that contains sufficient examples of both stress and non-stress states.

9.2.2 Introduction

Stress is the "health epidemic of the 21st century" [141]. Stress can be formally defined as the reaction people have when facing a situation bigger than their capacity to handle [35,196]. The concept of the term in healthcare is traced back to 1936 when Hans Selye defined it as the "nonspecific response of the body to any demand" after noticing that several patients with different disease diagnoses reported the same symptoms [180,196,197].

Stress can be considered a normal response to an unexpected situation, triggering the body's fight-or-flight response and allowing the individual to deal with the threatening circumstance. Physiologically speaking, the stress response is related to the autonomous nervous system (ANS), composed of the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The SNS activates the fight-or-flight response in response to stressors, signaling the adrenal glands to release hormones that will lead to several physiological changes (e.g., increased heartbeat, blood pressure and respiration rate, as well as an increase in glucose levels in the bloodstream). The goal at this stage is to generate energy that allows the body to deal with the unexpected threat or circumstances. Ideally, after the acute stressor is gone, the body returns to normal through the PNS, which typically has the opposing effect of the SNS [34,42,196].

However, constant exposure to stressors – a chronic stress response – in daily life can have severe negative consequences on the health of individuals, increasing the risk for hypertension, cardiovascular diseases, and stroke, among others [34,35,196]. Chronic stress can be severely debilitating, with over 25% of U.S. adults reporting such high levels of daily stress that they cannot function [33], and workplace stress cost over $300 billion USD annually in health costs, job performance and absenteeism [141,143]. It is estimated that workplace stress and lack of health insurance lead to 120,000 preventable deaths [144]. Stress also affects individuals in all walks of life: in Canada [31], individuals aged between 35 and 49 years had the largest percentage of reported high-level of daily perceived stress (27.8%), followed by individuals groups aged between 50-64 (22%) and 18-34 (21.9%). Individuals aged 12-17 and over 65 years had the lowest percentages, 14.5% and 10.9%, respectively. Recently, a survey from the American Psychological Association reported that nearly 80% of respondents perceived the COVID-19 pandemic as a source of stress in their life, with 67% experiencing increased stress because of the pandemic [32].

The goal of this study is to evaluate stress prediction in the context of a mobile health platform (MHP) developed to support public health surveillance efforts. This MHP collects data related to Apple Health (AH), a popular health data repository integrating health data from smart devices that are compatible with Apple operating systems. We will use objective sensor data collected from AH by the MHP to predict stress in a population, with the goal of providing public health agencies with a potential new tool to monitor stress and apply interventions.

9.2.2.1 Related Work on Stress Prediction

Due to its risk and consequences, and the development of smart technologies that can monitor health variables associated with stress, many studies have sought to use mobile and wearable devices to collect data and use ML to predict stress states, as shown in Table B6 in Appendix B. To classify the studies in Table B6, we use a nomenclature developed by Can et al. [169]:

- Laboratory-to-laboratory known context (LLKC): Data collection occurs in the laboratory environment and the stress labels are based on the known context of stressors applied in the same environment.
- Laboratory-to-laboratory self-report (LLSR): Data collection occurs in the laboratory environment and the stress labels are based on collected self-reports in the same environment.
- Daily-to-daily self-report (DDSR): Data collection occurs in the field and stress labels are collected on self-reports (as known context is not possible in the field)
- Laboratory-to-daily known context (LDKC): Data collection occurs in the laboratory environment with known context of stressors for training, and training occurs with data collected in the field.
- Laboratory-to-daily self-report (LDSR): Data collection occurs in the laboratory environment with self-report stress for training, and training occurs with data collected in the field.

As can be seen in Table B6, the studies vary in several ways, such as laboratory to real-world data, number of participants, duration to data collection protocol, method of measuring stress, models and variables used, and whether the model uses data from all participants or develops a model per participant. Further, in both the table and a survey conducted by Can et al. [35] of stress prediction studies in daily life scenarios with smart devices, successful and widely used methods for stress prediction include RFs and SVMs. The state-of-the-art for stress prediction in daily life seems to lie approximately between 60% and 80% [21], with real-life studies showing worse results.

For example, an LLKC/LLSR study trains an SVM on ECG and respiration data, with a 72% accuracy in real-life (compared to 90% in a laboratory setting) [166]. This study trains one general model using all participant data, using a leave-one-person-out (LOPO) validation method. Another study, this time DDSR, uses ECG, respiration, sleep, and galvanic skin response, obtaining 73% accuracy with SVM and 71% with RF [167]. This study also uses a LOPO method to train a general model. On the other hand, a DDSR study [46] that uses HRV, audio, physical activity and communication data trains a general model using LOPO with a 53% accuracy using Logistic Regression (LR) and 61% accuracy with individualized models for each user. Another study [195], DDSR, uses accelerometer data and finds a 71% accuracy for individualized models, with an accuracy of 52% for generalized (both using Naïve Bayes). To validate the generalized model, 5-fold cross-validation is used.

As can be seen by these examples and others in Table B6, there are many different ways of creating and testing the models. In the next section, we will describe our methods for data collection and model development. Of note, few studies leverage personal devices such as smartphones as smartwatches. In this manner, our pilot study is different from prior studies as it will focus on consumer-level, off-the-shelf devices and study their feasibility for data collection while proposing a mobile public health data collection system. To the best of our knowledge, this work and others using the same dataset [21,147] are some of the first studies to use these devices and related data (such as the Apple Watch ECG) for stress prediction with ML.

9.2.3 Methods

9.2.3.1 Recruitment and Study Protocol

For the participant sample, a power calculation was performed using $\alpha = 0.10$, $\beta = 0.20$ and $r = 0.40$ (estimated from the limited evidence on correlation between stress and HRV [152,198,199] and stress and sleep [129,200,201]) yielding a number of 37 participants. Looking at Table B6, this is larger than most reviewed studies. When considering these factors, and that ML models typically work better with more data, in this study, 45 participants were recruited. The smallest sample used ($n = 22$ as will be described below) is still larger than most studies in Table B6. Participants were recruited from the University of Waterloo and using Facebook (both regional groups as well as Ads) and Kijiji, a Canadian marketplace website. Participants were required to be near the Kitchener-Waterloo region in Ontario as devices were delivered in-person. $100.00 CAD was offered for two weeks of data collection with the following devices (shown per manufacturer):

- Apple: iPhone 8 with iOS 15.0 and Apple Watch Series 6 with watchOS 8.3.
- Withings: Withings Sleep, Withings Blood Pressure Monitor (BPM) Connect, Withings Thermos and Withings Body+
- Empatica: Empatica E4 wristband

Since the goal of the pilot study is to support public health monitoring with the MHP, most devices are personal, off-the-shelf, and consumer-level, with the exception of the Empatica E4. The inclusion of this device was to complement data if needed, since it is widely used in literature for stress prediction (Table B6). As described in subsequent sections, we conducted analyses with and without the Empatica E4 data to explore the feasibility of personal devices for stress prediction.

The iPhone contained the MHP mobile app installed. In addition to collecting sensor data, the MHP also contained modules for stress self-report questionnaires, described in subsequent sections. The architecture and interface of the MHP are described elsewhere.

Table 21 illustrates variables that can be collected by each device. The User Manual provided to participants is included in Appendix C, containing the instructions on how to install and use each of the devices during the study. In addition to the User Manual, a 1-hour video call was scheduled with each participant to provide an overview of the manual, answer any questions, and make sure devices were properly installed.

Figure 34 shows the study protocol. The Ecological Momentary Assessment (EMA) methodology was used to obtain self-reports close to real-world events [160]. In this manner, users collected data 6 times per day – once when they wake up, once at sleep, and the rest throughout the day at approximate 3-hour intervals. The times shown in Figure 1 are merely illustrative. Measures included all variables shown in Table 21 with the devices. Apple Watch and iPhone Steps, Apple Watch HR, and Empatica E4 data were collected continuously, and so do not require any action by participants.

Participants commonly found it difficult to take all measurements while proceeding with their daily life. When necessary, participants were asked to use devices for additional days if they were not able to complete measurements. Only 6 participants were able to finish the study in two weeks. More information on this is provided in the Results section. This study was approved by the University Waterloo Research Ethics Board (REB [43612]), and data collection occurred between December 2021 and December 2022.

Table 21: Variables Collected and Devices Used in Study

Variable List	Devices Distributed to Participants						
Name	**iPhone (with MHP Installed)**	**Apple Watch**	**Empatica E4**	**Withings Sleep**	**Withings Blood Pressure Monitor Connect**	**Withings Thermos**	**Withings Body +**
Weight (Kg)							
Steps							
Heart Rate, HR (bpm)							
Blood Pressure, BP (mmHg)							
Sleep							
Heart Rate Variability, HRV							
Electrocardiogram, ECG (mV)							
Temperature (celsius)							
Stress Self-Report							

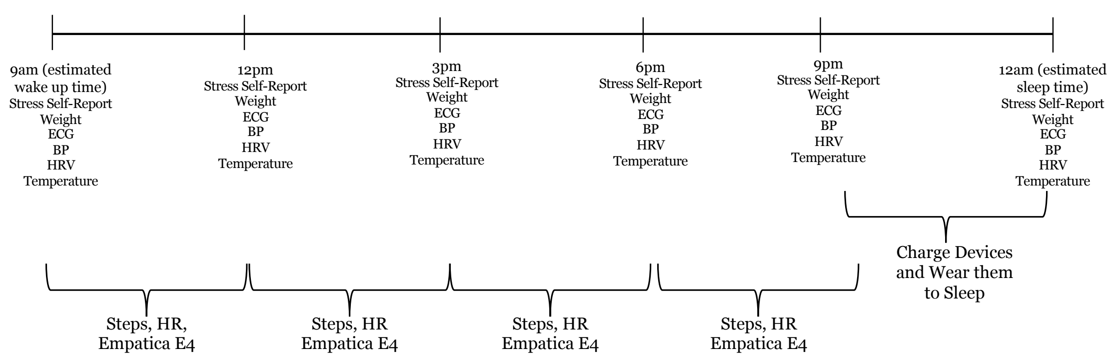


Figure 34: Study Protocol

9.2.3.2 Stress Self-Report

One of the challenges encountered when designing the study was a lack of validated stress questionnaires specifically for EMA, as most questionnaires have a validated period of days or weeks. To mitigate these issues, two questionnaires were used in this study: the widely validates stress subscale of the Depression, Anxiety, and Stress Scale (DASS-21) [128], and a single-item measure that lacks validation but was selected due to its simplicity and previous use in literature [129]. While the DASS-21 is usually applied over a week, there is promising evidence of using subscales of the DASS-21 with EMA [128]. In addition, the single-item measure was used successfully for stress quantification and is moderately correlated with robust stress questionnaires [129]. The stress-related questions are shown in Table B1 in Appendix B.

Following DASS-21 guidelines, question scores are multiplied by 2 and marked as stress if above 14 [162]. Single-Item scores are marked as stress if above 2. If the DASS-21 or single-item measure is marked as stress, the label for that questionnaire is stress. This label will serve as the ground truth for the ML models. For example, Figure A25 shows an example data point in the dataset. In this case, the DASS-21 score of all 7 questions is 7, or 14 when multiplied by 2. Since the value is not above 14, it is labelled as no stress. The single-item score is 3, which is labelled as stress since it is larger than 2. Since one of the scores is labelled as stress, the data point is classified as stress in the "stress_score" column, used as the model outcome.

9.2.3.3 Data Collection and Pre-Processing

As a first step in the data pre-processing stage, we looked at the missing data. For a feature, we evaluated if any participant had more than two data intervals missing, i.e., more than 6 hours between data collection. In the negative case, the average between the next and previous measurements was used to fill the gap. If positive (if two data points were missing continuously), k-nearest neighbors (kNN) was used to estimate the value based on the non-missing features through SciKit Learn's *KNNImputer* method (number of neighboring samples set to 5) [183]. Features were included if they had at least 30% of data for the specific user. This number was achieved by balancing the amount of data used with empirically observing KNNImputter behaviour for missing data. If a user had less than 30% of the data available for a feature, the user was removed from the specific dataset (the characteristics of each dataset's composition is detailed in the upcoming subsections and Results section).

Next, each collected variable was processed into features. An exhaustive list of all features derived from the study and their description is shown in Table B2 in Appendix B.

Steps We used the Apple Watch steps information to collect steps data. While the iPhone also collects steps information, it was not possible to integrate Apple Watch and iPhone data from HealthKit without duplication, and so only Apple Watch data was used. From the Apple Watch steps data we extracted the mean, maximum and minimum number of steps for the time interval in question (from the date of the previous data collection point to date of the current data collection point). This was done through an automated Python code which allowed the extraction of these values according to date and time. Of note, the *KNNImputer* was not used for missing steps data as it is possible the user did not walk during the period – in which case the step count is 0. Figure A26 in Appendix A shows a snapshot of the dataset with the 3 step features.

Heart Rate HR data was also measured with the Apple Watch. The device collects data throughout the day using green LED lights and light-sensitive photodiodes to measure the amount of blood in the wrist [22]. In this manner, heart rate Apple Watch data was passively and continuously collected over infrequent periods according to the device's system. Similarly to steps, mean, maximum and minimum heart rate for the time interval was extracted through a Python script. The heart rate was calculated as beats per minute. Figure A27 shows a snapshot of the Apple Watch HR features in the dataset.

In addition, when the user activates the Breathe or ECG app as requested during data collection (see Appendix C for User Manual), the device typically collected HR data in a millisecond interval. Therefore, a Short-Term HR feature was created considering these millisecond data close to the data collection date, extracting also the mean, maximum and minimum. Finally, data collected from the Apple Watch and for any other devices (e.g., Withings Sleep) during the night were also extracted and included as sleep features, as detailed in the section on sleep features.

Apple Watch ECG and Empatica data were processed with the Kubios Premium 3.5.0 software. Kubios' automatic beat correction algorithm was selected, and the software's automatic noise detection was set to medium. For Empatica, the raw interbeat interval data from the device was used. This file, named "BVP.csv" in the Empatica system, was downloaded directly from the Empatica web portal. For each data point, we uploaded into Kubios the respective file (BVP.csv from Empatica or a CSV file with 15360 voltage measurements and timestamps obtained from the ECG measurements in the Apple Watch). Kubios provided the mean, maximum, minimum, and standard deviation of HR for the time interval. More details on Empatica and ECG data extraction and Kubios are provided in the next subsection on heart rate variability.

Heart Rate Variability HRV is measured in AH as the SDNN (standard deviation of beat-to-beat measurements) metric through photoplethysmography (PPG) [22]. While this metric is collected throughout the day at random intervals and depending on user activity, users were also asked to have the Apple Watch Breathe app [202] open for 5 minutes as the final step of the data collection process, to forcefully trigger the HRV data collection. However, we found that SDNN data were collected per time interval without the need for the Breathe app 5-minute measurement, and these measurements have fewer missing data points. Since the passive monitoring also closely resembles a real-life deployment, the SDNN metric collected throughout the day without the Breathe was used, named HRV-1.

ECG data, composed of timestamps (30 seconds total) and voltage measurements on the Apple Watch ECG app (see User Manual in Appendix C), was processed using Kubios into HRV data. Each of the 15360 voltage and timestamp measurements comprising an ECG measure were saved to our database in JSON format, exported into a CSV file and ordered by time. Then, this file was uploaded to Kubios for feature extraction. To the best of our knowledge, this is one of the first works to use the Apple Watch ECG data in stress prediction.

According to the Task Force of The European Society of Cardiology and the North American Society of Pacing and Electrophysiology [153] indications, the following HRV features from the ECG were removed:

- pNN50 and the NN50, which are highly correlated to the RMSSD.
- TINN, HRV Tri Index, VLF and log measurements, which are not indicated for short recordings.

Empatica E4 data were also processed using the Kubios software, using the "BVP.csv" file provided by Empatica as previously mentioned. 10-minute intervals close to the data collection point were used to capture states close to the self-report. When possible, 5 minutes before and 5 minutes after the data collection point were used. We removed the same features as above with the exception of very low-frequency and log components that may capture more information in longer measurements.

Figure A28 shows a snapshot of the Apple Watch HRV-1 feature; Figure A29 shows a snapshot of ECG HRV features; and Figure A30 shows an example of Empatica HRV features.

Weight, Blood Pressure, Temperature For blood pressure, in addition to systolic and diastolic values, the mean arterial pressure was calculated with the formula: (sys + 3* dys)/3 [181,185]. Due to the size and weight of the Withings Body+ smart scale, participants were asked to take the weight measurements with the scale while at home and/or when they had access to the device. Since this reflects real-world deployment, we included weight data in this study despite large gaps, using the *KNNImputter* to fill these. Figure A31 shows a snapshot of temperature, weight and blood pressure features.

Sleep As mentioned previously, HR during the night (mean, maximum and minimum beats per minute) was calculated and included as a sleep feature, using the same Python script used to calculate these variables for steps and heart rate. In addition, the following sleep features from the Apple Watch and for the Withings Sleep device were calculated: Total Time Asleep, Number of Wake-Ups, Time Awake During Sleep , Total Time in Bed, and Percentage of Time Asleep While in Bed.

Withings Sleep provided additional information on Light, Deep and REM stages. While the Apple Watch recently enabled the collection of sleep stages information, this was not available at the time of the study [186].

Due to potential bidirectionality between sleep and mental health [53,187], features offset by 2/1 days before/after were created. As an example, if a data point was collected at time *t,* the *t-2* feature would place this value two days before, and the same with *t-1* (a day before), *t+1* (a day after)*,* and *t+2* (2 days after). Looking at Figure A32, for the data points on December 17th, for instance, the *t+2* features would place these on December 19th.

To avoid repetition of features, which could lead to a substantial increase in the dataset and poorer performance and results, initial RF models were used to calculate the feature importance (based on mean decrease in impurity [21]) of different sleep offset days. With this test, features from *t-2, t+2,* and *t* demonstrated good predictive value. Few features from other days were included, notably: *t+1* Apple Watch Mean HR, *t+1* Apple Watch Max HR, *t+1* Apple Watch Min HR, *t+1 Withings Total Time Asleep, t+1* Apple Watch Number of Wake-Ups*, t+1* Apple Watch Time Awake During Sleep *, t-1* Apple Watch Mean HR, *t-1* Apple Watch Max HR, t-*1* Apple Watch Min HR, *t-1 Withings Total Time Asleep, t-1* Apple Watch Total Time In Bed. More detail on the sleep features used can be found in Table B2 in Appendix B.

Since sleep data is collected at a different frequency than the EMA data, sleep features were included for the entire day, i.e., every data point of day *t* will have the same sleep feature.

9.2.3.4 Additional Feature Selection and Normalization

After pre-processing and cleaning the data, highly correlated features (with a Pearson correlation coefficient higher than 0.95) were removed. Data were also normalized for input in SVM models using *SciKit Learn's StandardScaler* method [188].

9.2.3.5 Analyses/Experiments

Given the number of devices and variety of data, as well as the different methods of testing and validating stress prediction models from the literature, several experiments were conducted with different subsets of the data. Further, due to different data collection frequencies, analyses were also conducted with and without sleep data.

Without sleep data, the following datasets were created:

- Dataset with all features, D (n = 22)
- Dataset with only ECG features, DECG (n = 42)
- Dataset with only Apple features, DA(n = 42)
- Dataset with only Withings Features, DW (n = 44)
- Dataset with Apple and Withings Features, DAW (n = 41)
- Dataset with Only Empatica Features, DE(n = 27).

Adding sleep features, the datasets used were:

- Sleep Dataset with only Apple features, SDA (n = 34)
- Sleep Dataset with only Withings Features, SDW (n = 34)
- Sleep Dataset with Withings and Apple Features, SDAW (n = 27)

- Sleep Dataset with Withings and Apple Only Sleep Features, SDS (n = 27)

DECG and DA contain the same participants, with additional Apple features in DA. As mentioned, if a participant possessed more than 30% of the data, this participant was removed from the dataset in question. For the datasets including sleep features, Empatica data was not included as this would result in very small datasets due to missing information (28 users had less than 30% of Empatica data due to connectivity problems or noise). Table B3 shows the participant characteristics for each dataset. More detail on how they differ is provided in the Results section.

These datasets were then evaluated with two different techniques:

- Generalized_Imb: since data was collected in real-life circumstances, it was common for users to have imbalance in the classes, resulting in some users having a predominance of *stress* or *no stress* classes that could affect results. In this manner, a variation of the LOPO approach (which is typically used for generalized models creation) was used. In our case, 80% of participants were selected for training and 20% for testing. This was chosen instead of a 70-30 split to provide more training data for smaller datasets (e.g., D with n = 22). 10-fold cross-validation is used on the training set to tune hyper-parameters. To avoid high variance in validation results depending on the selection of the participants for training, the *80-20 split - cross-validation - testing* procedure was repeated 50 times and the average of results used to evaluate the model.
- Individualized User-Specific Models (USM): separate model trained for each of the 45 users.

To further deal with class imbalance, the SMOTE (Synthetic Minority Over-sampling Technique) method on *SciKit Learn,* which upsamples the minority class [147,189], was used, and we conducted the analyses with the imbalanced classes as well as with balanced classes. SMOTE is applied only to the training sets to avoid overfitting [147]. This means the test set could potentially be imbalanced; for this reason, we report the f1-macro score, which considers both classes as being of equal importance.

Due to the relationship between stress and factors such as sex, age, income, work, and health [21], the following stratifications were made:

(1) Total: comprised of data from all participants in each dataset;
(2) Age: models in the age range of 18-24, 25-34 and 35-44 were trained. Models in the 45-64 and above were not included due to the scarcity of participants in this interval.
(3) Sex: models for male and female participants. We did not train a model for the participant that self-identified as gender fluid due to the scarcity of participants that self-identified as such.
(4) Income: models for participants with low socioeconomic status (SES) – earning less than CAD 30,000 – and participants belonging to middle and high SES. The CAD 30,000 cut-off point is based on an approximation of the Canadian tax cut-off for low-income populations [175].
(5) Profession: models for workers (full-time, part-time, and participants that are self-employed or classified as other) and students. We did not train a model for the retired participant as only one participant was in that category.
(6) Healthy: model removing participants that reported chronic diseases, illnesses, frequent alcohol or drug use, or prescription drugs.

Details on each dataset and its stratifications, including how the samples differ, are included in the Results section.

For each of these divisions, we trained the model with binary classification (*stress* vs *no stress*). While the SVM model performs many transformations to fit the data, feature importance for RF models was calculated using mean decrease in impurity. Figure 35 and Figure 36 show the process of obtaining the different datasets and training the Generalized_Imb and USM.

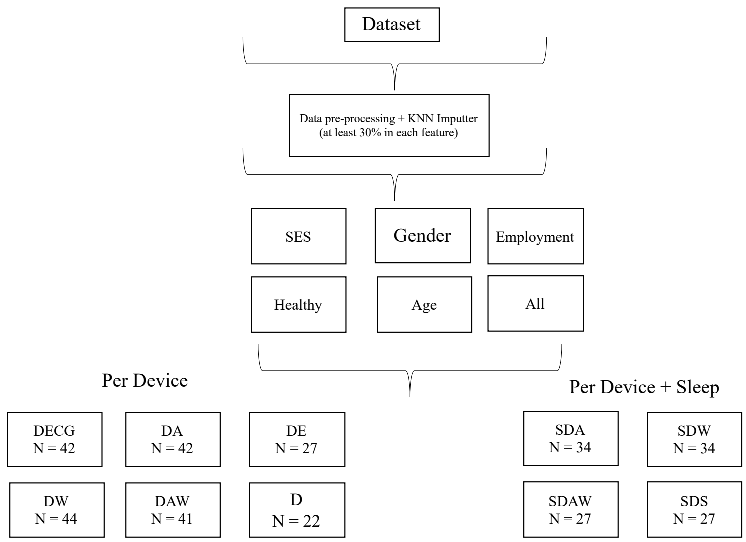


Figure 35: Division into Datasets per Device and Per Device + Sleep

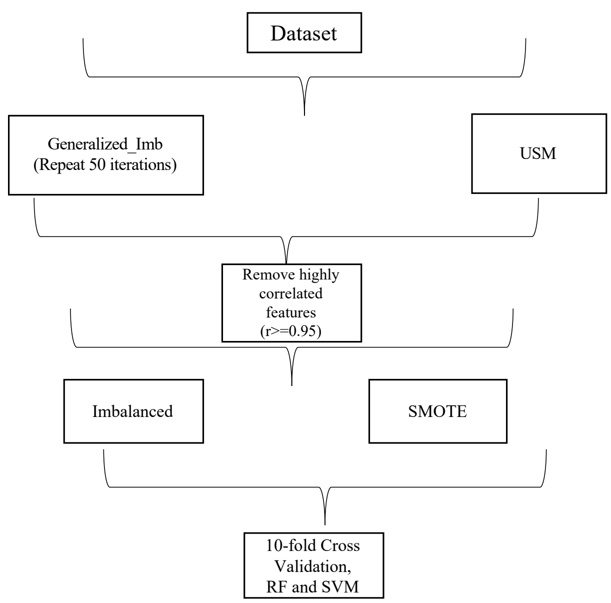


Figure 36: Training GENERALIZED_IMB and USM

9.2.4 Results

9.2.4.1 Descriptive Statistics

9.2.4.1.1 Population Data Characteristics

Table 22 describes the characteristics of the population. Most participants are in the younger range (87% aged 44 years or younger), female (67%), with low or medium income (84%), and workers (62%). A majority identified as white (33%), South Asian (24%) or Latin American (22%). Most participants (80%) did not have chronic disease or illnesses, use prescription drugs/alcohol or smoke. The average of days in the study for each participant was 17.1 (± 2.5), with an average of 78.91 (± 11.0) data points (rows in the dataset). It should be noted that the quality of the data did not differ between participants based on number of days; in fact, it was common that participants had datasets without many missing data instances when requested to extend the study due to difficulties in taking all 6 measurements for each day. Two categories had only one participant – one gender fluid participant and one participant above 65. We did not train Generalized_Imb models for these categories; rather, specific models were developed for each of them, as will be discussed in following sections and can be seen in Table 5.

When considering the entire sample of participants (n=45), 43% of answers were classified as stress (1539 data points), while the remaining 57% (2012 data points) were labelled as no stress. The proportion of data points labels is approximately maintained throughout the datasets: D (49% stress), DECG/DA/DAW/SDA (44% stress), DW (43% stress), DEmpatica (45% stress), SDAW/SDW/SDS (42% stress). Table B8 in Appendix B provides additional information on the percentage of data points labelled as stress for each division in the datasets.

Table 23 provides information on the percentage of data points labelled as stress for each division, considering all 45 participants. The total percentage of stressed labels considering the full data is 43%. Discounting categories with only one participant (which might not be representative), students had the highest percentage of stress labels at 50%, followed by younger populations aged 18-24 (48%) – which makes logical sense as most of this population is comprised of students –, healthy individuals (48%), and women (45%). Most additional categories had values between 43% and 41%, with the exception of men with only 37% of data points labelled as stress.

Looking at Table B8, male participants have a similar percentage of stress intervals in most non-sleep datasets when compared to the 43% stress labels from all 45 participants, varying between 44% and 39%. In sleep datasets, however, the percentage of stress labels drops, especially in SDAW and SDS at 18%. This is due to the smaller number of male participants included in these datasets due to missing data with only 6 men in SDAW and SDS (Table B3). On the other hand, women seem to be more stressed than men as reflected in most datasets with stressed labels varying between 52% and 45%.

In terms of income, individuals with low income had data points classified as stress, varying from 42% to 38% with the exception of DEmpatica with 47%. On the other hand, medium to high income individuals typically had higher values of stress (e.g., 52% in D, 49% in DECG/DA, 50% in DEmpatica), varying from 52% to 44%. Since all individuals with medium to high income except one are workers, this could likely be due to work stressors. When looking solely at workers, their stress values in the datasets varied from 46% (DEmpatica) to 37% (SDAW, SDS) with the exception of D, in which workers were stressed 54% of the time. Students were stressed approximately 48% of the time.

In terms of age, values ranged from 43% to 33% for ages 18-24, somewhat in contradiction to data from all 45 participants in Table 3. For ages 25-34, values ranged from 64% to 42% with consistently high values. For ages 35-44, values ranged from 41% to 23% (we report only values for 18-24, 25-34 and 35-44 due to the low number of participants in other categories). The big variations seen by age are likely due to the several options in this category (18-24, 25-34, 35-44, 45-64, Above 65). Therefore, removal or inclusion of participants by dataset could greatly affect the number of stress labels.

9.2.4.1.2 Quantification of Different Datasets

When looking at specific dataset characteristics (Table B3) when compared to the full study population (Table 22), we can see that, due to missing, inaccurate or noisy data, D removed several participants in the 18-24, 25-34 and 35-44 age ranges; halved the number of men and women and low and medium income; excluded high-income participants; included mostly full-time workers and students, with only 8 full-time workers compared to the full 21 in the study; and removed the majority of healthy participants, increasing by 7% the proportion of unhealthy participants in the study.

DECG/DA contain the same 42 participants. One participant each was removed from the 18-24 and 25-34 age categories, not changing the percentage of these categories when compared to all participants. The same can be said for male and female participants, with one men and two women removed. One participant with low income and two with medium income were removed, changing the proportion on the dataset slightly by 1% for low income and 2% for medium income. Only workers were removed (one full-time and two part-time), with the part-time proportion reducing by 4%. Finally, two healthy and one unhealthy participants were removed, with the proportion of healthy participants increasing 1% to the decrease in the proportion of unhealthy ones.

Similarly, DAW has one less participant from each age category used to build the models, with only the 25-34 category changing from 31% to 32%. 2 men and 2 women were removed. In terms of income, one low-income individual and 3 medium-income individuals were removed, with the medium-income proportion reducing to 37%. 2 full-time workers and 2 part-time workers were removed, with the proportion of part-time reducing from 11% to 7%. Finally, 3 healthy individuals and one unhealthy participant were removed, maintaining the same proportions.

DW only removed a 35-44 years old male healthy full-time worker with medium income, being very similar to data from all participants with n=44. On the other hand, DEmpatica had 18 participants removed, including: 6 participants aged 18-24 (proportion reduced from 29 to 26), 4 aged 25-24 (proportion increased from 31 to 37), and 6 aged 35-44 (proportion reduced from 27 to 22); 6 men (proportion reduced by only 1%) and 12 women, maintaining the same proportion; 6 low-income participants (proportion increased from 44 to 52), 9 medium income participants (proportion increased from 40 to 48) and all participants with high income or that did not wish to disclose; 12 full-time workers (proportion reduced from 47 to 33), 1 part-time worker (proportion increased from 11 to 15), and 5 students (proportion increased by 5%); 16 healthy participants (proportion decreased by 6%) and 2 unhealthy ones (proportion increased to 26).

When looking at the sleep datasets, SDA has 2 fewer 18-24 participants, 3 fewer 25-34 participants and 3 fewer 25-44, with proportions varying from 1% to 3%. 4 men and 7 women were removed, again with a slight 1% variation in proportions. 3 low, 4 medium and 4 high-income participants were removed, with big changes in the proportion of low-income (from 44% to 50%) and high-income (from 9% to 0%). 6 full-time and 4 part-time workers were removed as well as a retired participant, with an 8% reduction in the part-time proportion. Finally, 8 healthy and 2 unhealthy participants were removed, maintaining approximately similar proportions.

SDAW has 5 less 18-24, 5 less 25-24 and 5 less 35-44 participants, with very slight changes in proportion. 8 male participants were removed with a 9% decrease in proportion and 10 women were removed with a 7% increase. 7 low, 7 medium and 3 high-income participants, as well as one individual that did not disclose information, were removed, although the proportions remained similar with some variation. 9 full-time, 4 part-time, 4 students and 1 retired participant were removed, with only 1 remaining part-time participant (decrease from 11% to 4%). 16 healthy participants and 2 unhealthy ones were removed, with healthy proportion decreasing to 74%. SDS uses the same sample as SDAW.

When considering only Withings features in SDW, four 18-24, 4 25-34, and 2 35-44 participants were removed. 5 men and 6 women were removed, with the proportion of men decreasing from 31% to 26%. 5 low and 12 medium-income participants were removed as well as a participant that did not disclose information, with the proportion of medium-income dropping by 25. 5 full-time, 2 part-time, and 4 students were removed, maintaining similar proportions in the dataset. 8 healthy participants were removed, again maintaining similar proportions.

Table 22: Full Sample Participant Characteristics

Participants (N = 45)	*Frequency*	*Percentage*
Age		
18-24	13	29
25-34	14	31
35-44	12	27
45-64	5	11
Above 65	1	2
Sex/Gender		
Male	14	31
Female	30	67
Gender Fluid	1	2
SES		

Low (0-$30,000)	20	44
Medium ($30,000– $100,000)	18	40
High (Above $100,000)	4	9
Do not wish to disclose	3	7
Profession		
Full-time	21	47
Part-time	5	11
Student	16	36
Self-employed/Other	2	4
Retired	1	2
Ethnicity		
Black and Southeast Asian	1	2
Black or African American	3	7
Chinese	4	9
Indian	1	2
Latin American	10	22
South Asian	11	24
White	15	33
Health Status		
Healthy	36	80
Chronic Disease or Illness, Prescription Drug Use, Smoking or Alcohol	9	20

Table 23: % of Stress Datapoints Per Division, n = 45

Division	% Stress
Full	43

Gender Male	37
Gender Female	45
Gender fluid (n=1)	63
Income Low	41
Income Medium High	43
Employment Students	50
Employment Workers	41
Age 18-24	48
Age 25-34	42
Age 35-44	43
Age 45–64	42
Age Above 65 (n=1)	11
Healthy	48

9.2.4.2 Machine Learning Models

The following sections describe the results of the ML models. The focus of results reporting is the f1-macro metric, as accuracy may not reflect class imbalances in models trained without SMOTE. The f1-score is the harmonic mean between precision and recall, and the macro average treats both classes as equal. More detail on other metrics, including specifics for each prediction class, accuracy, f1-weighted, precision and recall can be found in Table B7 for Generalized_Imb and B9 for USM. The feature importance in each of the datasets are presented in Tables B20 to B29. Figures A16-A24 show the frequency of repeated top 10 important features for the stratifications in each dataset. DW was not included in these analyses as it only has 5 features.

9.2.4.2.1 Generalized_Imb

9.2.4.2.1.1 Without Sleep Data

In this section, we discuss the results of the Generalized_Imb models, illustrated in Table 24 with results for each dataset and each subset for imbalanced and SMOTE cases. The were usually around 50%, with stratification typically worsening the f1-macro scores. For imbalanced datasets, the SVM worked better than the RF model in most cases. Using the SMOTE technique to handle imbalanced data yielded better results for RF-SMOTE, although in general the f1-macro scores were still below or around 50%. The SVM-SMOTE technique yielded worse results in many cases.

Looking at the mix of different modalities of features in D in Table B20 there is a good balance between Apple ECG features (ECG_DC, ECG_AR_AbsolutePower_HF, ECG_Stress_Index, ECG_SDNN), Withings Features (MAP, Weight, dia) and Empatica features (Empatica_MSE2, Empatica_DC, Empatica_AR_LFHF) which is repeated among the different stratifications, although with different features appearing in different divisions. ECG_DC is present in several of the stratifications and repeated in Income, Age and All/Healthy (Figure A16). When considering the features from all datasets (Tables B20 to B29), ECG_DC, as well as Empatica_DC, are prominent in the datasets they are present. Withings features are also prominent and rank high among the most important features in datasets that mix different devices. In particular, weight is ranked first and repeated in DAW (Figure A19 and Table B23).

9.2.4.2.1.2 With Sleep Data

Integrating sleep data typically improved results, although they remained typically close to 50%. Stratifying the data yielded better results in several cases, especially for female and student participants. RF-SMOTE showed improvements in the metrics, with some subsets reaching the highest f1-macro score of the Generalized_Imb experiments of 58%. Much like in the previous case, SVM-SMOTE did not yield better results, sometimes causing worse metrics.

Regarding feature importance (Tables B26 to B29), sleep-related features typically rank among almost all the top 10 most important features, particularly Apple Watch sleep data, even in datasets that contain Withings sleep features such as SDAW and SDS. In SDS, containing only Withings features, non-sleep related data ranks high and repeats across stratifications (Figure A24 and Table B29).

The *t+1* and *t+2* Apple Watch Consolidated Time Awake During Sleep features, in addition to ranking high, also repeat among stratifications in SDA, SDS and SDAW.

9.2.4.2.2 USM

When creating the USM, 5 users were removed from the experiment as they had a Stress/No Stress ratio below 10%, which caused results to be poor due to lack of information on the *stress* class, or extremely high since the test dataset did not contain enough samples of the *stress* class for a non-biased testing process. Table 25 shows the characteristics of these 5 removed users: all healthy individuals, the majority (4 out of 5) men, full-time workers, South Asian and aged 35-44. Income varied from low to high.

User models for the remaining 40 participants were trained with all available features to each specific user, i.e., any feature that contained more than 30% of its data points. As mentioned, this choice was made to maximize the data used while still obtaining good empirical results from the *KNNImputter*. Sleep features were included when available. Empatica features were excluded from these analyses as only 27 users possessed a set with over 30% of data, and the features were shown not to have particularly high predictive power according to previous importance analyses.

Table 26 reports the f1-macro score for each user according to RF, SVM, RF-SMOTE, and SVM-SMOTE, the proportion of stress labels, and features that were excluded. The f1-macro averages for RF, SVM, RF-SMOTE, and SVM-SMOTE are, respectively: 59%, 52%, 62%, and 57%. Indeed, RF and RF-SMOTE have the best results. Individual f1-macros range from low (e.g., 25% with user 10) to very promising results (92% with user 1). 25 participants (63%) achieved metrics higher than 60%. Table B9 in Appendix B provides more details on precision, recall and accuracy metrics.

Of note, most of the missing features were sleep-related, whether from Apple Watch, Withings, or both (more on that in the Discussion section), which led to lower results from users without these features. The average of f1-macro without considering users with removed features removed, for RF, SVM, RF-SMOTE, and SVM-SMOTE, respectively (n = 24): 62%, 54%, 65%, 62%. RF and RF-SMOTE once again have the best results, which improved for all models.

Table 27 shows the averages of USM when stratified by gender, income, employment, age, and healthy participants. Typically, stratification seems to improve the results when considering USM, especially RF-SMOTE with results often above 65%. However, when the number of participants in a certain division is small (e.g., Gender Fluid, Age – 35-44, 45-64 and over 65), results are generally lower.

Table 24: Macro F1-Score Results for GENERALIZED_IMB

Dataset with all features (D, n = 22)

	RF	SVM	RF-SMOTE	SVM-SMOTE
All	0.49	0.52	0.5	0.47
Gender - Male	0.42	0.49	0.46	0.39
Gender - Female	0.51	0.51	0.51	0.48
Employment - Student	0.44	0.47	0.46	0.42
Employment - Worker	0.4	0.51	0.42	0.42
Income - Low	0.46	0.48	0.51	0.41
Income - Medium High	0.4	0.48	0.4	0.42
Age – 18-24	0.42	0.49	0.51	0.41
Age – 25-34	0.44	0.5	0.5	0.4
Age – 35-44	0.33	0.4	0.39	0.33
Healthy	0.51	0.54	0.53	0.44
Dataset with only ECG features (DECG, n = 42)				
All	0.5	0.5	0.52	0.5
Gender - Male	0.45	0.48	0.47	0.46
Gender - Female	0.52	0.51	0.53	0.49
Employment - Student	0.48	0.48	0.49	0.47
Employment - Worker	0.47	0.48	0.51	0.47
Income - Low	0.45	0.48	0.49	0.46
Income - Medium High	0.49	0.5	0.51	0.49
Age – 18-24	0.45	0.48	0.49	0.44
Age – 25-34	0.48	0.48	0.5	0.44
Age – 35-44	0.42	0.48	0.46	0.45
Healthy	0.49	0.5	0.51	0.49
Dataset with only Apple Features (DA, n = 42)				
All	0.5	0.51	0.53	0.41
Gender - Male	0.44	0.48	0..47	0.45
Gender - Female	0.5	0.51	0.51	0.45
Employment - Student	0.47	0.49	0..48	0.45
Employment - Worker	0.48	0.48	0.5	0.39

Income - Low	0.44	0.52	0.49	0.39
Income - Medium High	0.49	0.49	0.5	0.48
Age – 18-24	0.46	0.48	0.5	0.4
Age – 25-34	0.47	0.49	0.5	0.38
Age – 35-44	0.42	0.5	0.48	0.43
Healthy	0.48	0.5	0.53	0.38
Dataset with Apple and Withings Features (DAW, n = 41)				
All	0.47	0.51	0.51	0.38
Gender - Male	0.4	0.47	0.42	0.49
Gender - Female	0.45	0.48	0.47	0.39
Employment - Student	0.42	0.48	0.47	0.44
Employment - Worker	0.42	0.47	0.47	0.37
Income - Low	0.42	0.48	0.47	0.38
Income - Medium High	0.48	0.46	0.5	0.42
Age – 18-24	0.44	0.46	0.48	0.39
Age – 25-34	0.44	0.46	0.48	0.39
Age – 35-44	0.42	0.49	0.45	0.49
Healthy	0.46	0.52	0.51	0.37
Dataset with only Withings Features (DW, n = 44)				
All	0.46	0.47	0.46	0.49
Gender - Male	0.43	0.43	0.42	0.43
Gender - Female	0.44	0.47	0.44	0.48
Employment - Student	0.46	0.5	0.46	0.49
Employment - Worker	0.47	0.48	0.49	0.5
Income - Low	0.42	0.45	0.47	0.49
Income - Medium High	0.48	0.48	0.49	0.5
Age – 18-24	0.44	0.47	0.46	0.49
Age – 25-34	0.45	0.46	0.49	0.46
Age – 35-44	0.43	0.47	0.43	0.46
Healthy	0.34	0.46	0.48	0.46
Dataset with only Empatica Features (DEmpatica, n = 27)				

All	0.46	0.48	0.47	0.49
Gender - Male	0.42	0.56	0.5	0.51
Gender - Female	0.44	0.46	0.46	0.47
Employment - Student	0.46	0.47	0.49	0.48
Employment - Worker	0.42	0.44	0.42	0.43
Income - Low	0.45	0.47	0.51	0.49
Income - Medium High	0.37	0.42	0.38	0.42
Age – 18-24	0.41	0.46	0.43	0.46
Age – 25-34	0.44	0.44	0.46	0.44
Age – 35-44	0.44	0.48	0.45	0.46
Healthy	0.45	0.48	0.5	0.5
Sleep Dataset with only Apple Features (SDA, n = 34)				
All	0.52	0.54	0.54	0.47
Gender - Male	0.36	0.39	0.36	0.38
Gender - Female	0.55	0.54	0.55	0.49
Employment - Student	0.57	0.56	0.56	0.43
Employment - Worker	0.39	0.42	0.46	0.4
Income - Low	0.52	0.54	0.57	0.43
Income - Medium High	0.51	0.5	0.52	0.44
Age – 18-24	0.49	0.54	0.56	0.43
Age – 25-34	0.52	0.55	0.53	0.4
Age – 35-44	0.43	0.45	0.51	0.41
Healthy	0.49	0.53	0.45	0.44
Sleep Dataset with Apple and Withings Features (SDAW, n = 27)				
All	0.5	0.53	0.56	0.41
Gender - Male	0.45	0.45	0.46	0.45
Gender - Female	0.52	0.49	0.54	0.44
Employment - Student	0.55	0.53	0.57	0.48
Employment - Worker	0.4	0.46	0.54	0.38
Income - Low	0.51	0.56	0.55	0.39
Income - Medium High	0.51	0.51	0.55	0.36

Age – 18-24	0.54	0.58	0.56	0.41
Age – 25-34	0.57	0.51	0.57	0.43
Age – 35-44	0.44	0.48	0.49	0.44
Healthy	0.42	0.46	0.52	0.39
Sleep Dataset with Withings Features (SDW, n = 34)				
All	0.46	0.49	0.51	0.37
Gender - Male	0.48	0.47	0.51	0.37
Gender - Female	0.45	0.45	0.47	0.4
Employment - Student	0.49	0.52	0.51	0.49
Employment - Worker	0.43	0.44	0.48	0.37
Income - Low	0.43	0.44	0.51	0.39
Income - Medium High	0.42	0.47	0.46	0.41
Age – 18-24	0.5	0.5	0.51	0.38
Age – 25-34	0.48	0.47	0.5	0.4
Age – 35-44	0.44	0.4	0.58	0.38
Healthy	0.48	0.49	0.5	0.38
Sleep Dataset with Withings and Apple Only Sleep Features (SDS, n = 27)				
All	0.51	0.5	0.55	0.48
Gender - Male	0.45	0.45	0.45	0.44
Gender - Female	0.52	0.49	0.52	0.5
Employment - Student	0.55	0.52	0.57	0.52
Employment - Worker	0.4	0.47	0.47	0.39
Income - Low	0.52	0.53	0.58	0.5
Income - Medium High	0.52	0.51	0.54	0.48
Age – 18-24	0.5	0.53	0.53	0.5
Age – 25-34	0.52	0.49	0.52	0.52
Age – 35-44	0.44	0.44	0.5	0.44
Healthy	0.46	0.49	0.5	0.5

Table 25: Characteristics from 5 Users Removed from USM analyses

Age	Sex/Gender	Ethnicity	Employment	Income	Healthy
25-34	Male	South Asian	Full-Time	\$200,000 - \$249,999	Yes
35-44	Male	South Asian	Full-Time	Do not wish to disclose	Yes
35-44	Female	South Asian	Student	\$10,000 - \$29,999	Yes
35-44	Male	South Asian	Full-time	\$70,000 - \$79,999	Yes
35-44	Male	White	Full-Time	\$100,000 - \$124,999	Yes

Table 26: Results for USM

Users	RF	SVM	RF - SMOTE	SVM - SMOTE	% Stress	Features Removed
1	0.92	0.92	0.92	0.92	30%	
2	0.64	0.41	0.64	0.41	30%	
3	0.62	0.37	0.76	0.58	31%	
4	0.53	0.58	0.78	0.4	32%	
5	0.42	0.36	0.78	0.36	45%	AW Sleep
6	0.62	0.56	0.66	0.51	33%	Withings Sleep
7	0.41	0.41	0.55	0.41	27%	Withings Sleep
8	0.69	0.69	0.69	0.55	46%	
9	0.65	0.39	0.84	0.36	37%	AW Sleep, Withings Sleep
10	0.35	0.38	0.32	0.38	37%	Withings Sleep
11	0.8	0.54	0.8	0.72	44%	
12	0.91	0.45	0.84	0.73	19%	
13	0.54	0.41	0.65	0.79	31%	
14	0.52	0.56	0.41	0.45	20%	
15	0.76	0.81	0.7	0.81	43%	

16	0.44	0.44	0.56	0.44	25%	AW Sleep
17	0.72	0.63	0.67	0.75	34%	
18	0.43	0.42	0.38	0.43	22%	AW Sleep, Withings Sleep
19	0.61	0.76	0.61	0.76	43%	
20	0.61	0.5	0.67	0.58	45%	Withings Sleep
21	0.3	0.52	0.43	0.52	31%	ECG, AW Sleep
22	0.72	0.41	0.56	0.41	30%	
23	0.43	0.45	0.41	0.45	16%	AW Sleep
24	0.38	0.42	0.33	0.67	28%	
25	0.83	0.64	0.83	0.58	36%	AW Sleep
26	0.76	0.7	0.71	0.76	41%	
27	0.72	0.72	0.6	0.74	19%	
28	0.51	0.45	0.58	0.45	46%	ECG, AW Sleep, Withings Sleep
29	0.41	0.41	0.41	0.84	22%	
30	0.54	0.54	0.53	0.53	33%	
31	0.76	0.71	0.71	0.76	41%	Withings Sleep
32	0.59	0.55	0.71	0.66	33%	
33	0.41	0.43	0.73	0.43	25%	
34	0.42	0.38	0.68	0.47	35%	
35	0.52	0.35	0.59	0.66	47%	
36	0.71	0.38	0.62	0.38	36%	AW Sleep
37	0.33	0.38	0.38	0.42	33%	Temperature, Withings Sleep
38	0.69	0.72	0.65	0.74	32%	Withings Sleep
39	0.51	0.37	0.53	0.27	40%	
40	0.72	0.74	0.6	0.65	15%	ECG, Withings Sleep
Average	0.59	0.52	0.62	0.57		

Table 27: Results for USM, stratified

Stratification	**RF**	**SVM**	**RF - SMOTE**	**SVM - SMOTE**

Gender - Male (n = 10)	0.60	0.53	0.66	0.55
Gender - Female (n = 29)	0.58	0.52	0.60	0.57
Gender - Gender Fluid (n=1)	0.62	0.37	0.76	0.58
Income - Low (n = 19)	0.62	0.54	0.65	0.59
Income - Med High (n=17)	0.56	0.51	0.58	0.58
Employment - Student (n = 15)	0.63	0.55	0.64	0.65
Employment - Worker (n = 24)	0.57	0.51	0.62	0.52
Age - 18-24 (n = 13)	0.59	0.54	0.62	0.60
Age - 25-34 (n = 13)	0.64	0.58	0.65	0.62
Age - 35-44 (n = 8)	0.57	0.46	0.59	0.51
Age - 45-64 (n = 5)	0.47	0.45	0.65	0.49
Age - Over 65 (n = 1)	0.43	0.45	0.41	0.45
Healthy (n = 31)	0.58	0.53	0.62	0.56

9.2.5 Discussion

9.2.5.1 Model Results

Generalized_Imb models presented lower metrics, regardless of different datasets used, suggesting that this approach may not be ideal for training and testing models. On the other hand, user-specific results were very promising, with 63% of the 40 participants having an f1-score higher than 60%. The RF model typically overperformed the SVM, especially when applied with the SMOTE method for balancing classes. Therefore, predicting individualized models with within-individual data using RF models seems to be a promising approach for stress prediction in public health.

Of note, f1-scores are used to evaluate models to ensure we are taking into account the prediction of both stress and non-stress states, especially since, due to data collection in real-life environments, many participants did not have balanced stress and non-stress classes to test the models. However, accuracy is typically used to measure state-of-the-art in the literature. As mentioned, looking at Table B6, the state-of-the-art accuracy for stress prediction seems to lie between 60% and 80%, with lower results for models using real-world data. For evaluation purposes, Table 8 shows the accuracy results for each user (which can also be seen with additional metrics in Table B9). In this case, for RF, SVM, RF-SMOTE and SVM-SMOTE, the average accuracies are, respectively, 70%, 69%, 68%, 69%, with the RF model performing slightly better. Since accuracies may favor one class over another with better predictions, it makes sense that the balanced classes from the SMOTE method would not necessarily result in better results. When looking at the RF model, 33 out of 40 participants (82%) have accuracies higher than the bottom limit of the state-of-the-art of 60%. 22 participants (55%) have accuracies higher than 70%, and 6 participants (15%) performed better than the state-of-the-art. Therefore, considering within-individual data and USM, the majority of results are in line with or better than the state-of-the-art. Looking at Table B7 accuracies for Generalized_Imb models are usually above 60%, especially for sleep datasets, suggesting the models may predict one class better than another and that results could potentially be improved.

Table 28: Accuracy per User

Users	RF	SVM	RF-SMOTE	SVM-SMOTE	% Stress	Features Removed
1	0.94	0.94	0.94	0.94	30%	
2	0.75	0.7	0.7	0.7	30%	
3	0.71	0.59	0.76	0.65	31%	
4	0.67	0.73	0.8	0.67	32%	
5	0.43	0.57	0.43	0.57	45%	AW Sleep
6	0.71	0.59	0.71	0.53	33%	Withings Sleep
7	0.71	0.71	0.71	0.71	27%	Withings Sleep
8	0.69	0.69	0.69	0.56	46%	
9	0.71	0.64	0.86	0.57	37%	AW Sleep, Withings Sleep
10	0.54	0.62	0.46	0.62	37%	Withings Sleep
11	0.82	0.64	0.82	0.73	44%	
12	0.94	0.83	0.89	0.83	19%	
13	0.69	0.69	0.69	0.81	31%	
14	0.67	0.72	0.5	0.56	20%	
15	0.76	0.82	0.71	0.82	43%	
16	0.78	0.78	0.72	0.78	25%	AW Sleep
17	0.78	0.72	0.72	0.78	34%	

18	0.77	0.77	0.62	0.77	22%	AW Sleep, Withings Sleep
19	0.65	0.76	0.65	0.76	43%	
20	0.61	0.5	0.67	0.67	45%	Withings Sleep
21	0.43	0.64	0.5	0.64	31%	ECG, AW Sleep
22	0.76	0.71	0.59	0.71	30%	
23	0.76	0.82	0.71	0.82	16%	AW Sleep
24	0.61	0.72	0.5	0.72	28%	
25	0.87	0.67	0.87	0.6	36%	AW Sleep
26	0.76	0.71	0.71	0.76	41%	
27	0.88	0.88	0.76	0.82	19%	
28	0.54	0.46	0.62	0.46	46%	ECG, AW Sleep, Withings Sleep
29	0.71	0.71	0.71	0.88	22%	
30	0.61	0.61	0.56	0.56	33%	
31	0.76	0.71	0.71	0.76	41%	Withings Sleep
32	0.68	0.63	0.74	0.68	33%	
33	0.69	0.75	0.81	0.75	25%	
34	0.5	0.62	0.69	0.5	35%	
35	0.53	0.53	0.6	0.67	47%	
36	0.75	0.62	0.62	0.62	36%	AW Sleep
37	0.5	0.62	0.44	0.5	33%	Temperature, Withings Sleep
38	0.76	0.76	0.71	0.76	32%	Withings Sleep
39	0.65	0.59	0.59	0.59	40%	
40	0.89	0.72	0.78	0.83	15%	ECG, Withings Sleep
Average	**0.70**	**0.69**	**0.68**	**0.69**		

Still regarding USM, 4 participants had a very high f1-score above 80% - users 1,9, 25 and 12. All users were healthy, aged between 18-24 or 25-34. 2 were men and 2 women, with income varying from low (2 participants) to middle (1 participant) and high (1 participant). In particular, users 1, 25, 12 – who achieved the highest score of 92% for all models – had a robust, complete dataset, following the protocol strictly, with very few examples of missing data in the features and no missing Apple Watch sleep features. User 12 took all 6 measurements every day, necessitating only 14 days to complete the study; user 25 took 15 days; while user 1 was asked to wear the devices for 17 days to compensate for times in which not all measurements were taken. Interestingly when looking at user 12, only 23% of the data points were labelled as stress. However, the RF method without SMOTE for balancing classes actually performed better than RF-SMOTE (91% compared to 84%), suggesting that the classes do not necessarily need to be balanced so long as there are enough data points to provide information on the class for model training. User 9 seems to be an exception to the rule, with no sleep data and needing 20 days to complete the study and more instances of missing data. This user also had a relatively high proportion of stress points at 37%, which might account for the high f1-score values.

Most of the 15 users with an RF-SMOTE f1-score below 60% had at least one device sleep feature missing (Apple Watch, Withings, or both). Further, 4 of these participants were not considered healthy, with 3 of these not having any missing data, suggesting that health status may affect the capacity of USM to correctly identify stressed states. These participants greatly varied in terms of sex/gender, income, employment, and age. Further, these users' datasets were characterized by higher levels of missing data in features such as ECG and Blood Pressure that were applied to the *KNNImputter*, although other users that performed well also possessed data with these issues. Overall, suggests that sleep data is essential for good predictions, and models developed from participants without any chronic conditions, medication or drug/alcohol use will likelier be more accurate. Further, missing data imputation might behave differently depending on within-individual characteristics.

Also, of note, users 22, 25 and 35, which do not have sleep missing data and have generally complete datasets achieved promising metrics with other models other than the RF-SMOTE. For example, participant 35 had a 66% f1-score with SVM-SMOTE. Therefore, while RF-SMOTE consistently achieved higher values, that might not always be the case depending on individual differences. Interestingly, these 3 participants were all white females, although more research is needed to investigate these traits in terms of stress prediction.

To summarize findings on USM models, the results were promising, – specially using the RF method coupled with SMOTE for balancing classes – and could potentially be used for stress prediction. Sleep features seem to be essential for good prediction, and healthy individuals will likely achieve better results. Special care must be taken for individual differences, however, which might lead to worsening metrics, and evidence on the robustness of individual models should be collected before those models are applied in the real-world.

While USM models achieved results in line with the state-of-the-art, Generalized_Imb models had lower f1-scores, even when accounting for class imbalances using SMOTE. Seeing how USM models and their stratifications were mostly successful, it is likely that the models learn data for specific user metrics. Exploring the grouping of users according to characteristics such as gender, income, employment, age, or health, may also lead to improvement in performances, although that is not always the case and the number of users in each stratification should be high. New clusters of users could potentially be discovered using unsupervised learning approaches as well [77], generating more robust evidence on the effectiveness of individualized models.

9.2.5.2 Feature Importance

Apple Watch sleep features rank highly in importance when looking at Generalized_Imb. This is consistent with observations from the individualized models were users missing sleep data performed worse. Temperature, Weight, and Blood Pressure features from Withings also have high importance, as well as Apple Watch HRV data to a lesser extent. Offsets of the sleep data should also be considered, specifically, $t + 2$ and $t-2$. However, these results did not improve the Generalized_Imb models f1-scores in most cases.

In terms of non-sleep data, datasets that contained Withings data (D, DW, DAW) typically performed better. Coupled with the prevalence of Withings non-sleep features among the Apple sleep features in the sleep datasets, this suggests that using Withings devices to collect temperature, weight and blood pressure data would be an interesting avenue of research.

Empatica E4 data did not seem to be a good predictor regarding feature importance, as USM models without this data perform well, specially sleep-related models. Therefore, future studies can focus on the use of personal mobile and wearable devices.

The RF model performed better in general when compared to the SVM model, particularly on balanced datasets using the SMOTE method. If a system similar to the MHP is deployed in the real-world, datasets with more examples in each class could be used to train the model, therefore negating the need of generating synthetic health data which can potentially introduce biases.

9.2.5.3 Limitations and Future Directions

A limitation of the data collection method was the amount of missing data. Looking at Table 6, 18 out of 40 participants used in the USM models had over 70% of missing data for several features, in particular AW Sleep or Withings Sleep data, and these models typically performed worse than others, again suggesting the importance of sleep data. As previously mentioned, one of the main challenges with the data in this study was that, given the real-life nature of the design, several participants had imbalanced classes in their datasets. For this reason, when conducting the USM analyses, 5 users were removed, as their results were extremely good or poor due to the predominance of one class over the other. We used the SMOTE method to upsample the minority class for each user in our analyses, which typically generated better results, although any real-world deployments should be aware that large datasets should be collected to gather enough examples of each class.

The stress/no stress classes used as ground truth were derived from the DASS-21 and single-item measures, constituting a classification problem. Future work could further explore each of these datasets, creating models for each or even conducting ordinal classification for the numbered labels on the measures. In addition, developing USM models on controlled conditions (e.g., applying stressors in a lab and then applying the model to real-world data) would also solve the issue of imbalanced classes and, because of the noise reduction, could lead to improvements. Due to convenience sampling, a majority of participants were female, white, and young. Further studies should also explore a purposeful sampling of different participant traits for increased generalizability.

Further, despite the success of USM, more research should also be done on the Generalzed_Imb approach. As mentioned, the accuracy of Generalized_Imb models, especially for sleep datasets, was also in line with the state-of-the-art, suggesting these models are able to predict at least one class well. Conducting studies using this approach in more controlled conditions, as mentioned above, could lead to an improvement in results, as well as the use of more collapsed categories (for example, two age categories).

9.2.6 Conclusions

In this study, we developed RF and SVM models based on data from mobile and wearable devices collected with the MHP, a mobile app developed to support public health efforts. Two approaches were used, one that considers 80% of participants for training and 20% for testing and developing individualized models for each participant.

Several of the USM models were promising, and stratification according to characteristics such as gender, income, employment, age, and health, could further improve results. The Generalized_Imb approach had promising accuracy but worse f1-macro scores, requiring more validation. Therefore, the MHP seems to be a good approach for stress prediction in real-life when developing individualized models. Further evidence is needed on why models perform better on some individuals compared to others, and on the Generalized_Imb models. Lessons learned suggest that sleep data are extremely important, particularly from the Apple Watch, and that healthier participants are more likely to have good results. Future research should focus on collecting more data with purposeful sampling to develop improved models and further study stratification. Additional future work can involve further models based on different self-report metrics and conducting controlled stress experiments based on the data provided here for robustness.

9.3 Discussion

The accuracies of the "hybrid" approach (called Generalized_Imb in the paper), using 80% of participants for training and 20% for testing, were promising while the f1-macro scores were close to 50%, i.e., similar to chance. F1-macro scores, being the harmonic mean between precision and recall and considering both *stress* and *no stress* classes as having equal importance, can capture differences in prediction between the two classes better than accuracy. This suggests these models work well for one class, and future work (e.g., collecting more balanced and complete data in controlled conditions) could lead to improvements. In addition, the results of individualized user-specific models were in general promising.

Consistent with Chapter 8, sleep features were considered important for prediction, and users with these features missing performed worse than users with complete data.

The next chapter integrates the tools, models and results from the papers presented in this thesis (Chapter 4 to 9) into a final discussion, including implication of results for public health, and conclusions.

Chapter 10 - Discussion and Conclusions

After presenting the papers that comprise the research program, this chapter will discuss the research contributions, limitations, and future research derived from the results and conclusions from previous chapters.

10.1 Research Contributions

The work presented throughout the papers in this thesis on the development of a MHP and use of mobile and wearable technologies for public health has the potential to contribute to the fields of (a) public health surveillance, (b) stress prediction with ML. These contributions will be further detailed in the next sub-sections, as well as limitations of the research. Throughout this chapter, directions for future research are also highlighted.

10.1.1 Contributions to Public Health Surveillance

Public health surveillance is essential in collecting and analyzing data for decision-making and interventions that have the goal of improving the health of populations. In this research, I have discussed how smart technologies could potentially be used to complement public health studies, including in data collection and research, as well as opportunities, challenges, and limitations of using these devices in such a context (Chapter 4). The results show that smart technologies, using data from Apple Health, might be a promising tool for public health data collection and monitoring. Indeed, mobile and wearable devices can provide easy-to-use, faster and cost-effective tools to collect data continuously, passively, and in near real-time.

Equity must be considered when discussing public health data collection, which can be a major challenge as not everyone has equal access to and benefit from smart technologies. Age, income, location and ethnicity are major barriers to digital health equity in the Canadian context. As stated in Chapter 4, these issues should not hinder the application of mobile and wearable devices in health research and public health efforts, but rather *inform* it. This thesis showed how major Canadian surveys collect self-reported variables that might be complemented with objective sensor data from Apple Health and provided an overview of several works that use mobile technology as additional tools in health research. Public health agencies should be aware of current limitations in access to smart technologies and carefully ensure their protocols and studies consider these while also taking advantage of them when possible and applicable, creating best practices and guidelines for their use in public health programs. This is especially important as society is moving in a direction where barriers to technology access are decreasing, and hopefully will continue to do so.

In this context, Chapter 5 discussed the infrastructure of the MHP used as a data collection tool in the rest of the thesis, and which might be used as the basis for a future mobile surveillance application. Chapters 6, 7, 8 and 9 discussed the pilot study that uses the MHP to collect data and several analyses of these data for the quantification and prediction of stress.

Regarding the practical application of the MHP as a data collection tool in the pilot study, conducted under daily life conditions, several lessons were highlighted for public health, mainly in Chapter 8. First, to study the prevalence of a condition in a population with the use of supervised Machine Learning, it is essential that enough examples of each class involved in the prediction (e.g., *stress* and *no stress* for our case) be collected, which is challenging in real-world conditions. This means that large volumes of data should be gathered. To reduce the burden on users while ensuring scale, one approach could be asking users for few data points for a certain period over a long time, such as measurements at wake-up and before sleep every day for several weeks.

Public health analysts and researchers should also be mindful of missing or noisy data. In particular, data from the Empatica E4 device was severely affected by movement and lost Bluetooth connection, while the Withings Sleep device failed to integrate with the Apple Health app several times. Inaccurate data could also come from users forgetting to wear the device or wearing it improperly. While in the pilot study users were provided with an instruction manual and a video call was made to ensure everything was clear, possible real-world deployments of a mobile surveillance ecosystem should be proactive in detecting missing or inaccurate information, even providing actionable feedback to the user (e.g., a notification reminding users not to forget wearing their device and containing instructions on its proper use).

Given that Apple Watch sleep data was shown to be extremely important, future studies and research could prioritize this feature. Weight, temperature and blood pressure from Withings devices were also shown to be important features, which is particularly encouraging as these devices allow more flexible integration (e.g., they are also compatible with Android devices). Even if individuals do not have smart technologies to objectively collect this information, they might have devices that inform the data to them, such as regular weight scales. Therefore, future public health data collection efforts with mobile systems could allow for self-report of these data. It is important to note, however, that while this self-report may be collected more constantly than major surveys, it will still fall under the same limitations described in previous chapters for this type of data such as biases.

While devices that collect data in a similar manner to the ones used here should provide similar results, this is not a guarantee and, as mentioned in Chapter 4 comparison between devices is often difficult. In this manner, future studies should investigate other technologies for increased validity.

Still regarding the MHP, the data types that were collected in the pilot study were hardcoded into the app, i.e., queries and permissions for collecting these data were embedded in the source code. A potential real-world surveillance system should be flexible and allow customization according to different study objectives. In Apple systems, this could be done by hardcoding every data type and allowing researchers to add the data type they want to collect to their specific study application instance. Since the code for permission and queries will already be implemented, the data should be collected without issues.

While statistical methods did not reveal strong correlations, stress prediction model results were promising and often in line with state-of-the-art, which is especially encouraging considering models were built using data from personal, consumer-level devices as opposed to most research studies, which typically use research equipment (Table B6).

Different approaches for model training and evaluation yielded different results. In particular, using all data from participants to develop a generalized model or creating individualized user-specific models seem to be the best approaches. The current implication for public health agencies is that, if a potential system is put in place to predict stress in a population with smart technologies, large amounts of data from several participants should be collected for one generalized model, or models for each specific user should be developed, although both approaches need more validation. The latter approach had a lot of variation depending on specific users and data quality, and so should be deployed with great care and robust testing and validation. The fact that the models' performance drops with previously unseen data from several participants, as seen by the results of the Generalized_Imb models, limit their external validity and requires additional studies. However, accuracy metrics in this approach were often in line with state-of-the-art and showed promise. In short, mobile and wearable technologies might potentially be an asset for public health surveillance as results were promising and in-line with state-of-the-art. Further validation and testing studies are needed to apply stress prediction to mobile health data in order to study the prevalence of stress in the population. The next sub-section provides additional details on the stress prediction models and ML aspects of the study.

10.1.2 Contributions to Stress Prediction with ML

As detailed in Chapters 8 and 9, ML methods have been greatly used to process behavioural and physiological data in order to predict stress levels of participants. In the same chapters, we presented 3 distinct approaches to train, test and evaluate stress prediction models built with RF and SVM: Generalized models using data from all participants, with 80-20 train test split and 10-fold cross-validation; Generalized_imb models which use 80% of participants for training and 20% for testing with 10-fold cross-validation and 50 loops to average the results and decrease variance; and creating individualized, user-specific models.

The most successful approach was creating generalized models, which constantly performed in line with the state-of-the-art when looking at accuracy metrics. User-specific individualized models also typically achieved promising results, with a lot of variations among specific users depending on data quality and missing features. The Generalized_Imb approach performance decreased when looking into f1-scores, even when accounting for class imbalances using SMOTE. However, accuracy results were still in line with state-of-the-art, suggesting that these models can accurately predict at least one class and further validation could lead to improvements.

Seeing how often the “User” feature appears in several of the generalized datasets, and several user-specific models were very successful, it is likely that the models learn data for specific user metrics. Indeed, the Generalized_Imb approach that cross-validates by participant (thus, not using the “User” feature) shows worsened results, while the USM approach that uses data from each participant at a time has a lot of variance. This suggests that, while any potential public health efforts looking into predicting stress states should consider collecting large amounts of data to be used on one generalized model or develop a specific model for each user, both approaches have downsides – the first having reduced generalizability, and the second may have a lot of variance between subjects. Exploring the grouping of users according to characteristics such as gender, income, employment, age, or health, may also lead to improvement in performances, although that is not always the case and should be done in conjunction with careful planning and purposeful sampling. While the poor results of the Generalized_Imb approach limit the external generalizability and validity of the models, lessons can be learned for further validation and future research directions.

First, and encouragingly, Empatica E4 data did not seem to greatly affect the models, as generalized models and user-specific models without this data perform well, specially sleep-related models. Given this, data from personal consumer-level devices can be leveraged rather than having to resort to medical-grade wearables such as the Empatica E4. As discussed, Apple Watch sleep data was shown to generate particularly strong predictive features.

The RF models in general performed better than the SVM and having balanced classes can improve performance, further highlighting the need for large datasets to be collected without the need to generate synthetic data for balance.

The variance observed among the users for the user-specific models should give researchers pause before real-world deployment. How can public health agencies create individualized models and guarantee they will work for the majority of users, thus ensuring equity? While more validation is needed, an interesting future approach might be the use of reinforcement learning, a type of Machine Learning application in which the model learns from mistakes by trial and error through reward and punishment, maximizing the first while minimizing the other [191]. Feedback might be provided by the user of the smart technology. As a potential scenario, a user installs a system like the MHP on their device, which collects data and creates a user-specific model. This model then makes a prediction, and lets the user know of their condition. If the application asks feedback from the user on the accuracy of the prediction in the moment, the feedback can provide rewards or punishment to the classifier depending on the accuracy of the prediction made. This, in turn, might lead to more robust models.

In summary, the results of stress prediction for ML are encouraging, and often in line with the state-of-the-art. Given that state-of-the-art models are generally developed with research or medical-grade equipment, the good results of using mobile and wearable devices and the novel data types collected by these are very promising for future public health data collection initiatives, although more research and validation are needed for real-world deployment of similar systems.

10.2 Limitations and Future Directions

In this section some of the limitations in the research program are detailed, as well as approaches to address or mitigate them in the future.

10.2.1 Limitations of the MHP and Future Directions

As mentioned in the previous sub-section and throughout Chapters 7, 8 and 9, missing data was a major problem in this study. While missing data can come from users not wearing the device, in many instances the information was properly collected by the device and sent to Apple Health but could not be sent to the researcher's database due to technical issues (such as slow Internet connection). An approach to mitigate this in the future is the inclusion of backup features. Chapter 8 describes how such a feature could be implemented in the current iOS systems.

In our study, to handle this issue, data from the database was compared with manually exported Apple Health data to account for these types of missing data. A python script was created specifically to compare the data in Extensible Markup Language (XML) format exported from Apple Health with the data from our servers.

Even when considering errors in data sharing, there were still a lot of missing data to contend with, as evidenced by Chapters 8 and 9. This was mostly due to the burden of the study protocol on participants, who had difficulty using the devices during their daily routine or remembering to take every measurement. For future studies and real-world deployment, as we discussed, public health agencies should strive for collecting less data points over long periods to avoid placing a huge burden on users.

Apple guidelines provide different periods for background data delivery set up by HealthKit queries, as explained in Chapter 8 [139]. Further, background delivery is scheduled based on app usage, meaning that if the app is not constantly used, background data collection might not be triggered. To make matters worse, the iPhone's system may stop background queries. To circumvent these issues in the study, new queries are activated every time the app is opened. In case the app is used more constantly, or if users are meant to collect few data points over longer periods, different strategies for updating the data might be considered and implemented. To further engage users with the app, future research should also consider using the MHP system as a backend of a more integrated health management system. For example, an application that allows users to track their condition (whether stress or another) including improvements and downturns, and that also collects data similar to the MHP would further engage users and ensure they keep coming back to the application and system.

10.2.2 Limitations of the Pilot Study and Stress Prediction Models and Future Directions

While this pilot study is larger than most stress studies reviewed (see Table B6 in Chapter 9), 45 participants is still a small sample size when compared to traditional public health studies and initiatives. Therefore, stratified results in this thesis should be more indicative of future directions of research. On that note, due to convenience sampling, several participants were young, white, and female. Therefore, further studies should include more purposeful sampling in their design for different stratifications and increased generalizability. The fact that good results were found for stratifications with the generalized model approach, even with a relatively small dataset, is promising.

Participants found the data collection protocol burdensome. An alternative for future studies and real-world deployment would be collecting less data points for longer periods, reducing the burden on participants while collecting larger datasets. Further, since data was collected in real-life environments, subjective stress measures were used (see Chapters 6 to 9). However, different people may have different perceptions of stress. Indeed, "there is no uniform and universal relationship between a stressor and the stress response" [54]. In other words, different people might respond differently to the same stressors, which in turn may affect the labelled data. This could also explain why correlations in Chapter 6 were low. To mitigate this issue, we used validated questionnaires coupled with a single-item measure successfully used in literature and correlated with robust stress questionnaires. However, future work conducting experiments in more controlled conditions (e.g., applying stressors in a lab and collecting more objective stress measures like cortisol) would further increase the robustness of the evidence that mobile and wearable technologies can be used for stress prediction. Such experiments would also mitigate the issues of missing data and imbalanced classes, leading to improvements in models, although they would decrease the real-world validity of results.

As mentioned, complementary model training approaches could potentially be used to improve results, such as reinforcement learning to get feedback from users and reduce the variance in USM or gathering more context information for model development.

Nevertheless, the promising results found – especially considering the challenges of data collection "in the wild" and the use of novel, consumer-level devices not widely used in stress research – are encouraging and suggest that smart technologies have the potential to be effective tools in predicting stress, although care must be taken to address the issues and limitations presented in this work and what a real-world deployment would entail.

10.3 Conclusion

Mobile and wearable devices are in use by the majority of the population worldwide [203]. These devices have sensors that collect a variety of health variables which might potentially be used as a complement to traditional public health surveillance initiatives.

In this thesis, I have presented a potential new tool for data collection from smart technologies with the objective of providing a multi-modal prototype public health data collection and monitoring mobile ecosystem. The architecture and modelling of the MHP was presented, and the platform evaluated through a pilot study that used the MHP to collect stress-related data from mobile and wearable devices. These data, in turn, were used in the creation of a number of ML-based models to predict this condition among participants. Although more validation is needed, the results were promising, especially considering the models were built using data from personal devices.

The tools, ML-based models and results presented in this thesis represent a potential step towards the integration of mobile health technologies for public health. The ultimate goal of this work is to provide public health agencies with possible new directions, tools and methods to improve the lives of populations.

Bibliography

1. Association CPH. What is public health? [Internet]. [cited 2023 Jan 18]. Available from: https://www.cpha.ca/what-public-health
2. Kjellsson G, Clarke P, Gerdtham UG. Forgetting to remember or remembering to forget: A study of the recall period length in health care survey questions. J Health Econ [Internet]. 2014;35(1):34–46. Available from: http://dx.doi.org/10.1016/j.jhealeco.2014.01.007
3. Clarke PM, Fiebig DG, Gerdtham UG. Optimal recall length in survey design. J Health Econ. 2008;27(5):1275–84.
4. Canada S. Canadian Community Health Survey - Annual Component (CCHS). 2022.
5. Day B, Langlois R, Tremblay M, Knoppers BM. Canadian Health Measures Survey: ethical, legal and social issues. Health Rep. 2007;18 Suppl(82):37–51.
6. Canada S. Canadian Health Measures Survey (CHMS) [Internet]. 2023 [cited 2023 Jan 23]. Available from: https://www.statcan.gc.ca/en/survey/household/5071
7. NHS. Health Survey for England 2016. 2016;(December 2016). Available from: http://www.hscic.gov.uk/catalogue/PUB19295/HSE2014-Sum-bklet.pdf
8. Brener ND, Billy JOG, Grady WR. Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. J Adolesc Heal. 2003;33(6):436–57.
9. Lofters A, Vahabi M, Glazier RH. The validity of self-reported cancer screening history and the role of social disadvantage in Ontario, Canada. BMC Public Health. 2015;15(1):1–10.
10. Johnson TP. Handbook of Health Survey Methods. 2015. 816 p.
11. Gordis L. Epidemiology. (5th ed.). Elsevier Ltd; 2014.
12. Gibson TB, Ph D. How Accurate are Self-Reports? An Analysis of Self-Reported Healthcare Utilization and Absence When Compared to Administrative Data. J Occup Environ Med Coll Occup Environ Med. 2010;51(7):786–96.
13. Bhandari A, Wagner T. Self-reported utilization of health care services: Improving measurement and accuracy. Med Care Res Rev. 2006;63(2):217–35.

14. Winckers ANE, Mackenbach JD, Compernolle S, Nicolaou M, van der Ploeg HP, De Bourdeaudhuij I, et al. Educational differences in the validity of self-reported physical activity. BMC Public Health [Internet]. 2015;15(1):1299. Available from: http://www.biomedcentral.com/1471-2458/15/1299

15. Mc Dowell CP, Carlin A, Capranica L, Dillon C, Harrington JM, Lakerveld J, et al. Associations of self-reported physical activity and depression in 10,000 Irish adults across harmonised datasets: A DEDIPAC-study. BMC Public Health. 2018;18(1):1–8.

16. Barrett MA, Humblet O, Hiatt RA, Adler NE. Big Data and Disease Prevention: From Quantified Self to Quantified Communities . Big Data. 2013;1(3):168–75.

17. Jajosky RA, Groseclose SL. Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. BMC Public Health. 2004;4:1–9.

18. Bojovic O, Medenica M, Zivkovic D, Rakocevic B, Trajkovic G, Kisic-Tepavcevic D, et al. Factors associated with patient and health system delays in diagnosis and treatment of tuberculosis in Montenegro, 2015–2016. PLoS One. 2018;13(3):2015–6.

19. Delay to Mental Health Inquiry report | Health Central [Internet]. [cited 2019 Apr 29]. Available from: https://healthcentral.nz/delay-to-mental-health-inquiry-report/

20. Lovett L. Apple's ECG feature can detect A-Fib with high HR thanks to new FDA clearance [Internet]. MobiHealthNews. 2020 [cited 2022 Aug 26]. Available from: https://www.mobihealthnews.com/news/apples-ecg-feature-can-detect-fib-high-hr-thanks-new-fda-clearance

21. Velmovitsky PE, Alencar P, Leatherdale ST, Cowan D, Morita PP. Using apple watch ECG data for heart rate variability monitoring and stress prediction: A pilot study. Front Digit Heal. 2022;4(December):1–19.

22. Apple. Monitor your heart rate with Apple Watch - Apple Support [Internet]. [cited 2023 Jan 6]. Available from: https://support.apple.com/en-us/HT204666

23. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data. 2016;3:1–9.

24. Shephard RJ. The Objective Monitoring of Physical Activity: Contributions of Accelerometry to Epidemiology, Exercise Science and Rehabilitation. Object Monit Phys Act Contrib Accelerometry to Epidemiol Exerc Sci Rehabil. 2016;383.

25. Sahu KS, Oteomo A, Morita PP. Improving Public Health Surveillance methods via Smart Home technologies. Online J Public Health Inform. 2019;11(1):9–10.
26. Hirten RP, Danieletto M, Tomalin L, Choi KH, Zweig M, Golden E, et al. Use of physiological data from a wearable device to identify SARS-CoV-2 infection and symptoms and predict COVID-19 diagnosis: Observational study. J Med Internet Res. 2021;23(2).
27. Statista. • Smartphone users in Canada 2018-2024 | Statista [Internet]. [cited 2020 Sep 9]. Available from: https://www.statista.com/statistics/467190/forecast-of-smartphone-users-in-canada/
28. Statista. Wearables - Canada [Internet]. [cited 2018 Dec 13]. Available from: https://www.statista.com/outlook/319/108/wearables/canada
29. Apple. HealthKit [Internet]. [cited 2018 Oct 12]. Available from: https://developer.apple.com/healthkit/
30. Apple. ResearchKit and CareKit - Apple [Internet]. Apple. 2018 [cited 2018 Oct 12]. Available from: https://www.apple.com/researchkit/
31. Canada S. Perceived life stress, by age group [Internet]. 2018. Available from: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310009604
32. Association AP. Stress in America - A National Mental Health Crisis. 2020.
33. Association AP. More than a quarter of U.S. adults say they're so stressed they can't function [Internet]. American Psychologi Association. 2022 [cited 2023 Feb 8]. Available from: https://www.apa.org/news/press/releases/2022/10/multiple-stressors-no-function
34. Association AP. Stress effects on the body [Internet]. 2018 [cited 2022 Aug 10]. Available from: https://www.apa.org/topics/stress/body
35. Can YS, Arnrich B, Ersoy C. Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. J Biomed Inform [Internet]. 2019;92(February):103139. Available from: https://doi.org/10.1016/j.jbi.2019.103139
36. Mayo Clinic Staff. Stress symptoms: Effects on your body and behavior [Internet]. Clinic, Mayo. [cited 2023 Feb 8]. Available from: https://www.mayoclinic.org/healthy-lifestyle/stress-management/in-depth/stress-symptoms/art-20050987
37. StatisticsCanada. Canadian Health Measures Survey (Cycle 6) - Household Questionnaire

[Internet]. [cited 2023 Jan 23]. Available from: https://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&a=1&&lang=en&Item_Id=1202489

38. StatisticsCanada. Canadian Health Measures Survey (Cycle 6) - Clinic Questionnaire [Internet]. [cited 2023 Jan 23]. Available from: https://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&a=1&&lang=en&Item_Id=1202488

39. Physical Activity, Sedentary Behaviour, and Sleep Indicator Framework | Public Health Infobase - Canada.ca [Internet]. [cited 2019 Apr 29]. Available from: https://infobase.phac-aspc.gc.ca/pass-apcss/

40. Withings. Withings | Smart Scales, Watches and Health Monitors [Internet]. [cited 2022 Apr 27]. Available from: https://www.withings.com/ua/en/

41. Apple. Xcode | Apple Developer Documentation [Internet]. [cited 2023 Jan 24]. Available from: https://developer.apple.com/documentation/xcode

42. Setz C, Arnrich B, Schumm J, Marca R La, Tr G, Ehlert U. Discriminating stress from cognitive load using a wearable EDA device. Technology. 2010;14(2):410–7.

43. Mate G, M.D. When the Body Says No: The Cost of Hidden Stress. 2011;320. Available from: http://books.google.com/books?id=8hjoyCFTu6IC&pgis=1

44. Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and heart rate variability: A meta-analysis and review of the literature. Psychiatry Investig. 2018;15(3):235–45.

45. Jin CW, Oravecz Z, Osotsi A. Predicting stress in teens from wearable device data using machine learning methods. medRxiv. 2020;

46. Muaremi A, Arnrich B, Tröster G. Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep. Bionanoscience. 2013;3(2):172–83.

47. Stults-Kolehmainen MA, Sinha R. The effects of stress on physical activity and exercise. Vol. 44, Sports Medicine. 2014. 81–121 p.

48. Föhr T, Tolvanen A, Myllymäki T, Järvelä-Reijonen E, Peuhkuri K, Rantala S, et al. Physical activity, heart rate variability–based stress and recovery, and subjective stress during a 9-month study period. Scand J Med Sci Sport. 2017;27(6):612–21.

49. Kivimäki M, Head J, Ferrie JE, Shipley MJ, Brunner E, Vahtera J, et al. Work stress,

weight gain and weight loss: Evidence for bidirectional effects of job strain on body mass index in the Whitehall II study. Int J Obes. 2006;30(6):982–7.

50. Ayanian JZ, Block JP, He Y, Zaslavsky AM, Ding L. Psychosocial stress and change in weight among US adults. Am J Epidemiol. 2009;170(2):181–92.

51. Gasperin D, Netuveli G, Dias-da-Costa JS, Pattussi MP. Effect of psychological stress on blood pressure increase: A meta-analysis of cohort studies. Cad Saude Publica. 2009;25(4):715–26.

52. Muaremi A, Bexheti A, Gravenhorst F, Arnrich B, Troster G. Monitoring the impact of stress on the sleep patterns of pilgrims using wearable sensors. In: 2014 IEEE-EMBS International Conference on Biomedical and Health Informatics, BHI 2014. 2014. p. 185–8.

53. Alvaro PK, Roberts RM, Harris JK. A Systematic Review Assessing Bidirectionality between Sleep Disturbances, Anxiety, and Depression. Sleep. 2013;36(7):1059–68.

54. Maté G. When the Body Says No: Stress and The Mind/Body Unity [Internet]. 2015 [cited 2023 Apr 17]. Available from: https://drgabormate.com/wp-content/uploads/2015/04/when-the-body-says-no.pdf

55. Mooney SJ, Pejaver V. Big Data in Public Health: Terminology, Machine Learning, and Privacy. Ssrn. 2018;

56. Yuan M, Boston-Fisher N, Luo Y, Verma A, Buckeridge DL. A systematic review of aberration detection algorithms used in public health surveillance. J Biomed Inform [Internet]. 2019;94(April):103181. Available from: https://doi.org/10.1016/j.jbi.2019.103181

57. King RC, Villeneuve E, White RJ, Sherratt RS, Holderbaum W, Harwin WS. Application of data fusion techniques and technologies for wearable health monitoring. Med Eng Phys. 2017;42:1–12.

58. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? J Mach Learn Res. 2014;15:3133–81.

59. Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD. An introduction to decision tree modeling. J Chemom. 2004;18(6):275–85.

60. Tomaschek F, Hendrix P, Baayen RH. Strategies for addressing collinearity in multivariate linguistic data. J Phon [Internet]. 2018;71:249–67. Available from: https://doi.org/10.1016/j.wocn.2018.09.004

61. Maxhuni A, Hernandez-Leal P, Sucar LE, Osmani V, Morales EF, Mayora O. Stress modelling and prediction in presence of scarce data. J Biomed Inform [Internet]. 2016;63:344–56. Available from: http://dx.doi.org/10.1016/j.jbi.2016.08.023

62. Strobl C, Malley J, Tutz G. An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. Psychol Methods. 2009;14(4):323–48.

63. Hackeling G. Mastering Machine Learning with scikit-learn. 2014. 238 p.

64. Sedig K, Ola O. The Challenge of Big Data in Public Health: An Opportunity for Visual Analytics. Online J Public Health Inform [Internet]. 2014 Feb 5 [cited 2019 Feb 20];5(3):223. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24678376

65. Tomaschek F, Hendrix P, Baayen RH. Strategies for addressing collinearity in multivariate linguistic data. J Phon. 2018;71:249–67.

66. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC Bioinformatics. 2008;9:1–11.

67. Noble WS. What is a support vector machine? Nat Biotechnol. 2006;24(12):1565–7.

68. Ontario PH. Health Equity [Internet]. 2023. Available from: https://www.publichealthontario.ca/en/health-topics/health-equity

69. Mohadisdudis HM, Ali NM. A study of smartphone usage and barriers among the elderly. Proc - 2014 3rd Int Conf User Sci Eng Exp Eng Engag i-USEr 2014. 2015;109–14.

70. Canada D. Digital equity | Deloitte Canada [Internet]. [cited 2023 Apr 12]. Available from: https://www2.deloitte.com/ca/en/pages/future-of-canada-center/articles/digital-equity.html

71. Canada G of. High-speed Internet for all Canadians [Internet]. 2023. Available from: High-speed Internet for all Canadians

72. Soucie JM. Public health surveillance and data collection: general principles and impact on hemophilia care. Hematology. 2012 Apr;17 Suppl 1(0 1):S144-6.

73. Section 1: Evaluation of the surveillance function at the Public Health Agency of Canada

– Introduction - Canada.ca [Internet]. Public Health Agency of Canada. 2013 [cited 2018 Oct 12]. Available from: https://www.canada.ca/en/public-health/corporate/mandate/about-agency/office-evaluation/evaluation-reports/evaluation-surveillance-function-public-health-agency-canada/introduction.html#a1-1

74. Lee L, Teutsch S, Thacker S, Louis M. Principles and Practice of Public Health Surveillance. (3th ed.). Oxford University Press; 2010.

75. Tremblay M. The Need for Directly Measured Health Data in Canada. Can J Public Heal. 2004;95(3):165–6.

76. de Arriba-Perez F, Caeiro-Rodriguez M, Santos-Gago JM. Collection and processing of data from wrist wearable devices in heterogeneous and multiple-user scenarios. Sensors (Switzerland). 2016;16(9).

77. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. npj Digit Med [Internet]. 2019;2(1). Available from: http://dx.doi.org/10.1038/s41746-019-0121-1

78. Velmovitsky PE, Bevilacqua T, Alencar P, Cowan D, Morita PP. Convergence of Precision Medicine and Public Health Into Precision Public Health: Toward a Big Data Perspective. Front Public Heal [Internet]. 2021;9. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85104520090&doi=10.3389%2Ffpubh.2021.561873&partnerID=40&md5=4aad837e6d9a72f32e9b72b236a179b2

79. Kaihlanen AM, Virtanen L, Buchert U, Safarov N, Valkonen P, Hietapakka L, et al. Towards digital health equity - a qualitative study of the challenges experienced by vulnerable groups in using digital health services in the COVID-19 era. BMC Health Serv Res [Internet]. 2022;22(1):1–12. Available from: https://doi.org/10.1186/s12913-022-07584-4

80. Canadian Community Health Survey (CCHS) - 2019 [Internet]. [cited 2020 Sep 29]. Available from: https://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&lang=en&Item_Id=1207185

81. Ma BD, Ng SL, Schwanen T, Zacharias J, Zhou M, Kawachi I, et al. Pokémon GO and

physical activity in Asia: multilevel study. J Med Internet Res. 2018;20(6):1–12.

82. Rashed EA, Hirata A. Infectivity upsurge by COVID-19 viral variants in japan: Evidence from deep learning modeling. Int J Environ Res Public Health. 2021;18(15).

83. ecobee. ecobee Launches ‘Donate Your Data,’ Connects Customers to Leading Energy Researchers to Address Climate Change [Internet]. [cited 2020 May 31]. Available from: https://www.ecobee.com/en-ca/newsroom/press-releases/ecobee-launches-donate-your-data/

84. Ecobee. Donate your Data | ecobee | Smart Home Technology [Internet]. [cited 2019 Apr 29]. Available from: https://www.ecobee.com/donateyourdata/

85. Kaur J, Sahu KS, Oetomo A, Morita P. A Smart Thermostat-Based Population-Level Behavioural Changes During the COVID-19 Pandemic in the United States: A Proposed Study. DigiBiom 2022 - Proc 2022 Emerg Devices Digit Biomarkers. 2022;7–12.

86. Jalali N, Sahu KS, Oetomo A, Morita PP. Usability of Smart Home Thermostat to Evaluate the Impact of Weekdays and Seasons on Sleep Patterns and Indoor Stay: Observational Study. JMIR mHealth uHealth. 2022;10(4):1–18.

87. Sahu KS, Oetomo A, Morita P. Smart Monitoring of Population Health Risk Behaviour. Proc Int Symp Hum Factors Ergon Heal Care. 2019;8(1):76–80.

88. Lab UHT. Share Your Data [Internet]. 2023. Available from: https://shareyourdata.ubilab.ca/

89. Velmovitsky PE, Alencar P, Leatherdale ST, Cowan D, Morita PP. Towards real-time public health: A novel mobile health monitoring system. In: IEEE International Conference on Big Data (Big Data). IEEE; 2022. p. 6049–51.

90. Velmovitsky PE, Alencar P, Leatherdale ST, Cowan D, Morita PP. A novel mobile platform for stress prediction. In: DigiBiom ’22: Proceedings of the 2022 Workshop on Emerging Devices for Digital Biomarkers. 2022. p. 18–23.

91. Hart TL, Ainsworth BE, Tudor-Locke C. Objective and subjective measures of sedentary behavior and physical activity. Med Sci Sports Exerc. 2011;43(3):449–56.

92. Hagstromer M, Ainsworth BE, Oja P, Sjostrom M. Comparison of a subjective and an objective measure of physical activity in a population sample. J Phys Act Heal. 2010;7(4):541–50.

93. Bunn JA, Navalta JW, Fountaine CJ, Reece JD. Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015-2017. Int J Exerc Sci [Internet]. 2017;11(7):503–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29541338%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5841672
94. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of fitbit flex for step count, moderate to vigorous physical activity and activity energy expenditure. PLoS One. 2016;11(9):1–14.
95. Hernando D, Roca S, Sancho J, Alesanco Á, Bailón R. Validation of the apple watch for heart rate variability measurements during relax and mental stress in healthy subjects. Sensors (Switzerland). 2018;18(8).
96. Roomkham S, Hittle M, Cheung J, Lovell D, Mignot E, Perrin D. Sleep monitoring with the Apple Watch: comparison to a clinically validated actigraph. F1000Research. 2019;8:754.
97. Colvonen PJ, DeYoung PN, Bosompra NOA, Owens RL. Limiting racial disparities and bias for wearable devices in health science research. Sleep. 2020;43(10):1–3.
98. Statista. Wearables : Apple users in Statista Consumer Insights : Global Consumer Survey Brand Report. 2023;(February).
99. Tien NH. International distribution policy comparative analysis between samsung and apple Change management View project International distribution policy comparative analysis between samsung and apple. Int J Res Mark Manag Sales [Internet]. 2019;1(2):1–4. Available from: https://www.researchgate.net/publication/338719419
100. Statista. Apple's global brand value from 2006 to 2022 [Internet]. 2023 [cited 2023 Apr 12]. Available from: https://www.statista.com/statistics/326052/apple-brand-value/#:~:text=Apple's brand value has seen,impressive 947 billion U.S. dollars.
101. Statista. Wearables : Fitbit users in Statista Consumer Insights : Global Consumer Survey Brand Report. 2023;(February).
102. Statista. Fitbit - statistics and facts [Internet]. 2023. Available from: https://www.statista.com/topics/2595/fitbit/#topicOverview
103. Statista. Wearables : Samsung users in Statista Consumer Insights : Global Consumer

Survey Brand Report. 2023;(February).

104. Canada ice of the PC of. Personal Health Information Privacy Laws Across the Provinces: What Works, What Doesn't and What Can we Do About It? [Internet]. 2006 [cited 2023 May 23]. Available from: https://www.priv.gc.ca/en/opc-news/speeches/2006/sp-d_060508_pk/

105. Thorogood A, Simkevitz H, Phillips M, Dove ES, Joly Y. Protecting the Privacy of Canadians' Health Information in the Cloud. Can J Law Technol. 2017;14(1).

106. Personal Information Protection and Electronic Documents Act (PIPEDA). 2000.

107. PIPEDA in brief - Office of the Privacy Commissioner of Canada [Internet]. [cited 2018 Nov 15]. Available from: https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda_brief/

108. Summary of privacy laws in Canada - Office of the Privacy Commissioner of Canada [Internet]. [cited 2020 Jan 23]. Available from: https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/02_05_d_15/#heading-0-0-3-1

109. Government of Ontario. Personal Health Information Protection Act , 2004 : An Overview. Access [Internet]. 2004;(November):1–63. Available from: http://www.health.gov.on.ca/english/providers/legislation/priv_legislation/overview_leg.pdf

110. Cavoukian A. Frequently Asked Questions: Personal Health Information Protection Act. 2004;(December):1–42. Available from: https://www.ipc.on.ca/images/Resources/hfaq-e.pdf%5Cnhttp://www.ipc.on.ca/images/resources/hguide-e.pdf

111. Gandarilla H. Safeguarding Personal Health Information: Case Study. Adv Intell Syst Comput. 2018;738(1):3–6.

112. HIPAA Compliance Checklist [Internet]. [cited 2020 Jan 22]. Available from: https://www.hipaajournal.com/hipaa-compliance-checklist/

113. Velmovitsky PE, Miranda PADSES, Vaillancourt H, Donovska T, Teague J, Morita PP. A Blockchain-Based Consent Platform for Active Assisted Living: Modeling Study and Conceptual Framework. J Med Internet Res [Internet]. 2020 Dec;22(12):e20832. Available from: http://www.ncbi.nlm.nih.gov/pubmed/33275111

114. Micheli M, Ponti M, Craglia M, Suman AB. Emerging models of data governance in the age of datafication: https://doi.org/101177/2053951720948087 [Internet]. 2020;7(2). Available from: https://journals.sagepub.com/doi/full/10.1177/2053951720948087

115. Confessore NC. Cambridge Analytica and Facebook: The Scandal and the Fallout So Far - The New York Times [Internet]. The New York Times. 2018 [cited 2023 May 24]. Available from: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

116. Canada fice of the PC of. Who to contact with concerns about the protection of your personal health information [Internet]. 2017 [cited 2023 May 23]. Available from: https://www.priv.gc.ca/en/privacy-topics/health-genetic-and-other-body-information/02_05_d_70_phi/

117. Canada fice of the PC of. Who to contact with concerns about the protection of your personal health information. 2017.

118. Canada S. Access to the Internet in Canada, 2020 [Internet]. 2021 [cited 2023 Apr 12]. Available from: https://www150.statcan.gc.ca/n1/daily-quotidien/210531/dq210531d-eng.htm

119. Statista. Number of internet users in Canada 2023 | Statista [Internet]. 2023 [cited 2023 Apr 12]. Available from: https://www.statista.com/statistics/243808/number-of-internet-users-in-canada/

120. Faverio M. Share of those 65 and older who are tech users has grown in the past decade. Pew Research Center. 2022.

121. (NCE) A-WN of C of E. COVID-19 has significantly increased the use of many technologies among older Canadians: poll Français [Internet]. Newswire. 2020. Available from: https://www.newswire.ca/news-releases/covid-19-has-significantly-increased-the-use-of-many-technologies-among-older-canadians-poll-865959030.html#:~:text=A new AGE-WELL poll,83%25) use it daily.

122. (NCE) A-WN of C of E. COVID-19 has significantly increased the use of many technologies among older Canadians: poll Français. Newswire. 2020.

123. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health. 2021;21(1):1–12.

124. Canada G of. High-speed Internet for all Canadians. 2023.

125. Latif S, Qadir J, Farooq S, Imran MA. How 5G Wireless (and Concomitant Technologies) will Revolutionize Healthcare? Futur Internet [Internet]. 2017;9(4). Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85037742584&doi=10.3390%2Ffi9040093&partnerID=40&md5=be53347b26cb53c7b7cb64e11226968f

126. Foundation C. What is Public Health? [Internet]. 2019 [cited 2019 Aug 8]. Available from: https://www.cdcfoundation.org/what-public-health

127. Apple. Taking an ECG with the ECG app on Apple Watch Series 4, Series 5, or Series 6 [Internet]. [cited 2022 Aug 26]. Available from: https://support.apple.com/en-ca/HT208955

128. Sala M, Brosof LC, Rosenfield D, Fernandez KC, Levinson CA. Stress is associated with exercise differently among individuals with higher and lower eating disorder symptoms: An ecological momentary assessment study. Int J Eat Disord. 2017;50(12):1413–20.

129. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. UbiComp 2014 - Proc 2014 ACM Int Jt Conf Pervasive Ubiquitous Comput. 2014;3–14.

130. Apple. Swift - Apple Developer [Internet]. [cited 2023 Jan 24]. Available from: https://developer.apple.com/swift/

131. Apple. SwiftUI Overview - Xcode - Apple Developer [Internet]. [cited 2023 Jan 24]. Available from: https://developer.apple.com/xcode/swiftui/

132. McDonough JE. Singleton Design Pattern. In: Object-Oriented Design with ABAP: A Practical Approach [Internet]. Berkeley, CA: Apress; 2017. p. 137–45. Available from: https://doi.org/10.1007/978-1-4842-2838-8_11

133. Apple. HKQuery [Internet]. 2023 [cited 2023 Apr 10]. Available from: https://developer.apple.com/documentation/healthkit/hkquery

134. Apple. HKStatisticsQuery | Apple Developer Documentation [Internet]. 2023 [cited 2023 May 25]. Available from: https://developer.apple.com/documentation/healthkit/hkstatisticsquery

135. Apple. HKSourceQuery | Apple Developer Documentation [Internet]. 2023 [cited 2023 May 25]. Available from: https://developer.apple.com/documentation/healthkit/hksourcequery

136. Apple. HKActivitySummaryQuery | Apple Developer Documentation [Internet]. 2023 [cited 2023 May 25]. Available from: https://developer.apple.com/documentation/healthkit/hkactivitysummaryquery

137. Apple. HKAnchoredObjectQuery | Apple Developer Documentation [Internet]. 2023 [cited 2023 May 25]. Available from: https://developer.apple.com/documentation/healthkit/hkanchoredobjectquery

138. Apple. HKObserverQuery | Apple Developer Documentation [Internet]. 2023 [cited 2023 May 25]. Available from: https://developer.apple.com/documentation/healthkit/hkobserverquery

139. Apple. Background Delivery [Internet]. [cited 2022 Apr 26]. Available from: https://developer.apple.com/documentation/healthkit/hkhealthstore/1614175-enablebackgrounddelivery

140. Acharya UR, Joseph KP, Kannathal N, Lim CM, Suri JS. Heart rate variability: A review. Vol. 44, Medical and Biological Engineering and Computing. 2006. p. 1031–51.

141. Fink G. Stress: The Health Epidemic of the 21st Century [Internet]. SciTech Connect. 2016 [cited 2023 Jan 24]. Available from: https://scitechconnect.elsevier.com/stress-health-epidemic-21st-century/

142. Statista. Percentage of U.S. adults aged 40 years and older who reported having high stress levels in 2020, by age. Statista. 2021.

143. Lowell Um. Financial Costs of Job Stress | Total Worker Health for Employers | CPH-NEW | Research | [Internet]. [cited 2023 Jan 24]. Available from: https://www.uml.edu/research/cph-new/worker/stress-at-work/financial-costs.aspx

144. Goh J, Pfeffer J, Zenios SA. Reducing the health toll from U.S. workplace stress. Behav Sci Policy. 2019;5(1):1–13.

145. Use Apple Watch to practice mindfulness - Apple Support (CA) [Internet]. [cited 2022 Apr 25]. Available from: https://support.apple.com/en-ca/guide/watch/apd371dfe3d7/watchos

146. Baek HJ, Cho CH, Cho J, Woo JM. Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability. Telemed e-Health. 2015 May 1;21(5):404–14.

147. Benchekroun M, Velmovitsky PE, Istrate D, Zalc V, Morita PP, Lenne D. Cross Dataset Analysis for Generalizability of HRV-Based Stress Detection Models. Sensors. 2023;1–15.

148. McCraty R, Atkinson M, Tomasino D, Stuppy WP. Analysis of twenty-four hour heart rate variability in patients with panic disorder. Biol Psychol. 2001;56(2):131–50.

149. Hong SJ, Lee D, Park J, Namkoong K, Lee J, Jang DP, et al. Altered heart rate variability during gameplay in internet gaming disorder: The impact of situations during the game. Front Psychiatry. 2018;9(SEP):1–7.

150. Seipäjärvi SM, Tuomola A, Juurakko J, Rottensteiner M, Rissanen APE, Kurkela JLO, et al. Measuring psychosocial stress with heart rate variability-based methods in different health and age groups. Physiol Meas. 2022;43(5).

151. Martinez GJ, Grover T, Mattingly SM, Mark G, D'Mello S, Aledavood T, et al. Alignment between Heart Rate Variability from Fitness Trackers and Perceived Stress: Perspectives from a Large-Scale in Situ Longitudinal Study of Information Workers. JMIR Hum Factors. 2022;9(3):1–21.

152. Da Silva AGCB, Araujo DN, Costa ACM, Dias BAL, Fregonezi GADF, Dias FAL. Increase in perceived stress is correlated to lower heart rate variability in healthy young subjects. Acta Sci - Heal Sci. 2015;37(1):7–10.

153. Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology. Heart rate variability - Standards of measurement, physiological interpretation, and clinical use. Eur Heart J. 1996;17:354–281.

154. Benchekroun M, Chevallier B, Zalc V, Istrate D, Lenne D, Vera N, et al. Analysis of the Impact of Inter-Beat-Interval Interpolation on real-time HRV Feature Estimation for e-Health Applications [Internet]. Available from: https://hal.archives-ouvertes.fr/hal-03501191

155. Shaffer F, Meehan ZM, Zerr CL. A Critical Review of Ultra-Short-Term Heart Rate Variability Norms Research. Front Neurosci. 2020;14(November):1–11.

156. Munoz ML, Van Roon A, Riese H, Thio C, Oostenbroek E, Westrik I, et al. Validity of (Ultra-)Short recordings for heart rate variability measurements. PLoS One. 2015;10(9):1–15.

157. Seshadri DR, Bittel B, Browsky D, Penny H, Drummond CK, Desai, et al. Accuracy of Apple Watch for Detection of Atrial Fibrillation. Circulation. 2020;141:702–3.

158. Saghir N, Aggarwal A, Soneji N, Valencia V, Rodgers G, Kurian T. A comparison of manual electrocardiographic interval and waveform analysis in lead 1 of 12-lead ECG and Apple Watch ECG: A validation study. Cardiovasc Digit Heal J [Internet]. 2020;1(1):30–6. Available from: https://doi.org/10.1016/j.cvdhj.2020.07.002

159. Shaffer F, Ginsberg JP. An Overview of Heart Rate Variability Metrics and Norms. Front Public Heal. 2017;5(September):1–17.

160. Moskowitz DS, Young SN. Ecological momentary assessment: What it is and why it is a method of the future in clinical psychopharmacology. J Psychiatry Neurosci. 2006;31(1):13–20.

161. Kubios. Kubios User Guide [Internet]. 2022 [cited 2022 Aug 26]. Available from: https://www.kubios.com/downloads/Kubios_HRV_Users_Guide.pdf

162. Motor Accident Insurance Comission. Dass21 questionnaire.

163. Statista. Percentage of Canadians who perceived quite a lot of life stress from 2003 to 2020 [Internet]. 2020 [cited 2022 Aug 10]. Available from: https://www.statista.com/statistics/434081/share-of-canadians-perceiving-their-life-stress-as-quite-a-lot/

164. Statista. Apple watch users worldwide 2020 [Internet]. [cited 2022 Aug 26]. Available from: https://www.statista.com/statistics/1221051/apple-watch-users-worldwide/

165. Morita PP. Design of mobile health technology. In: Design for Health. Elsevier; 2020. p. 87–102.

166. Hovsepian K, Al'absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. CStress: Towards a gold standard for continuous stress assessment in the mobile environment. In: UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015. p. 493–504.

167. Muaremi A, Bexheti A, Gravenhorst F, Arnrich B, Troster G. Monitoring the impact of

stress on the sleep patterns of pilgrims using wearable sensors. 2014 IEEE-EMBS Int Conf Biomed Heal Informatics, BHI 2014. 2014;185–8.

168. Gjoreski M, Gjoreski H, Luštrek M, Gams M. Continuous stress detection using a wrist device - in laboratory and real life. UbiComp 2016 Adjun - Proc 2016 ACM Int Jt Conf Pervasive Ubiquitous Comput. 2016;1185–93.

169. Can YS, Gokay D, Kılıç DR, Ekiz D, Chalabianloo N, Ersoy C. How laboratory experiments can be exploited for monitoring stress in the wild: A bridge between laboratory and daily life. Sensors (Switzerland). 2020;20(3).

170. Apple. Instructions for Use: ECG App [Internet]. [cited 2022 Aug 26]. Available from: https://www.apple.com/legal/privacy/Instructions-for-Use-ECG-App.pdf

171. Kubios. Kubios – Heart rate variability – Kubios HRV [Internet]. [cited 2022 Apr 25]. Available from: https://www.kubios.com/

172. Frankenhaeuser M. Stress and gender. Eur Rev. 1996 Oct;4(4):313–27.

173. Higher stress among minority and low-income populations can lead to health disparities, says report [Internet]. [cited 2022 Sep 22]. Available from: https://www.apa.org/news/press/releases/2018/01/stress-minority-income

174. Lavretsky H, Newhouse PA. Stress, Inflammation, and Aging. Am J Geriatr Psychiatry. 2012 Sep;20(9):729–33.

175. Low income cut-offs (LICOs) before and after tax by community size and family size, in current dollars [Internet]. [cited 2022 Sep 22]. Available from: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1110024101

176. Bauer A, Kantelhardt JW, Barthel P, Schneider R, Mäkikallio T, Ulm K, et al. Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: cohort study. Lancet. 2006;367(9523):1674–81.

177. Bublitz FM, Oetomo A, Sahu KS, Kuang A, Fadrique LX, Velmovitsky PE, et al. Disruptive technologies for environment and health research: An overview of artificial intelligence, blockchain, and internet of things. Int J Environ Res Public Health. 2019;16(20):1–24.

178. Sahu KS, Oetomo A, Mendonca RN, Fadrique L, Morita P. Leveraging Smart-Home Technology for Real-Time Monitoring of Health Behaviour at a Population Level

[Internet]. [cited 2019 Apr 29]. Available from: http://www.healthdatachallenge.com/resource/ubilab-3/

179. Katapally T. The Smart Platform [Internet]. 2019 [cited 2023 Apr 13]. Available from: https://tarunkatapally.ca/the-smart-platform/

180. Tan SY, Yip A. Hans Selye (1907-1982): Founder of the stress theory. Singapore Med J. 2018;59(4):170–1.

181. Akmandor AO, Jha NK. Keep the Stress Away with SoDA: Stress Detection and Alleviation System. IEEE Trans Multi-Scale Comput Syst. 2017;3(4):269–82.

182. Liao CY, Chen RC, Tai SK. Emotion stress detection using EEG signal and deep learning technologies. Proc 4th IEEE Int Conf Appl Syst Innov 2018, ICASI 2018. 2018;(2):90–3.

183. Learn S. sklearn.impute.KNNImputer — [Internet]. 2023 [cited 2023 Apr 13]. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.impute.KNNImputer.html

184. Castaldo R, Montesions L, Melillo P, Massaro S, Pecchia L. To What Extent Can we Shorten HRV Analysis in Wearable Sensing? A Case Study on Mental Stress Detection. In: EMBEC & NBC 2017: Joint Conference of the European Medical and Biological Engineering Conference (EMBEC) and the Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC [Internet]. 2017. Available from: http://link.springer.com/10.1007/978-981-10-5122-7

185. Center LN. Calculating the mean arterial pressure (MAP). 2011.

186. Apple. HealthKit - Sleep. 2023.

187. LaVoy EC, Palmer CA, So C, Alfano CA. Bidirectional relationships between sleep and biomarkers of stress and immunity in youth. Int J Psychophysiol [Internet]. 2020;158:331–9. Available from: https://www.sciencedirect.com/science/article/pii/S0167876020302555

188. Standard Scaler [Internet]. SciKit Learn. 2023 [cited 2023 Jul 14]. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

189. Learn S. SMOTE [Internet]. 2023 [cited 2023 Jul 14]. Available from: https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html

190. Apple. HealthKit - Sleep [Internet]. 2023 [cited 2023 Apr 10]. Available from:

https://developer.apple.com/documentation/healthkit/hkcategoryvaluesleepanalysis

191. Kaelbling LP, Littman ML, Moore AW. Reinforcement Learning: A Survey. J Artif Intell Res. 1996;4:237–85.

192. Apple. Information Property List [Internet]. 2023. Available from: https://developer.apple.com/library/archive/documentation/General/Reference/InfoPlistKeyReference/Articles/AboutInformationPropertyListFiles.html

193. Labs B. BfB Labs [Internet]. 2023 [cited 2023 Apr 10]. Available from: BfB Labs

194. Network UH. Medly - A Companion for Heart Care. 2023; Available from: https://medly.ca/wp-content/uploads/2019/11/Medly-Program-Overview-v1.0.pdf

195. Garcia-Ceja E, Osmani V, Mayora O. Automatic Stress Detection in Working Environments from Smartphones' Accelerometer Data: A First Step. IEEE J Biomed Heal Informatics. 2016;20(4):1053–60.

196. Mate G, M.D. When the Body Says No: The Cost of Hidden Stress. 2011;320.

197. Selye H. The evolution of the stress concept. JIntAcadPrevMed. 1975;2(1):1–21.

198. Low A, McCraty R. Heart Rate Variability: New Perspectives on Assessment of Stress and Health Risk at the Workplace. Hear Mind. 2018;2:16–27.

199. Uusitalo A, Mets T, Martinmäki K, Mauno S, Kinnunen U, Rusko H. Heart rate variability related to effort at work. Appl Ergon [Internet]. 2011;42(6):830–8. Available from: http://dx.doi.org/10.1016/j.apergo.2011.01.005

200. Barutcu Atas D, Aydin Sunbul E, Velioglu A, Tuglular S. The association between perceived stress with sleep quality, insomnia, anxiety and depression in kidney transplant recipients during Covid-19 pandemic. Vol. 16, PloS one. 2021. p. e0248117.

201. Kashani M, Eliasson A, Vernalis M. Perceived stress correlates with disturbed sleep: A link connecting stress and cardiovascular disease. Stress. 2012;15(1):45–51.

202. Apple. Breathe app [Internet]. [cited 2021 Apr 15]. Available from: https://support.apple.com/en-ca/HT206999

203. BankMyCell. How Many Smartphones Are in the World? [Internet]. 2021 [cited 2023 Apr 17]. Available from: https://www.bankmycell.com/blog/how-many-phones-are-in-the-world

204. Choi J, Ahmed B, Gutierrez-Osuna R. Development and evaluation of an ambulatory

stress monitor based on wearable sensors. IEEE Trans Inf Technol Biomed. 2012;16(2):279–86.

205. Sano A, Picard RW. Stress recognition using wearable sensors and mobile phones. Proc - 2013 Hum Assoc Conf Affect Comput Intell Interact ACII 2013. 2013;671–6.

206. Can YS, Chalabianloo N, Ekiz D, Ersoy C. Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. Sensors (Switzerland). 2019;19(8).

207. Vanitha V, Krishnan P. Real time stress detection system based on EEG signals. Biomed Res. 2016;2016(Special Issue 2):S271–5.

208. Sandulescu V, Andrews S, Ellis D, Bellotto N, Mozos OM. Stress detection using wearable physiological sensors. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2015;9107:526–32.

Appendix A – Additional Support Figures

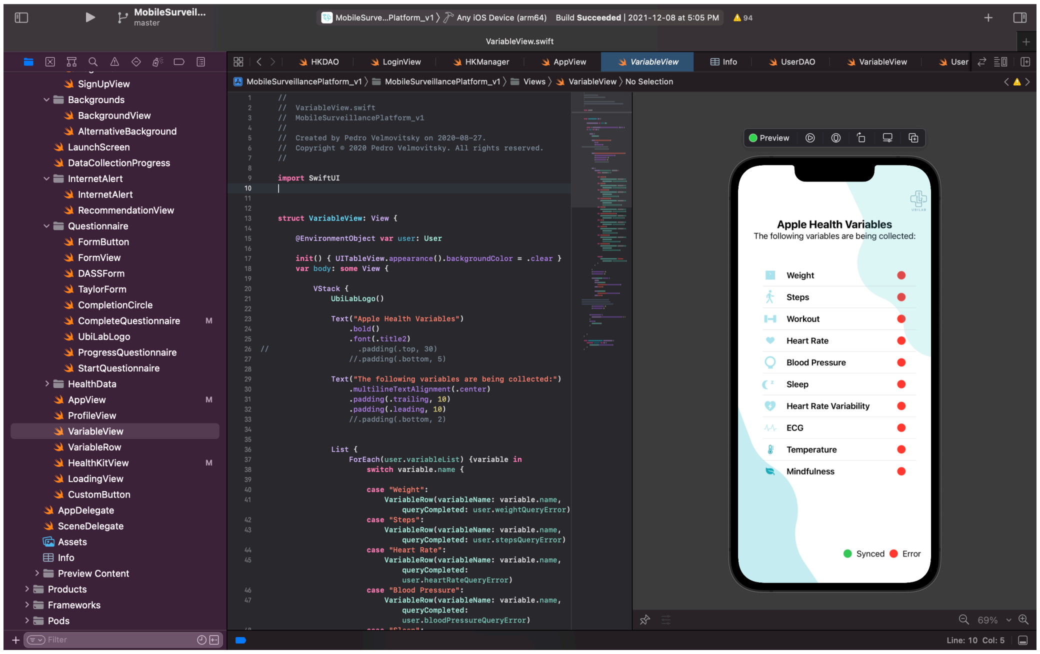


Figure A1: XCode's Interface

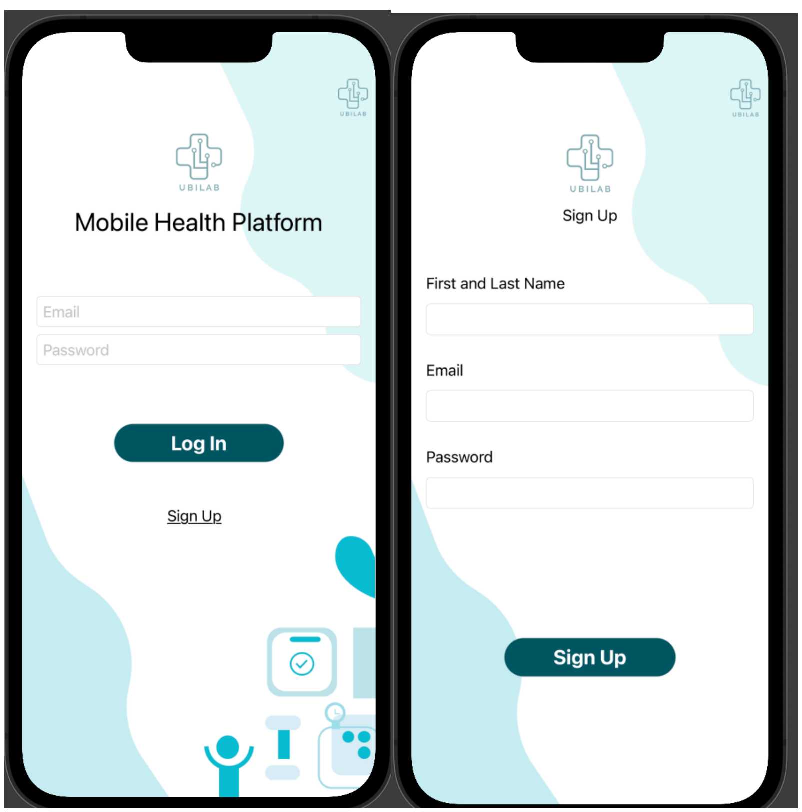


Figure A2: Login and Sign-Up Screens

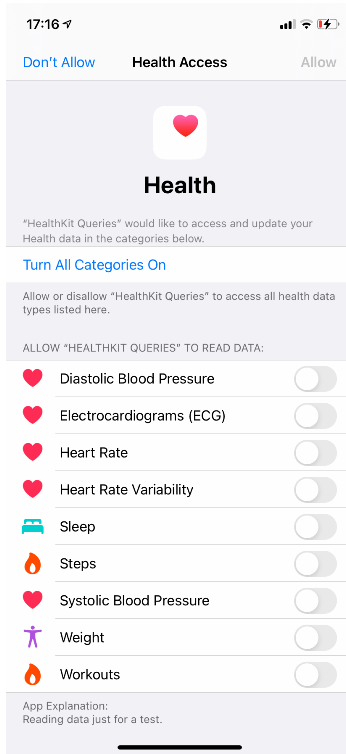


Figure A3: Apple Health Consent Screen

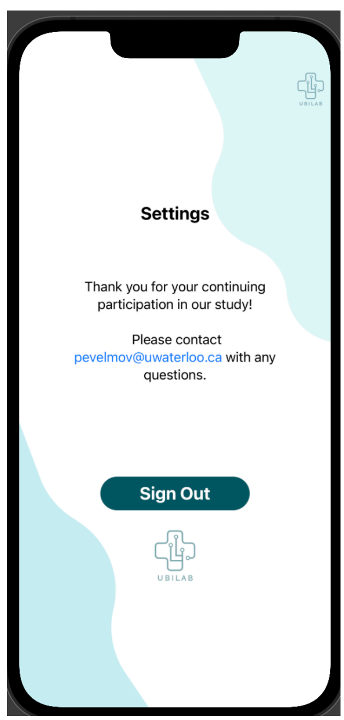


Figure A4: Profile View with Settings

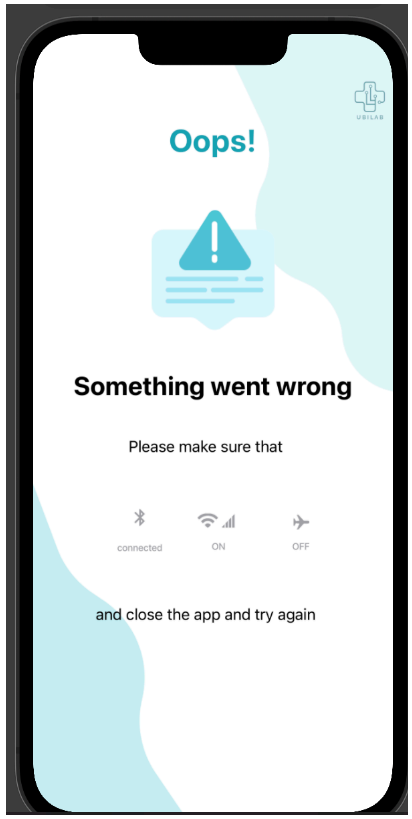


Figure A5: Internet Alert View

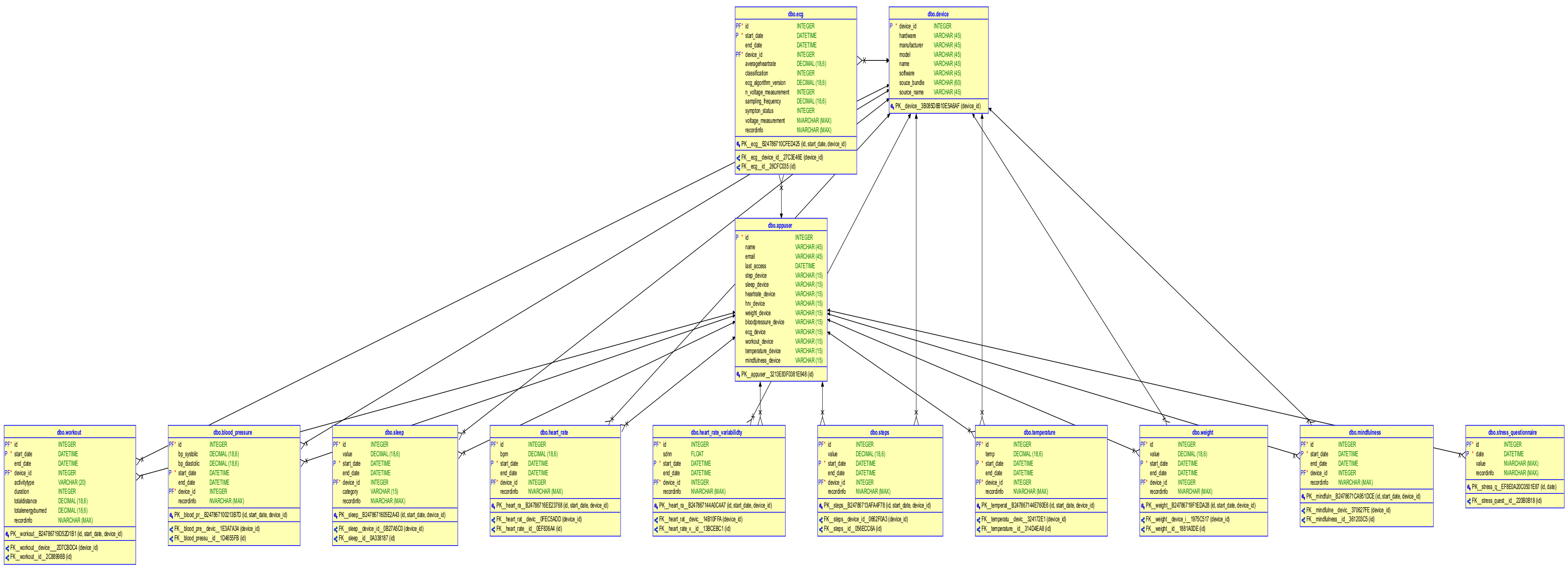


Figure A6: Database Structure

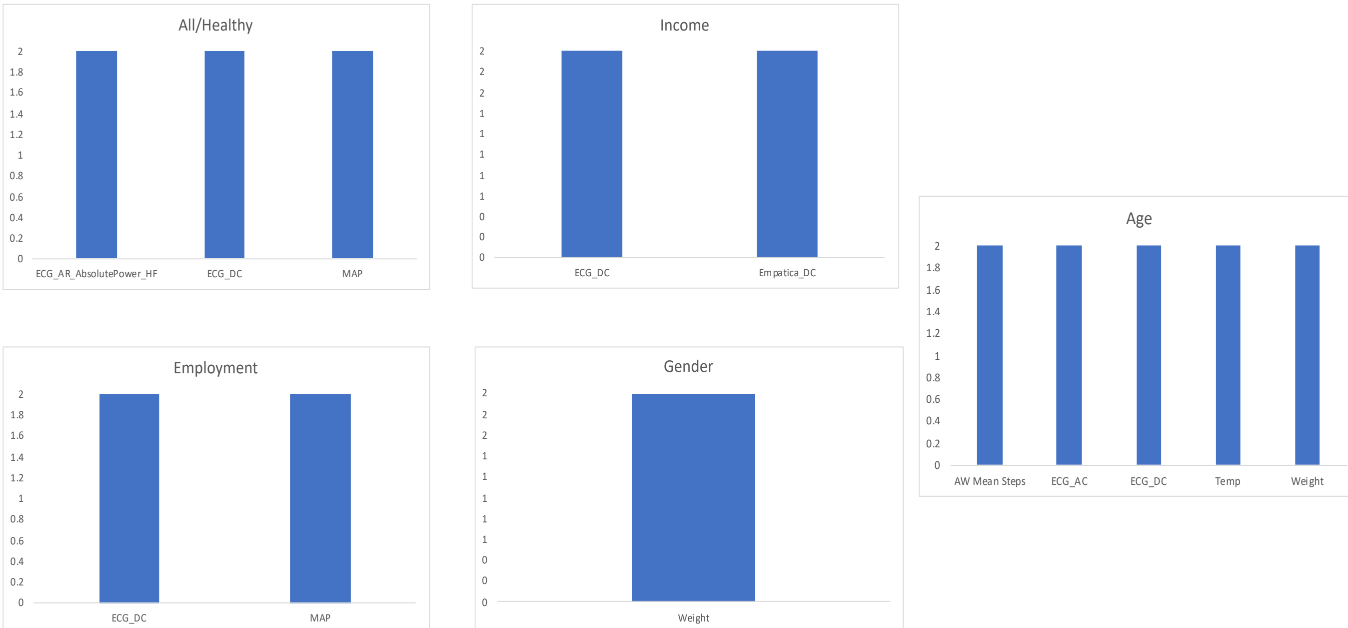


Figure A7: Frequency of top 10 features in each stratification, dataset D

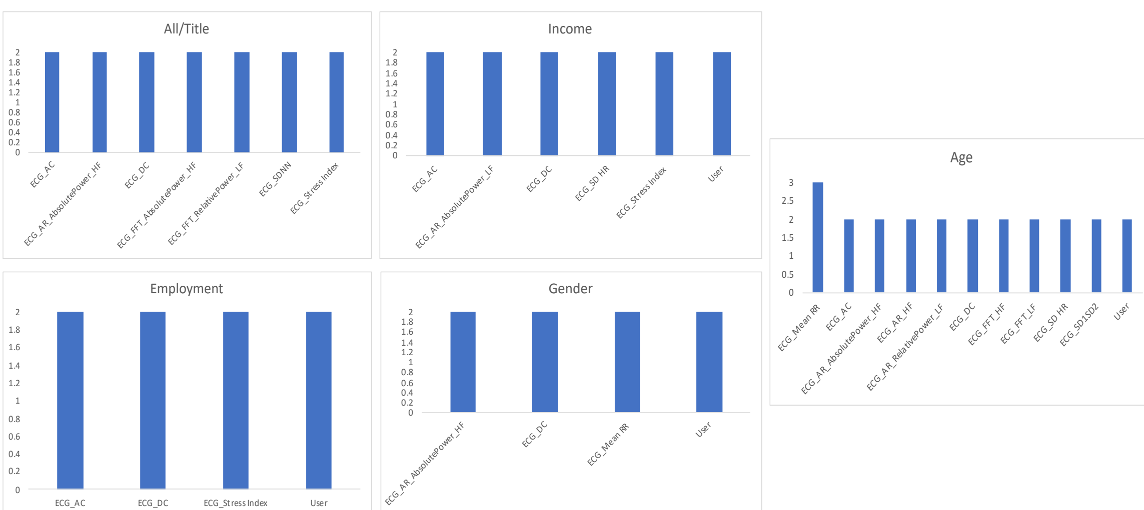


Figure A8: Frequency of top 10 features in each stratification, dataset DECG

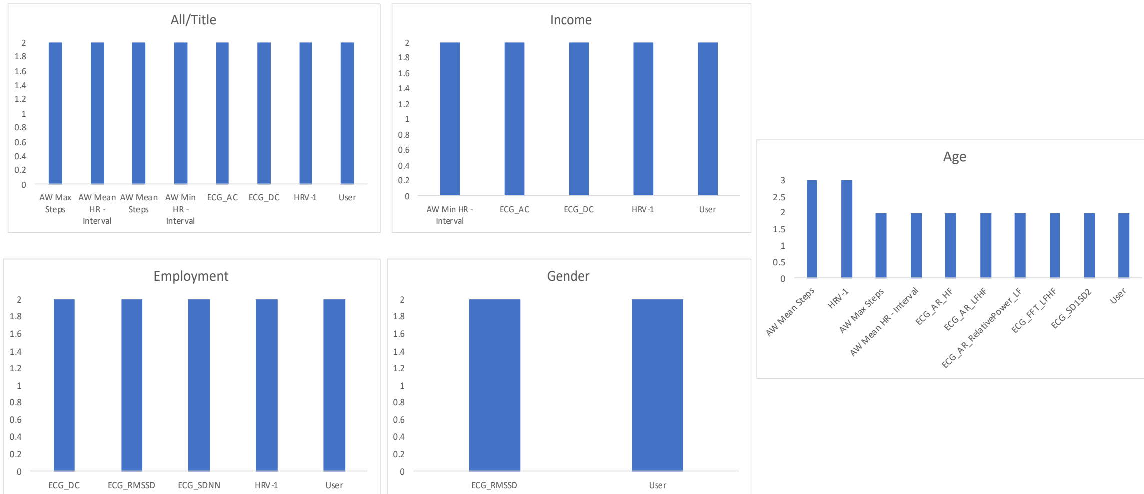


Figure A9: Frequency of top 10 features in each stratification, dataset DA

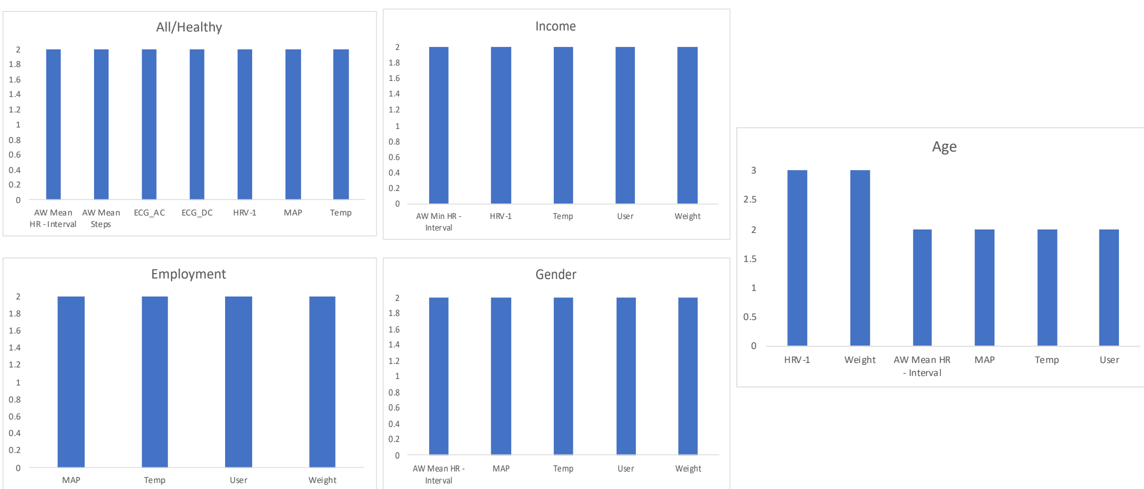


Figure A10: Frequency of top 10 features in each stratification, dataset DAW

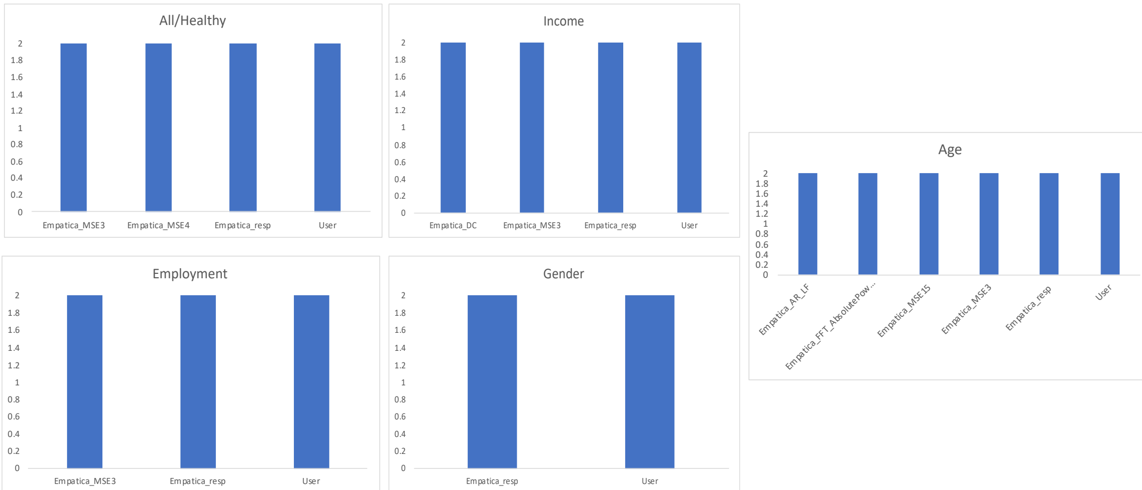


Figure A11: Frequency of top 10 features in each stratification, dataset DEmpatica

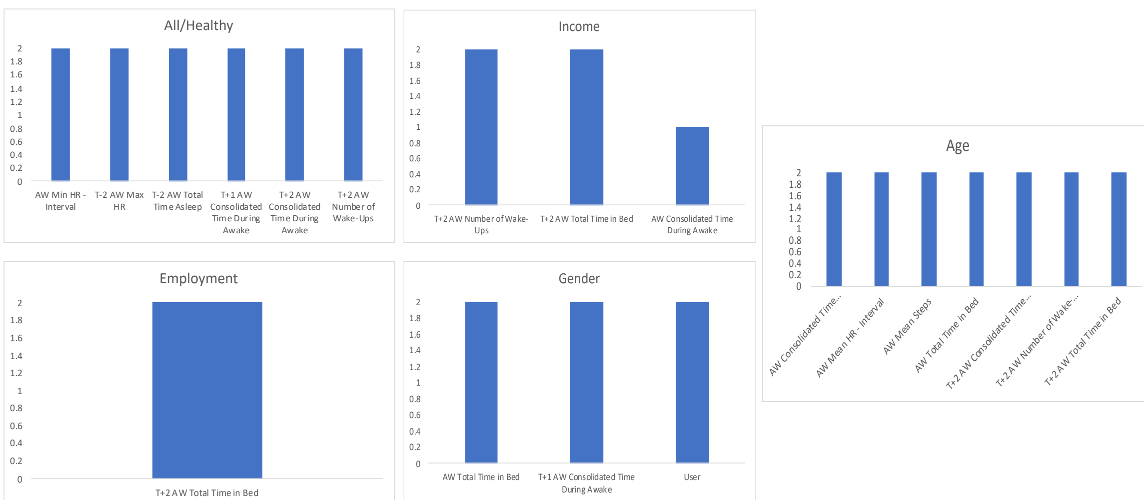


Figure A12: Frequency of top 10 features in each stratification, dataset SDA

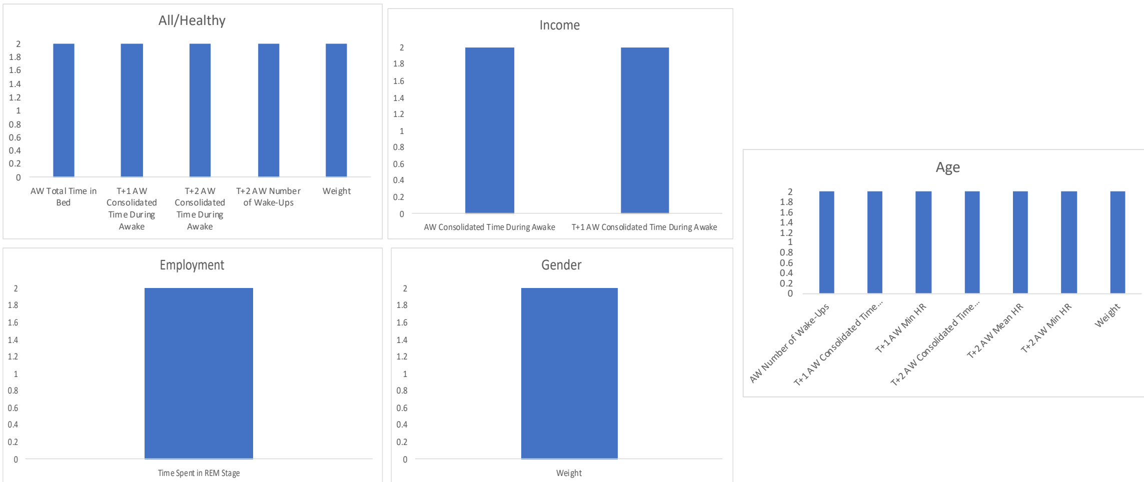


Figure A13: Frequency of top 10 features in each stratification, dataset SDAW

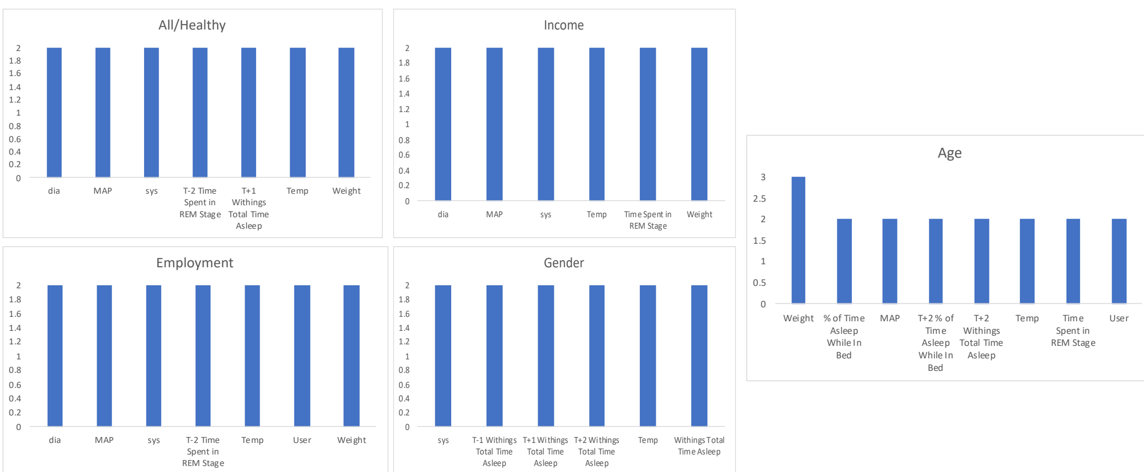


Figure A14: Frequency of top 10 features in each stratification, dataset SDW

Figure A15: Frequency of top 10 features in each stratification, dataset SDS (no features repeated for Gender)

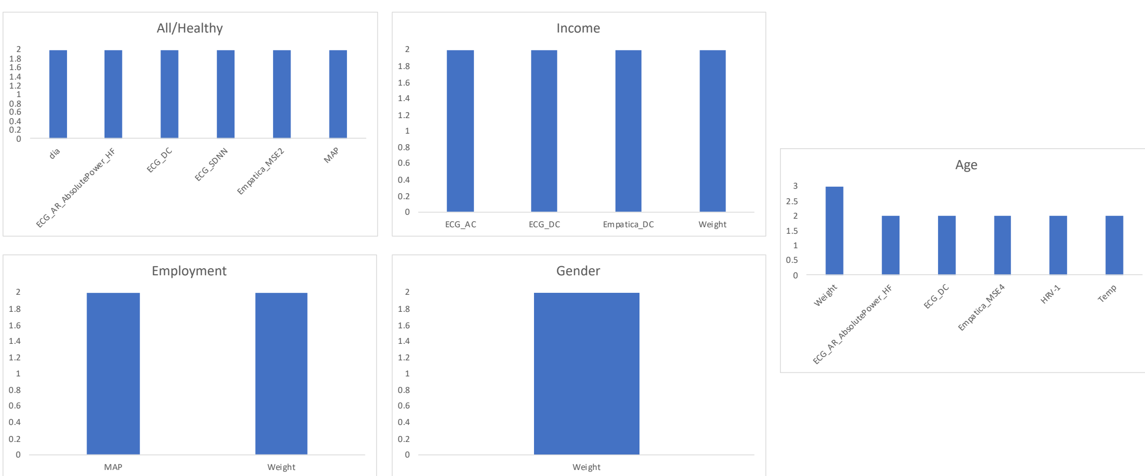


Figure A16: Generalized_Imb - Frequency of top 10 features in each stratification, dataset D

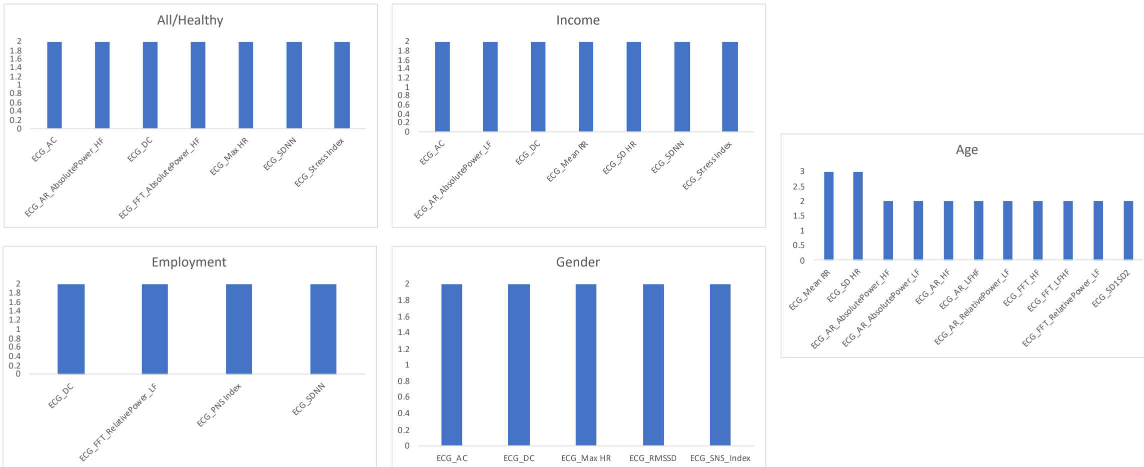


Figure A17: Generalized_Imb - Frequency of top 10 features in each stratification, dataset DECG

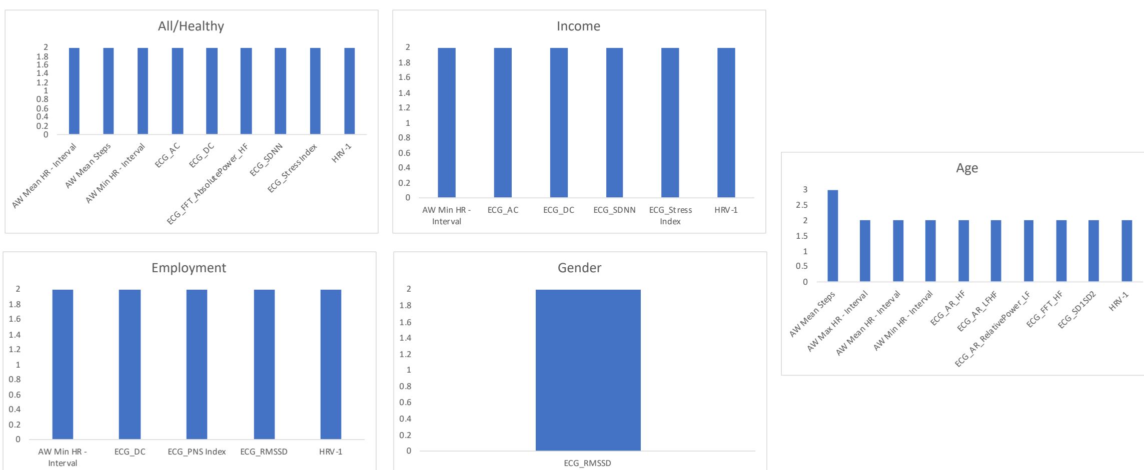


Figure A18: Generalized_Imb - Frequency of top 10 features in each stratification, dataset DA

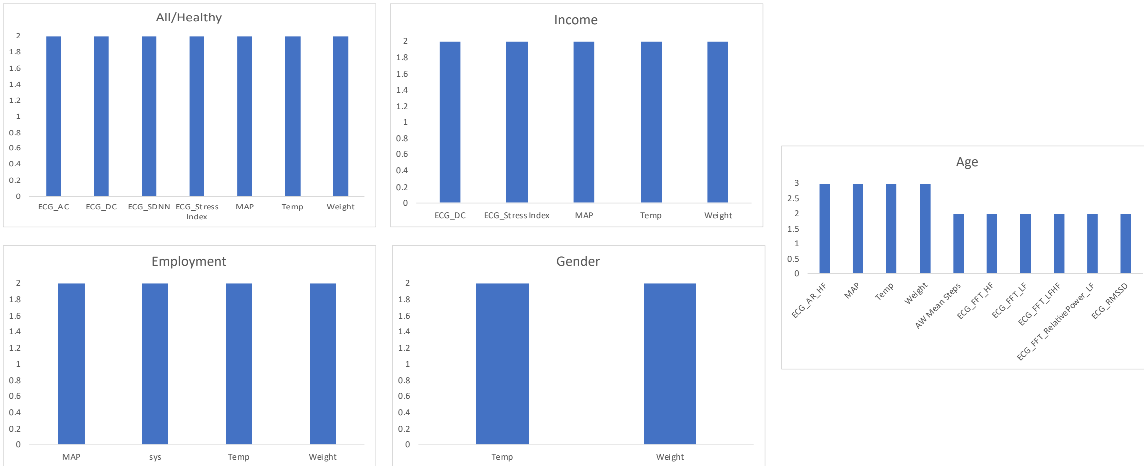


Figure A19: Generalized_Imb - Frequency of top 10 features in each stratification, dataset DAW

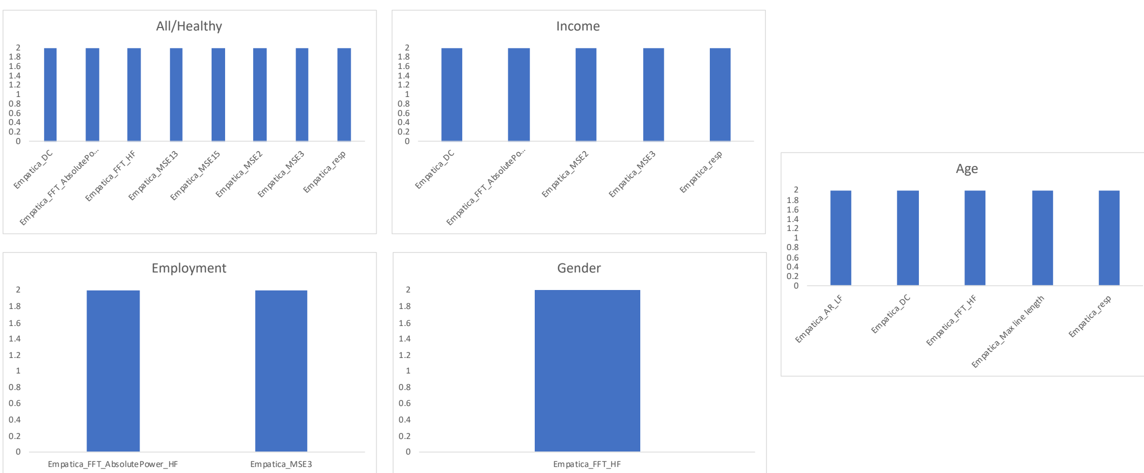


Figure A20: Generalized_Imb - Frequency of top 10 features in each stratification, dataset DEmpatica

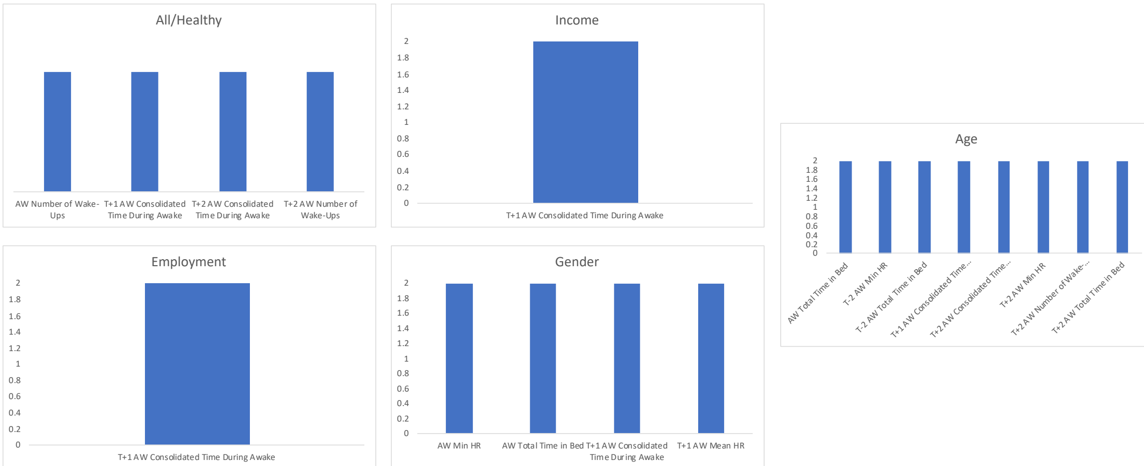


Figure A21: Generalized_Imb - Frequency of top 10 features in each stratification, dataset SDA

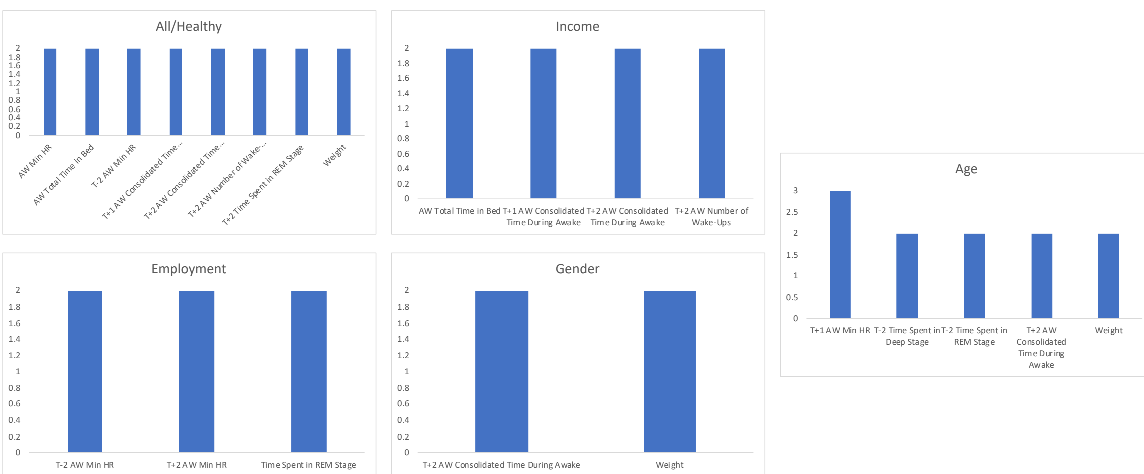


Figure A22: Generalized_Imb - Frequency of top 10 features in each stratification, dataset SDAW

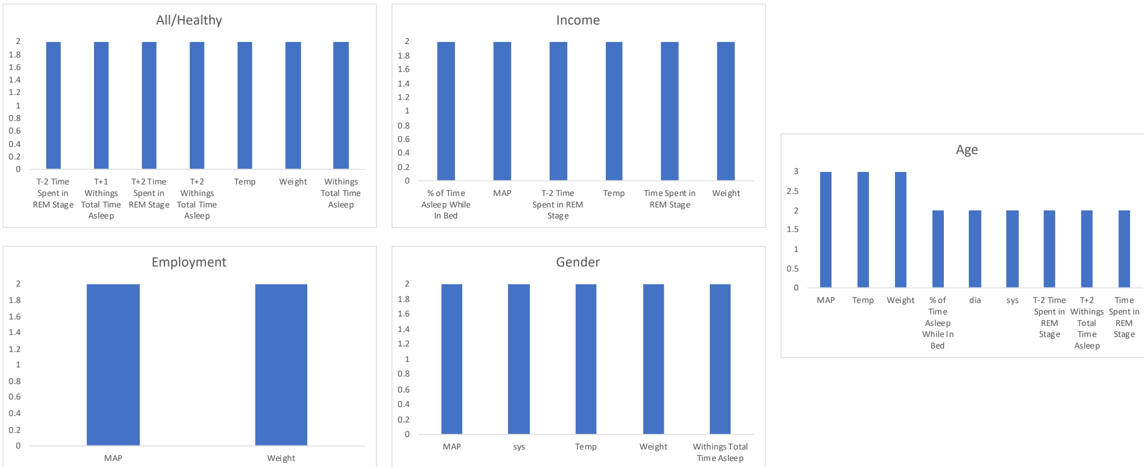


Figure A23: Generalized_Imb - Frequency of top 10 features in each stratification, dataset SDW

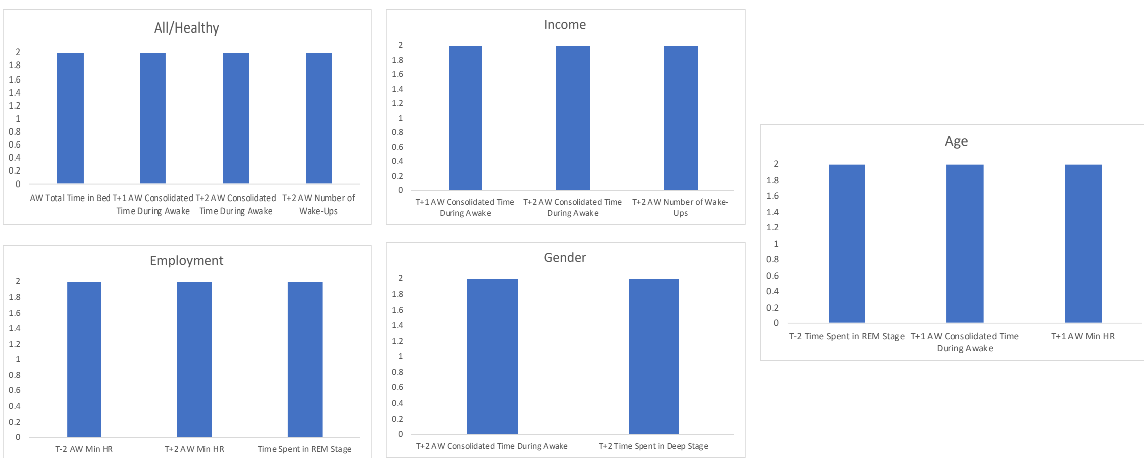


Figure A24: Generalized_Imb - Frequency of top 10 features in each stratification, dataset SDS

DASS-21	DASS-21 by 2	Single-Item	DASS-21 Label	Single-Item Label	stress_score
7	14	3	no stress	stress	stress

Figure A25: Example of Stress Scoring in the Dataset

date	Apple Watch Mean Steps	Apple Watch Max Steps	Apple Watch Min Steps
2022-07-19 17:23	323	73	3
2022-07-20 5:43	902	286	2
2022-07-20 12:02	4336	544	2
2022-07-20 15:01	890	354	2
2022-07-20 18:22	664	192	2
2022-07-20 22:47	289	89	4
2022-07-21 5:57	144	41	2
2022-07-21 10:24	2949	266	3
2022-07-21 12:05	732	158	8
2022-07-21 15:14	4383	960	12
2022-07-21 18:33	868	232	2
2022-07-21 22:43	190	154	11
2022-07-22 6:20	160	35	1
2022-07-22 9:49	3089	863	4
2022-07-22 12:13	149	64	2

Figure A26: Snapshot of Steps Features in Dataset

date	Apple Watch Mean HR	Apple Watch Max HR	Apple Watch Min HR	Short Term Mean	Short Term Max	Short Term Min
2022-03-25 18:12	74.08333333	96	66	67.81818182	81	66
2022-03-26 7:29	74.83991789	130	48	71.19444444	77	68
2022-03-26 10:43	78.96395833	128	65.8329	74.78571429	78	71
2022-03-26 13:58	73.56785714	103	65	75.1	80	71
2022-03-26 17:34	95.57547	130	69	67.13333333	81	61
2022-03-26 21:32	77.50740597	116	61	80.125	116	78
2022-03-27 9:01	69.96650663	83	60	70.85	88	61
2022-03-27 12:13	75.06995488	116	62	66.75	116	60

Figure A27: Snapshot of Apple Watch Heart Rate Features in Dataset

date	HRV-1
2021-12-22 18:33	97.8822632
2021-12-24 11:42	78.0749166
2021-12-24 15:18	58.26757
2021-12-24 18:06	70.4980155
2021-12-24 21:07	82.7284611
2021-12-25 0:11	83.4258374

Figure A28: HRV-1 Feature in Dataset

date	ECG_SDNN	ECG_RMSSD	ECG_DC	ECG_Dcmod	AC	Acmod	ECG_FFT_LF	ECG_FFT_HF
2021-12-22 18:33	54.9752	41.1561	12.1341	29.6926	-40.1031	-52.1308	0.083333	0.15
2021-12-24 11:42	63.2904	54.1436	72.2826	64.2351	-58.9332	-70.0911	0.07	0.16
2021-12-24 15:18	32.8672	33.7018	29.7489	29.4675	-55.0323	-58.6914	0.086667	0.233333
2021-12-24 18:06	24.1772	20.7515	19.4846	21.2471	-20.8294	-25.9608	0.093333	0.21
2021-12-24 21:07	38.9913	38.4	35.6271	39.0835	-42.6263	-51.8975	0.083333	0.333333
2021-12-25 0:11	77.8256	101.1473	112.7574	128.3948	-105.5618	-108.4131	0.053333	0.246667
2021-12-25 11:10	94.6957	89.1805	128.9606	109.2013	-121.3333	-94.085	0.15	0.17
2021-12-25 18:49	62.3767	67.7229	79.7015	82.3133	-89.253	-82.9041	0.076667	0.25
2021-12-25 21:26	19.9138	8.8184	13.5127	11.802	-13.7764	-12.8101	0.073333	0.3

Figure A29: Snapshot of ECG HRV Features in Dataset

date	Empatica_SDNN	Empatica_RMSSD	Empatica_DC	Empatica_Dcmod	Empatica_AC	Empatica_Acmod	Empatica_FFT_VLF
2021-12-24 15:18	54.5943	62.6102	48.5108	66.5029	-54.2499	-72.9754	0.03
2021-12-24 18:06	92.3005	103.7627	123.7033	132.6077	-110.6131	-106.4192	0.04
2021-12-24 21:07	58.5437	62.3022	37.9067	52.8152	-38.3686	-61.6199	0.033333
2021-12-25 0:11	74.3692	87.4479	52.6365	93.042	-45.6713	-83.2	0.04
2021-12-25 11:10	61.2419	69.3109	67.4478	75.2529	-67.4712	-83.2621	0.04
2021-12-25 18:49	36.166	28.9106	24.6657	28.1102	-29.5699	-34.843	0.036667
2021-12-25 21:26	76.8693	74.3898	43.0551	65.7467	-52.2285	-76.1086	0.036667
2021-12-26 0:28	26.4727	28.6203	21.5175	29.6758	-23.6592	-34.6064	0.04
2021-12-26 11:22	49.9801	50.4043	40.4033	52.314	-40.1525	-54.1422	0.033333
2021-12-26 14:59	41.9048	44.2596	40.619	46.3529	-40.2333	-51.2704	0.036667
2021-12-26 19:52	33.68	36.1314	28.4687	38.4335	-28.9325	-39.7022	0.033333
2021-12-26 21:31	46.5444	50.1325	28.9476	49.4571	-31.2287	-52.8229	0.04

Figure A30: Snapshot of Empatica HRV Features in Dataset

date	Temp	Weight	sys	dia	MAP
2021-12-24 15:18	36.863003	70.939336	109	61	97.3333333
2021-12-24 18:06	36.203003	70.819	117	72	111
2021-12-24 21:07	37.201	71.527	122	85	125.666667
2021-12-25 0:11	36.857002	71.256004	113	66	103.666667
2021-12-25 11:10	35.729	69.212006	140	82	128.666667
2021-12-25 18:49	37.088001	70.044006	128	65	107.666667
2021-12-25 21:26	36.922001	70.184006	126	69	111

Figure A31: Snapshot of Temperature, Weight and Blood Pressure Features in Dataset

date	Withings Total Time Asleep (Min)	Apple Watch Total Time Asleep (Min)	Apple Watch Number of Wake-Ups	Apple Watch Consolidated Time During Awake (Min)
2021-12-16 7:58	458	437	11	36
2021-12-16 11:14	458	437	11	36
2021-12-16 14:49	458	437	11	36
2021-12-16 17:38	458	437	11	36
2021-12-16 20:35	458	437	11	36
2021-12-16 23:06	458	437	11	36
2021-12-17 7:44	393	373	11	48
2021-12-17 10:16	393	373	11	48
2021-12-17 14:58	393	373	11	48
2021-12-17 17:44	393	373	11	48
2021-12-17 20:38	393	373	11	48
2021-12-17 22:13	393	373	11	48
2021-12-18 7:55	484	471	8	26
2021-12-18 11:31	484	471	8	26
2021-12-18 14:31	484	471	8	26
2021-12-18 18:46	484	471	8	26
2021-12-18 21:06	484	471	8	26

Figure A32: Snapshot of Sleep Features in Dataset

Appendix B – Additional Support Tables

Table B1: Questionnaire Questions

Questionnaire	Question	Answer Options
DASS-21	I found it hard to wind down;	0 – Not at all 1 – To Some Degree 2 – To a Considerable Degree 3 – Very Much
DASS-21	I felt that I was using a lot of nervous energy;	0 – Not at all 1 – To Some Degree 2 – To a Considerable Degree 3 – Very Much
DASS-21	I found myself getting agitated;	0 – Not at all 1 – To Some Degree 2 – To a Considerable Degree 3 – Very Much
DASS-21	I found it difficult to relax;	0 – Not at all 1 – To Some Degree 2 – To a Considerable Degree 3 – Very Much
DASS-21	I tended to over-react to situations;	0 – Not at all 1 – To Some Degree 2 – To a Considerable Degree 3 – Very Much
DASS-21	I was intolerant of anything that kept me from getting on with what I was doing;	0 – Not at all 1 – To Some Degree 2 – To a Considerable Degree 3 – Very Much
DASS-21	I felt that I was rather touchy;	0 – Not at all 1 – To Some Degree 2 – To a Considerable Degree 3 – Very Much
Single-Item	Right now, I am…	1 – Feeling Great 2 – Feeling Good 3 – A little stressed 4 – Definitely Stressed 5 – Stressed Out

Table B2: Features Used in the Study

Manufacturer	**Variable**	**Feature**	Description (unit)	Dataset
Apple	Steps	Apple Watch Mean Steps	Mean of steps for the time interval	D, DA, DAW, SDA,

				SDAW
Apple	Steps	Apple Watch Max Steps	Maximum of steps for the time interval	D, DA, DAW, SDA, SDAW
Apple	Steps	Apple Watch Min Steps	Minimum HR for the time interval (bpm)	D, DA, DAW, SDA, SDAW
Apple	HR	Apple Watch Mean HR - Interval	Mean HR for the time interval (bpm)	D, DA, DAW, SDA, SDAW
Apple	HR	Apple Watch Max HR – Interval	Maximum HR for the time interval (bpm)	D, DA, DAW, SDA, SDAW
Apple	HR	Apple Watch Min HR - Interval	Minimum of steps for the time interval	D, DA, DAW, SDA, SDAW
Apple	HR	Short Term Mean	Mean HR for the millisecond time interval close to data collection (bpm)	D, DA, DAW, SDA, SDAW
Apple	HR	Short Term Max	Maximum HR for the millisecond time interval close to data collection (bpm)	D, DA, DAW, SDA, SDAW
Apple	HR	Short Term Min	Minimum HR for the millisecond time interval close to data collection (bpm)	D, DA, DAW, SDA, SDAW
Apple	HR	ECG_Mean HR	Mean of heart rate from ECG(ms)	D, DECG, DA, DAW, SDA, SDAW
Apple	HR	ECG_ SD HR	Standard deviation of instantaneous heart rate from ECG (1/min)	D, DECG, DA, DAW, SDA, SDAW
Apple	HR	ECG_Min HR	Minimum instantaneous heart rate calculated using 5 beat moving average from ECG(1/min)	D, DECG, DA, DAW, SDA, SDAW
Apple	HR	ECG_Max HR	Maximum instantaneous heart rate calculated using 5 beat moving average from ECG (1/min)	D, DECG, DA, DAW, SDA, SDAW
Empatica	HR	Empatica_Mean HR	Mean of heart rate from Empatica device (ms)	D, DEmpatica
Empatica	HR	Empatica_ SD HR	Standard deviation of instantaneous heart rate	D, DEmpatica

			from Empatica device (1/min)	
Empatica	HR	Empatica_Min HR	Minimum instantaneous heart rate calculated using 5 beat moving average from Empatica device (1/min)	D, DEmpatica
Empatica	HR	Empatica_Max HR	Maximum instantaneous heart rate calculated using 5 beat moving average from Empatica device (1/min)	D, DEmpatica
Apple	HRV	HRV-1	Heart rate variability collected as SDNN with the Apple Watch	D, DA, DAW, SDA, SDAW
Apple	HRV	ECG_PNS Index	Parasympathetic nervous system activity compared to normal resting values	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_SNS Index	Sympathetic nervous system activity compared to normal resting values	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_Stress Index	Square root of Baevsky's stress index	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_Mean RR	Mean of R-R intervals (ms)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_SDNN	Standard deviation of R-R intervals (ms)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_RMSSD	Square root of the mean squared differences between successive RR intervals f(ms)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_DC	Heart rate deceleration capacity (ms)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_DCMod	Modified DC computer as a two-point difference (ms)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_AC	Heart rate acceleration capacity (ms)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_ACMod	Modified AC computer as a two-point difference (ms)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_FFT LF	Fast Fourier Transform Low Frequency band components	D, DECG, DA, DAW,

			(Hz)	SDA, SDAW
Apple	HRV	ECG_FFT HF	Fast Fourier Transform High Frequency band components (Hz)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_AR LF	Autoregressive Low Frequency band components (Hz)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_AR HF	Autoregressive High Frequency band components (Hz)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_FFT Absolute Power LF	Fast Fourier Transform Absolute Power of Low Frequency band components (ms2)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_FFT Absolute Power HF	Fast Fourier Transform Absolute Power of High Frequency band components (ms2)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_AR Absolute Power LF	Autoregressive Absolute Power of Low Frequency band components (ms2)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_AR Absolute Power HF	Autoregressive Absolute Power of High Frequency band components (ms2)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_FFT Relative Power LF	Fast Fourier Transform Relative Power of Low Frequency band components (%)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_FFT Relative Power HF	Fast Fourier Transform Relative Power of High Frequency band components (%)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_AR Relative Power LF	Autoregressive Relative Power of Low Frequency band components (%)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_AR Relative Power HF	Autoregressive Relative Power of High Frequency band components (%)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_FFT Normalized Power LF	Fast Fourier Transform Normalized Power of Low Frequency band components (n.u)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_FFT Normalized Power HF	Fast Fourier Transform Normalized Power of High Frequency band components (n.u)	D, DECG, DA, DAW, SDA, SDAW

Apple	HRV	ECG_FFT Total Power	Fast Fourier Transform Total Power (ms2)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_FFT LF/HF	Fast Fourier Transform ratio between low and high frequency	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_AR Normalized Power LF	Autoregressive Normalized Power of Low Frequency band components (n.u)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_AR Normalized Power HF	Autoregressive Normalized Power of High Frequency band components (n.u)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_AR Total Power	Autoregressive Total Power (ms2)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_AR LF/HF	Autoregressive ratio between low and high frequency	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_SD1	The standard deviation perpendicular to the line-of-identity in Poincaré plot (ms)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_SD2	The standard deviation along the line-of-identity in Poincaré plot (ms)	D, DECG, DA, DAW, SDA, SDAW
Apple	HRV	ECG_SD2/SD1	Ratio between SD2 and SD1 (ms)	D, DECG, DA, DAW, SDA, SDAW
Empatica	HRV	Empatica_PNS Index	Parasympathetic nervous system activity compared to normal resting values	D, DEmpatica
Empatica	HRV	Empatica_SNS Index	Sympathetic nervous system activity compared to normal resting values	D, DEmpatica
Empatica	HRV	Empatica_Stress Index	Square root of Baevsky's stress index	D, DEmpatica
Empatica	HRV	Empatica_Mean RR	Mean of R-R intervals (ms)	D, DEmpatica
Empatica	HRV	Empatica_SDNN	Standard deviation of R-R intervals (ms)	D, DEmpatica
Empatica	HRV	Empatica_RMSSD	Square root of the mean squared differences between successive RR intervals (ms)	D, DEmpatica
Empatica	HRV	Empatica_DC	Heart rate deceleration capacity (ms)	D, DEmpatica
Empatica	HRV	Empatica_DCMod	Modified DC computer as a	D,

			two-point difference (ms)	DEmpatica
Empatica	HRV	Empatica_AC	Heart rate acceleration capacity (ms)	D, DEmpatica
Empatica	HRV	Empatica_ACMod	Modified AC computer as a two-point difference (ms)	D, DEmpatica
Empatica	HRV	Empatica_FFT VLF	Fast Fourier Transform Very Low Frequency band components (Hz)	D, DEmpatica
Empatica	HRV	Empatica_FFT LF	Fast Fourier Transform Low Frequency band components (Hz)	D, DEmpatica
Empatica	HRV	Empatica_FFT HF	Fast Fourier Transform High Frequency band components (Hz)	D, DEmpatica
Empatica	HRV	Empatica_AR_VLF	Autoregressive Very Low Frequency band components (Hz)	D, DEmpatica
Empatica	HRV	Empatica_AR LF	Autoregressive Low Frequency band components (Hz)	D, DEmpatica
Empatica	HRV	Empatica_AR HF	Autoregressive High Frequency band components (Hz)	D, DEmpatica
Empatica	HRV	Empatica_FFT Absolute Power VLF	Fast Fourier Transform Absolute Power of Very Low Frequency band components (ms2)	D, DEmpatica
Empatica	HRV	Empatica_FFT Absolute Power LF	Fast Fourier Transform Absolute Power of Low Frequency band components (ms2)	D, DEmpatica
Empatica	HRV	Empatica_FFT Absolute Power HF	Fast Fourier Transform Absolute Power of High Frequency band components (ms2)	D, DEmpatica
Empatica	HRV	Empatica_AR Absolute Power VLF	Autoregressive Absolute Power of Very Low Frequency band components (ms2)	D, DEmpatica
Empatica	HRV	Empatica_AR Absolute Power LF	Autoregressive Absolute Power of Low Frequency band components (ms2)	D, DEmpatica
Empatica	HRV	Empatica_AR Absolute Power HF	Autoregressive Absolute Power of High Frequency band components (ms2)	D, DEmpatica
Empatica	HRV	Empatica_FFT	Fast Fourier Transform	D,

		Relative Power VLF	Relative Power of Very Low Frequency band components (%)	DEmpatica
Empatica	HRV	Empatica_FFT Relative Power LF	Fast Fourier Transform Relative Power of Low Frequency band components (%)	D, DEmpatica
Empatica	HRV	Empatica_FFT Relative Power HF	Fast Fourier Transform Relative Power of High Frequency band components (%)	D, DEmpatica
Empatica	HRV	Empatica_AR Relative Power VLF	Autoregressive Relative Power of Very Low Frequency band components (%)	D, DEmpatica
Empatica	HRV	Empatica_AR Relative Power LF	Autoregressive Relative Power of Low Frequency band components (%)	D, DEmpatica
Empatica	HRV	Empatica_AR Relative Power HF	Autoregressive Relative Power of High Frequency band components (%)	D, DEmpatica
Empatica	HRV	Empatica_FFT Normalized Power LF	Fast Fourier Transform Normalized Power of Low Frequency band components (n.u)	D, DEmpatica
Empatica	HRV	Empatica_FFT Normalized Power HF	Fast Fourier Transform Normalized Power of High Frequency band components (n.u)	D, DEmpatica
Empatica	HRV	Empatica_FFT Total Power	Fast Fourier Transform Total Power (ms2)	D, DEmpatica
Empatica	HRV	Empatica_FFT LF/HF	Fast Fourier Transform ratio between low and high frequency	D, DEmpatica
Empatica	HRV	Empatica_resp	Respiration rate (Hz)	D, DEmpatica
Empatica	HRV	Empatica_AR Normalized Power LF	Autoregressive Normalized Power of Low Frequency band components (n.u)	D, DEmpatica
Empatica	HRV	Empatica_AR Normalized Power HF	Autoregressive Normalized Power of High Frequency band components (n.u)	D, DEmpatica
Empatica	HRV	Empatica_AR Total Power	Autoregressive Total Power (ms2)	D, DEmpatica
Empatica	HRV	Empatica_AR LF/HF	Autoregressive ratio between low and high	D, DEmpatica

			frequency	
Empatica	HRV	Empatica_SD1	The standard deviation perpendicular to the line-of-identity in Poincaré plot (ms)	D, DEmpatica
Empatica	HRV	Empatica_SD2	The standard deviation along the line-of-identity in Poincaré plot (ms)	D, DEmpatica
Empatica	HRV	Empatica_SD2/SD1	Ratio between SD2 and SD1 (ms)	D, DEmpatica
Empatica	HRV	Empatica_ApEn	Approximate entropy	D, DEmpatica
Empatica	HRV	Empatica_SampEn	Sample entropy	D, DEmpatica
Empatica	HRV	Empatica_alpha1	In detrended fluctuation, short term fluctuation slope	D, DEmpatica
Empatica	HRV	Empatica_alpha2	In detrended fluctuation, long term fluctuation slope	D, DEmpatica
Empatica	HRV	Empatica_D2	Correlation dimension	D, DEmpatica
Empatica	HRV	Empatica_Mean line length	Mean line length of the recurrent plot analysis	D, DEmpatica
Empatica	HRV	Empatica_Max line length	Max line length of the recurrent plot analysis	D, DEmpatica
Empatica	HRV	Empatica_REC	Recurrence rate of the recurrent plot analysis	D, DEmpatica
Empatica	HRV	Empatica_DET	Determinism of the recurrent plot analysis	D, DEmpatica
Empatica	HRV	Empatica_Shannon	Shannon entropy of the recurrent plot analysis	D, DEmpatica
Empatica	HRV	Empatica_MSE1	Multiscale entropy for scale factor 1	D, DEmpatica
Empatica	HRV	Empatica_MSE2	Multiscale entropy for scale factor 2	D, DEmpatica
Empatica	HRV	Empatica_MSE3	Multiscale entropy for scale factor 3	D, DEmpatica
Empatica	HRV	Empatica_MSE4	Multiscale entropy for scale factor 4	D, DEmpatica
Empatica	HRV	Empatica_MSE5	Multiscale entropy for scale factor 5	D, DEmpatica
Empatica	HRV	Empatica_MSE6	Multiscale entropy for scale factor 6	D, DEmpatica
Empatica	HRV	Empatica_MSE7	Multiscale entropy for scale factor 7	D, DEmpatica
Empatica	HRV	Empatica_MSE8	Multiscale entropy for scale factor 8	D, DEmpatica

Empatica	HRV	Empatica_MSE9	Multiscale entropy for scale factor 9	D, DEmpatica
Empatica	HRV	Empatica_MSE10	Multiscale entropy for scale factor 10	D, DEmpatica
Empatica	HRV	Empatica_MSE11	Multiscale entropy for scale factor 11	D, DEmpatica
Empatica	HRV	Empatica_MSE12	Multiscale entropy for scale factor 12	D, DEmpatica
Empatica	HRV	Empatica_MSE13	Multiscale entropy for scale factor 13	D, DEmpatica
Empatica	HRV	Empatica_MSE14	Multiscale entropy for scale factor 14	D, DEmpatica
Empatica	HRV	Empatica_MSE15	Multiscale entropy for scale factor 15	D, DEmpatica
Empatica	HRV	Empatica_MSE16	Multiscale entropy for scale factor 16	D, DEmpatica
Empatica	HRV	Empatica_MSE17	Multiscale entropy for scale factor 17	D, DEmpatica
Empatica	HRV	Empatica_MSE18	Multiscale entropy for scale factor 18	D, DEmpatica
Empatica	HRV	Empatica_MSE19	Multiscale entropy for scale factor 19	D, DEmpatica
Empatica	HRV	Empatica_MSE20	Multiscale entropy for scale factor 20	D, DEmpatica
Withings	Temperature	Temp	Temperature (Celsius)	DAW, DW, SDAW, SDW
Withings	Weight	Weight	Weight (kg)	DAW, DW, SDAW, SDW
Withings	Systolic blood pressure	sys	Systolic Blood Pressure (mmHg)	DAW, DW, SDAW, SDW
Withings	Diastolic blood pressure	dia	Diastolic Blood Pressure (mmHg)	DAW, DW, SDAW, SDW
Withings	Mean arterial pressure	MAP	Mean Arterial Pressure calculated as (sys + 3* dys)/3 (mmHg)	DAW, DW, SDAW, SDW
Apple	Sleep/Steps	Apple Watch Mean Steps (also included offset by: *t+2, t-2*)	Mean of steps for the night from last measure of previous day to first measure of day	SDA, SDAW, SDS
Apple	Sleep/Steps	Apple Watch Max Steps (also included offset by: *t+2, t-2*)	Maximum of steps from last measure of previous day to first measure of day	SDA, SDAW, SDS

Apple	Sleep/Steps	Apple Watch Min Steps (also included offset by: *t+2, t-2*)	Minimum HR for the time interval from last measure of previous day to first measure of day (bpm)	SDA, SDAW, SDS
Apple	Sleep/HR	Apple Watch Mean HR – Interval (also included offset by: *t+2, t-2, t+1, t-1*)	Mean HR from last measure of previous day to first measure of day (bpm)	SDA, SDAW, SDS
Apple	Sleep/HR	Apple Watch Max HR – Interval (also included offset by: *t+2, t-2, t+1, t-1*)	Maximum HR from last measure of previous day to first measure of day (bpm)	SDA, SDAW, SDS
Apple	Sleep/HR	Apple Watch Min HR – Interval (also included offset by: *t+2, t-2, t+1, t-1*)	Minimum of steps from last measure of previous day to first measure of day (bpm)	SDA, SDAW, SDS
Apple	Sleep	Apple Watch Total Time Asleep (also included offset by: *t+2, t-2*)	Total time asleep calculated with the Apple Watch (min)	SDA, SDAW, SDS
Apple	Sleep	Apple Watch Number of Wake-Ups (also included offset by: *t+2, t-2, t+1*)	Number of wake-ups in the night calculated with the Apple Watch	SDA, SDAW, SDS
Apple	Sleep	Apple Watch Consolidated Time Awake During Sleep (also included offset by: *t+2, t-2, t+1*)	Aggregated time duration of wake-ups (min) calculated with the Apple Watch	SDA, SDAW, SDS
Apple	Sleep	Apple Watch Total Time In Bed (also included offset by: *t+2, t-2, t-1*)	Total time spent in bed, awake or asleep (min) calculated with the Apple Watch	SDA, SDAW, SDS
Apple	Sleep	Apple Watch % of Time Asleep While In Bed (also included offset by: *t+2, t-2*)	Percentage of time spent asleep compared to total time in bed calculated with the Apple Watch	SDA, SDAW, SDS
Withings	Sleep	Withings Total Time Asleep (also included offset by: *t+2, t-2, t+1, t-1*)	Total time asleep calculated with Withings Sleep (min)	SDW, SDAW, SDS
Withings	Sleep	Withings Number of Wake-Ups (also included offset by:	Number of wake-ups in the night calculated with the Withings Sleep	SDW, SDAW, SDS

		t+2, t-2)		
Withings	Sleep	Withings Consolidated Time Awake During Sleep (also included offset by: *t+2, t-2*)	Aggregated time duration of wake-ups calculated with the Withings Sleep (min)	SDW, SDAW, SDS
Withings	Sleep	Withings Total Time In Bed (also included offset by: *t+2, t-2*)	Total time spent in bed, awake or asleep calculated with the Apple Watch (min)	SDW, SDAW, SDS
Withings	Sleep	Withings % of Time Asleep While In Bed (also included offset by: *t+2, t-2*)	Percentage of time spent asleep compared to total time in bed calculated with the Withings Sleep (min)	SDW, SDAW, SDS
Withings	Sleep	Total Time Spent in Light Stage (also included offset by: *t+2, t-2*)	Time spent in light sleep stage calculated with Withings Sleep (min)	SDW, SDAW, SDS
Withings	Sleep	Total Time Spent in Deep Stage (also included offset by: *t+2, t-2*)	Time spent in deep sleep stage calculated with Withings Sleep (min)	SDW, SDAW, SDS
Withings	Sleep	Total Time Spent in REM Stage (also included offset by: *t+2, t-2*)	Time spent in REM sleep stage calculated with Withings Sleep (min)	SDW, SDAW, SDS

Table B3: Participant Characteristics in Each Dataset

<u>Dataset with All Features (D), N = 22</u>

Participants	*Frequency*	*Percentage*
Age		
18-24	5	23.
25-34	9	41
35-44	5	23
Sex/Gender		
Male	7	32
Female	14	64
Gender Fluid	1	5
SES		
Low (0-$30,000)	12	54
Medium ($30,000– $100,000)	10	46
High (Above $100,000)	0	0
Do not wish to disclose	0	0

Profession		
Full-time	8	36
Part-time	1	5
Student	11	50
Self-employed/Other	1	5
Retired	1	5
Health Status		
Healthy	16	73
Chronic Disease or Illness, Prescription Drug Use, Smoking or Alcohol	6	27

<u>Dataset with Only ECG Features (DECG) / Dataset with Only Apple Features (DA), N = 42</u>

Age		
18-24	12	29
25-34	13	31
35-44	12	29
Sex/Gender		
Male	13	31
Female	28	67
Gender Fluid	1	2
SES		
Low (0-$30,000)	19	45
Medium ($30,000– $100,000)	6	38
High (Above $100,000)	4	10
Do not wish to disclose	3	7
Profession		
Full-time	20	48
Part-time	3	7
Student	16	38
Self-employed/Other	2	4
Retired	1	2
Health Status		
Healthy	34	81
Chronic Disease or Illness, Prescription Drug Use, Smoking or Alcohol	8	19

Dataset with Apple and Withings Features (DAW), N = 41

Age		
18-24	12	29
25-34	13	32
35-44	11	27
Sex/Gender		
Male	12	29
Female	28	68
Gender Fluid	1	2
SES		
Low (0-$30,000)	19	46
Medium ($30,000– $100,000)	15	37
High (Above $100,000)	4	10
Do not wish to disclose	3	7
Profession		
Full-time	19	46
Part-time	3	7
Student	16	39
Self-employed/Other	2	4
Retired	1	2
Health Status		
Healthy	33	80
Chronic Disease or Illness, Prescription Drug Use, Smoking or Alcohol	8	20

Dataset with Only Withings Features (DW), N = 44

Age		
18-24	13	30
25-34	14	32
35-44	11	25
Sex/Gender		
Male	13	30
Female	30	68
Gender Fluid	1	2
SES		

Low (0-$30,000)	20	45
Medium ($30,000– $100,000)	17	39
High (Above $100,000)	4	9
Do not wish to disclose	3	7
Profession		
Full-time	20	45
Part-time	5	11
Student	16	36
Self-employed/Other	2	5
Retired	1	3
Health Status		
Healthy	35	80
Chronic Disease or Illness, Prescription Drug Use, Smoking or Alcohol	9	20

Dataset with Only Empatica Features (DEmpatica), N = 27

Age		
18-24	7	26
25-34	10	37
35-44	6	22
Sex/Gender		
Male	8	30
Female	18	67
Gender Fluid	1	4
SES		
Low (0-$30,000)	14	52
Medium ($30,000– $100,000)	13	48
High (Above $100,000)	0	0
Do not wish to disclose	0	0
Profession		
Full-time	9	33
Part-time	4	15
Student	11	41
Self-employed/Other	2	8
Retired	1	4
Health Status		
Healthy	20	74
Chronic Disease or Illness,	7	26

Prescription Drug Use, Smoking or Alcohol		

<u>Sleep Dataset with only Apple Features (SDA), N = 34</u>

Age		
18-24	11	32
25-34	11	32
35-44	9	26
Sex/Gender		
Male	10	29
Female	23	68
Gender Fluid	1	3
SES		
Low (0-$30,000)	17	50
Medium ($30,000– $100,000)	14	41
High (Above $100,000)	0	0
Do not wish to disclose	3	9
Profession		
Full-time	15	44
Part-time	1	3
Student	16	47
Self-employed/Other	2	6
Retired	0	0
Health Status		
Healthy	28	82
Chronic Disease or Illness, Prescription Drug Use, Smoking or Alcohol	7	18

<u>Sleep Dataset with Apple and Withings Features (SDAW), N = 27</u>

Age		
18-24	8	30
25-34	9	33
35-44	7	26
Sex/Gender		
Male	6	22

Female	20	74
Gender Fluid	1	4
SES		
Low (0-$30,000)	13	48
Medium ($30,000– $100,000)	11	41
High (Above $100,000)	1	4
Do not wish to disclose	2	7
Profession		
Full-time	12	44
Part-time	1	4
Student	12	44
Self-employed/Other	2	8
Retired	0	0
Health Status		
Healthy	20	74
Chronic Disease or Illness, Prescription Drug Use, Smoking or Alcohol	7	26
Classes		
Stress	894	42
No Stress	1245	58

<u>Sleep Dataset with Withings Features (SDW), N = 34</u>

Age		
18-24	9	26
25-34	10	29
35-44	10	29
Sex/Gender		
Male	9	26
Female	24	71
Gender Fluid	1	3
SES		
Low (0-$30,000)	15	44
Medium ($30,000– $100,000)	6	38
High (Above $100,000)	4	12
Do not wish to disclose	2	6
Profession		
Full-time	16	47
Part-time	3	9

Student	12	35
Self-employed/Other	2	6
Retired	1	3
Health Status		
Healthy	8	82
Chronic Disease or Illness, Prescription Drug Use, Smoking or Alcohol	28	18

Sleep Dataset with Withings and Apple Only Sleep Features (SDS), N = 27

Age		
18-24	8	30
25-34	9	33
35-44	7	26
Sex/Gender		
Male	6	22
Female	20	74
Gender Fluid	1	4
SES		
Low (0-$30,000)	13	48
Medium ($30,000– $100,000)	11	41
High (Above $100,000)	1	4
Do not wish to disclose	2	7
Profession		
Full-time	12	44
Part-time	1	4
Student	12	44
Self-employed/Other	2	8
Retired	0	0
Health Status		
Healthy	7	26
Chronic Disease or Illness, Prescription Drug Use, Smoking or Alcohol	20	74

Table B4: Precision, Recall, Accuracy, F1-Score for Non-Sleep Datasets, Generalized

D	RF	SVM	Support

	Items	*Precision*	*Recall*	*F1-Score*	*Precision*	*Recall*	*F1-Score*	
Complete Dataset	*No Stress*	0.67	0.67	0.67	0.66	0.65	0.66	185
	Stress	0.65	0.65	0.65	0.64	0.65	0.64	175
	Accuracy			0.66			0.65	360
	Weighted Average	0.66	0.66	0.66	0.65	0.65	0.65	360
	Macro Average	0.66	0.66	0.66	0.65	0.65	0.65	360
Complete Dataset (SMOTE)	*No Stress*	0.70	0.67	0.68	0.67	0.63	0.65	185
	Stress	0.66	0.69	0.68	0.63	0.67	0.65	175
	Accuracy			0.68			0.65	360
	Weighted Average	0.68	0.68	0.68	0.65	0.65	0.65	360
	Macro Average	0.68	0.68	0.68	0.65	0.65	0.65	360
Gender - Male	*No Stress*	0.69	0.85	0.76	0.71	0.82	0.76	67
	Stress	0.72	0.50	0.59	0.71	0.58	0.64	52
	Accuracy			0.70			0.71	119
	Weighted Average	0.70	0.70	0.69	0.71	0.71	0.71	119
	Macro Average	0.70	0.68	0.68	0.71	0.70	0.70	119
	No Stress	0.66	0.76	0.71	0.67	0.73	0.70	67

Gender – Male (SMOTE)	*Stress*	0.62	0.50	0.55	0.61	0.54	0.57	52
	Accuracy			0.65			0.65	119
	Weighted Average	0.64	0.65	0.64	0.64	0.65	0.64	119
	Macro Average	0.64	0.63	0.63	0.64	0.63	0.64	119
Gender – Female	*No Stress*	0.76	0.52	0.62	0.64	0.46	0.54	108
	Stress	0.66	0.85	0.74	0.61	0.76	0.68	118
	Accuracy			0.69			0.62	226
	Weighted Average	0.71	0.69	0.68	0.62	0.62	0.61	226
	Macro Average	0.71	0.68	0.68	0.62	0.61	0.61	226
Gender – Female (SMOTE)	*No Stress*	0.68	0.56	0.62	0.66	0.56	0.61	108
	Stress	0.65	0.75	0.70	0.65	0.74	0.69	118
	Accuracy			0.66			0.65	226
	Weighted Average	0.67	0.66	0.66	0.66	0.65	0.65	226
	Macro Average	0.67	0.66	0.66	0.66	0.65	0.65	226
Employment - Student	*No Stress*	0.65	0.55	0.60	0.66	0.60	0.63	53
	Stress	0.57	0.67	0.61	0.59	0.64	0.61	63

	Accuracy			0.61			0.62	116
	Weighted Average	0.61	0.61	0.61	0.63	0.62	0.62	116
	Macro Average	0.61	0.61	0.61	0.62	0.62	0.62	116
Employment – Student (SMOTE)	*No Stress*	0.65	0.54	0.59	0.66	0.59	0.63	53
	Stress	0.56	0.67	0.61	0.59	0.65	0.62	63
	Accuracy			0.60			0.62	116
	Weighted Average	0.61	0.60	0.60	0.63	0.62	0.62	116
	Macro Average	0.60	0.60	0.60	0.62	0.62	0.62	116
Employment – Worker	*No Stress*	0.63	0.51	0.57	0.58	0.57	0.57	76
	Stress	0.64	0.74	0.68	0.63	0.65	0.64	88
	Accuracy			0.63			0.61	164
	Weighted Average	0.63	0.63	0.63	0.61	0.61	0.61	164
	Macro Average	0.63	0.63	0.62	0.61	0.61	0.61	164
Employment – Worker (SMOTE)	*No Stress*	0.61	0.59	0.60	0.55	0.54	0.54	76
	Stress	0.66	0.67	0.66	0.61	0.61	0.61	88
	Accuracy			0.63			0.58	164

	Weighted Average	0.63	0.63	0.63	0.58	0.58	0.58	164
	Macro Average	0.63	0.63	0.63	0.58	0.58	0.58	164
Income - Low	*No Stress*	0.67	0.74	0.70	0.68	0.68	0.68	112
	Stress	0.59	0.50	0.54	0.57	0.57	0.57	82
	Accuracy			0.64			0.63	194
	Weighted Average	0.63	0.64	0.63	0.63	0.63	0.63	194
	Macro Average	0.63	0.62	0.62	0.63	0.63	0.63	194
Income – Low (SMOTE)	*No Stress*	0.68	0.68	0.68	0.64	0.91	0.75	112
	Stress	0.57	0.57	0.57	0.71	0.30	0.43	82
	Accuracy			0.63			0.65	194
	Weighted Average	0.63	0.63	0.63	0.67	0.65	0.62	194
	Macro Average	0.63	0.63	0.63	0.68	0.61	0.59	194
Income – Medium High	*No Stress*	0.82	0.55	0.66	0.82	0.49	0.62	73
	Stress	0.72	0.90	0.80	0.70	0.91	0.79	94
	Accuracy			0.75			0.73	167
	Weighted Average	0.76	0.75	0.74	0.75	0.73	0.72	167

	Macro Average	0.77	0.73	0.73	0.76	0.70	0.70	167
Income – Medium High (SMOTE)	*No Stress*	0.71	0.62	0.66	0.70	0.60	0.65	73
	Stress	0.73	0.81	0.77	0.72	0.80	0.76	94
	Accuracy			0.72			0.71	167
	Weighted Average	0.72	0.72	0.72	0.71	0.71	0.71	167
	Macro Average	0.72	0.71	0.71	0.71	0.70	0.70	167
Age – 18-24	*No Stress*	0.67	0.94	0.78	0.65	1.00	0.79	52
	Stress	0.57	0.14	0.23	0.00	0.00	0.00	28
	Accuracy			0.66			0.65	80
	Weighted Average	0.64	0.66	0.59	0.42	0.65	0.51	80
	Macro Average	0.62	0.54	0.51	0.33	0.50	0.39	80
Age – 18-24 (SMOTE)	*No Stress*	0.67	0.81	0.73	0.65	1.00	0.79	52
	Stress	0.41	0.25	0.31	0.00	0.00	0.00	28
	Accuracy			0.61			0.65	80
	Weighted Average	0.58	0.61	0.58	0.42	0.65	0.51	80
	Macro Average	0.54	0.53	0.52	0.33	0.50	0.39	80

Age 25-34	*No Stress*	0.56	0.27	0.36	0.56	0.41	0.47	56
	Stress	0.67	0.88	0.76	0.71	0.81	0.76	97
	Accuracy			0.65			0.67	153
	Weighted Average	0.63	0.65	0.62	0.65	0.67	0.65	153
	Macro Average	0.62	0.57	0.56	0.63	0.61	0.62	153
Age 25-34 (SMOTE)	*No Stress*	0.50	0.45	0.47	0.40	0.07	0.12	56
	Stress	0.70	0.74	0.72	0.64	0.94	0.76	97
	Accuracy			0.63			0.62	153
	Weighted Average	0.63	0.63	0.63	0.55	0.62	0.53	153
	Macro Average	0.60	0.59	0.60	0.52	0.50	0.44	153
Age 35-44	*No Stress*	0.60	0.78	0.68	0.63	0.89	0.74	46
	Stress	0.44	0.25	0.32	0.62	0.25	0.36	32
	Accuracy			0.56			0.63	78
	Weighted Average	0.54	0.56	0.53	0.62	0.63	0.58	78
	Macro Average	0.52	0.52	0.50	0.62	0.57	0.55	78
Age 35-44 (SMOTE)	*No Stress*	0.62	0.78	0.69	0.67	0.80	0.73	46

	Stress	0.50	0.31	0.38	0.61	0.44	0.51	32
	Accuracy			0.59			0.65	78
	Weighted Average	0.57	0.59	0.57	0.65	0.65	0.64	78
	Macro Average	0.56	0.55	0.54	0.64	0.62	0.62	78
Age 45-64	*No Stress*	0.65	0.61	0.63	0.75	0.67	0.71	18
	Stress	0.59	0.62	0.61	0.67	0.75	0.71	16
	Accuracy			0.62			0.71	34
	Weighted Average	0.62	0.62	0.62	0.71	0.71	0.71	34
	Macro Average	0.62	0.62	0.62	0.71	0.71	0.71	34
Age 45-64 (SMOTE)	*No Stress*	0.65	0.61	0.63	0.71	0.67	0.69	18
	Stress	0.59	0.62	0.61	0.65	0.69	0.67	16
	Accuracy			0.62			0.68	34
	Weighted Average	0.62	0.62	0.62	0.68	0.68	0.68	34
	Macro Average	0.62	0.62	0.62	0.68	0.68	0.68	34
Healthy	*No Stress*	0.68	0.78	0.72	0.69	0.73	0.71	139
	Stress	0.69	0.57	0.63	0.67	0.62	0.65	120

	Accuracy			0.68			0.68	259
	Weighted Average	0.68	0.68	0.68	0.68	0.68	0.68	259
	Macro Average	0.68	0.68	0.68	0.68	0.68	0.68	259
Healthy - SMOTE	*No Stress*	0.69	0.71	0.70	0.64	0.66	0.65	139
	Stress	0.65	0.62	0.64	0.59	0.57	0.58	120
	Accuracy			0.67			0.62	259
	Weighted Average	0.67	0.67	0.67	0.62	0.62	0.62	259
	Macro Average	0.67	0.67	0.67	0.62	0.62	0.62	259
DECG	**RF**				**SVM**			
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	**Support**
Complete Dataset	*No Stress*	0.64	0.74	0.69	0.58	0.66	0.62	373
	Stress	0.59	0.48	0.53	0.48	0.40	0.44	294
	Accuracy	-	-	0.63	-	-	0.55	667
	Weighted Average	0.62	0.61	0.61	0.54	0.55	0.54	667
	Macro Average	0.62	0.63	0.62	0.53	0.53	0.53	667
Complete Dataset (SMOTE)	*No Stress*	0.65	0.66	0.66	0.59	0.76	0.67	373
	Stress	0.56	0.55	0.56	0.52	0.34	0.41	294

	Accuracy	-	-	0.61	-	-	0.57	667
	Weighted Average	0.61	0.61	0.61	0.56	0.57	0.55	667
	Macro Average	0.61	0.61	0.61	0.56	0.55	0.54	667
Gender - Male	*No Stress*	0.65	0.73	0.69	0.65	0.70	0.68	114
	Stress	0.58	0.49	0.53	0.56	0.50	0.53	86
	Accuracy			0.62			0.61	200
	Weighted Average	0.62	0.62	0.62	0.61	0.61	0.61	200
	Macro Average	0.61	0.61	0.61	0.60	0.60	0.60	200
Gender – Male (SMOTE)	*No Stress*	0.67	0.67	0.67	0.65	0.64	0.65	114
	Stress	0.56	0.57	0.57	0.53	0.55	0.54	86
	Accuracy			0.62			0.60	200
	Weighted Average	0.63	0.62	0.63	0.60	0.60	0.60	200
	Macro Average	0.62	0.62	0.62	0.59	0.59	0.59	200
Gender – Female	*No Stress*	0.65	0.69	0.67	0.63	0.65	0.64	247
	Stress	0.60	0.55	0.57	0.56	0.53	0.55	204
	Accuracy			0.63			0.60	451

	Weighted Average	0.63	0.63	0.63	0.60	0.60	0.60	451
	Macro Average	0.62	0.62	0.62	0.59	0.59	0.59	451
Gender – Female (SMOTE)	*No Stress*	0.66	0.64	0.65	0.61	0.58	0.59	247
	Stress	0.58	0.60	0.59	0.52	0.55	0.53	204
	Accuracy			0.62			0.57	451
	Weighted Average	0.62	0.62	0.62	0.57	0.57	0.57	451
	Macro Average	0.62	0.62	0.62	0.56	0.56	0.56	451
Employment - Student	*No Stress*	0.69	0.58	0.63	0.61	0.64	0.63	135
	Stress	0.60	0.71	0.65	0.58	0.55	0.57	121
	Accuracy			0.64			0.60	256
	Weighted Average	0.65	0.64	0.64	0.60	0.60	0.60	256
	Macro Average	0.65	0.64	0.64	0.60	0.60	0.60	256
Employment – Student (SMOTE)	*No Stress*	0.66	0.59	0.62	0.62	0.64	0.63	135
	Stress	0.59	0.67	0.63	0.58	0.56	0.57	121
	Accuracy			0.62			0.60	256
	Weighted Average	0.63	0.62	0.62	0.60	0.60	0.60	256

	Macro Average	0.63	0.63	0.62	0.60	0.60	0.60	256
Employment – Worker	*No Stress*	0.63	0.75	0.68	0.61	0.77	0.68	224
	Stress	0.56	0.43	0.49	0.54	0.36	0.44	171
	Accuracy			0.61			0.59	395
	Weighted Average	0.60	0.59	0.59	0.58	0.59	0.57	395
	Macro Average	0.60	0.61	0.60	0.58	0.57	0.56	395
Employment – Worker (SMOTE)	*No Stress*	0.64	0.64	0.64	0.57	0.81	0.67	224
	Stress	0.53	0.54	0.53	0.44	0.20	0.27	171
	Accuracy			0.59			0.54	395
	Weighted Average	0.60	0.59	0.60	0.51	0.54	0.50	395
	Macro Average	0.59	0.59	0.59	0.51	0.50	0.47	395
Income - Low	*No Stress*	0.63	0.79	0.70	0.64	0.67	0.65	224
	Stress	0.47	0.29	0.36	0.45	0.43	0.44	171
	Accuracy			0.59			0.57	395
	Weighted Average	0.57	0.59	0.57	0.57	0.57	0.57	395
	Macro Average	0.55	0.54	0.53	0.55	0.55	0.55	395

Income – Low (SMOTE)	*No Stress*	0.63	0.65	0.64	0.60	0.82	0.69	224
	Stress	0.43	0.41	0.42	0.38	0.17	0.24	171
	Accuracy			0.56			0.56	395
	Weighted Average	0.55	0.56	0.55	0.51	0.56	0.51	395
	Macro Average	0.53	0.53	0.53	0.49	0.49	0.46	395
Income – Medium High	*No Stress*	0.73	0.58	0.65	0.64	0.57	0.60	166
	Stress	0.64	0.77	0.70	0.60	0.67	0.63	159
	Accuracy			0.68			0.62	325
	Weighted Average	0.69	0.68	0.67	0.62	0.62	0.61	325
	Macro Average	0.68	0.68	0.67	0.62	0.62	0.61	325
Income – Medium High (SMOTE)	*No Stress*	0.70	0.63	0.66	0.62	0.49	0.55	166
	Stress	0.65	0.72	0.68	0.56	0.68	0.62	159
	Accuracy			0.67			0.58	325
	Weighted Average	0.68	0.67	0.67	0.59	0.59	0.58	325
	Macro Average	0.68	0.67	0.67	0.59	0.58	0.58	325
Age – 18-24	*No Stress*	0.68	0.84	0.75	0.67	0.71	0.69	166

	Stress	0.61	0.39	0.47	0.50	0.46	0.48	159
	Accuracy			0.66			0.61	325
	Weighted Average	0.65	0.66	0.65	0.61	0.61	0.61	325
	Macro Average	0.65	0.61	0.61	0.59	0.58	0.58	325
Age – 18-24 (SMOTE)	*No Stress*	0.66	0.66	0.66	0.63	0.85	0.72	166
	Stress	0.47	0.47	0.47	0.48	0.22	0.30	159
	Accuracy			0.59			0.61	325
	Weighted Average	0.59	0.59	0.59	0.57	0.61	0.56	325
	Macro Average	0.57	0.57	0.57	0.56	0.54	0.51	325
Age 25-34	*No Stress*	0.63	0.35	0.45	0.58	0.38	0.46	91
	Stress	0.64	0.85	0.73	0.64	0.80	0.71	124
	Accuracy			0.64			0.62	215
	Weighted Average	0.63	0.60	0.59	0.62	0.62	0.61	215
	Macro Average	0.63	0.60	0.59	0.61	0.59	0.59	215
Age 25-34 (SMOTE)	*No Stress*	0.54	0.48	0.51	0.43	0.44	0.44	91
	Stress	0.65	0.70	0.67	0.59	0.58	0.58	124

	Accuracy			0.61			0.52	215
	Weighted Average	0.60	0.61	0.61	0.52	0.52	0.52	215
	Macro Average	0.60	0.59	0.59	0.51	0.51	0.51	215
Age 35-44	*No Stress*	0.67	0.84	0.75	0.66	0.69	0.67	113
	Stress	0.57	0.34	0.42	0.46	0.42	0.44	71
	Accuracy			0.65			0.59	184
	Weighted Average	0.63	0.65	0.62	0.58	0.59	0.58	184
	Macro Average	0.62	0.59	0.58	0.56	0.56	0.56	184
Age 35-44 (SMOTE)	*No Stress*	0.69	0.72	0.70	0.62	0.83	0.71	113
	Stress	0.52	0.49	0.51	0.42	0.20	0.27	71
	Accuracy			0.63			0.59	184
	Weighted Average	0.63	0.63	0.63	0.55	0.59	0.54	184
	Macro Average	0.61	0.60	0.61	0.52	0.51	0.49	184
Healthy	*No Stress*	0.64	0.75	0.69	0.62	0.72	0.66	308
	Stress	0.55	0.42	0.48	0.50	0.39	0.44	224
	Accuracy			0.61			0.58	532

	Weighted Average	0.60	0.61	0.60	0.57	0.58	0.57	532
	Macro Average	0.59	0.58	0.58	0.56	0.56	0.55	532
Healthy (SMOTE)	*No Stress*	0.68	0.68	0.68	0.64	0.65	0.64	308
	Stress	0.56	0.55	0.55	0.50	0.49	0.50	224
	Accuracy			0.63			0.58	532
	Weighted Average	0.63	0.63	0.63	0.58	0.58	0.58	532
	Macro Average	0.62	0.62	0.62	0.57	0.57	0.57	532
DA	**RF**				**SVM**			
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	**Support**
Complete Dataset	*No Stress*	0.66	0.72	0.69	0.62	0.73	0.67	373
	Stress	0.60	0.53	0.56	0.56	0.44	0.49	294
	Accuracy	-	-	0.65	-	-	0.60	667
	Weighted Average	0.63	0.64	0.63	0.60	0.60	0.59	667
	Macro Average	0.63	0.628	0.63	0.59	0.58	0.58	667
Complete Dataset (SMOTE)	*No Stress*	0.69	0.64	0.66	0.59	0.61	0.60	373
	Stress	0.58	0.63	0.60	0.48	0.46	0.47	294
	Accuracy	-	-	0.63	-	-	0.54	667

	Weighted Average	0.64	0.63	0.64	0.54	0.54	0.54	667
	Macro Average	0.63	0.63	0.63	0.53	0.53	0.53	667
Gender - Male	*No Stress*	0.67	0.76	0.71	0.65	0.71	0.68	114
	Stress	0.61	0.50	0.55	0.57	0.50	0.53	86
	Accuracy			0.65			0.62	200
	Weighted Average	0.65	0.65	0.64	0.62	0.62	0.62	200
	Macro Average	0.64	0.63	0.63	0.61	0.61	0.61	200
Gender – Male (SMOTE)	*No Stress*	0.68	0.72	0.70	0.63	0.64	0.64	114
	Stress	0.60	0.56	0.58	0.52	0.51	0.51	86
	Accuracy			0.65			0.58	200
	Weighted Average	0.65	0.65	0.65	0.58	0.58	0.58	200
	Macro Average	0.64	0.64	0.64	0.58	0.58	0.58	200
Gender – Female	*No Stress*	0.62	0.68	0.65	0.57	0.63	0.60	247
	Stress	0.56	0.50	0.53	0.49	0.43	0.46	204
	Accuracy			0.60			0.54	451
	Weighted Average	0.60	0.60	0.60	0.54	0.54	0.54	451

	Macro Average	0.59	0.59	0.59	0.53	0.53	0.53	451
Gender – Female (SMOTE)	*No Stress*	0.64	0.56	0.60	0.63	0.57	0.60	247
	Stress	0.54	0.62	0.58	0.53	0.59	0.56	204
	Accuracy			0.59			0.58	451
	Weighted Average	0.60	0.59	0.59	0.58	0.58	0.58	451
	Macro Average	0.59	0.59	0.59	0.58	0.58	0.58	451
Employment - Student	*No Stress*	0.71	0.58	0.64	0.57	0.55	0.56	135
	Stress	0.61	0.74	0.67	0.52	0.54	0.53	121
	Accuracy			0.65			0.54	256
	Weighted Average	0.66	0.65	0.65	0.54	0.54	0.54	256
	Macro Average	0.66	0.66	0.65	0.54	0.54	0.54	256
Employment – Student (SMOTE)	*No Stress*	0.71	0.52	0.60	0.60	0.60	0.60	135
	Stress	0.59	0.77	0.67	0.55	0.55	0.55	121
	Accuracy			0.64			0.58	256
	Weighted Average	0.65	0.64	0.63	0.58	0.58	0.58	256
	Macro Average	0.65	0.64	0.63	0.58	0.58	0.58	256

Employment – Worker	*No Stress*	0.66	0.75	0.71	0.62	0.65	0.63	224
	Stress	0.61	0.50	0.55	0.51	0.49	0.50	171
	Accuracy			0.64			0.58	395
	Weighted Average	0.64	0.64	0.64	0.58	0.58	0.58	395
	Macro Average	0.63	0.63	0.63	0.57	0.57	0.57	395
Employment – Worker (SMOTE)	*No Stress*	0.66	0.66	0.66	0.60	0.62	0.61	224
	Stress	0.56	0.56	0.56	0.48	0.45	0.46	171
	Accuracy			0.62			0.55	395
	Weighted Average	0.62	0.62	0.62	0.55	0.55	0.55	395
	Macro Average	0.61	0.61	0.61	0.54	0.54	0.54	395
Income - Low	*No Stress*	0.64	0.84	0.73	0.63	0.73	0.68	224
	Stress	0.52	0.27	0.36	0.44	0.32	0.37	171
	Accuracy			0.62			0.57	395
	Weighted Average	0.59	0.62	0.58	0.55	0.57	0.56	395
	Macro Average	0.58	0.56	0.54	0.53	0.53	0.52	395
Income – Low (SMOTE)	*No Stress*	0.67	0.69	0.68	0.64	0.67	0.66	224

	Stress	0.50	0.47	0.48	0.45	0.42	0.44	171
	Accuracy			0.61			0.57	395
	Weighted Average	0.60	0.61	0.60	0.57	0.57	0.57	395
	Macro Average	0.58	0.58	0.58	0.55	0.55	0.55	395
Income – Medium High	*No Stress*	0.69	0.63	0.66	0.60	0.64	0.62	166
	Stress	0.65	0.70	0.67	0.59	0.55	0.57	159
	Accuracy			0.67			0.60	325
	Weighted Average	0.67	0.67	0.67	0.60	0.60	0.60	325
	Macro Average	0.67	0.67	0.67	0.60	0.60	0.60	325
Income – Medium High (SMOTE)	*No Stress*	0.68	0.61	0.65	0.60	0.61	0.60	166
	Stress	0.64	0.70	0.67	0.59	0.58	0.58	159
	Accuracy			0.66			0.59	325
	Weighted Average	0.66	0.66	0.66	0.59	0.59	0.59	325
	Macro Average	0.66	0.66	0.66	0.59	0.59	0.59	325
Age – 18-24	*No Stress*	0.64	0.81	0.72	0.66	0.66	0.66	166
	Stress	0.50	0.29	0.37	0.46	0.46	0.46	159

	Accuracy			0.61			0.58	325
	Weighted Average	0.59	0.61	0.58	0.58	0.58	0.58	325
	Macro Average	0.57	0.55	0.54	0.56	0.56	0.56	325
Age – 18-24 (SMOTE)	*No Stress*	0.63	0.65	0.64	0.65	0.63	0.64	166
	Stress	0.43	0.40	0.41	0.45	0.47	0.46	159
	Accuracy			0.56			0.57	325
	Weighted Average	0.55	0.56	0.56	0.57	0.57	0.57	325
	Macro Average	0.53	0.53	0.53	0.55	0.55	0.55	325
Age 25-34	*No Stress*	0.65	0.35	0.46	0.58	0.34	0.43	91
	Stress	0.64	0.86	0.74	0.63	0.82	0.71	124
	Accuracy			0.65			0.62	215
	Weighted Average	0.65	0.65	0.62	0.61	0.62	0.59	215
	Macro Average	0.65	0.61	0.60	0.61	0.58	0.57	215
Age 25-34 (SMOTE)	*No Stress*	0.60	0.52	0.56	0.50	0.46	0.48	91
	Stress	0.68	0.75	0.71	0.63	0.66	0.64	124
	Accuracy			0.65			0.58	215

	Weighted Average	0.65	0.65	0.65	0.57	0.58	0.57	215
	Macro Average	0.64	0.63	0.63	0.56	0.56	0.56	215
Age 35-44	*No Stress*	0.67	0.89	0.77	0.66	0.82	0.74	113
	Stress	0.65	0.31	0.42	0.55	0.34	0.42	71
	Accuracy			0.67			0.64	184
	Weighted Average	0.66	0.67	0.63	0.60	0.58	0.58	184
	Macro Average	0.66	0.60	0.59	0.60	0.64	0.61	184
Age 35-44 (SMOTE)	*No Stress*	0.73	0.75	0.74	0.63	0.69	0.66	113
	Stress	0.58	0.55	0.57	0.42	0.35	0.38	71
	Accuracy			0.67			0.56	184
	Weighted Average	0.67	0.67	0.67	0.55	0.56	0.55	184
	Macro Average	0.65	0.65	0.65	0.52	0.52	0.52	184
Healthy	*No Stress*	0.65	0.79	0.71	0.62	0.77	0.69	308
	Stress	0.59	0.42	0.49	0.54	0.36	0.43	224
	Accuracy			0.63			0.60	532
	Weighted Average	0.63	0.63	0.62	0.59	0.60	0.58	532

	Macro Average	0.62	0.60	0.60	0.58	0.57	0.56	532
Healthy (SMOTE)	*No Stress*	0.68	0.71	0.69	0.65	0.57	0.61	308
	Stress	0.57	0.54	0.56	0.49	0.58	0.53	224
	Accuracy			0.64			0.57	532
	Weighted Average	0.63	0.64	0.64	0.58	0.57	0.57	532
	Macro Average	0.63	0.62	0.62	0.57	0.57	0.57	532
DAW	**RF**				**SVM**			
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	***Support***
Complete Dataset	*No Stress*	0.69	0.82	0.75	0.64	0.69	0.66	367
	Stress	0.70	0.52	0.60	0.55	0.50	0.53	284
	Accuracy	-	-	0.69	-	-	0.61	651
	Weighted Average	0.69	0.69	0.68	0.60	0.60	0.60	651
	Macro Average	0.69	0.67	0.67	0.60	0.61	0.60	651
Complete Dataset (SMOTE)	*No Stress*	0.72	0.77	0.74	0.64	0.69	0.67	367
	Stress	0.67	0.60	0.64	0.56	0.51	0.53	284
	Accuracy	-	-	0.70	-	-	0.61	651
	Weighted Average	0.70	0.70	0.70	0.60	0.60	0.60	651

	Macro Average	0.69	0.69	0.69	0.61	0.61	0.61	651
Gender - Male	*No Stress*	0.74	0.92	0.82	0.74	0.80	0.77	109
	Stress	0.81	0.52	0.63	0.67	0.59	0.62	75
	Accuracy			0.76			0.71	184
	Weighted Average	0.77	0.76	0.74	0.71	0.71	0.71	184
	Macro Average	0.77	0.72	0.73	0.70	0.69	0.70	184
Gender – Male (SMOTE)	*No Stress*	0.75	0.82	0.78	0.72	0.83	0.77	109
	Stress	0.69	0.60	0.64	0.68	0.53	0.60	75
	Accuracy			0.73			0.71	184
	Weighted Average	0.73	0.73	0.72	0.70	0.68	0.68	184
	Macro Average	0.72	0.71	0.71	0.70	0.71	0.70	184
Gender – Female	*No Stress*	0.68	0.79	0.73	0.65	0.70	0.68	247
	Stress	0.68	0.54	0.60	0.61	0.55	0.58	204
	Accuracy			0.68			0.63	451
	Weighted Average	0.68	0.68	0.67	0.63	0.63	0.63	451
	Macro Average	0.68	0.67	0.67	0.63	0.63	0.63	451

Gender – Female (SMOTE)	*No Stress*	0.70	0.68	0.69	0.61	0.67	0.64	247
	Stress	0.62	0.64	0.63	0.55	0.49	0.51	204
	Accuracy			0.66			0.59	451
	Weighted Average	0.66	0.66	0.66	0.58	0.58	0.58	451
	Macro Average	0.66	0.66	0.66	0.58	0.59	0.58	451
Employment - Student	*No Stress*	0.69	0.61	0.65	0.66	0.66	0.66	135
	Stress	0.62	0.70	0.66	0.62	0.63	0.63	121
	Accuracy			0.65			0.64	256
	Weighted Average	0.66	0.65	0.65	0.64	0.64	0.64	256
	Macro Average	0.66	0.65	0.65	0.64	0.64	0.64	256
Employment – Student (SMOTE)	*No Stress*	0.69	0.52	0.59	0.62	0.69	0.65	135
	Stress	0.58	0.74	0.65	0.61	0.55	0.57	121
	Accuracy			0.62			0.62	256
	Weighted Average	0.64	0.62	0.62	0.62	0.62	0.61	256
	Macro Average	0.63	0.63	0.62	0.62	0.61	0.61	256
Employment – Worker	*No Stress*	0.69	0.81	0.75	0.68	0.74	0.71	219

	Stress	0.66	0.50	0.57	0.59	0.51	0.55	160
	Accuracy			0.68			0.65	379
	Weighted Average	0.68	0.68	0.67	0.64	0.65	0.64	379
	Macro Average	0.68	0.66	0.66	0.64	0.63	0.63	379
Employment – Worker (SMOTE)	*No Stress*	0.71	0.74	0.73	0.67	0.74	0.71	219
	Stress	0.63	0.59	0.61	0.59	0.51	0.55	160
	Accuracy			0.68			0.64	379
	Weighted Average	0.68	0.68	0.68	0.63	0.63	0.63	379
	Macro Average	0.67	0.67	0.67	0.64	0.64	0.64	379
Income - Low	*No Stress*	0.64	0.85	0.73	0.67	0.79	0.73	162
	Stress	0.50	0.25	0.33	0.53	0.38	0.44	101
	Accuracy			0.62			0.63	263
	Weighted Average	0.59	0.62	0.58	0.62	0.63	0.62	263
	Macro Average	0.57	0.55	0.53	0.60	0.58	0.58	263
Income – Low (SMOTE)	*No Stress*	0.68	0.74	0.71	0.66	0.73	0.69	162
	Stress	0.52	0.45	0.48	0.48	0.39	0.43	101

	Accuracy			0.63			0.60	263
	Weighted Average	0.62	0.63	0.62	0.59	0.60	0.59	263
	Macro Average	0.60	0.59	0.59	0.57	0.56	0.56	263
Income – Medium High	*No Stress*	0.75	0.68	0.71	0.66	0.68	0.67	161
	Stress	0.68	0.76	0.72	0.64	0.62	0.63	148
	Accuracy			0.72			0.65	309
	Weighted Average	0.72	0.72	0.72	0.65	0.65	0.65	309
	Macro Average	0.72	0.72	0.72	0.65	0.65	0.65	309
Income – Medium High (SMOTE)	*No Stress*	0.75	0.65	0.70	0.69	0.68	0.68	161
	Stress	0.67	0.76	0.71	0.66	0.67	0.66	148
	Accuracy			0.71			0.67	309
	Weighted Average	0.71	0.71	0.70	0.67	0.67	0.67	309
	Macro Average	0.71	0.71	0.71	0.67	0.67	0.67	309
Age – 18-24	*No Stress*	0.68	0.85	0.75	0.67	0.79	0.72	113
	Stress	0.60	0.36	0.45	0.54	0.39	0.45	72
	Accuracy			0.66			0.63	185

	Weighted Average	0.65	0.66	0.64	0.62	0.63	0.62	185
	Macro Average	0.64	0.61	0.60	0.60	0.59	0.59	185
Age – 18-24 (SMOTE)	*No Stress*	0.71	0.74	0.72	0.68	0.79	0.73	113
	Stress	0.56	0.51	0.54	0.56	0.42	0.48	72
	Accuracy			0.65			0.64	185
	Weighted Average	0.65	0.65	0.65	0.63	0.64	0.63	185
	Macro Average	0.63	0.63	0.63	0.62	0.60	0.60	185
Age 25-34	*No Stress*	0.73	0.52	0.61	0.57	0.58	0.58	91
	Stress	0.71	0.86	0.78	0.69	0.68	0.68	124
	Accuracy			0.72			0.64	215
	Weighted Average	0.72	0.72	0.71	0.64	0.64	0.64	215
	Macro Average	0.72	0.69	0.69	0.63	0.63	0.63	215
Age 25-34 (SMOTE)	*No Stress*	0.64	0.54	0.59	0.54	0.45	0.49	91
	Stress	0.70	0.78	0.74	0.64	0.72	0.68	124
	Accuracy			0.68			0.60	215
	Weighted Average	0.68	0.68	0.67	0.60	0.60	0.60	215

	Macro Average	0.67	0.66	0.66	0.59	0.58	0.58	215
Age 35-44	*No Stress*	0.71	0.91	0.79	0.76	0.83	0.80	108
	Stress	0.66	0.32	0.43	0.64	0.53	0.58	60
	Accuracy			0.70			0.73	168
	Weighted Average	0.69	0.70	0.66	0.72	0.73	0.72	168
	Macro Average	0.68	0.61	0.61	0.70	0.68	0.69	168
Age 35-44 (SMOTE)	*No Stress*	0.75	0.84	0.79	0.69	0.81	0.74	108
	Stress	0.63	0.48	0.55	0.49	0.33	0.40	60
	Accuracy			0.71			0.64	168
	Weighted Average	0.70	0.71	0.70	0.61	0.64	0.62	168
	Macro Average	0.69	0.66	0.67	0.59	0.57	0.57	168
Healthy	*No Stress*	0.67	0.82	0.74	0.68	0.79	0.73	303
	Stress	0.62	0.42	0.50	0.61	0.46	0.52	213
	Accuracy			0.66			0.66	516
	Weighted Average	0.65	0.66	0.64	0.65	0.65	0.64	516
	Macro Average	0.64	0.62	0.62	0.64	0.63	0.63	516

Healthy (SMOTE)	*No Stress*	0.67	0.70	0.68	0.66	0.66	0.66	303
	Stress	0.54	0.51	0.53	0.52	0.52	0.52	213
	Accuracy			0.62			0.60	516
	Weighted Average	0.62	0.62	0.62	0.59	0.59	0.59	516
	Macro Average	0.61	0.60	0.60	0.60	0.60	0.60	516

DW	**RF**				**SVM**			
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	***Support***
Complete Dataset	*No Stress*	0.72	0.81	0.76	0.66	0.76	0.71	398
	Stress	0.69	0.57	0.62	0.60	0.48	0.54	297
	Accuracy	-	-	0.71	-	-	0.64	695
	Weighted Average	0.70	0.71	0.70	0.64	0.64	0.64	695
	Macro Average	0.70	0.69	0.69	0.63	0.62	0.62	695
Complete Dataset (SMOTE)	*No Stress*	0.72	0.68	0.70	0.71	0.64	0.67	398
	Stress	0.60	0.64	0.62	0.57	0.65	0.61	297
	Accuracy	-	-	0.66	-	-	0.64	695
	Weighted Average	0.66	0.66	0.66	0.65	0.64	0.65	695
	Macro Average	0.67	0.66	0.66	0.64	0.65	0.64	695

Gender - Male	*No Stress*	0.71	0.87	0.78	0.72	0.89	0.80	124
	Stress	0.67	0.42	0.51	0.71	0.45	0.56	77
	Accuracy			0.70			0.72	201
	Weighted Average	0.69	0.70	0.68	0.72	0.72	0.70	201
	Macro Average	0.69	0.64	0.65	0.72	0.67	0.68	201
Gender – Male (SMOTE)	*No Stress*	0.71	0.74	0.72	0.72	0.69	0.70	124
	Stress	0.55	0.51	0.53	0.53	0.57	0.55	77
	Accuracy			0.65			0.64	201
	Weighted Average	0.65	0.65	0.65	0.65	0.64	0.64	201
	Macro Average	0.63	0.62	0.63	0.63	0.63	0.63	201
Gender – Female	*No Stress*	0.65	0.76	0.70	0.61	0.72	0.66	251
	Stress	0.63	0.50	0.56	0.55	0.43	0.48	203
	Accuracy			0.65			0.59	454
	Weighted Average	0.64	0.65	0.64	0.58	0.59	0.58	454
	Macro Average	0.64	0.63	0.63	0.58	0.57	0.57	454
	No Stress	0.67	0.72	0.69	0.65	0.67	0.66	251

Gender – Female (SMOTE)	*Stress*	0.62	0.56	0.59	0.58	0.56	0.57	203
	Accuracy			0.65			0.62	454
	Weighted Average	0.65	0.65	0.65	0.62	0.62	0.62	454
	Macro Average	0.64	0.64	0.64	0.62	0.62	0.62	454
Employment - Student	*No Stress*	0.70	0.61	0.65	0.68	0.65	0.67	135
	Stress	0.62	0.70	0.66	0.63	0.66	0.65	121
	Accuracy			0.66			0.66	256
	Weighted Average	0.66	0.66	0.66	0.66	0.66	0.66	256
	Macro Average	0.66	0.65	0.65	0.66	0.66	0.66	256
Employment – Student (SMOTE)	*No Stress*	0.68	0.58	0.63	0.66	0.52	0.58	135
	Stress	0.60	0.70	0.65	0.57	0.70	0.63	121
	Accuracy			0.64			0.61	256
	Weighted Average	0.64	0.64	0.64	0.61	0.61	0.60	256
	Macro Average	0.64	0.64	0.64	0.62	0.61	0.60	256
Employment – Worker	*No Stress*	0.69	0.85	0.76	0.68	0.79	0.73	249
	Stress	0.68	0.47	0.55	0.61	0.46	0.52	174

	Accuracy			0.69			0.65	423
	Weighted Average	0.69	0.69	0.68	0.64	0.63	0.63	423
	Macro Average	0.69	0.66	0.66	0.65	0.65	0.64	423
Employment – Worker (SMOTE)	*No Stress*	0.70	0.74	0.72	0.71	0.63	0.67	249
	Stress	0.60	0.55	0.57	0.54	0.63	0.59	174
	Accuracy			0.66			0.63	423
	Weighted Average	0.65	0.65	0.65	0.63	0.63	0.63	423
	Macro Average	0.66	0.66	0.66	0.63	0.63	0.63	423
Income - Low	*No Stress*	0.66	0.77	0.71	0.67	0.74	0.71	189
	Stress	0.52	0.39	0.44	0.52	0.44	0.48	121
	Accuracy			0.62			0.62	310
	Weighted Average	0.61	0.62	0.61	0.60	0.59	0.59	310
	Macro Average	0.59	0.58	0.58	0.61	0.62	0.62	310
Income – Low (SMOTE)	*No Stress*	0.65	0.61	0.63	0.67	0.57	0.61	189
	Stress	0.45	0.49	0.47	0.46	0.57	0.51	121
	Accuracy			0.56			0.57	310

	Weighted Average	0.57	0.56	0.57	0.56	0.57	0.56	310
	Macro Average	0.55	0.55	0.55	0.59	0.57	0.57	310
Income – Medium High	*No Stress*	0.75	0.74	0.74	0.71	0.73	0.72	182
	Stress	0.70	0.72	0.71	0.67	0.66	0.66	157
	Accuracy			0.73			0.69	339
	Weighted Average	0.73	0.73	0.73	0.69	0.69	0.69	339
	Macro Average	0.73	0.73	0.73	0.65	0.65	0.65	339
Income – Medium High (SMOTE)	*No Stress*	0.73	0.67	0.70	0.72	0.68	0.70	182
	Stress	0.65	0.71	0.68	0.65	0.69	0.67	157
	Accuracy			0.69			0.69	339
	Weighted Average	0.69	0.69	0.69	0.69	0.69	0.69	339
	Macro Average	0.69	0.69	0.69	0.69	0.69	0.69	339
Age – 18-24	*No Stress*	0.64	0.89	0.74	0.65	0.82	0.73	122
	Stress	0.52	0.20	0.29	0.51	0.30	0.38	76
	Accuracy			0.62			0.62	198
	Weighted Average	0.59	0.62	0.57	0.60	0.62	0.59	198

	Macro Average	0.58	0.54	0.51	0.58	0.56	0.55	198
Age – 18-24 (SMOTE)	*No Stress*	0.64	0.69	0.66	0.64	0.61	0.62	122
	Stress	0.42	0.37	0.39	0.42	0.46	0.44	76
	Accuracy			0.57			0.55	198
	Weighted Average	0.55	0.57	0.56	0.53	0.53	0.53	198
	Macro Average	0.53	0.53	0.53	0.56	0.55	0.55	198
Age 25-34	*No Stress*	0.65	0.35	0.46	0.58	0.34	0.43	91
	Stress	0.64	0.86	0.74	0.63	0.82	0.71	124
	Accuracy			0.65			0.62	215
	Weighted Average	0.65	0.65	0.62	0.61	0.62	0.59	215
	Macro Average	0.65	0.61	0.60	0.61	0.58	0.57	215
Age 25-34 (SMOTE)	*No Stress*	0.60	0.52	0.56	0.50	0.46	0.48	91
	Stress	0.68	0.75	0.71	0.63	0.66	0.64	124
	Accuracy			0.65			0.58	215
	Weighted Average	0.65	0.65	0.65	0.57	0.58	0.57	215
	Macro Average	0.64	0.63	0.63	0.56	0.56	0.56	215

Age 35-44	*No Stress*	0.71	0.87	0.78	0.72	0.88	0.79	108
	Stress	0.61	0.37	0.46	0.64	0.38	0.48	60
	Accuracy			0.69			0.70	168
	Weighted Average	0.68	0.69	0.67	0.69	0.70	0.68	168
	Macro Average	0.66	0.62	0.62	0.68	0.63	0.64	168
Age 35-44 (SMOTE)	*No Stress*	0.74	0.79	0.76	0.69	0.69	0.69	108
	Stress	0.57	0.50	0.53	0.44	0.43	0.44	60
	Accuracy			0.68			0.60	168
	Weighted Average	0.68	0.68	0.68	0.60	0.60	0.60	168
	Macro Average	0.65	0.64	0.65	0.56	0.56	0.56	168
Healthy	*No Stress*	0.68	0.77	0.72	0.66	0.80	0.72	327
	Stress	0.58	0.47	0.52	0.56	0.38	0.45	220
	Accuracy			0.65			0.63	547
	Weighted Average	0.64	0.65	0.64	0.62	0.63	0.61	547
	Macro Average	0.63	0.62	0.62	0.61	0.59	0.58	547
Healthy (SMOTE)	*No Stress*	0.69	0.68	0.68	0.66	0.65	0.66	327

	Stress	0.54	0.55	0.54	0.49	0.50	0.50	220
	Accuracy			0.63			0.59	547
	Weighted Average	0.63	0.63	0.63	0.59	0.59	0.59	547
	Macro Average	0.61	0.62	0.61	0.58	0.58	0.58	547
DEmpatica	**RF**				**SVM**			
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	**Support**
Complete Dataset	*No Stress*	0.64	0.79	0.71	0.63	0.67	0.65	238
	Stress	0.65	0.47	0.55	0.57	0.53	0.55	197
	Accuracy	-	-	0.65	-	-	0.61	435
	Weighted Average	0.65	0.63	0.63	0.60	0.61	0.61	435
	Macro Average	0.65	0.65	0.64	0.60	0.60	0.60	435
Complete Dataset (SMOTE)	*No Stress*	0.68	0.69	0.68	0.64	0.67	0.65	238
	Stress	0.62	0.60	0.61	0.58	0.55	0.56	197
	Accuracy	-	-	0.65	-	-	0.61	435
	Weighted Average	0.65	0.65	0.65	0.61	0.61	0.61	435
	Macro Average	0.65	0.65	0.65	0.61	0.61	0.61	435
Gender - Male	*No Stress*	0.69	0.87	0.77	0.74	0.83	0.78	82

	Stress	0.67	0.41	0.51	0.68	0.56	0.61	54
	Accuracy			0.68			0.72	136
	Weighted Average	0.68	0.68	0.66	0.72	0.72	0.71	136
	Macro Average	0.68	0.64	0.64	0.71	0.69	0.70	136
Gender – Male (SMOTE)	*No Stress*	0.73	0.78	0.75	0.66	0.91	0.77	82
	Stress	0.62	0.56	0.59	0.70	0.30	0.42	54
	Accuracy			0.69			0.67	136
	Weighted Average	0.69	0.69	0.69	0.68	0.61	0.59	136
	Macro Average	0.68	0.67	0.67	0.68	0.61	0.59	136
Gender – Female	*No Stress*	0.70	0.66	0.68	0.66	0.72	0.69	145
	Stress	0.66	0.70	0.68	0.67	0.60	0.63	137
	Accuracy			0.68			0.66	282
	Weighted Average	0.68	0.68	0.68	0.66	0.66	0.66	282
	Macro Average	0.68	0.68	0.68	0.66	0.66	0.66	282
Gender – Female (SMOTE)	*No Stress*	0.70	0.62	0.66	0.66	0.70	0.68	145
	Stress	0.64	0.72	0.68	0.66	0.61	0.64	137

	Accuracy			0.67			0.66	282
	Weighted Average	0.67	0.67	0.67	0.62	0.66	0.66	282
	Macro Average	0.67	0.67	0.67	0.62	0.66	0.66	282
Employment - Student	*No Stress*	0.71	0.66	0.68	0.70	0.62	0.66	96
	Stress	0.64	0.69	0.66	0.62	0.69	0.65	84
	Accuracy			0.67			0.66	180
	Weighted Average	0.67	0.67	0.67	0.66	0.66	0.66	180
	Macro Average	0.67	0.67	0.67	0.66	0.66	0.66	180
Employment – Student (SMOTE)	*No Stress*	0.72	0.66	0.69	0.64	0.77	0.70	96
	Stress	0.65	0.71	0.68	0.66	0.50	0.57	84
	Accuracy			0.68			0.64	180
	Weighted Average	0.69	0.68	0.68	0.65	0.64	0.64	180
	Macro Average	0.68	0.69	0.68	0.65	0.64	0.63	180
Employment – Worker	*No Stress*	0.66	0.73	0.70	0.67	0.77	0.72	122
	Stress	0.64	0.56	0.60	0.67	0.55	0.61	103
	Accuracy			0.65			0.67	225

	Weighted Average	0.65	0.65	0.65	0.67	0.66	0.66	225
	Macro Average	0.65	0.65	0.65	0.67	0.67	0.67	225
Employment – Worker (SMOTE)	*No Stress*	0.67	0.69	0.68	0.66	0.75	0.70	122
	Stress	0.62	0.60	0.61	0.65	0.54	0.59	103
	Accuracy			0.65			0.66	225
	Weighted Average	0.65	0.65	0.65	0.66	0.65	0.65	225
	Macro Average	0.65	0.65	0.65	0.66	0.66	0.65	225
Income - Low	*No Stress*	0.64	0.82	0.72	0.65	0.70	0.67	131
	Stress	0.55	0.33	0.41	0.50	0.44	0.47	89
	Accuracy			0.62			0.60	220
	Weighted Average	0.60	0.62	0.59	0.59	0.60	0.59	220
	Macro Average	0.59	0.57	0.56	0.57	0.57	0.57	220
Income – Low (SMOTE)	*No Stress*	0.66	0.63	0.64	0.64	0.86	0.74	131
	Stress	0.48	0.52	0.50	0.59	0.29	0.39	89
	Accuracy			0.58			0.63	220
	Weighted Average	0.59	0.58	0.58	0.62	0.63	0.60	220

	Macro Average	0.57	0.57	0.57	0.62	0.58	0.56	220
Income – Medium High	*No Stress*	0.64	0.66	0.65	0.67	0.62	0.64	107
	Stress	0.65	0.63	0.64	0.65	0.69	0.67	108
	Accuracy			0.65			0.66	215
	Weighted Average	0.65	0.65	0.65	0.66	0.66	0.66	215
	Macro Average	0.65	0.65	0.65	0.66	0.66	0.66	215
Income – Medium High (SMOTE)	*No Stress*	0.68	0.64	0.66	0.67	0.62	0.64	107
	Stress	0.67	0.70	0.68	0.65	0.69	0.67	108
	Accuracy			0.67			0.66	215
	Weighted Average	0.67	0.67	0.67	0.66	0.66	0.66	215
	Macro Average	0.67	0.67	0.67	0.66	0.66	0.66	215
Age – 18-24	*No Stress*	0.68	0.95	0.79	0.69	0.77	0.72	69
	Stress	0.50	0.10	0.17	0.39	0.30	0.34	30
	Accuracy			0.67			0.61	99
	Weighted Average	0.62	0.67	0.58	0.59	0.61	0.60	99
	Macro Average	0.59	0.53	0.48	0.54	0.53	0.53	99

Age – 18-24 (SMOTE)	*No Stress*	0.69	0.67	0.68	0.66	0.80	0.72	69
	Stress	0.38	0.40	0.39	0.29	0.17	0.21	30
	Accuracy			0.58			0.59	99
	Weighted Average	0.53	0.53	0.53	0.54	0.59	0.55	99
	Macro Average	0.58	0.58	0.58	0.48	0.48	0.47	99
Age 25-34	*No Stress*	0.79	0.44	0.56	0.64	0.52	0.57	71
	Stress	0.70	0.92	0.79	0.70	0.79	0.74	100
	Accuracy			0.72			0.68	171
	Weighted Average	0.74	0.72	0.70	0.67	0.68	0.67	171
	Macro Average	0.75	0.68	0.68	0.67	0.68	0.67	171
Age 25-34 (SMOTE)	*No Stress*	0.63	0.56	0.60	0.63	0.41	0.50	71
	Stress	0.71	0.77	0.74	0.66	0.83	0.74	100
	Accuracy			0.68			0.65	171
	Weighted Average	0.68	0.68	0.68	0.65	0.65	0.64	171
	Macro Average	0.67	0.67	0.67	0.65	0.62	0.62	171
Age 35-44	*No Stress*	0.65	0.95	0.77	0.72	0.88	0.79	59

	Stress	0.70	0.19	0.30	0.71	0.46	0.56	37
	Accuracy			0.66			0.72	96
	Weighted Average	0.67	0.66	0.59	0.72	0.72	0.70	96
	Macro Average	0.68	0.57	0.54	0.72	0.67	0.68	96
Age 35-44 (SMOTE)	*No Stress*	0.67	0.83	0.74	0.67	0.75	0.70	59
	Stress	0.57	0.35	0.43	0.50	0.41	0.45	37
	Accuracy			0.65			0.61	96
	Weighted Average	0.63	0.65	0.62	0.60	0.61	0.61	96
	Macro Average	0.62	0.59	0.59	0.58	0.61	0.61	168
Healthy	*No Stress*	0.65	0.81	0.72	0.66	0.76	0.71	186
	Stress	0.61	0.41	0.59	0.58	0.45	0.51	135
	Accuracy			0.65			0.63	321
	Weighted Average	0.64	0.64	0.63	0.63	0.63	0.62	321
	Macro Average	0.63	0.61	0.61	0.62	0.61	0.61	321
Healthy (SMOTE)	*No Stress*	0.69	0.70	0.69	0.66	0.66	0.66	186
	Stress	0.58	0.56	0.57	0.53	0.53	0.53	135

	Accuracy			0.64			0.60	321
	Weighted Average	0.64	0.64	0.64	0.60	0.60	0.60	321
	Macro Average	0.63	0.63	0.63	0.59	0.60	0.60	321

Table B5: Precision, Recall, F1-Score, Accuracy for Sleep Datasets, Generalized

SDA	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
Complete Dataset	*No Stress*	0.72	0.78	0.75	0.71	0.65	0.68	309
	Stress	0.68	0.61	0.65	0.59	0.65	0.62	238
	Accuracy	-	-	0.71	-	-	0.65	547
	Weighted Average	0.71	0.71	0.71	0.65	0.65	0.65	547
	Macro Average	0.70	0.70	0.70	0.65	0.65	0.65	547
Complete Dataset (SMOTE)	*No Stress*	0.77	0.76	0.76	0.68	0.82	0.75	309
	Stress	0.69	0.71	0.70	0.69	0.50	0.58	238
	Accuracy	-	-	0.73	-	-	0.68	547
	Weighted Average	0.74	0.73	0.74	0.68	0.68	0.67	547
	Macro Average	0.73	0.73	0.73	0.68	0.66	0.66	547
Gender - Male	*No Stress*	0.72	0.85	0.78	0.72	0.92	0.81	102

	Stress	0.58	0.39	0.47	0.69	0.33	0.45	54
	Accuracy			0.69			0.72	156
	Weighted Average	0.68	0.69	0.67	0.71	0.72	0.69	156
	Macro Average	0.65	0.62	0.63	0.71	0.63	0.63	156
Gender – Male (SMOTE)	*No Stress*	0.75	0.80	0.77	0.73	0.93	0.82	102
	Stress	0.57	0.48	0.52	0.72	0.33	0.46	54
	Accuracy			0.69			0.72	156
	Weighted Average	0.68	0.69	0.69	0.72	0.72	0.69	156
	Macro Average	0.66	0.64	0.65	0.72	0.63	0.64	156
Gender – Female	*No Stress*	0.72	0.70	0.71	0.68	0.64	0.66	186
	Stress	0.69	0.71	0.70	0.64	0.68	0.66	173
	Accuracy			0.71			0.66	359
	Weighted Average	0.71	0.71	0.71	0.66	0.66	0.66	359
	Macro Average	0.71	0.71	0.71	0.66	0.66	0.66	359
Gender – Female (SMOTE)	*No Stress*	0.73	0.70	0.71	0.64	0.75	0.69	186
	Stress	0.69	0.72	0.70	0.67	0.54	0.60	173

	Accuracy			0.71			0.65	359
	Weighted Average	0.71	0.71	0.71	0.65	0.65	0.64	359
	Macro Average	0.71	0.71	0.71	0.65	0.65	0.64	359
Employment - Student	*No Stress*	0.73	0.65	0.69	0.75	0.58	0.65	135
	Stress	0.65	0.73	0.69	0.62	0.79	0.70	121
	Accuracy			0.69			0.68	256
	Weighted Average	0.69	0.69	0.69	0.69	0.68	0.67	256
	Macro Average	0.69	0.69	0.69	0.69	0.68	0.67	256
Employment – Student (SMOTE)	*No Stress*	0.74	0.64	0.69	0.70	0.66	0.68	135
	Stress	0.65	0.75	0.70	0.64	0.69	0.66	121
	Accuracy			0.69			0.67	256
	Weighted Average	0.70	0.69	0.69	0.67	0.67	0.67	256
	Macro Average	0.70	0.69	0.69	0.67	0.67	0.67	256
Employment – Worker	*No Stress*	0.71	0.84	0.77	0.71	0.78	0.75	169
	Stress	0.65	0.47	0.54	0.59	0.50	0.55	107
	Accuracy			0.70			0.67	276

	Weighted Average	0.69	0.70	0.68	0.67	0.67	0.67	276
	Macro Average	0.68	0.65	0.66	0.65	0.64	0.65	276
Employment – Worker (SMOTE)	*No Stress*	0.74	0.77	0.76	0.68	0.86	0.76	169
	Stress	0.61	0.58	0.60	0.62	0.37	0.47	107
	Accuracy			0.70			0.67	276
	Weighted Average	0.69	0.70	0.69	0.66	0.67	0.65	276
	Macro Average	0.68	0.67	0.68	0.65	0.62	0.61	276
Income - Low	*No Stress*	0.74	0.81	0.78	0.68	0.70	0.69	161
	Stress	0.67	0.57	0.62	0.53	0.52	0.53	108
	Accuracy			0.72			0.62	269
	Weighted Average	0.71	0.72	0.71	0.62	0.62	0.62	269
	Macro Average	0.71	0.69	0.70	0.61	0.61	0.61	269
Income – Low (SMOTE)	*No Stress*	0.77	0.76	0.77	0.70	0.90	0.79	161
	Stress	0.65	0.67	0.66	0.74	0.43	0.54	108
	Accuracy			0.72			0.71	269
	Weighted Average	0.73	0.72	0.73	0.72	0.71	0.69	269

	Macro Average	0.71	0.72	0.71	0.72	0.66	0.66	269
Income – Medium High	*No Stress*	0.76	0.75	0.76	0.73	0.75	0.74	122
	Stress	0.73	0.74	0.74	0.72	0.70	0.71	111
	Accuracy			0.75			0.73	233
	Weighted Average	0.75	0.75	0.75	0.73	0.73	0.73	233
	Macro Average	0.75	0.75	0.75	0.72	0.72	0.72	233
Income – Medium High (SMOTE)	*No Stress*	0.76	0.69	0.72	0.69	0.68	0.68	122
	Stress	0.69	0.76	0.72	0.65	0.66	0.65	111
	Accuracy			0.72			0.67	233
	Weighted Average	0.72	0.72	0.72	0.67	0.67	0.67	233
	Macro Average	0.72	0.72	0.72	0.67	0.67	0.67	233
Age – 18-24	*No Stress*	0.74	0.83	0.78	0.73	0.65	0.69	103
	Stress	0.68	0.57	0.62	0.55	0.64	0.59	69
	Accuracy			0.72			0.65	172
	Weighted Average	0.72	0.70	0.70	0.64	0.64	0.64	172
	Macro Average	0.71	0.70	0.70	0.66	0.65	0.65	172

Age – 18-24 (SMOTE)	*No Stress*	0.77	0.75	0.76	0.69	0.83	0.75	103
	Stress	0.64	0.67	0.65	0.64	0.43	0.52	69
	Accuracy			0.72			0.67	172
	Weighted Average	0.72	0.72	0.72	0.67	0.67	0.66	172
	Macro Average	0.70	0.71	0.71	0.66	0.63	0.64	172
Age 25-34	*No Stress*	0.65	0.53	0.59	0.58	0.41	0.48	68
	Stress	0.73	0.82	0.77	0.68	0.81	0.74	105
	Accuracy			0.71			0.65	173
	Weighted Average	0.70	0.71	0.70	0.64	0.65	0.64	173
	Macro Average	0.69	0.67	0.68	0.63	0.61	0.61	173
Age 25-34 (SMOTE)	*No Stress*	0.59	0.65	0.62	0.67	0.41	0.51	68
	Stress	0.76	0.71	0.74	0.69	0.87	0.77	105
	Accuracy			0.69			0.69	173
	Weighted Average	0.68	0.68	0.68	0.68	0.69	0.67	173
	Macro Average	0.69	0.69	0.69	0.68	0.64	0.64	173
Age 35-44	*No Stress*	0.78	0.95	0.85	0.79	0.84	0.81	98

	Stress	0.76	0.37	0.50	0.57	0.49	0.53	43
	Accuracy			0.77			0.73	141
	Weighted Average	0.77	0.77	0.75	0.72	0.73	0.72	141
	Macro Average	0.77	0.66	0.68	0.68	0.66	0.67	141
Age 35-44 (SMOTE)	*No Stress*	0.80	0.82	0.81	0.75	0.91	0.82	98
	Stress	0.56	0.53	0.55	0.59	0.30	0.40	43
	Accuracy			0.73			0.72	141
	Weighted Average	0.73	0.73	0.73	0.70	0.72	0.69	141
	Macro Average	0.68	0.68	0.68	0.67	0.61	0.61	141
Healthy	*No Stress*	0.75	0.83	0.79	0.71	0.74	0.73	254
	Stress	0.70	0.60	0.65	0.60	0.56	0.58	174
	Accuracy			0.73			0.67	428
	Weighted Average	0.73	0.73	0.73	0.66	0.65	0.65	428
	Macro Average	0.73	0.71	0.72	0.67	0.67	0.67	428
Healthy - SMOTE	*No Stress*	0.77	0.77	0.77	0.71	0.70	0.70	254
	Stress	0.66	0.66	0.66	0.57	0.58	0.58	174

	Accuracy			0.72			0.65	428
	Weighted Average	0.72	0.72	0.72	0.65	0.65	0.65	428
	Macro Average	0.71	0.71	0.71	0.64	0.64	0.64	428
SDAW	**RF**				**SVM**			
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
Complete Dataset	*No Stress*	0.75	0.82	0.79	0.72	0.80	0.76	249
	Stress	0.72	0.63	0.67	0.67	0.58	0.62	179
	Accuracy	-	-	0.74	-	-	0.70	428
	Weighted Average	0.74	0.74	0.74	0.70	0.70	0.70	428
	Macro Average	0.74	0.72	0.73	0.70	0.69	0.69	428
Complete Dataset (SMOTE)	*No Stress*	0.78	0.75	0.76	0.76	0.76	0.76	249
	Stress	0.67	0.70	0.68	0.66	0.66	0.66	179
	Accuracy	-	-	0.73	-	-	0.72	428
	Weighted Average	0.73	0.73	0.73	0.72	0.72	0.72	428
	Macro Average	0.72	0.72	0.72	0.71	0.71	0.71	428
Gender - Male	*No Stress*	0.83	0.93	0.88	0.83	0.88	0.85	73
	Stress	0.29	0.12	0.17	0.25	0.19	0.21	16

	Accuracy			0.79			0.75	89
	Weighted Average	0.73	0.79	0.75	0.73	0.75	0.74	89
	Macro Average	0.56	0.53	0.53	0.54	0.53	0.53	89
Gender – Male (SMOTE)	*No Stress*	0.83	0.88	0.85	0.82	0.99	0.89	73
	Stress	0.25	0.19	0.21	0.00	0.00	0.00	16
	Accuracy			0.75			0.81	89
	Weighted Average	0.73	0.75	0.74	0.67	0.81	0.73	89
	Macro Average	0.54	0.53	0.53	0.41	0.49	0.45	89
Gender – Female	*No Stress*	0.70	0.73	0.72	0.68	0.67	0.68	166
	Stress	0.70	0.68	0.69	0.66	0.66	0.66	157
	Accuracy			0.70			0.67	323
	Weighted Average	0.70	0.70	0.70	0.67	0.67	0.67	323
	Macro Average	0.70	0.70	0.70	0.67	0.67	0.67	323
Gender – Female (SMOTE)	*No Stress*	0.71	0.72	0.72	0.68	0.63	0.65	166
	Stress	0.70	0.69	0.70	0.64	0.69	0.66	157
	Accuracy			0.71			0.66	323

	Weighted Average	0.71	0.71	0.71	0.66	0.66	0.66	323
	Macro Average	0.71	0.71	0.71	0.66	0.66	0.66	323
Employment - Student	*No Stress*	0.75	0.75	0.75	0.74	0.66	0.70	99
	Stress	0.73	0.73	0.73	0.67	0.75	0.71	92
	Accuracy			0.74			0.70	191
	Weighted Average	0.74	0.74	0.74	0.71	0.70	0.70	191
	Macro Average	0.74	0.74	0.74	0.70	0.70	0.70	191
Employment – Student (SMOTE)	*No Stress*	0.76	0.73	0.74	0.75	0.74	0.74	99
	Stress	0.72	0.75	0.73	0.72	0.74	0.73	92
	Accuracy			0.74			0.74	191
	Weighted Average	0.74	0.74	0.74	0.74	0.74	0.74	191
	Macro Average	0.74	0.74	0.74	0.74	0.74	0.74	191
Employment – Worker	*No Stress*	0.74	0.84	0.79	0.74	0.78	0.76	151
	Stress	0.64	0.49	0.56	0.58	0.52	0.55	87
	Accuracy			0.71			0.68	238
	Weighted Average	0.71	0.71	0.70	0.68	0.68	0.68	238

	Macro Average	0.69	0.67	0.67	0.66	0.65	0.65	238
Employment – Worker (SMOTE)	*No Stress*	0.77	0.81	0.79	0.73	0.73	0.73	151
	Stress	0.64	0.59	0.61	0.53	0.53	0.53	87
	Accuracy			0.73			0.66	238
	Weighted Average	0.72	0.73	0.72	0.66	0.66	0.66	238
	Macro Average	0.70	0.70	0.70	0.63	0.63	0.63	238
Income - Low	*No Stress*	0.80	0.82	0.81	0.78	0.76	0.77	119
	Stress	0.69	0.67	0.68	0.62	0.64	0.63	73
	Accuracy			0.76			0.71	192
	Weighted Average	0.76	0.76	0.76	0.72	0.71	0.71	192
	Macro Average	0.75	0.74	0.74	0.70	0.70	0.70	192
Income – Low (SMOTE)	*No Stress*	0.83	0.75	0.79	0.80	0.73	0.76	119
	Stress	0.65	0.75	0.70	0.61	0.70	0.65	73
	Accuracy			0.75			0.72	192
	Weighted Average	0.76	0.75	0.75	0.73	0.72	0.72	192
	Macro Average	0.74	0.75	0.74	0.71	0.71	0.71	192

Income – Medium High	*No Stress*	0.71	0.75	0.73	0.72	0.75	0.73	110
	Stress	0.64	0.77	0.70	0.66	0.62	0.64	86
	Accuracy			0.68			0.69	196
	Weighted Average	0.69	0.68	0.67	0.69	0.69	0.69	196
	Macro Average	0.68	0.68	0.67	0.69	0.69	0.69	196
Income – Medium High (SMOTE)	*No Stress*	0.70	0.73	0.71	0.71	0.73	0.72	110
	Stress	0.63	0.60	0.62	0.64	0.63	0.64	86
	Accuracy			0.67			0.68	196
	Weighted Average	0.67	0.67	0.67	0.68	0.68	0.68	196
	Macro Average	0.67	0.67	0.67	0.68	0.68	0.68	196
Age – 18-24	*No Stress*	0.79	0.79	0.79	0.80	0.73	0.76	66
	Stress	0.72	0.72	0.72	0.68	0.76	0.72	50
	Accuracy			0.76			0.74	116
	Weighted Average	0.76	0.76	0.76	0.75	0.74	0.74	116
	Macro Average	0.75	0.75	0.75	0.74	0.74	0.74	116
Age – 18-24 (SMOTE)	*No Stress*	0.78	0.74	0.76	0.79	0.76	0.78	66

	Stress	0.68	0.72	0.70	0.70	0.74	0.72	50
	Accuracy			0.73			0.75	116
	Weighted Average	0.74	0.73	0.73	0.75	0.75	0.75	116
	Macro Average	0.73	0.73	0.73	0.75	0.75	0.75	116
Age 25-34	*No Stress*	0.70	0.63	0.66	0.64	0.57	0.60	67
	Stress	0.71	0.78	0.74	0.67	0.74	0.70	80
	Accuracy			0.71			0.66	147
	Weighted Average	0.71	0.71	0.71	0.66	0.66	0.66	147
	Macro Average	0.71	0.70	0.71	0.66	0.65	0.65	147
Age 25-34 (SMOTE)	*No Stress*	0.68	0.63	0.65	0.64	0.67	0.66	67
	Stress	0.71	0.75	0.73	0.71	0.69	0.70	80
	Accuracy			0.69			0.68	147
	Weighted Average	0.69	0.69	0.69	0.68	0.68	0.68	147
	Macro Average	0.69	0.69	0.69	0.68	0.68	0.68	147
Age 35-44	*No Stress*	0.81	0.95	0.88	0.84	0.89	0.86	82
	Stress	0.64	0.28	0.39	0.55	0.44	0.49	25

	Accuracy			0.79			0.79	107
	Weighted Average	0.77	0.79	0.76	0.77	0.79	0.78	107
	Macro Average	0.72	0.62	0.63	0.69	0.67	0.68	107
Age 35-44 (SMOTE)	*No Stress*	0.83	0.88	0.85	0.78	0.98	0.87	82
	Stress	0.50	0.40	0.44	0.60	0.12	0.20	25
	Accuracy			0.77			0.78	107
	Weighted Average	0.75	0.77	0.76	0.74	0.78	0.71	107
	Macro Average	0.66	0.64	0.65	0.69	0.55	0.53	107
Healthy	*No Stress*	0.75	0.85	0.80	0.76	0.75	0.75	195
	Stress	0.67	0.52	0.58	0.58	0.60	0.59	114
	Accuracy			0.73			0.69	309
	Weighted Average	0.72	0.73	0.72	0.69	0.69	0.69	309
	Macro Average	0.71	0.68	0.69	0.67	0.67	0.67	309
Healthy (SMOTE)	*No Stress*	0.79	0.74	0.77	0.79	0.76	0.77	195
	Stress	0.60	0.66	0.63	0.61	0.65	0.63	114
	Accuracy			0.71			0.72	309

	Weighted Average	0.72	0.71	0.71	0.72	0.72	0.72	309
	Macro Average	0.69	0.70	0.70	0.70	0.70	0.70	309
SDW		**RF**			**SVM**			
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
Complete Dataset	*No Stress*	0.73	0.76	0.74	0.62	0.73	0.67	307
	Stress	0.65	0.62	0.64	0.56	0.44	0.49	228
	Accuracy	-	-	0.70	-	-	0.60	535
	Weighted Average	0.70	0.70	0.70	0.60	0.60	0.59	535
	Macro Average	0.69	0.69	0.69	0.59	0.58	0.58	535
Complete Dataset (SMOTE)	*No Stress*	0.72	0.74	0.73	0.73	0.74	0.73	307
	Stress	0.64	0.61	0.62	0.64	0.63	0.63	228
	Accuracy	-	-	0.68	-	-	0.69	535
	Weighted Average	0.68	0.68	0.68	0.69	0.69	0.69	535
	Macro Average	0.68	0.67	0.67	0.68	0.68	0.68	535
Gender - Male	*No Stress*	0.80	0.96	0.88	0.83	0.93	0.88	85
	Stress	0.90	0.58	0.71	0.84	0.67	0.74	48
	Accuracy			0.83			0.83	133

	Weighted Average	0.84	0.83	0.82	0.84	0.83	0.83	133
	Macro Average	0.85	0.77	0.79	0.84	0.80	0.81	133
Gender – Male (SMOTE)	*No Stress*	0.83	0.91	0.87	0.86	0.86	0.86	85
	Stress	0.80	0.67	0.73	0.75	0.75	0.75	48
	Accuracy			0.82			0.82	133
	Weighted Average	0.82	0.82	0.82	0.82	0.82	0.82	133
	Macro Average	0.81	0.79	0.80	0.80	0.80	0.80	133
Gender – Female	*No Stress*	0.70	0.75	0.73	0.68	0.73	0.70	212
	Stress	0.67	0.61	0.64	0.64	0.59	0.61	174
	Accuracy			0.69			0.67	386
	Weighted Average	0.69	0.69	0.69	0.66	0.67	0.66	386
	Macro Average	0.69	0.68	0.68	0.66	0.66	0.66	386
Gender – Female (SMOTE)	*No Stress*	0.70	0.71	0.71	0.70	0.64	0.67	212
	Stress	0.64	0.64	0.64	0.60	0.66	0.63	174
	Accuracy			0.68			0.65	386
	Weighted Average	0.68	0.68	0.68	0.65	0.65	0.65	386

	Macro Average	0.67	0.67	0.67	0.65	0.65	0.65	386
Employment - Student	*No Stress*	0.74	0.68	0.71	0.74	0.75	0.74	99
	Stress	0.68	0.74	0.71	0.73	0.72	0.72	92
	Accuracy			0.71			0.73	191
	Weighted Average	0.71	0.71	0.71	0.73	0.73	0.73	191
	Macro Average	0.71	0.71	0.71	0.73	0.73	0.73	191
Employment – Student (SMOTE)	*No Stress*	0.73	0.67	0.69	0.75	0.73	0.74	99
	Stress	0.67	0.73	0.70	0.72	0.74	0.73	92
	Accuracy			0.70			0.73	191
	Weighted Average	0.70	0.70	0.73	0.73	0.73	0.73	191
	Macro Average	0.70	0.70	0.73	0.73	0.73	0.73	191
Employment – Worker	*No Stress*	0.72	0.76	0.74	0.71	0.74	0.72	195
	Stress	0.62	0.57	0.60	0.59	0.55	0.57	133
	Accuracy			0.69			0.66	328
	Weighted Average	0.68	0.69	0.68	0.66	0.66	0.66	328
	Macro Average	0.67	0.67	0.67	0.65	0.65	0.65	328

Employment – Worker (SMOTE)	*No Stress*	0.74	0.74	0.74	0.71	0.72	0.72	195
	Stress	0.62	0.61	0.61	0.58	0.56	0.57	133
	Accuracy			0.69			0.66	328
	Weighted Average	0.69	0.69	0.69	0.66	0.66	0.66	328
	Macro Average	0.68	0.68	0.68	0.64	0.64	0.64	328
Income - Low	*No Stress*	0.73	0.85	0.78	0.74	0.82	0.78	144
	Stress	0.66	0.49	0.56	0.64	0.53	0.58	88
	Accuracy			0.71			0.71	232
	Weighted Average	0.70	0.71	0.70	0.70	0.71	0.70	232
	Macro Average	0.70	0.67	0.67	0.69	0.68	0.68	232
Income – Low (SMOTE)	*No Stress*	0.76	0.78	0.77	0.76	0.76	0.76	144
	Stress	0.62	0.60	0.61	0.61	0.61	0.61	88
	Accuracy			0.71			0.70	232
	Weighted Average	0.71	0.71	0.71	0.70	0.70	0.70	232
	Macro Average	0.69	0.69	0.69	0.68	0.68	0.68	232
Income – Medium High	*No Stress*	0.78	0.65	0.71	0.79	0.69	0.74	149

	Stress	0.65	0.78	0.71	0.68	0.78	0.72	125
	Accuracy			0.71			0.73	274
	Weighted Average	0.72	0.71	0.71	0.74	0.73	0.73	274
	Macro Average	0.72	0.72	0.71	0.73	0.73	0.73	274
Income – Medium High (SMOTE)	*No Stress*	0.79	0.64	0.71	0.80	0.66	0.73	149
	Stress	0.65	0.79	0.71	0.67	0.80	0.73	125
	Accuracy			0.71			0.73	274
	Weighted Average	0.73	0.71	0.71	0.74	0.73	0.73	274
	Macro Average	0.72	0.72	0.71	0.73	0.73	0.73	274
Age – 18-24	*No Stress*	0.75	0.66	0.70	0.82	0.66	0.73	80
	Stress	0.60	0.69	0.65	0.64	0.80	0.71	59
	Accuracy			0.68			0.72	139
	Weighted Average	0.69	0.68	0.68	0.74	0.72	0.72	139
	Macro Average	0.67	0.68	0.67	0.73	0.73	0.72	139
Age – 18-24 (SMOTE)	*No Stress*	0.73	0.64	0.68	0.75	0.64	0.69	80
	Stress	0.58	0.68	0.63	0.59	0.71	0.65	59

	Accuracy			0.65			0.67	139
	Weighted Average	0.67	0.65	0.66	0.68	0.67	0.67	139
	Macro Average	0.65	0.66	0.65	0.67	0.67	0.67	139
Age 25-34	*No Stress*	0.70	0.69	0.70	0.68	0.82	0.74	72
	Stress	0.76	0.77	0.76	0.83	0.69	0.75	90
	Accuracy			0.73			0.75	162
	Weighted Average	0.73	0.73	0.73	0.76	0.75	0.75	162
	Macro Average	0.73	0.73	0.73	0.75	0.75	0.75	162
Age 25-34 (SMOTE)	*No Stress*	0.63	0.79	0.70	0.67	0.83	0.74	72
	Stress	0.79	0.63	0.70	0.83	0.67	0.74	90
	Accuracy			0.70			0.74	162
	Weighted Average	0.72	0.70	0.70	0.76	0.74	0.74	162
	Macro Average	0.71	0.71	0.70	0.75	0.75	0.74	162
Age 35-44	*No Stress*	0.78	0.91	0.84	0.78	0.91	0.84	99
	Stress	0.74	0.50	0.60	0.75	0.52	0.61	52
	Accuracy			0.77			0.77	151

	Weighted Average	0.76	0.77	0.75	0.77	0.77	0.76	151
	Macro Average	0.76	0.70	0.72	0.77	0.71	0.73	151
Age 35-44 (SMOTE)	*No Stress*	0.79	0.88	0.83	0.78	0.77	0.78	99
	Stress	0.71	0.56	0.62	0.57	0.60	0.58	52
	Accuracy			0.77			0.71	151
	Weighted Average	0.76	0.77	0.76	0.71	0.71	0.71	151
	Macro Average	0.75	0.72	0.73	0.68	0.68	0.68	151
Healthy	*No Stress*	0.75	0.86	0.80	0.75	0.80	0.78	243
	Stress	0.73	0.56	0.63	0.66	0.59	0.62	157
	Accuracy			0.74			0.72	400
	Weighted Average	0.74	0.74	0.74	0.72	0.72	0.72	400
	Macro Average	0.74	0.71	0.72	0.71	0.70	0.70	400
Healthy (SMOTE)	*No Stress*	0.75	0.80	0.78	0.80	0.75	0.77	243
	Stress	0.66	0.59	0.62	0.65	0.71	0.67	157
	Accuracy			0.72			0.73	400
	Weighted Average	0.72	0.72	0.72	0.74	0.73	0.73	400

	Macro Average	0.71	0.70	0.70	0.72	0.73	0.72	400
SDS	**RF**				**SVM**			
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	***Support***
Complete Dataset	*No Stress*	0.73	0.81	0.77	0.73	0.78	0.76	249
	Stress	0.69	0.59	0.64	0.66	0.60	0.63	179
	Accuracy	-	-	0.72	-	-	0.71	428
	Weighted Average	0.72	0.72	0.72	0.70	0.71	0.70	428
	Macro Average	0.71	0.70	0.70	0.70	0.69	0.69	428
Complete Dataset (SMOTE)	*No Stress*	0.78	0.69	0.73	0.78	0.72	0.75	249
	Stress	0.63	0.73	0.68	0.65	0.71	0.68	179
	Accuracy	-	-	0.71	-	-	0.72	428
	Weighted Average	0.72	0.71	0.71	0.72	0.72	0.72	428
	Macro Average	0.70	0.71	0.70	0.71	0.72	0.71	428
Gender - Male	*No Stress*	0.81	0.86	0.83	0.81	0.85	0.83	73
	Stress	0.09	0.06	0.07	0.08	0.06	0.07	16
	Accuracy			0.72			0.71	89
	Weighted Average	0.68	0.72	0.70	0.68	0.71	0.69	89

	Macro Average	0.45	0.46	0.45	0.44	0.46	0.45	89
Gender – Male (SMOTE)	*No Stress*	0.81	0.82	0.82	0.84	0.59	0.69	73
	Stress	0.13	0.12	0.13	0.21	0.50	0.30	16
	Accuracy			0.70			0.57	89
	Weighted Average	0.69	0.70	0.69	0.73	0.57	0.62	89
	Macro Average	0.47	0.47	0.47	0.53	0.54	0.49	89
Gender – Female	*No Stress*	0.72	0.72	0.72	0.71	0.74	0.72	160
	Stress	0.71	0.70	0.70	0.71	0.68	0.70	153
	Accuracy			0.71			0.71	313
	Weighted Average	0.71	0.71	0.71	0.71	0.71	0.71	313
	Macro Average	0.71	0.71	0.71	0.71	0.71	0.71	313
Gender – Female (SMOTE)	*No Stress*	0.72	0.72	0.72	0.71	0.72	0.72	160
	Stress	0.71	0.71	0.71	0.71	0.69	0.70	153
	Accuracy			0.71			0.71	313
	Weighted Average	0.71	0.71	0.71	0.71	0.71	0.71	313
	Macro Average	0.71	0.71	0.71	0.71	0.71	0.71	313

Employment - Student	*No Stress*	0.76	0.71	0.73	0.75	0.74	0.74	99
	Stress	0.71	0.76	0.73	0.72	0.74	0.73	92
	Accuracy			0.73			0.74	191
	Weighted Average	0.73	0.73	0.73	0.74	0.74	0.74	191
	Macro Average	0.73	0.73	0.73	0.74	0.74	0.74	191
Employment – Student (SMOTE)	*No Stress*	0.75	0.66	0.70	0.77	0.72	0.74	99
	Stress	0.67	0.76	0.71	0.72	0.77	0.74	92
	Accuracy			0.71			0.74	191
	Weighted Average	0.71	0.71	0.71	0.75	0.74	0.74	191
	Macro Average	0.71	0.71	0.71	0.74	0.74	0.74	191
Employment – Worker	*No Stress*	0.73	0.85	0.79	0.73	0.79	0.76	151
	Stress	0.64	0.45	0.53	0.58	0.51	0.54	87
	Accuracy			0.71			0.68	238
	Weighted Average	0.70	0.71	0.69	0.68	0.68	0.68	238
	Macro Average	0.68	0.65	0.66	0.66	0.65	0.65	238
	No Stress	0.73	0.72	0.72	0.73	0.68	0.70	151

Employment – Worker (SMOTE)	*Stress*	0.53	0.55	0.54	0.50	0.56	0.53	87
	Accuracy			0.66			0.63	238
	Weighted Average	0.66	0.66	0.66	0.65	0.63	0.64	238
	Macro Average	0.63	0.63	0.63	0.61	0.62	0.62	238
Income - Low	*No Stress*	0.76	0.78	0.77	0.78	0.78	0.78	125
	Stress	0.63	0.60	0.61	0.64	0.64	0.64	78
	Accuracy			0.71			0.72	203
	Weighted Average	0.71	0.71	0.71	0.72	0.72	0.72	203
	Macro Average	0.69	0.69	0.69	0.71	0.71	0.71	203
Income – Low (SMOTE)	*No Stress*	0.81	0.76	0.79	0.82	0.74	0.78	125
	Stress	0.65	0.72	0.68	0.64	0.74	0.69	78
	Accuracy			0.74			0.74	203
	Weighted Average	0.75	0.74	0.75	0.75	0.74	0.74	203
	Macro Average	0.73	0.74	0.73	0.73	0.74	0.73	203
Income – Medium High	*No Stress*	0.71	0.69	0.70	0.70	0.71	0.71	110
	Stress	0.62	0.64	0.63	0.62	0.62	0.62	86

	Accuracy			0.67			0.67	196
	Weighted Average	0.67	0.67	0.67	0.67	0.67	0.67	196
	Macro Average	0.66	0.67	0.66	0.66	0.66	0.66	196
Income – Medium High (SMOTE)	*No Stress*	0.69	0.65	0.67	0.70	0.68	0.69	110
	Stress	0.58	0.62	0.60	0.61	0.63	0.62	86
	Accuracy			0.64			0.66	196
	Weighted Average	0.64	0.64	0.64	0.66	0.66	0.66	196
	Macro Average	0.63	0.64	0.63	0.65	0.65	0.65	196
Age – 18-24	*No Stress*	0.72	0.62	0.67	0.66	0.61	0.63	71
	Stress	0.58	0.69	0.63	0.54	0.60	0.57	55
	Accuracy			0.65			0.60	126
	Weighted Average	0.66	0.65	0.65	0.61	0.60	0.60	126
	Macro Average	0.65	0.66	0.65	0.60	0.60	0.60	126
Age – 18-24 (SMOTE)	*No Stress*	0.74	0.59	0.66	0.67	0.62	0.64	71
	Stress	0.58	0.73	0.65	0.55	0.60	0.57	55
	Accuracy			0.65			0.61	126

	Weighted Average	0.67	0.65	0.65	0.63	0.62	0.61	126
	Macro Average	0.66	0.66	0.65	0.62	0.61	0.61	126
Age 25-34	*No Stress*	0.67	0.58	0.62	0.67	0.58	0.62	67
	Stress	0.69	0.76	0.72	0.69	0.76	0.72	80
	Accuracy			0.68			0.68	147
	Weighted Average	0.68	0.68	0.68	0.68	0.68	0.68	147
	Macro Average	0.68	0.67	0.67	0.68	0.67	0.67	147
Age 25-34 (SMOTE)	*No Stress*	0.67	0.63	0.65	0.67	0.64	0.66	67
	Stress	0.70	0.74	0.72	0.71	0.74	0.72	80
	Accuracy			0.69			0.69	147
	Weighted Average	0.69	0.69	0.69	0.69	0.69	0.69	147
	Macro Average	0.68	0.68	0.68	0.69	0.69	0.69	147
Age 35-44	*No Stress*	0.81	0.87	0.84	0.81	0.84	0.83	82
	Stress	0.42	0.32	0.46	0.41	0.36	0.38	25
	Accuracy			0.74			0.73	107
	Weighted Average	0.72	0.74	0.73	0.72	0.73	0.72	107

	Macro Average	0.61	0.59	0.60	0.61	0.60	0.60	107
Age 35-44 (SMOTE)	*No Stress*	0.84	0.80	0.82	0.80	0.60	0.69	82
	Stress	0.43	0.48	0.45	0.28	0.52	0.37	25
	Accuracy			0.73			0.58	107
	Weighted Average	0.74	0.73	0.73	0.68	0.58	0.61	107
	Macro Average	0.63	0.64	0.64	0.54	0.56	0.53	107
Healthy	*No Stress*	0.71	0.71	0.71	0.67	0.82	0.74	55
	Stress	0.75	0.75	0.75	0.81	0.66	0.73	65
	Accuracy			0.73			0.73	120
	Weighted Average	0.73	0.73	0.73	0.74	0.74	0.73	120
	Macro Average	0.73	0.73	0.73	0.75	0.73	0.73	120
Healthy (SMOTE)	*No Stress*	0.67	0.82	0.74	0.68	0.82	0.74	55
	Stress	0.81	0.66	0.73	0.81	0.68	0.74	65
	Accuracy			0.73			0.74	120
	Weighted Average	0.75	0.73	0.73	0.75	0.74	0.74	120
	Macro Average	0.74	0.74	0.73	0.75	0.75	0.74	120

Table B6: Review of Stress Prediction ML Studies

Study	Year	Type	Device	Variables	Size	Period	Ground Truth	Model	G/I	Accuracy	Validation
46	2013	DDSR	Wahoo chest belt	Audio, Physical Activity, HRV, Communication	35	4 Months	PANAS self-report questionnaire + audio self-report + stress self-assessment before sleep	LR	G/I	Generalized: 53% Individualized: 61%	Generalized Leave One Person Out Individualized Leave One Day Out
42	2010	LLKC	Emotion Board	EDA	33	N/A	Known context	LDA, SVM, NCC	G	83% (LDA), ˜81% (SVM)	Leave One Person Out
45	2020	DDSR	Empatica E4	EDA, BVP, Acc, HR, Skin Temperature	6	4 Weeks	Button Press from device indicating a stressful event	RF, SVM	G	87.4% (RF), 82.1% (SVM) - calculated as AUC	10-fold Cross Validation, testing on 10% validation set
204	2012	LLSR	Polar WearLink+, SA9311M, AgCI electrodes	HRV, Respiration, EDA	10	N/A	7-point LIKERT scale on perceived stress levels	LR	I	81%	Leave One Person Out

169	2020	Mix of different modalities	Empatica E4	EDA, HRV, HR	14	1 week	DDSR: PSS-5 and EMA LDSR: PSS-5 and known context.	MLP, RF, kNN, SVM, LR	G	DDSR: 68% (SVM), 64% (LR), 60% (kNN), 52% (RF) LDSR: 74.6% (RF), 73.8% (LR), 73.4% (SVM), kNN AND MLP (72.2%)	10-fold CV with data from all users 80% training, 20% testing
167	2014	DDSR	BioHarness 3.0 + , Empatica E4	ECG, HRV, Resp, Temp, GSR, Posture, Accelerometer, Sleep	10	18 days	2 questions before sleep on how the participant felt during the day	SVM, LR, kNN, RF, NN	G	73% (SVM), 71% (RF), NN (˜63%), kNN (˜60%), LR (52%)	Leave One Person Out
181	2017	LLKC	NeuLog (ECG, GSR, Resp), Contec PM50 (BP), Kenek Edge pulse oximeter (SpO2)	ECG, Resp, GSR, BP, SpO2	32	N/A	Known context	SVM, kNN	G/I	Generalized: 89.2% (kNN), 83.1% (SVM) Individualized: 94.5% (kNN), 86.7% (SVM)	Train, Test, Validation

166	2015	LLKC and LLSR	Sensor suite similar to the BioHarness	ECG, Resp	20	1 week	Known context in lab, PSS-5 with EMA in the field	SVM	G	LLKC: 90% LLSR: 72%	Leave One Person Out
168	2016	LLKC and LLSR	Empatica E4	BVP, HR, HRV, ST and GSR	5	55 days	Known context and Short STAY-Y Questionnaire in the laboratory, 4-6 random EMA prompts asking users for the duration and the level of stress in the field	RF for laboratory stress followed by SVM for real-life data	I	92% (SVM)	Leave One Person Out
182	2018	LLKC	NeuroSky Mindwave Mobile	EEG	7	N/A	Known context while listening to music (meditation and attention instead of stress/non-stress)	Neural Network	G	0.60 (Attention), 0.01 (Meditation) – calculated as F1-score	Train, Test and Validation

205	2013	LLSR	Androind Phone + Wristworn Sensor	ACC, Skin Conductance, Phone Usage	18	5 days	PSS, PSQI, Big Five Inventory Personality Test	SVM, kNN	G	Over 75% (with a number of trained model and feature combinations)	10-fold CV with data from all users 90% training, 10% testing
195	2015	DDSR	Samsung Galaxy SIII Mini	ACC	30	8 weeks	OLBI + 3 EMA prompts per day asking users to rate stress on a 1-5 scale	Naïve Bayes, DT, Ordinal Naïve Bayes	G/I	Generalized: Accuracy - 52% MAE – 0.83 RMSE – 1 Individualized: Accuracy - 71% MAE – 0.66 RMSE – 0.96	5-fold CV for Individualized, Leave One Person Out for Generalized
206	2019	DDSR and DDKC (in a structured programming contest)	Samsung Gear S1, S2 and S3, Empatica E4	HRV, ACC, EDA	21	9 days	Frustration collected in the NASA-TLX questionnaire + 0-100 scale with question, 3 class stress	PCA + LDA, SVM, LR, RF, Multilayer Perceptrom,	G/I	Generalized: 88.20% (RF) with DDKC, 86.38% with DDSR Individualized: 97.92% (RF)	Train amd Test for Individualized, 10-fold CV for Generalized
184	2018	LLKC	BioPatch M3	HRV	128	N/A	Known Context (, Stroop Colour Word Test, videogames)	SVM	G	64%	10-fold CV
207	2016	LLKC	Emotiv Epoch headset	EEG	6	N/A	Mathematical Questions, self-report with the NASA-TLX	SVM, LDA, QDA, kNN	G	89% (SVM)	10-fold CV

208	2015	LLKC	BioNomadix model BN-PPGED	EDA, HRV	5	N/A	State Trait Anxiety Inventory, Triet Social Stress Test	SVM	I	For each participant: 78.90%, 73.26%, 83.08%, 82.82%, 76.83%	Train and Test (75% and 25%)

Table B7: Precision, Recall, F1-Score, Accuracy for Datasets, Generalized_Imb

D (n = 22)	**RF**			**SVM**		
Items	**F1-Weighted**	**F1-Macro**	***Accuracy***	**F1- Weighted**	**F1-Macro**	***Accuracy***
Full	0.48	0.49	0.49	0.53	0.52	0.53
Gender Male	0.43	0.42	0.48	0.51	0.49	0.52
Gender Female	0.53	0.51	0.55	0.52	0.51	0.52
Income Low	0.50	0.46	0.54	0.52	0.48	0.54
Income Medium High	0.43	0.40	0.50	0.50	0.48	0.51
Employment Students	0.44	0.44	0.46	0.47	0.47	0.47
Employment Workers	0.40	0.40	0.45	0.52	0.51	0.52
Age 18-24	0.54	0.42	0.64	0.60	0.49	0.64
Age 25-34	0.53	0.44	0.62	0.56	0.50	0.57
Age 35-44	0.37	0.33	0.39	0.50	0.40	0.47
Healthy	0.52	0.51	0.53	0.55	0.54	0.54
D SMOTE (n = 22)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1- Weighted**	**F1-Macro**	***Accuracy***
Full	0.50	0.50	0.50	0.46	0.47	0.50
Gender Male	0.47	0.46	0.50	0.39	0.39	0.49
Gender Female	0.53	0.51	0.55	0.48	0.48	0.50
Income Low	0.54	0.51	0.54	0.46	0.41	0.58
Income Medium High	0.43	0.40	0.48	0.45	0.42	0.51
Employment Students	0.46	0.46	0.48	0.41	0.42	0.48
Employment Workers	0.42	0.42	0.45	0.42	0.42	0.50
Age 18-24	0.62	0.51	0.65	0.54	0.41	0.66
Age 25-34	0.56	0.50	0.58	0.51	0.40	0.64
Age 35-44	0.44	0.39	0.42	0.42	0.33	0.48
Healthy	0.54	0.53	0.55	0.44	0.44	0.51
DECG (n = 42)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-weighted**	**F1-Macro**	***Accuracy***
Full	0.51	0.50	0.52	0.52	0.50	0.53
Gender Male	0.48	0.45	0.50	0.51	0.48	0.52
Gender Female	0.53	0.52	0.53	0.52	0.51	0.52
Income Low	0.50	0.45	0.54	0.52	0.48	0.53
Income Medium High	0.49	0.49	0.50	0.50	0.50	0.50
Employment Students	0.47	0.48	0.48	0.48	0.48	0.49
Employment Workers	0.49	0.47	0.52	0.50	0.48	0.51
Age 18-24	0.48	0.45	0.51	0.51	0.48	0.52
Age 25-34	0.51	0.48	0.55	0.52	0.48	0.54
Age 35-44	0.50	0.42	0.55	0.54	0.48	0.54
Healthy	0.51	0.49	0.53	0.52	0.50	0.53

DECG SMOTE (n = 42)	RF			SVM		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.52	0.52	0.52	0.52	0.50	0.54
Gender Male	0.51	0.47	0.51	0.50	0.46	0.55
Gender Female	0.53	0.53	0.53	0.50	0.49	0.53
Income Low	0.52	0.49	0.52	0.52	0.46	0.58
Income Medium High	0.52	0.51	0.52	0.49	0.49	0.50
Employment Students	0.49	0.49	0.50	0.47	0.47	0.49
Employment Workers	0.52	0.51	0.52	0.48	0.47	0.53
Age 18-24	0.51	0.49	0.51	0.50	0.44	0.57
Age 25-34	0.53	0.50	0.54	0.48	0.44	0.55
Age 35-44	0.53	0.46	0.54	0.53	0.45	0.60
Healthy	0.53	0.51	0.52	0.51	0.49	0.55
DA (n = 42)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.51	0.50	0.53	0.52	0.51	0.52
Gender Male	0.48	0.44	0.51	0.51	0.48	0.51
Gender Female	0.50	0.50	0.51	0.52	0.51	0.53
Income Low	0.49	0.44	0.54	0.55	0.52	0.56
Income Medium High	0.49	0.49	0.50	0.50	0.49	0.49
Employment Students	0.46	0.47	0.48	0.49	0.49	0.49
Employment Workers	0.49	0.48	0.52	0.50	0.48	0.51
Age 18-24	0.49	0.46	0.52	0.52	0.48	0.54
Age 25-34	0.49	0.47	0.53	0.52	0.49	0.54
Age 35-44	0.46	0.42	0.54	0.54	0.50	0.56
Healthy	0.50	0.48	0.53	0.52	0.50	0.53
DA SMOTE (n = 42)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.53	0.53	0.53	0.43	0.41	0.53
Gender Male	0.50	0.47	0.50	0.48	0.45	0.50
Gender Female	0.51	0.51	0.51	0.46	0.45	0.51
Income Low	0.52	0.49	0.53	0.46	0.39	0.59
Income Medium High	0.51	0.50	0.51	0.48	0.48	0.49
Employment Students	0.48	0.48	0.49	0.45	0.45	0.49
Employment Workers	0.51	0.50	0.51	0.41	0.39	0.52
Age 18-24	0.52	0.50	0.52	0.46	0.40	0.58
Age 25-34	0.52	0.50	0.52	0.41	0.38	0.51
Age 35-44	0.51	0.48	0.53	0.47	0.43	0.55
Healthy	0.52	0.49	0.53	0.42	0.38	0.55
DAW (n = 31)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.49	0.47	0.52	0.52	0.51	0.53
Gender Male	0.44	0.40	0.49	0.52	0.47	0.53
Gender Female	0.46	0.45	0.47	0.50	0.48	0.50

Income Low	0.49	0.42	0.56	0.53	0.48	0.55
Income Medium High	0.48	0.48	0.49	0.47	0.46	0.47
Employment Students	0.46	0.46	0.48	0.49	0.49	0.50
Employment Workers	0.45	0.42	0.50	0.49	0.47	0.50
Age 18-24	0.48	0.44	0.54	0.49	0.46	0.52
Age 25-34	0.48	0.44	0.54	0.49	0.46	0.52
Age 35-44	0.50	0.42	0.58	0.56	0.49	0.58
Age 45-64	0.39	0.37	0.39	0.48	0.43	0.49
Healthy	0.49	0.46	0.54	0.54	0.52	0.55
DAW SMOTE (n = 31)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.52	0.51	0.52	0.41	0.38	0.55
Gender Male	0.46	0.42	0.48	0.53	0.49	0.55
Gender Female	0.48	0.47	0.48	0.40	0.39	0.49
Income Low	0.51	0.47	0.53	0.48	0.38	0.62
Income Medium High	0.51	0.50	0.51	0.42	0.42	0.45
Employment Students	0.48	0.47	0.49	0.44	0.44	0.48
Employment Workers	0.48	0.47	0.49	0.42	0.37	0.55
Age 18-24	0.51	0.48	0.52	0.46	0.39	0.58
Age 25-34	0.51	0.48	0.52	0.46	0.39	0.58
Age 35-44	0.52	0.45	0.55	0.57	0.49	0.60
Healthy	0.53	0.51	0.53	0.42	0.37	0.56
DW (n = 44)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.48	0.46	0.51	0.49	0.47	0.49
Gender Male	0.49	0.43	0.52	0.48	0.43	0.49
Gender Female	0.44	0.44	0.45	0.49	0.47	0.49
Income Low	0.46	0.42	0.52	0.49	0.45	0.54
Income Medium High	0.48	0.48	0.51	0.49	0.48	0.50
Employment Students	0.47	0.46	0.48	0.51	0.50	0.52
Employment Workers	0.50	0.47	0.53	0.50	0.48	0.51
Age 18-24	0.49	0.44	0.56	0.51	0.47	0.53
Age 25-34	0.46	0.45	0.50	0.48	0.46	0.50
Age 35-44	0.49	0.43	0.52	0.52	0.47	0.52
Healthy	0.34	0.35	0.39	0.49	0.46	0.53
DW SMOTE (n = 44)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.48	0.46	0.48	0.50	0.49	0.50
Gender Male	0.46	0.42	0.48	0.46	0.43	0.46
Gender Female	0.45	0.44	0.45	0.50	0.48	0.49
Income Low	0.49	0.46	0.50	0.52	0.49	0.52
Income Medium High	0.50	0.49	0.51	0.50	0.50	0.50
Employment Students	0.47	0.47	0.48	0.49	0.49	0.50
Employment Workers	0.50	0.48	0.51	0.49	0.48	0.49
Age 18-24	0.50	0.46	0.52	0.52	0.49	0.52

Age 25-34	0.50	0.49	0.52	0.48	0.46	0.49
Age 35-44	0.48	0.43	0.51	0.51	0.46	0.50
Healthy	0.49	0.48	0.50	0.48	0.46	0.48
DEmpatica (n = 27)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.46	0.46	0.48	0.49	0.48	0.51
Gender Male	0.46	0.42	0.53	0.58	0.56	0.60
Gender Female	0.44	0.44	0.46	0.46	0.46	0.48
Income Low	0.50	0.45	0.57	0.51	0.47	0.53
Income Medium High	0.36	0.37	0.40	0.42	0.42	0.44
Employment Students	0.47	0.46	0.50	0.47	0.47	0.48
Employment Workers	0.42	0.42	0.45	0.45	0.44	0.46
Age 18-24	0.54	0.41	0.64	0.56	0.46	0.60
Age 25-34	0.47	0.44	0.53	0.49	0.44	0.54
Age 35-44	0.44	0.40	0.50	0.53	0.48	0.55
Healthy	0.48	0.45	0.54	0.50	0.48	0.54
DEmpatica SMOTE (n = 27)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.47	0.47	0.49	0.49	0.49	0.51
Gender Male	0.53	0.50	0.54	0.53	0.51	0.55
Gender Female	0.46	0.46	0.49	0.47	0.47	0.48
Income Low	0.54	0.51	0.54	0.52	0.49	0.52
Income Medium High	0.38	0.38	0.40	0.42	0.42	0.44
Employment Students	0.50	0.49	0.52	0.48	0.48	0.49
Employment Workers	0.42	0.42	0.43	0.43	0.43	0.45
Age 18-24	0.51	0.43	0.52	0.56	0.46	0.60
Age 25-34	0.49	0.46	0.51	0.50	0.44	0.54
Age 35-44	0.46	0.45	0.46	0.50	0.46	0.49
Healthy	0.51	0.50	0.51	0.52	0.50	0.53

SDA (n = 34)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.54	0.52	0.56	0.56	0.54	0.57
Gender Male	0.43	0.36	0.54	0.45	0.39	0.53
Gender Female	0.55	0.55	0.56	0.54	0.54	0.54
Income Low	0.57	0.52	0.60	0.57	0.54	0.58
Income Medium High	0.52	0.51	0.53	0.50	0.50	0.51
Employment Students	0.57	0.57	0.59	0.56	0.56	0.57
Employment Workers	0.45	0.39	0.56	0.47	0.42	0.53
Age 18-24	0.52	0.49	0.57	0.56	0.54	0.57
Age 25-34	0.57	0.52	0.61	0.58	0.55	0.60
Age 35-44	0.59	0.43	0.69	0.59	0.45	0.65
Healthy	0.50	0.45	0.55	0.56	0.53	0.58
SDA SMOTE (n = 34)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***

Full	0.55	0.54	0.55	0.50	0.47	0.56
Gender Male	0.42	0.36	0.49	0.47	0.38	0.60
Gender Female	0.56	0.55	0.57	0.49	0.49	0.51
Income Low	0.59	0.57	0.61	0.48	0.43	0.59
Income Medium High	0.54	0.52	0.55	0.43	0.44	0.49
Employment Students	0.59	0.58	0.60	0.50	0.51	0.54
Employment Workers	0.49	0.46	0.54	0.47	0.40	0.59
Age 18-24	0.58	0.56	0.60	0.48	0.43	0.58
Age 25-34	0.57	0.53	0.58	0.46	0.40	0.57
Age 35-44	0.63	0.51	0.68	0.59	0.41	0.70
Healthy	0.52	0.50	0.53	0.49	0.44	0.59
SDAW (n = 27)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.54	0.50	0.57	0.55	0.53	0.56
Gender Male	0.74	0.45	0.81	0.71	0.45	0.74
Gender Female	0.53	0.52	0.54	0.49	0.49	0.50
Income Low	0.57	0.51	0.63	0.59	0.56	0.61
Income Medium High	0.53	0.51	0.55	0.52	0.51	0.54
Employment Students	0.55	0.55	0.57	0.53	0.53	0.53
Employment Workers	0.49	0.40	0.60	0.53	0.46	0.58
Age 18-24	0.58	0.54	0.61	0.58	0.58	0.59
Age 25-34	0.61	0.57	0.62	0.52	0.51	0.54
Age 35-44	0.69	0.44	0.78	0.67	0.48	0.71
Healthy	0.49	0.42	0.58	0.52	0.46	0.57
SDAW SMOTE (n = 27)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.58	0.56	0.59	0.47	0.41	0.59
Gender Male	0.71	0.46	0.74	0.75	0.45	0.83
Gender Female	0.55	0.54	0.57	0.43	0.44	0.47
Income Low	0.59	0.55	0.61	0.49	0.39	0.63
Income Medium High	0.57	0.55	0.58	0.37	0.36	0.50
Employment Students	0.59	0.57	0.60	0.47	0.48	0.51
Employment Workers	0.54	0.48	0.57	0.50	0.38	0.63
Age 18-24	0.61	0.56	0.64	0.42	0.41	0.52
Age 25-34	0.62	0.57	0.65	0.43	0.43	0.50
Age 35-44	0.70	0.49	0.76	0.69	0.44	0.78
Healthy	0.56	0.52	0.58	0.48	0.39	0.62
SDW (n = 34)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.50	0.46	0.57	0.52	0.49	0.54
Gender Male	0.57	0.48	0.63	0.56	0.47	0.61
Gender Female	0.46	0.45	0.49	0.46	0.45	0.51
Income Low	0.52	0.43	0.62	0.53	0.44	0.61
Income Medium High	0.42	0.42	0.45	0.48	0.47	0.49
Employment Students	0.50	0.49	0.53	0.52	0.52	0.53

Employment Workers	0.57	0.43	0.55	0.49	0.44	0.55
Age 18-24	0.54	0.50	0.59	0.52	0.50	0.54
Age 25-34	0.49	0.48	0.52	0.48	0.47	0.49
Age 35-44	0.53	0.44	0.60	0.50	0.40	0.56
Healthy	0.61	0.48	0.66	0.54	0.49	0.57
SDW SMOTE (n = 34)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.54	0.51	0.55	0.44	0.37	0.58
Gender Male	0.60	0.51	0.63	0.44	0.37	0.54
Gender Female	0.48	0.47	0.49	0.42	0.40	0.52
Income Low	0.57	0.51	0.60	0.50	0.39	0.64
Income Medium High	0.47	0.46	0.48	0.41	0.41	0.46
Employment Students	0.52	0.51	0.55	0.49	0.49	0.51
Employment Workers	0.51	0.48	0.53	0.44	0.37	0.58
Age 18-24	0.56	0.51	0.59	0.42	0.38	0.54
Age 25-34	0.51	0.50	0.53	0.40	0.40	0.48
Age 35-44	0.67	0.58	0.66	0.48	0.38	0.59
Healthy	0.54	0.50	0.56	0.47	0.38	0.62
SDS (n = 27)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.54	0.51	0.57	0.52	0.50	0.53
Gender Male	0.75	0.45	0.82	0.74	0.45	0.82
Gender Female	0.52	0.52	0.54	0.49	0.49	0.49
Income Low	0.59	0.52	0.65	0.57	0.53	0.59
Income Medium High	0.54	0.52	0.57	0.52	0.51	0.54
Employment Students	0.56	0.55	0.58	0.52	0.52	0.53
Employment Workers	0.49	0.40	0.60	0.53	0.47	0.58
Age 18-24	0.52	0.50	0.56	0.54	0.53	0.54
Age 25-34	0.57	0.52	0.59	0.49	0.49	0.52
Age 35-44	0.69	0.44	0.78	0.68	0.44	0.76
Healthy	0.47	0.46	0.50	0.49	0.49	0.51
SDS SMOTE (n = 27)	**RF**			**SVM**		
Items	**F1-Weighted**	***F1-Macro***	***Accuracy***	**F1-Weighted**	**F1-Macro**	***Accuracy***
Full	0.57	0.55	0.57	0.50	0.48	0.54
Gender Male	0.66	0.45	0.67	0.72	0.44	0.78
Gender Female	0.53	0.52	0.55	0.50	0.50	0.51
Income Low	0.62	0.58	0.63	0.54	0.50	0.58
Income Medium High	0.57	0.54	0.59	0.48	0.48	0.51
Employment Students	0.58	0.57	0.60	0.53	0.52	0.54
Employment Workers	0.51	0.47	0.52	0.49	0.39	0.62
Age 18-24	0.56	0.53	0.59	0.50	0.50	0.53
Age 25-34	0.59	0.52	0.62	0.54	0.52	0.55
Age 35-44	0.67	0.50	0.69	0.69	0.44	0.77
Healthy	0.52	0.50	0.54	0.52	0.50	0.52

Table B8: Percentage of Stress in Each Division, Each Dataset

D (n = 22)	**% Stress**
Full	49
Gender Male	44
Gender Female	52
Income Low	42
Income Medium High	56
Employment Students	47
Employment Workers	54
Age 18-24	35
Age 25-34	64
Age 35-44	41
Healthy	46
DECG (n = 42)	% Stress
Full	44
Gender Male	43
Gender Female	45
Income Low	39
Income Medium High	49
Employment Students	47
Employment Workers	43
Age 18-24	39
Age 25-34	42
Age 35-44	38
Healthy	42
DA (n = 42)	
Items	
Full	44
Gender Male	43
Gender Female	45
Income Low	39
Income Medium High	49
Employment Students	47
Employment Workers	43
Age 18-24	39
Age 25-34	42
Age 35-44	38
Healthy	42
DAW (n = 41)	44
Full	44

Gender Male	41
Gender Female	45
Income Low	39
Income Medium High	48
Employment Students	47
Employment Workers	42
Age 18-24	39
Age 25-34	42
Age 35-44	36
Healthy	41
DW (n = 44)	
Full	43
Gender Male	39
Gender Female	45
Income Low	39
Income Medium High	46
Employment Students	47
Employment Workers	41
Age 18-24	39
Age 25-34	42
Age 35-44	36
Healthy	40
DEmpatica (n = 27)	
Full	45
Gender Male	40
Gender Female	49
Income Low	47
Income Medium High	50
Employment Students	47
Employment Workers	46
Age 18-24	33
Age 25-34	59
Age 35-44	39
Healthy	42
SDA (n = 27)	
Full	44
Gender Male	35
Gender Female	48
Income Low	40
Income Medium High	48
Employment Students	47
Employment Workers	39
Age 18-24	40
Age 25-34	61
Age 35-44	31

Healthy	41
SDAW (n = 27)	
Full	42
Gender Male	18
Gender Female	49
Income Low	38
Income Medium High	44
Employment Students	48
Employment Workers	37
Age 18-24	43
Age 25-34	55
Age 35-44	23
Healthy	37
SDW (n = 34)	
Full	42
Gender Male	36
Gender Female	45
Income Low	38
Income Medium High	46
Employment Students	48
Employment Workers	41
Age 18-24	42
Age 25-34	55
Age 35-44	35
Healthy	39
SDS (n = 27)	
Full	42
Gender Male	18
Gender Female	49
Income Low	38
Income Medium High	44
Employment Students	48
Employment Workers	37
Age 18-24	43
Age 25-34	45
Age 35-44	23
Healthy	54

Table B9: Precision, Recall, F1-Score and Accuracy for Each User

User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
	No Stress	0.92	1.00	0.96	0.92	1.00	0.96	11

1	*Stress*	1.00	0.80	0.89	1.00	0.80	0.89	5
	Accuracy	-	-	0.94	-	-	0.94	16
	Weighted Average	0.94	0.94	0.94	0.94	0.94	0.94	16
	Macro Average	0.96	0.90	0.92	0.96	0.90	0.92	16
1 - SMOTE	*No Stress*	0.92	1.00	0.96	0.92	1.00	0.96	11
	Stress	1.00	0.80	0.89	1.00	0.80	0.89	5
	Accuracy	-	-	0.94	-	-	0.94	16
	Weighted Average	0.94	0.94	0.94	0.94	0.94	0.94	16
	Macro Average	0.96	0.90	0.92	0.96	0.90	0.92	16
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
2	*No Stress*	0.76	0.93	0.84	0.70	1.00	0.82	14
	Stress	0.67	0.33	0.44	0.00	0.00	0.00	6
	Accuracy	-	-	0.75	-	-	0.70	20
	Weighted Average	0.74	0.75	0.72	0.49	0.70	0.58	20
	Macro Average	0.72	0.63	0.64	0.35	0.50	0.41	20
2 - SMOTE	*No Stress*	0.79	0.79	0.79	0.70	1.00	0.82	14
	Stress	0.50	0.50	0.50	0.00	0.00	0.00	6

	Accuracy	-	-	0.70	-	-	0.70	20
	Weighted Average	0.70	0.70	0.70	0.49	0.70	0.58	20
	Macro Average	0.64	0.64	0.64	0.35	0.50	0.41	20
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
3	*No Stress*	0.77	0.83	0.80	0.67	0.83	0.74	12
	Stress	0.50	0.40	0.44	0.00	0.00	0.00	5
	Accuracy	-	-	0.71	-	-	0.59	17
	Weighted Average	0.69	0.71	0.70	0.47	0.59	0.52	17
	Macro Average	0.63	0.62	0.62	0.33	0.42	0.37	17
3 - SMOTE	*No Stress*	0.83	0.83	0.83	0.75	0.75	0.75	12
	Stress	0.60	0.60	0.60	0.40	0.40	0.40	5
	Accuracy	-	-	0.76	-	-	0.65	17
	Weighted Average	0.76	0.76	0.76	0.65	0.65	0.65	17
	Macro Average	0.72	0.72	0.76	0.57	0.57	0.58	17
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
4	*No Stress*	0.69	0.90	0.78	0.71	1.00	0.83	10
	Stress	0.50	0.20	0.29	1.00	0.20	0.33	5

	Accuracy	-	-	0.67	-	-	0.73	15
	Weighted Average	0.63	0.67	0.62	0.81	0.73	0.67	15
	Macro Average	0.60	0.55	0.53	0.86	0.60	0.58	15
4 - SMOTE	*No Stress*	0.89	0.80	0.84	0.67	1.00	0.80	10
	Stress	0.67	0.80	0.73	0.00	0.00	0.00	5
	Accuracy	-	-	0.80	-	-	0.67	15
	Weighted Average	0.81	0.80	0.80	0.44	0.67	0.53	15
	Macro Average	0.78	0.80	0.78	0.33	0.50	0.40	15
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
5	*No Stress*	0.33	0.33	0.33	0.00	0.00	0.00	6
	Stress	0.50	0.50	0.50	0.57	1.00	0.73	8
	Accuracy	-	-	0.43	-	-	0.57	14
	Weighted Average	0.43	0.43	0.43	0.33	0.57	0.42	14
	Macro Average	0.42	0.42	0.42	0.29	0.50	0.36	14
5 - SMOTE	*No Stress*	0.33	0.33	0.33	0.00	0.00	0.00	6
	Stress	0.50	0.50	0.50	0.57	1.00	0.73	8
	Accuracy	-	-	0.43	-	-	0.57	14

	Weighted Average	0.43	0.43	0.43	0.33	0.57	0.42	14
	Macro Average	0.42	0.80	0.78	0.29	0.50	0.36	14
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
6	*No Stress*	0.71	0.91	0.80	0.70	0.64	0.67	11
	Stress	0.67	0.33	0.44	0.43	0.50	0.46	6
	Accuracy	-	-	0.71	-	-	0.59	17
	Weighted Average	0.70	0.71	0.67	0.60	0.59	0.59	17
	Macro Average	0.69	0.62	0.62	0.56	0.57	0.56	17
6 - SMOTE	*No Stress*	0.75	0.82	0.78	0.67	0.55	0.60	11
	Stress	0.60	0.50	0.55	0.38	0.50	0.43	6
	Accuracy	-	-	0.71	-	-	0.53	17
	Weighted Average	0.70	0.71	0.70	0.56	0.53	0.54	17
	Macro Average	0.68	0.66	0.66	0.52	0.52	0.51	17
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
7	*No Stress*	0.71	1.00	0.83	0.71	1.00	0.83	12
	Stress	0.00	0.00	0.00	0.00	0.00	0.00	5
	Accuracy	-	-	0.71	-	-	0.71	17

	Weighted Average	0.50	0.71	0.58	0.50	0.71	0.58	17
	Macro Average	0.35	0.50	0.41	0.35	0.50	0.41	17
7 - SMOTE	*No Stress*	0.73	0.92	0.81	0.71	1.00	0.83	12
	Stress	0.50	0.20	0.29	0.00	0.00	0.00	5
	Accuracy	-	-	0.71	-	-	0.71	17
	Weighted Average	0.66	0.71	0.66	0.50	0.71	0.58	17
	Macro Average	0.62	0.56	0.55	0.35	0.50	0.41	17
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
8	*No Stress*	0.62	0.71	0.67	0.62	0.71	0.67	7
	Stress	0.75	0.67	0.71	0.75	0.67	0.71	9
	Accuracy	-	-	0.69	-	-	0.69	16
	Weighted Average	0.70	0.69	0.69	0.70	0.69	0.69	16
	Macro Average	0.69	0.69	0.69	0.69	0.69	0.69	16
8 - SMOTE	*No Stress*	0.62	0.71	0.67	0.50	0.43	0.46	7
	Stress	0.75	0.67	0.71	0.60	0.67	0.63	9
	Accuracy	-	-	0.69	-	-	0.56	16

	Weighted Average	0.70	0.69	0.69	0.56	0.56	0.56	16
	Macro Average	0.69	0.69	0.69	0.55	0.55	0.55	16
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
9	*No Stress*	0.67	0.40	0.50	0.00	0.00	0.00	5
	Stress	0.73	0.89	0.80	0.64	1.00	0.78	9
	Accuracy	-	-	0.71	-	-	0.64	14
	Weighted Average	0.71	0.71	0.69	0.41	0.64	0.50	14
	Macro Average	0.70	0.64	0.65	0.32	0.50	0.39	14
9 - SMOTE	*No Stress*	0.80	0.80	0.80	0.00	0.00	0.00	5
	Stress	0.89	0.89	0.89	0.62	0.89	0.73	9
	Accuracy	-	-	0.86	-	-	0.57	14
	Weighted Average	0.86	0.86	0.86	0.40	0.57	0.47	14
	Macro Average	0.84	0.84	0.84	0.31	0.44	0.36	14
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
10	*No Stress*	0.58	0.88	0.70	0.62	1.00	0.76	8
	Stress	0.00	0.00	0.00	0.00	0.00	0.00	5
	Accuracy	-	-	0.54	-	-	0.62	13

	Weighted Average	0.36	0.54	0.43	0.38	0.62	0.47	13
	Macro Average	0.29	0.44	0.35	0.31	0.50	0.38	13
10 - SMOTE	*No Stress*	0.55	0.75	0.63	0.62	1.00	0.76	8
	Stress	0.00	0.00	0.00	0.00	0.00	0.00	5
	Accuracy	-	-	0.46	-	-	0.62	13
	Weighted Average	0.34	0.46	0.39	0.38	0.62	0.47	13
	Macro Average	0.27	0.38	0.32	0.31	0.50	0.38	13
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
11	*No Stress*	0.75	1.00	0.86	0.60	1.00	0.75	6
	Stress	1.00	0.60	0.75	1.00	0.20	0.33	5
	Accuracy	-	-	0.82	-	-	0.64	11
	Weighted Average	0.86	0.82	0.81	0.78	0.64	0.56	11
	Macro Average	0.88	0.80	0.80	0.80	0.60	0.54	11
11 - SMOTE	*No Stress*	0.75	1.00	0.86	0.71	0.83	0.77	6
	Stress	1.00	0.60	0.75	0.75	0.60	0.67	5
	Accuracy	-	-	0.82	-	-	0.73	11

	Weighted Average	0.86	0.82	0.81	0.73	0.73	0.72	11
	Macro Average	0.88	0.80	0.80	0.73	0.72	0.72	11
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
12	*No Stress*	1.00	0.93	0.97	0.83	1.00	0.91	15
	Stress	0.75	1.00	0.86	0.00	0.00	0.00	3
	Accuracy	-	-	0.94	-	-	0.83	18
	Weighted Average	0.96	0.94	0.95	0.69	0.83	0.76	18
	Macro Average	0.88	0.97	0.91	0.42	0.50	0.45	18
12 - SMOTE	*No Stress*	1.00	0.87	0.93	0.93	0.87	0.90	15
	Stress	0.60	1.00	0.75	0.50	0.67	0.57	3
	Accuracy	-	-	0.89	-	-	0.83	18
	Weighted Average	0.93	0.89	0.90	0.86	0.83	0.84	18
	Macro Average	0.88	0.93	0.84	0.71	0.77	0.73	18
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
13	*No Stress*	0.71	0.91	0.80	0.69	1.00	0.81	11
	Stress	0.50	0.20	0.29	0.00	0.00	0.00	5
	Accuracy	-	-	0.69	-	-	0.69	16

	Weighted Average	0.65	0.69	0.64	0.47	0.69	0.56	16
	Macro Average	0.61	0.55	0.54	0.34	0.50	0.41	16
13 - SMOTE	*No Stress*	0.80	0.73	0.76	0.90	0.82	0.86	11
	Stress	0.50	0.60	0.55	0.67	0.80	0.73	5
	Accuracy	-	-	0.69	-	-	0.81	16
	Weighted Average	0.71	0.69	0.69	0.83	0.81	0.82	16
	Macro Average	0.65	0.66	0.65	0.78	0.81	0.79	16
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
14	*No Stress*	0.25	0.25	0.25	0.33	0.25	0.29	4
	Stress	0.79	0.79	0.79	0.80	0.86	0.83	14
	Accuracy	-	-	0.67	-	-	0.72	18
	Weighted Average	0.67	0.67	0.67	0.70	0.72	0.71	18
	Macro Average	0.52	0.52	0.52	0.57	0.55	0.56	18
14 - SMOTE	*No Stress*	0.14	0.25	0.18	0.17	0.25	0.20	4
	Stress	0.73	0.57	0.64	0.75	0.64	0.69	14
	Accuracy	-	-	0.50	-	-	0.56	18

User	Items	RF Precision	RF Recall	RF F1-Score	SVM Precision	SVM Recall	SVM F1-Score	Support
	Weighted Average	0.60	0.50	0.54	0.62	0.56	0.58	18
	Macro Average	0.44	0.41	0.41	0.46	0.45	0.45	18
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
15	*No Stress*	0.80	0.80	0.80	0.82	0.90	0.86	10
	Stress	0.71	0.71	0.71	0.83	0.71	0.77	7
	Accuracy	-	-	0.76	-	-	0.82	17
	Weighted Average	0.76	0.76	0.76	0.82	0.82	0.82	17
	Macro Average	0.76	0.76	0.76	0.83	0.81	0.81	17
15 - SMOTE	*No Stress*	0.78	0.70	0.74	0.82	0.90	0.86	10
	Stress	0.62	0.71	0.67	0.83	0.71	0.77	7
	Accuracy	-	-	0.71	-	-	0.82	17
	Weighted Average	0.71	0.71	0.71	0.82	0.82	0.82	17
	Macro Average	0.70	0.71	0.70	0.83	0.81	0.81	17
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
16	*No Stress*	0.78	1.00	0.88	0.78	1.00	0.88	14
	Stress	0.00	0.00	0.00	0.00	0.00	0.00	4
	Accuracy	-	-	0.78	-	-	0.78	18

	Weighted Average	0.60	0.78	0.68	0.60	0.78	0.68	18
	Macro Average	0.39	0.50	0.44	0.39	0.50	0.44	18
16 - SMOTE	*No Stress*	0.80	0.86	0.83	0.78	1.00	0.88	14
	Stress	0.33	0.25	0.29	0.00	0.00	0.00	4
	Accuracy	-	-	0.72	-	-	0.78	18
	Weighted Average	0.70	0.72	0.71	0.60	0.78	0.68	18
	Macro Average	0.57	0.55	0.56	0.39	0.50	0.44	18
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
17	*No Stress*	0.75	0.50	0.60	0.67	0.33	0.44	6
	Stress	0.79	0.92	0.85	0.73	0.92	0.81	12
	Accuracy	-	-	0.78	-	-	0.72	18
	Weighted Average	0.77	0.78	0.76	0.71	0.72	0.69	18
	Macro Average	0.77	0.71	0.72	0.70	0.62	0.63	18
17 - SMOTE	*No Stress*	0.60	0.50	0.55	0.67	0.67	0.67	6
	Stress	0.77	0.83	0.80	0.83	0.83	0.83	12
	Accuracy	-	-	0.72	-	-	0.78	18

	Weighted Average	0.71	0.72	0.72	0.78	0.78	0.78	18
	Macro Average	0.68	0.67	0.67	0.75	0.75	0.75	18
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
18	*No Stress*	0.77	1.00	0.87	0.77	1.00	0.87	10
	Stress	0.00	0.00	0.00	0.00	0.00	0.00	3
	Accuracy	-	-	0.77	-	-	0.77	13
	Weighted Average	0.59	0.77	0.67	0.59	0.77	0.67	13
	Macro Average	0.38	0.50	0.43	0.38	0.50	0.42	13
18 - SMOTE	*No Stress*	0.73	0.80	0.76	0.77	1.00	0.87	10
	Stress	0.00	0.00	0.00	0.00	0.00	0.00	3
	Accuracy	-	-	0.62	-	-	0.77	13
	Weighted Average	0.56	0.62	0.59	0.59	0.77	0.67	13
	Macro Average	0.36	0.40	0.38	0.38	0.50	0.43	13
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
19	*No Stress*	0.60	0.43	0.50	0.71	0.71	0.71	7
	Stress	0.67	0.80	0.73	0.80	0.80	0.80	10
	Accuracy	-	-	0.65	-	-	0.76	17

	Weighted Average	0.64	0.65	0.63	0.76	0.76	0.76	17
	Macro Average	0.63	0.61	0.61	0.76	0.76	0.76	17
19 - SMOTE	*No Stress*	0.60	0.43	0.50	0.71	0.71	0.71	7
	Stress	0.67	0.80	0.73	0.80	0.80	0.80	10
	Accuracy	-	-	0.65	-	-	0.76	17
	Weighted Average	0.64	0.65	0.63	0.76	0.76	0.76	17
	Macro Average	0.63	0.61	0.61	0.76	0.76	0.76	17
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
20	*No Stress*	0.67	0.60	0.63	0.56	0.50	0.53	10
	Stress	0.56	0.62	0.59	0.44	0.50	0.47	8
	Accuracy	-	-	0.61	-	-	0.50	18
	Weighted Average	0.62	0.61	0.61	0.51	0.50	0.50	18
	Macro Average	0.61	0.61	0.61	0.50	0.50	0.50	18
20 - SMOTE	*No Stress*	0.70	0.70	0.70	0.62	1.00	0.77	10
	Stress	0.62	0.62	0.62	1.00	0.25	0.40	8
	Accuracy	-	-	0.67	-	-	0.67	18

	Weighted Average	0.66	0.66	0.66	0.79	0.67	0.61	18
	Macro Average	0.67	0.67	0.67	0.81	0.62	0.58	18
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
21	*No Stress*	0.60	0.60	0.60	0.73	0.80	0.76	10
	Stress	0.00	.00	.00	0.33	0.25	0.29	4
	Accuracy	-	-	0.43	-	-	0.64	14
	Weighted Average	0.43	0.43	0.43	0.61	0.64	0.63	14
	Macro Average	0.30	0.30	0.30	0.53	0.53	0.52	14
21 - SMOTE	*No Stress*	0.67	0.60	0.63	0.73	0.80	0.76	10
	Stress	0.20	0.25	0.22	0.33	0.25	0.29	4
	Accuracy	-	-	0.50	-	-	0.64	14
	Weighted Average	0.53	0.50	0.51	0.61	0.64	0.63	14
	Macro Average	0.43	0.42	0.43	0.53	0.53	0.52	14
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
22	*No Stress*	0.83	0.83	0.83	0.71	1.00	0.83	12
	Stress	0.60	0.60	0.60	0.00	0.00	0.00	5
	Accuracy	-	-	0.76	-	-	0.71	17

	Weighted Average	0.76	0.76	0.76	0.50	0.71	0.58	17
	Macro Average	0.72	0.72	0.72	0.35	0.50	0.41	17
22 - SMOTE	*No Stress*	0.78	0.58	0.67	0.71	1.00	0.83	12
	Stress	0.38	0.60	0.46	0.00	0.00	0.00	5
	Accuracy	-	-	0.59	-	-	0.71	17
	Weighted Average	0.66	0.59	0.61	0.50	0.71	0.58	17
	Macro Average	0.58	0.59	0.56	0.35	0.50	0.41	17
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
23	*No Stress*	0.81	0.93	0.87	0.82	1.00	0.90	14
	Stress	0.00	0.00	0.00	0.00	0.00	0.00	3
	Accuracy	-	-	0.76	-	-	0.82	17
	Weighted Average	0.67	0.76	0.71	0.68	0.82	0.74	17
	Macro Average	0.41	0.46	0.43	0.41	0.50	0.45	17
23 - SMOTE	*No Stress*	0.80	0.86	0.83	0.82	1.00	0.90	14
	Stress	0.00	0.00	0.00	0.00	0.00	0.00	3
	Accuracy	-	-	0.71	-	-	0.82	17

	Weighted Average	0.66	0.71	0.68	0.68	0.82	0.74	17
	Macro Average	0.40	0.43	0.41	0.41	0.50	0.45	17
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
24	*No Stress*	0.69	0.85	0.76	0.72	1.00	0.84	13
	Stress	0.00	0.00	0.00	0.00	0.00	0.00	5
	Accuracy	-	-	0.61	-	-	0.72	18
	Weighted Average	0.50	0.61	0.55	0.52	0.72	0.61	18
	Macro Average	0.34	0.42	0.38	0.36	0.50	0.42	18
24 - SMOTE	*No Stress*	0.64	0.69	0.67	0.83	0.77	0.80	13
	Stress	0.00	0.00	0.00	0.50	0.60	0.55	5
	Accuracy	-	-	0.50	-	-	0.72	18
	Weighted Average	0.46	0.50	0.48	0.74	0.72	0.73	18
	Macro Average	0.32	0.35	0.33	0.67	0.68	0.67	18
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
25	*No Stress*	1.00	0.60	0.75	0.50	0.60	0.55	5
	Stress	0.83	1.00	0.91	0.78	0.70	0.74	10
	Accuracy	-	-	0.87	-	-	0.67	15

	Weighted Average	0.89	0.87	0.86	0.69	0.67	0.67	15
	Macro Average	0.92	0.80	0.83	0.64	0.65	0.64	15
25 - SMOTE	*No Stress*	1.00	0.60	0.75	0.43	0.60	0.50	5
	Stress	0.83	1.00	0.91	0.75	0.60	0.67	10
	Accuracy	-	-	0.87	-	-	0.60	15
	Weighted Average	0.89	0.87	0.86	0.64	0.60	0.61	15
	Macro Average	0.92	0.80	0.83	0.59	0.60	0.58	15
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
26	*No Stress*	0.67	0.86	0.75	0.62	0.71	0.67	7
	Stress	0.88	0.70	0.78	0.78	0.70	0.74	10
	Accuracy	-	-	0.76	-	-	0.71	17
	Weighted Average	0.79	0.76	0.77	0.71	0.71	0.71	17
	Macro Average	0.77	0.78	0.76	0.70	0.71	0.70	17
26 - SMOTE	*No Stress*	0.60	0.86	0.71	0.67	0.86	0.75	7
	Stress	0.86	0.60	0.71	0.88	0.70	0.78	10
	Accuracy	-	-	0.71	-	-	0.76	17

User	Items	RF Precision	RF Recall	RF F1-Score	SVM Precision	SVM Recall	SVM F1-Score	Support
	Weighted Average	0.75	0.71	0.71	0.79	0.76	0.77	17
	Macro Average	0.73	0.73	0.71	0.77	0.78	0.76	17
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
27	*No Stress*	1.00	0.33	0.50	1.00	0.33	0.50	3
	Stress	0.88	1.00	0.93	0.88	1.00	0.93	14
	Accuracy	-	-	0.88	-	-	0.88	17
	Weighted Average	0.90	0.88	0.86	0.90	0.88	0.86	17
	Macro Average	0.94	0.67	0.72	0.94	0.67	0.72	17
27 - SMOTE	*No Stress*	0.33	0.33	0.33	0.00	0.00	0.00	3
	Stress	0.86	0.86	0.86	0.82	1.00	0.90	14
	Accuracy	-	-	0.76	-	-	0.82	17
	Weighted Average	0.76	0.76	0.76	0.68	0.82	0.74	17
	Macro Average	0.60	0.60	0.60	0.41	0.50	0.74	17
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
28	*No Stress*	0.50	0.33	0.40	0.40	0.33	0.36	6
	Stress	0.56	0.71	0.63	0.50	0.57	0.53	7
	Accuracy	-	-	0.54	-	-	0.46	13

	Weighted Average	0.53	0.54	0.52	0.45	0.46	0.46	13
	Macro Average	0.53	0.52	0.51	0.45	0.45	0.45	13
28 - SMOTE	*No Stress*	0.67	0.33	0.44	0.40	0.33	0.36	6
	Stress	0.60	0.86	0.71	0.50	0.57	0.53	7
	Accuracy	-	-	0.62	-	-	0.46	13
	Weighted Average	0.63	0.62	0.59	0.45	0.46	0.46	13
	Macro Average	0.63	0.60	0.58	0.45	0.45	0.45	13
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
29	*No Stress*	0.00	0.00	0.00	0.00	0.00	0.00	4
	Stress	0.75	0.92	0.83	0.75	0.92	0.83	13
	Accuracy	-	-	0.71	-	-	0.71	17
	Weighted Average	0.57	0.71	0.63	0.57	0.71	0.63	17
	Macro Average	0.38	0.46	0.41	0.38	0.46	0.41	17
29 - SMOTE	*No Stress*	0.00	0.00	0.00	0.75	0.75	0.75	4
	Stress	0.75	0.92	0.83	0.92	0.92	0.92	13
	Accuracy	-	-	0.71	-	-	0.88	17

	Weighted Average	0.57	0.71	0.63	0.88	0.88	0.88	17
	Macro Average	0.38	0.46	0.41	0.84	0.84	0.84	17
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
30	*No Stress*	0.40	0.33	0.36	0.40	0.33	0.36	6
	Stress	0.69	0.75	0.72	0.69	0.75	0.72	12
	Accuracy	-	-	0.61	-	-	0.61	18
	Weighted Average	0.59	0.61	0.60	0.59	0.61	0.60	18
	Macro Average	0.55	0.54	0.54	0.55	0.54	0.54	18
30 - SMOTE	*No Stress*	0.38	0.50	0.43	0.38	0.50	0.43	6
	Stress	0.70	0.58	0.64	0.70	0.58	0.64	12
	Accuracy	-	-	0.56	-	-	0.56	18
	Weighted Average	0.59	0.56	0.57	0.59	0.56	0.57	18
	Macro Average	0.54	0.54	0.53	0.54	0.54	0.53	18
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
31	*No Stress*	0.67	0.86	0.75	0.62	0.71	0.67	7
	Stress	0.88	0.70	0.78	0.78	0.70	0.74	10
	Accuracy	-	-	0.76	-	-	0.71	17

	Weighted Average	0.79	0.76	0.77	0.71	0.71	0.71	17
	Macro Average	0.77	0.78	0.76	0.70	0.71	0.70	17
31 - SMOTE	*No Stress*	0.60	0.86	0.71	0.67	0.86	0.75	7
	Stress	0.86	0.60	0.71	0.88	0.70	0.78	10
	Accuracy	-	-	0.71	-	-	0.76	17
	Weighted Average	0.75	0.71	0.71	0.79	0.76	0.77	17
	Macro Average	0.73	0.73	0.71	0.77	0.78	0.76	17
User	**RF**			**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
32	*No Stress*	0.73	0.85	0.79	0.71	0.77	0.74	13
	Stress	0.50	0.33	0.40	0.40	0.33	0.36	6
	Accuracy	-	-	0.68	-	-	0.63	19
	Weighted Average	0.66	0.68	0.66	0.62	0.63	0.62	19
	Macro Average	0.62	0.59	0.59	0.56	0.55	0.55	19
32 - SMOTE	*No Stress*	0.83	0.77	0.80	0.82	0.69	0.75	13
	Stress	0.57	0.67	0.62	0.50	0.67	0.57	6
	Accuracy	-	-	0.74	-	-	0.68	19

	Weighted Average	0.75	0.74	0.74	0.72	0.68	0.66	19
	Macro Average	0.70	0.72	0.71	0.66	0.68	0.66	19
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
33	*No Stress*	0.73	0.92	0.81	0.75	1.00	0.86	12
	Stress	0.00	0.00	0.00	0.00	0.00	0.00	4
	Accuracy	-	-	0.69	-	-	0.75	16
	Weighted Average	0.55	0.69	0.61	0.56	0.75	0.64	16
	Macro Average	0.37	0.46	0.41	0.38	0.50	0.43	16
33 - SMOTE	*No Stress*	0.85	0.92	0.88	0.75	1.00	0.86	12
	Stress	0.67	0.50	0.57	0.00	0.00	0.00	4
	Accuracy	-	-	0.81	-	-	0.75	16
	Weighted Average	0.80	0.81	0.80	0.56	0.75	0.64	16
	Macro Average	0.76	0.71	0.73	0.38	0.50	0.43	16
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
34	*No Stress*	0.25	0.17	0.20	0.00	0.00	0.00	6
	Stress	0.58	0.70	0.64	0.62	1.00	0.77	10
	Accuracy	-	-	0.50	-	-	0.62	16

	Weighted Average	0.46	0.50	0.47	0.39	0.62	0.48	16
	Macro Average	0.42	0.43	0.42	0.31	0.50	0.38	16
34 - SMOTE	*No Stress*	0.57	0.67	0.62	0.33	0.33	0.33	6
	Stress	0.78	0.70	0.74	0.60	0.60	0.60	10
	Accuracy	-	-	0.69	-	-	0.50	16
	Weighted Average	0.70	0.69	0.69	0.50	0.50	0.50	16
	Macro Average	0.67	0.68	0.68	0.47	0.47	0.47	16
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
35	*No Stress*	0.56	0.62	0.59	0.53	1.00	0.70	8
	Stress	0.50	0.43	0.46	0.00	0.00	0.00	7
	Accuracy	-	-	0.53	-	-	0.53	15
	Weighted Average	0.53	0.53	0.53	0.28	0.53	0.37	15
	Macro Average	0.53	0.53	0.52	0.27	0.50	0.35	15
35 - SMOTE	*No Stress*	0.60	0.75	0.67	0.67	0.75	0.71	8
	Stress	0.60	0.43	0.50	0.67	0.57	0.62	7
	Accuracy	-	-	0.60	-	-	0.67	15

	Weighted Average	0.60	0.59	0.58	0.67	0.67	0.66	15
	Macro Average	0.60	0.60	0.59	0.67	0.66	0.66	15
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
36	*No Stress*	0.75	0.90	0.82	0.62	1.00	0.77	10
	Stress	0.75	0.50	0.60	0.00	0.00	0.00	6
	Accuracy	-	-	0.75	-	-	0.62	16
	Weighted Average	0.75	0.75	0.74	0.39	0.62	0.48	16
	Macro Average	0.75	0.70	0.71	0.31	0.50	0.38	16
36 - SMOTE	*No Stress*	0.75	0.60	0.67	0.62	1.00	0.77	10
	Stress	0.50	0.67	0.57	0.00	0.00	0.00	6
	Accuracy	-	-	0.62	-	-	0.62	16
	Weighted Average	0.66	0.62	0.63	0.39	0.62	0.48	16
	Macro Average	0.62	0.63	0.62	0.31	0.50	0.38	16
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
37	*No Stress*	0.00	0.00	0.00	0.00	0.00	0.00	5
	Stress	0.62	0.73	0.67	0.67	0.91	0.77	11
	Accuracy	-	-	0.50	-	-	0.62	16

	Weighted Average	0.42	0.50	0.46	0.46	0.62	0.53	16
	Macro Average	0.31	0.36	0.33	0.33	0.45	0.38	16
37 - SMOTE	*No Stress*	0.17	0.20	0.18	0.20	0.20	0.20	5
	Stress	0.60	0.55	0.57	0.64	0.64	0.64	11
	Accuracy	-	-	0.44	-	-	0.50	16
	Weighted Average	0.46	0.44	0.45	0.50	0.50	0.50	16
	Macro Average	0.38	0.37	0.38	0.42	0.42	0.42	16
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
38	*No Stress*	0.75	0.43	0.55	0.67	0.57	0.62	7
	Stress	0.76	0.93	0.84	0.80	0.86	0.83	14
	Accuracy	-	-	0.76	-	-	0.76	21
	Weighted Average	0.76	0.76	0.74	0.76	0.76	0.76	21
	Macro Average	0.76	0.68	0.69	0.73	0.71	0.72	21
38 - SMOTE	*No Stress*	0.60	0.43	0.50	0.62	0.71	0.67	7
	Stress	0.75	0.86	0.80	0.85	0.79	0.81	14
	Accuracy	-	-	0.71	-	-	0.76	21

	Weighted Average	0.70	0.71	0.70	0.77	0.76	0.77	21
	Macro Average	0.8	0.64	0.65	0.74	0.75	0.74	21
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
39	*No Stress*	0.62	1.00	0.77	0.59	1.00	0.74	10
	Stress	1.00	0.14	0.25	0.00	0.00	0.00	7
	Accuracy	-	-	0.65	-	-	0.59	17
	Weighted Average	0.78	0.65	0.56	0.35	0.59	0.44	17
	Macro Average	0.81	0.57	0.51	0.29	0.50	0.37	17
39 - SMOTE	*No Stress*	0.62	0.80	0.70	0.59	1.00	0.74	10
	Stress	0.50	0.29	0.36	0.00	0.00	0.00	7
	Accuracy	-	-	0.59	-	-	0.59	17
	Weighted Average	0.57	0.59	0.56	0.35	0.59	0.44	17
	Macro Average	0.56	0.54	0.53	0.29	0.50	0.37	17
User	**RF**				**SVM**			**Support**
	Items	***Precision***	***Recall***	***F1-Score***	***Precision***	***Recall***	***F1-Score***	
40	*No Stress*	0.88	1.00	0.94	0.86	0.80	0.83	15
	Stress	1.00	0.33	0.50	0.25	0.33	0.29	3
	Accuracy	-	-	0.89	-	-	0.72	18

	Weighted Average	0.90	0.89	0.86	0.76	0.72	0.74	18
	Macro Average	0.94	0.67	0.72	0.55	0.57	0.56	18
40- SMOTE	*No Stress*	0.87	0.87	0.87	0.88	0.93	0.90	15
	Stress	0.33	0.33	0.33	0.50	0.33	0.40	3
	Accuracy	-	-	0.78	-	-	0.83	18
	Weighted Average	0.78	0.78	0.78	0.81	0.83	0.82	18
	Macro Average	0.60	0.60	0.60	0.69	0.63	0.65	18

Table B10: Generalized Model - RF Feature Importance, D

All	
Feature	**Values**
MAP	0.018
ECG_DC	0.016
User	0.015
ECG_AR_AbsolutePower_HF	0.014
ECG_AR_AbsolutePower_LF	0.014
dia	0.013
HRV-1	0.013
Empatica_AR_RelativePower_LF	0.013
Empatica_AR_LFHF	0.012
ECG_Stress Index	0.012
Gender - Male	
Feature	**Values**
MAP	0.032
AW Min HR - Interval	0.027
Empatica_SampEn	0.021
Empatica_MSE13	0.018
Empatica_MSE17	0.017

sys	0.016
Empatica_MSE3	0.016
Empatica_MSE2	0.016
Short Term Min	0.016
Weight	0.016
Gender - Female	
Feature	**Values**
User	0.038
Weight	0.022
ECG_SNS_Index	0.020
ECG_DC	0.019
ECG_SDNN	0.019
ECG_AC	0.016
ECG_RMSSD	0.014
Empatica_FFT_LFHF	0.014
ECG_FFT_AbsolutePower_HF	0.014
ECG_AR_AbsolutePower_HF	0.013
Income - Low	
Feature	**Values**
ECG_Dcmod	0.019
Empatica_MSE2	0.019
MAP	0.018
ECG_RMSSD	0.018
User	0.017
ECG_SDNN	0.017
ECG_Stress Index	0.016
Empatica_MSE12	0.016
Empatica_DC	0.015
ECG_DC	0.014
Income Medium High	
Feature	**Values**
ECG_DC	0.021
ECG_SD1SD2	0.016
ECG_AC	0.015
dia	0.014
Empatica_DC	0.014
Empatica_FFT_RelativePower_VLF	0.014
HRV-1	0.014
ECG_AR_AbsolutePower_HF	0.013

Empatica_FFT_AbsolutePower_HF	0.013
Empatica_FFT_HF	0.012
Employment Students	
Feature	**Values**
User	0.042
ECG_RMSSD	0.023
MAP	0.022
ECG_Dcmod	0.021
ECG_Stress Index	0.021
ECG_DC	0.020
ECG_FFT_TotalPower	0.019
ECG_FFT_AbsolutePower_LF	0.018
ECG_SNS_Index	0.017
ECG_AR_AbsolutePower_HF	0.017
Employment Workers	
Feature	**Values**
MAP	0.027
Weight	0.017
sys	0.015
AW Max Steps	0.015
HRV-1	0.014
Empatica_FFT_RelativePower_VLF	0.014
Empatica_MSE11	0.014
ECG_DC	0.014
Empatica_Max HR	0.014
AW Mean Steps	0.014
Age 18-24	
Feature	**Values**
Weight	0.026
sys	0.024
Empatica_Mean line length	0.018
Empatica_MSE2	0.016
ECG_FFT_RelativePower_LF	0.015
Empatica_Min HR	0.015
AW Max Steps	0.015
AW Mean Steps	0.014
ECG_FFT_LF	0.014
ECG_FFT_AbsolutePower_LF	0.014
Age 25-34	

Feature	Values
MAP	0.021
HRV-1	0.021
Weight	0.021
Empatica_MSE4	0.016
Empatica_FFT_RelativePower_VLF	0.015
Temp	0.014
ECG_SD HR	0.014
ECG_DC	0.013
ECG_AR_AbsolutePower_LF	0.013
ECG_AC	0.013
Age 35-44	
Feature	**Values**
Temp	0.019
User	0.018
ECG_AR_LFHF	0.016
ECG_AC	0.015
Empatica_MSE15	0.015
AW Mean Steps	0.015
ECG_AR_RelativePower_LF	0.015
Empatica_FFT_HF	0.015
Empatica_alpha2	0.014
ECG_DC	0.014
Healthy	
Feature	**Values**
MAP	0.023
ECG_DC	0.017
ECG_AR_HF	0.017
ECG_AR_AbsolutePower_HF	0.015
Empatica_MSE2	0.014
ECG_FFT_AbsolutePower_HF	0.014
AW Min HR	0.014
Empatica_MSE3	0.014
sys	0.013
ECG_SDNN	0.013

Table B11: Generalized Model - RF Feature Importance, DECG

All	
Feature	**Values**
User	0.11
ECG_DC	0.049
ECG_AR_AbsolutePower_HF	0.044
ECG_Max HR	0.043
ECG_FFT_AbsolutePower_HF	0.041
ECG_SDNN	0.040
ECG_AC	0.040
ECG_Stress Index	0.040
ECG_FFT_RelativePower_LF	0.039
ECG_SNS_Index	0.038
Gender - Male	
Feature	**Values**
User	0.137
ECG_Mean RR	0.064
ECG_Mean HR	0.053
ECG_PNS Index	0.052
ECG_AR_HF	0.049
ECG_SD1SD2	0.042
ECG_RMSSD	0.040
ECG_AR_AbsolutePower_LF	0.037
ECG_DC	0.036
ECG_AR_AbsolutePower_HF	0.036
Gender - Female	
Feature	**Values**
User	0.067
ECG_DC	0.044
ECG_AC	0.044
ECG_AR_AbsolutePower_HF	0.044
ECG_Stress Index	0.044
ECG_FFT_AbsolutePower_HF	0.043
ECG_SDNN	0.042
ECG_Max HR	0.042
ECG_Mean RR	0.041
ECG_SNS_Index	0.041
Income - Low	

Feature	Values
User	0.045
ECG_Stress Index	0.042
ECG_DC	0.042
ECG_AC	0.041
ECG_AR_AbsolutePower_HF	0.041
ECG_Dcmod	0.039
ECG_RMSSD	0.039
ECG_SD HR	0.039
ECG_AR_AbsolutePower_LF	0.039
ECG_FFT_LFHF	0.039
Income Medium High	
Feature	**Values**
User	0.158
ECG_DC	0.052
ECG_AC	0.045
ECG_AR_AbsolutePower_LF	0.041
ECG_SD1SD2	0.040
ECG_SDNN	0.038
ECG_Mean RR	0.038
ECG_Stress Index	0.036
ECG_FFT_AbsolutePower_HF	0.034
ECG_SD HR	0.034
Employment Students	
Feature	**Values**
User	0.094
ECG_AR_AbsolutePower_HF	0.047
ECG_RMSSD	0.042
ECG_DC	0.041
ECG_SD HR	0.040
ECG_Stress Index	0.040
ECG_Dcmod	0.040
ECG_FFT_RelativePower_LF	0.039
ECG_FFT_LFHF	0.038
ECG_AC	0.037
Employment Workers	
Feature	**Values**
User	0.110
ECG_SDNN	0.044

ECG_AR_AbsolutePower_LF	0.041
ECG_AC	0.040
ECG_SD1SD2	0.039
ECG_Max HR	0.039
ECG_DC	0.039
ECG_Mean RR	0.038
ECG_PNS Index	0.038
ECG_Stress Index	0.037
Age 18-24	
Feature	**Values**
ECG_AR_HF	0.049
ECG_SD1SD2	0.045
ECG_FFT_HF	0.045
ECG_SD HR	0.043
ECG_Mean RR	0.042
ECG_FFT_AbsolutePower_LF	0.041
ECG_Mean HR	0.041
ECG_AR_AbsolutePower_HF	0.041
ECG_AC	0.040
ECG_FFT_LF	0.040
Age 25-34	
Feature	**Values**
User	0.048
ECG_DC	0.042
ECG_AR_LFHF	0.042
ECG_AR_AbsolutePower_HF	0.042
ECG_FFT_LF	0.042
ECG_AR_AbsolutePower_LF	0.042
ECG_Mean RR	0.041
ECG_FFT_LFHF	0.040
ECG_SNS_Index	0.040
ECG_AR_RelativePower_LF	0.040
Age 35-44	
Feature	**Values**
User	0.083
ECG_SD1SD2	0.047
ECG_AR_HF	0.046
ECG_Mean RR	0.039
ECG_AR_RelativePower_LF	0.038

ECG_SD HR	0.038
ECG_AC	0.038
ECG_FFT_RelativePower_LF	0.037
ECG_DC	0.036
ECG_FFT_HF	0.036
Healthy	
Feature	**Values**
User	0.053
ECG_AC	0.044
ECG_DC	0.043
ECG_AR_AbsolutePower_HF	0.041
ECG_FFT_AbsolutePower_HF	0.039
ECG_SDNN	0.039
ECG_SD1SD2	0.039
ECG_FFT_LFHF	0.038
ECG_FFT_RelativePower_LF	0.038
ECG_Stress Index	0.038

Table B12: Generalized Model - RF Feature Importance, DA

All	
Feature	**Values**
User	0.052
AW Mean Steps	0.033
HRV-1	0.033
ECG_AC	0.033
AW Mean HR - Interval	0.033
AW Min HR - Interval	0.032
ECG_DC	0.031
ECG_SDNN	0.030
AW Max Steps	0.030
ECG_Max HR	0.029
Gender - Male	
Feature	**Values**
User	0.091
AW Min HR - Interval	0.046
ECG_Mean RR	0.036
ECG_Mean HR	0.036

AW Mean HR - Interval	0.035
HRV-1	0.034
ECG_AR_HF	0.031
ECG_SD1SD2	0.031
ECG_PNS Index	0.029
ECG_RMSSD	0.027
Gender - Female	
Feature	**Values**
User	0.053
ECG_SDNN	0.036
ECG_DC	0.036
ECG_FFT_AbsolutePower_HF	0.036
ECG_AC	0.035
ECG_Stress Index	0.034
ECG_AR_AbsolutePower_HF	0.034
AW Mean Steps	0.032
ECG_RMSSD	0.031
ECG_SD HR	0.031
Income - Low	
Feature	**Values**
AW Mean Steps	0.035
User	0.034
HRV-1	0.032
ECG_AC	0.032
ECG_Stress Index	0.032
ECG_DC	0.030
AW Min HR - Interval	0.030
AW Max Steps	0.030
ECG_RMSSD	0.029
AW Mean HR	0.029
Income Medium High	
Feature	**Values**
User	0.075
ECG_DC	0.035
AW Mean HR - Interval	0.034
ECG_AC	0.032
HRV-1	0.031
Short Term Min	0.030
ECG_SD1SD2	0.030

AW Min HR - Interval	0.029
AW Max HR - Interval	0.028
ECG_SD HR	0.028
Employment Students	
Feature	**Values**
User	0.116
ECG_RMSSD	0.041
ECG_Stress Index	0.038
ECG_DC	0.036
ECG_SDNN	0.036
ECG_Dcmod	0.036
ECG_AR_AbsolutePower_HF	0.033
HRV-1	0.031
ECG_AC	0.030
ECG_SD HR	0.029
Employment Workers	
Feature	**Values**
User	0.103
AW Mean HR - Interval	0.039
HRV-1	0.034
AW Min HR - Interval	0.031
ECG_SDNN	0.030
AW Mean Steps	0.029
ECG_AR_AbsolutePower_LF	0.029
ECG_RMSSD	0.028
ECG_DC	0.028
AW Max Steps	0.028
Age 18-24	
Feature	**Values**
AW Mean Steps	0.048
ECG_AR_HF	0.041
ECG_FFT_HF	0.038
AW Max HR	0.033
HRV-1	0.032
ECG_FFT_LFHF	0.031
AW Max Steps	0.031
ECG_SD1SD2	0.031
AW Mean HR - Interval	0.030
ECG_FFT_RelativePower_LF	0.030

Age 25-34	
Feature	**Values**
HRV-1	0.041
User	0.040
AW Mean HR - Interval	0.038
ECG_AR_LFHF	0.031
ECG_DC	0.031
AW Mean Steps	0.030
ECG_SD HR	0.030
ECG_AR_AbsolutePower_HF	0.030
ECG_AR_RelativePower_LF	0.029
ECG_FFT_LFHF	0.029
Age 35-44	
Feature	**Values**
User	0.068
AW Max HR - Interval	0.036
ECG_SD1SD2	0.033
ECG_AR_HF	0.032
ECG_AR_RelativePower_LF	0.032
HRV-1	0.031
Short Term Min	0.030
AW Max Steps	0.029
ECG_AR_LFHF	0.029
AW Mean Steps	0.029
Healthy	
Feature	**Values**
User	0.040
HRV-1	0.034
AW Mean Steps	0.034
AW Mean HR - Interval	0.033
ECG_AC	0.033
ECG_DC	0.031
AW Max HR - Interval	0.030
ECG_SD1SD2	0.030
AW Max Steps	0.030
AW Min HR - Interval	0.030

Table B13: Generalized Model - RF Feature Importance, DAW

All	
Feature	**Values**
Weight	0.047
User	0.042
MAP	0.034
Temp	0.031
AW Mean Steps	0.029
HRV-1	0.028
AW Mean HR - Interval	0.028
ECG_DC	0.026
dia	0.026
ECG_AC	0.025
Gender - Male	
Feature	**Values**
Weight	0.058
User	0.051
AW Min HR - Interval	0.040
Temp	0.038
MAP	0.037
AW Mean HR - Interval	0.030
ECG_SD1SD2	0.027
ECG_AR_HF	0.027
dia	0.026
sys	0.026
Gender - Female	
Feature	**Values**
Weight	0.041
User	0.039
Temp	0.030
MAP	0.030
HRV-1	0.028
ECG_AC	0.028
AW Mean Steps	0.028
AW Mean HR - Interval	0.027
ECG_DC	0.026
ECG_Stress Index	0.026
Income - Low	

Feature	Values
Weight	0.038
Temp	0.035
AW Min HR - Interval	0.028
MAP	0.028
AW Mean Steps	0.027
ECG_AC	0.026
User	0.026
ECG_SD HR	0.026
HRV-1	0.026
ECG_FFT_AbsolutePower_HF	0.026
Income Medium High	
Feature	**Values**
User	0.074
Weight	0.064
Temp	0.041
AW Mean HR - Interval	0.036
dia	0.034
sys	0.031
HRV-1	0.027
AW Min HR - Interval	0.026
ECG_SD1SD2	0.025
AW Max HR - Interval	0.024
Employment Students	
Feature	**Values**
Weight	0.110
User	0.074
ECG_RMSSD	0.033
MAP	0.032
Temp	0.032
ECG_Stress Index	0.032
ECG_Dcmod	0.029
ECG_DC	0.029
ECG_SD HR	0.027
ECG_AR_AbsolutePower_HF	0.027
Employment Workers	
Feature	**Values**
User	0.065
Weight	0.053

Temp	0.042
AW Mean HR - Interval	0.036
MAP	0.033
AW Min HR - Interval	0.030
HRV-1	0.026
ECG_SD1SD2	0.025
sys	0.025
dia	0.025
Age 18-24	
Feature	**Values**
Weight	0.061
AW Mean Steps	0.036
ECG_AR_HF	0.032
Temp	0.031
ECG_FFT_HF	0.028
HRV-1	0.028
ECG_FFT_LFHF	0.027
ECG_FFT_RelativePower_LF	0.027
AW Max Steps	0.026
AW Mean HR - Interval	0.025
Age 25-34	
Feature	**Values**
Weight	0.063
Temp	0.032
HRV-1	0.031
MAP	0.031
AW Mean HR - Interval	0.031
AW Min HR - Interval	0.030
ECG_DC	0.028
ECG_AR_AbsolutePower_HF	0.027
User	0.026
ECG_FFT_LF	0.025
Age 35-44	
Feature	**Values**
Weight	0.053
User	0.051
ECG_SD1SD2	0.036
ECG_AR_RelativePower_LF	0.033
dia	0.031

MAP	0.031
sys	0.030
ECG_AR_LFHF	0.028
HRV-1	0.027
AW Max HR - Interval	0.026
Healthy	
Feature	**Values**
Weight	0.054
Temp	0.035
MAP	0.032
User	0.027
AW Mean Steps	0.026
HRV-1	0.026
ECG_AC	0.026
AW Mean HR - Interval	0.025
sys	0.025
ECG_DC	0.025

Table B14: Generalized Model - RF Feature Importance, DW

All	
Feature	**Values**
Weight	0.251
User	0.204
Temp	0.160
MAP	0.151
dia	0.118
Gender - Male	
Feature	**Values**
Weight	0.295
User	0.201
Temp	0.147
MAP	0.128
dia	0.122
Gender - Female	
Feature	**Values**
Weight	0.25
User	0.18
Temp	0.18

MAP	0.15
sys	0.12
Income - Low	
Feature	**Values**
Weight	0.235
Temp	0.193
MAP	0.166
User	0.139
dia	0.138
Income Medium High	
Feature	**Values**
Weight	0.299
User	0.278
Temp	0.160
dia	0.139
sys	0.125
Employment Students	
Feature	**Values**
Weight	0.293
User	0.222
Temp	0.145
MAP	0.144
dia	0.101
Employment Workers	
Feature	**Values**
Weight	0.252
User	0.237
Temp	0.154
MAP	0.137
dia	0.122
Age 18-24	
Feature	**Values**
Weight	0.295
Temp	0.207
MAP	0.151
dia	0.135
sys	0.127
Age 25-34	

Feature	Values
Weight	0.295
Temp	0.207
MAP	0.151
dia	0.135
sys	0.127
Age 35-44	
Feature	**Values**
Weight	0.231
MAP	0.173
Temp	0.168
sys	0.151
dia	0.146
Healthy	
Feature	**Values**
Weight	0.248
Temp	0.198
MAP	0.172
sys	0.139
dia	0.137

Table B15: Generalized Model - RF Feature Importance, DEmpatica

All	
Feature	**Values**
User	0.046
Empatica_resp	0.026
Empatica_MSE3	0.021
Empatica_ApEn	0.020
Empatica_FFT_RelativePower_VLF	0.020
Empatica_Max HR	0.018
Empatica_SD HR	0.018
Empatica_MSE4	0.018
Empatica_FFT_LF	0.018
Empatica_D2	0.017
Gender - Male	
Feature	**Values**
Empatica_MSE13	0.037

Empatica_resp	0.034
Empatica_MSE3	0.028
Empatica_MSE15	0.026
Empatica_MSE2	0.024
Empatica_MSE9	0.023
Empatica_MSE11	0.021
Empatica_FFT_LF	0.020
User	0.019
Empatica_MSE8	0.018
Gender - Female	
Feature	**Values**
User	0.056
Empatica_NN50	0.025
Empatica_MSE18	0.021
Empatica_TINN	0.019
Empatica_DC	0.019
Empatica_SampEn	0.019
Empatica_SD1SD2	0.019
Empatica_AR_RelativePower_VLF	0.018
Empatica_MSE19	0.018
Empatica_resp	0.017
Income - Low	
Feature	**Values**
Empatica_DET	0.024
Empatica_DC	0.024
User	0.024
Empatica_REC	0.021
Empatica_MSE15	0.020
Empatica_AR_AbsolutePower_VLF	0.020
Empatica_SampEn	0.020
Empatica_MSE3	0.020
Empatica_MSE2	0.019
Empatica_resp	0.019
Income Medium High	
Feature	**Values**
User	0.0512
Empatica_resp	0.0252
Empatica_DC	0.0212

Empatica_FFT_RelativePower_VLF	0.0210
Empatica_MSE12	0.0202
Empatica_AR_LF	0.0193
Empatica_MSE3	0.0191
Empatica_FFT_AbsolutePower_HF_log	0.0189
Empatica_FFT_AbsolutePower_HF	0.0175
Empatica_MSE14	0.0175
Employment Students	
Feature	**Values**
User	0.066
Empatica_resp	0.023
Empatica_DC	0.022
Empatica_SampEn	0.021
Empatica_FFT_HF	0.021
Empatica_REC	0.019
Empatica_MSE3	0.019
Empatica_FFT_RelativePower_LF	0.019
Empatica_MSE12	0.019
Empatica_AR_AbsolutePower_VLF	0.019
Employment Workers	
Feature	**Values**
Empatica_MSE13	0.023
User	0.021
Empatica_FFT_RelativePower_VLF	0.020
Empatica_MSE15	0.020
Empatica_Min HR	0.019
Empatica_MSE3	0.019
Empatica_ApEn	0.018
Empatica_MSE18	0.018
Empatica_Mean RR	0.018
Empatica_resp	0.018
Age 18-24	
Feature	**Values**
Empatica_MSE4	0.028
Empatica_MSE17	0.026
Empatica_FFT_AbsolutePower_VLF_log	0.024
Empatica_AR_LF	0.020
Empatica_FFT_AbsolutePower_VLF	0.020

Empatica_FFT_LF	0.019
Empatica_MSE15	0.019
Empatica_D2	0.019
Empatica_MSE9	0.019
Empatica_FFT_RelativePower_VLF	0.019
Age 25-34	
Feature	**Values**
User	0.096
Empatica_MSE6	0.024
Empatica_AR_LF	0.023
Empatica_resp	0.022
Empatica_MSE8	0.021
Empatica_MSE19	0.020
Empatica_FFT_LFHF	0.019
Empatica_SampEn	0.019
Empatica_MSE18	0.018
Empatica_MSE3	0.018
Age 35-44	
Feature	**Values**
User	0.032
Empatica_MSE3	0.031
Empatica_MSE15	0.028
Empatica_FFT_HF	0.027
Empatica_resp	0.025
Empatica_MSE5	0.021
Empatica_MSE16	0.020
Empatica_MSE13	0.019
Empatica_MSE12	0.018
Empatica_FFT_AbsolutePower_VLF_log	0.018
Healthy	
Feature	**Values**
Empatica_resp	0.030
Empatica_MSE3	0.028
Empatica_MSE2	0.022
Empatica_MSE15	0.022
User	0.021
Empatica_FFT_HF	0.020
Empatica_MSE4	0.019
Empatica_MSE13	0.019

Empatica_MSE17	0.018
Empatica_MSE8	0.018

Table B16: Generalized Model - RF Feature Importance, SDA

All	
Feature	**Values**
T+2 AW Consolidated Time Awake During Sleep	0.023
T+1 AW Consolidated Time Awake During Sleep	0.023
T+2 AW Number of Wake-Ups	0.021
AW Consolidated Time Awake During Sleep	0.018
AW Mean HR - Interval	0.017
AW Min HR - Interval	0.017
T-2 AW Mean HR	0.017
AW Total Time in Bed	0.017
T-2 AW Max HR	0.017
T-2 AW Total Time Asleep	0.016
Gender - Male	
Feature	**Values**
AW Min HR - Interval	0.042
AW Total Time in Bed	0.024
Short Term Min	0.023
User	0.023
T-2 AW Total Time Asleep	0.022
T+1 AW Consolidated Time Awake During Sleep	0.021
AW Mean HR - Interval	0.020
T-2 AW Total Time in Bed	0.020
ECG_AR_HF	0.019
T+2 AW Max HR	0.019
Gender - Female	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.030
T+2 AW Consolidated Time Awake During Sleep	0.024
User	0.020
AW Total Time in Bed	0.019
T+2 AW Number of Wake-Ups	0.019
AW Total Time Asleep	0.018
ECG_Stress Index	0.017
T-1 AW Total Time in Bed	0.017

AW Consolidated Time Awake During Sleep	0.017
AW Mean Steps	0.016
Income - Low	
Feature	**Values**
T+2 AW Consolidated Time Awake During Sleep	0.034
T+2 AW Number of Wake-Ups	0.033
T+1 AW Consolidated Time Awake During Sleep	0.031
AW Total Time in Bed	0.021
AW Consolidated Time Awake During Sleep	0.021
T+2 AW Total Time in Bed	0.020
T-1 Total Time in Bed	0.020
T-2 AW Number of Wake-Ups	0.019
T+1 AW Number of Wake-Ups	0.019
T-2 Consolidated Time Awake During Sleep	0.018
Income Medium High	
Feature	**Values**
User	0.029
AW Min HR	0.023
Short Term Min	0.022
T+2 AW Total Time in Bed	0.021
T-2 AW Mean HR	0.021
AW Min HR - Interval	0.020
T+2 AW Min HR	0.020
T+1 AW Consolidated Time Awake During Sleep	0.020
AW Mean HR - Interval	0.020
T+2 AW Number of Wake-Ups	0.019
Employment Students	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.035
T+2 AW Number of Wake-Ups	0.035
T+2 AW Consolidated Time Awake During Sleep	0.029
T+1 AW Number of Wake-Ups	0.022
AW Consolidated Time Awake During Sleep	0.021
AW Number of Wake-Ups	0.017
T+2 AW Total Time in Bed	0.017
T-2 AW Number of Wake-Ups	0.017
ECG_FFT_RelativePower_LF	0.016
ECG_FFT_LFHF	0.016

Employment Workers	
Feature	**Values**
User	0.028
AW Min HR - Interval	0.024
AW Total Time in Bed	0.021
AW Mean HR - Interval	0.020
T+1 AW Max HR	0.019
Short Term Min	0.019
T+2 AW Total Time in Bed	0.019
T-2 AW Mean HR	0.019
AW Total Time Asleep	0.018
T+2 AW Mean HR	0.017
Age 18-24	
Feature	**Values**
T+2 AW Number of Wake-Ups	0.024
T+1 AW Consolidated Time Awake During Sleep	0.024
T+2 AW Consolidated Time Awake During Sleep	0.023
AW Total Time in Bed	0.021
AW Mean Steps	0.020
AW % of Time Asleep While In Bed	0.019
T+2 AW % of Time Asleep While In Bed	0.019
T+2 AW Total Time in Bed	0.018
AW Consolidated Time Awake During Sleep	0.017
T-2 AW Min HR	0.017
Age 25-34	
Feature	**Values**
T+2 AW Consolidated Time Awake During Sleep	0.027
T+2 AW Number of Wake-Ups	0.026
T-2 AW Total Time Asleep	0.025
T-2 AW Total Time in Bed	0.024
AW Consolidated Time Awake During Sleep	0.022
T+2 AW Total Time in Bed	0.022
AW Total Time in Bed	0.022
AW Mean HR - Interval	0.021
HRV-1	0.021
T-1 AW Total Time in Bed	0.020
Age 35-44	
Feature	**Values**
AW Min HR	0.030

T+1 AW Mean HR	0.026
T-2 AW % of Time Asleep While In Bed	0.024
AW Min HR - Interval	0.024
Short Term Min	0.022
AW Mean HR - Interval	0.020
AW Min Steps	0.019
T+2 AW Mean HR	0.019
AW Mean Steps	0.018
T-1 AW Min HR	0.018
Healthy	
Feature	**Values**
T+2 AW Consolidated Time Awake During Sleep	0.021
T-2 AW Max HR	0.020
T+1 AW Consolidated Time Awake During Sleep	0.020
AW Mean Steps	0.019
T+1 AW Max HR	0.017
AW Min HR - Interval	0.017
AW Total Time Asleep	0.017
T-2 AW Total Time Asleep	0.017
AW Max Steps	0.017
T+2 AW Number of Wake-Ups	0.016

Table B17: Generalized Model - RF Feature Importance, SDAW

All	
Feature	**Values**
T+2 AW Number of Wake-Ups	0.020
T+1 AW Consolidated Time Awake During Sleep	0.019
Weight	0.017
T+2 AW Consolidated Time Awake During Sleep	0.017
Temp	0.015
AW Total Time in Bed	0.014
MAP	0.014
T+2 AW Min HR	0.014
AW Min HR - Interval	0.013
T+1 AW Number of Wake-Ups	0.013
Gender - Male	
Feature	**Values**
T+1 AW Max HR	0.030

T-2 AW Total Time Asleep	0.024
T-2 AW Max HR	0.021
AW Max HR	0.020
ECG_FFT_RelativePower_HF	0.018
Weight	0.018
MAP	0.018
T-1 AW Max HR	0.017
sys	0.016
Short Term Mean	0.016
Gender - Female	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.022
AW Total Time in Bed	0.018
Weight	0.017
Temp	0.017
T+2 AW Number of Wake-Ups	0.015
T+2 AW Consolidated Time Awake During Sleep	0.015
ECG_Stress Index	0.014
AW Min HR - Interval	0.014
AW Min HR	0.014
ECG_SNS_Index	0.013
Income - Low	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.033
T+2 AW Consolidated Time Awake During Sleep	0.028
T+1 AW Number of Wake-Ups	0.027
T+2 AW Number of Wake-Ups	0.024
AW Consolidated Time Awake During Sleep	0.016
T-2 AW Number of Wake-Ups	0.015
T+2 AW % of Time Asleep While In Bed	0.014
T+2 Time Spent in REM Stage	0.014
AW Min HR - Interval	0.014
T-2 Time Spent in Light Stage	0.013
Income Medium High	
Feature	**Values**
AW Min HR	0.027
T+2 AW Min HR	0.018
T-1 AW Total Time in Bed	0.017

T-2 AW Min HR	0.017
AW Total Time in Bed	0.016
Temp	0.016
T+1 AW Consolidated Time Awake During Sleep	0.016
T-2 AW Mean HR	0.015
AW Consolidated Time Awake During Sleep	0.014
ECG_AR_AbsolutePower_HF	0.014
Employment Students	
Feature	**Values**
T+2 AW Consolidated Time Awake During Sleep	0.031
T+2 AW Number of Wake-Ups	0.027
T+2 Time Spent in REM Stage	0.026
T+1 AW Consolidated Time Awake During Sleep	0.025
T+1 AW Number of Wake-Ups	0.023
T+2 AW Min HR	0.019
Time Spent in REM Stage	0.018
AW Number of Wake-Ups	0.017
Weight	0.013
T-2 AW Min HR	0.012
Employment Workers	
Feature	**Values**
AW Total Time in Bed	0.018
AW Mean HR - Interval	0.017
AW Min HR	0.017
Temp	0.016
AW Mean Steps	0.015
AW Min HR - Interval	0.015
T-1 AW Total Time in Bed	0.014
Time Spent in REM Stage	0.014
HRV-1	0.013
sys	0.013
Age 18-24	
Feature	**Values**
Weight	0.025
T+2 Time Spent in REM Stage	0.025
T+2 AW Consolidated Time Awake During Sleep	0.024
T+2 AW Min HR	0.023
T+1 AW Min HR	0.019

T+1 AW Consolidated Time Awake During Sleep	0.016
Time Spent in REM Stage	0.016
AW Number of Wake-Ups	0.015
ECG_FFT_RelativePower_LF	0.015
T-1 AW Min HR	0.015
Age 25-34	
Feature	**Values**
T+2 AW Consolidated Time Awake During Sleep	0.030
T+2 AW Number of Wake-Ups	0.028
Weight	0.022
T+2 AW Mean HR	0.018
T+1 AW Min HR	0.018
T+2 AW Min HR	0.018
T-2 % of Time Asleep While In Bed	0.017
T+1 AW Consolidated Time Awake During Sleep	0.017
% of Time Asleep While In Bed	0.016
AW Number of Wake-Ups	0.015
Age 35-44	
Feature	**Values**
T+2 AW Mean HR	0.018
T+1 AW Mean HR	0.018
T+2 Time Spent in Light Stage	0.018
T-1 Withings Total Time Asleep	0.016
AW Max HR - Interval	0.016
Temp	0.015
T-2 Time Spent in Deep Stage	0.015
HRV-1	0.015
AW Mean HR - Interval	0.015
AW Min HR	0.014
Healthy	
Feature	**Values**
T+2 AW Number of Wake-Ups	0.020
AW Total Time in Bed	0.019
Weight	0.018
T+1 AW Consolidated Time Awake During Sleep	0.018
T+2 AW Consolidated Time Awake During Sleep	0.017
User	0.016
T-2 AW Max HR	0.015
T-2 AW Total Time in Bed	0.015

T-2 AW Min HR	0.014
T-2 Time Spent in REM Stage	0.014

Table B18: Generalized Model - RF Feature Importance, SDW

All	
Feature	**Values**
Weight	0.089
Temp	0.082
MAP	0.068
dia	0.063
sys	0.062
T-2 Time Spent in REM Stage	0.034
T+2 Time Spent in REM Stage	0.031
Withings Total Time Asleep	0.030
T+1 Withings Total Time Asleep	0.029
Time Spent in REM Stage	0.029
Gender - Male	
Feature	**Values**
T-1 Withings Total Time Asleep	0.058
T+2 Withings Total Time Asleep	0.056
T-2 Withings Total Time Asleep	0.055
Temp	0.049
MAP	0.048
T+1 Withings Total Time Asleep	0.045
sys	0.044
T-2 Time Spent in Light Stage	0.044
T+2 Time Spent in Light Stage	0.043
Withings Total Time Asleep	0.043
Gender - Female	
Feature	**Values**
Weight	0.067
User	0.053
T+1 Withings Total Time Asleep	0.053
Temp	0.044
T+2 Withings Total Time Asleep	0.042
Withings Total Time Asleep	0.042
T-2 Time Spent in REM Stage	0.039

sys	0.037
T+2 Time Spent in REM Stage	0.036
T-1 Withings Total Time Asleep	0.035
Income - Low	
Feature	**Values**
Weight	0.066
Temp	0.062
MAP	0.055
sys	0.050
dia	0.049
T+2 Time Spent in REM Stage	0.043
Time Spent in Light Stage	0.040
T-2 Time Spent in REM Stage	0.039
Time Spent in REM Stage	0.035
Total Time In Bed	0.034
Income Medium High	
Feature	**Values**
Temp	0.086
Weight	0.071
MAP	0.065
sys	0.060
dia	0.060
User	0.052
Time Spent in REM Stage	0.042
T+1 Withings Total Time Asleep	0.033
T+2 Time Spent in Light Stage	0.028
T+2 Withings Total Time Asleep	0.028
Employment Students	
Feature	**Values**
T+2 Time Spent in REM Stage	0.070
Weight	0.067
Time Spent in REM Stage	0.055
Temp	0.053
MAP	0.047
T-2 Time Spent in REM Stage	0.046
User	0.040
sys	0.040
dia	0.039

T+2 % of Time Asleep While In Bed	0.036
Employment Workers	
Feature	**Values**
Temp	0.088
Weight	0.071
MAP	0.068
dia	0.067
sys	0.064
T+1 Withings Total Time Asleep	0.035
User	0.034
T-1 Withings Total Time Asleep	0.032
T-2 Time Spent in REM Stage	0.029
T+2 Withings Total Time Asleep	0.029
Age 18-24	
Feature	**Values**
Weight	0.098
Temp	0.078
MAP	0.072
sys	0.066
dia	0.065
T+2 Time Spent in REM Stage	0.049
Time Spent in REM Stage	0.044
% of Time Asleep While In Bed	0.033
T+2 Withings Total Time Asleep	0.031
T+2 % of Time Asleep While In Bed	0.030
Age 25-34	
Feature	**Values**
Weight	0.103
T-2 % of Time Asleep While In Bed	0.057
T+2 % of Time Asleep While In Bed	0.050
T-2 Time Spent in Deep Stage	0.049
% of Time Asleep While In Bed	0.049
User	0.046
Time Spent in REM Stage	0.046
Temp	0.038
MAP	0.038
T+2 Time Spent in Deep Stage	0.034
Age 35-44	

Feature	Values
T+2 Withings Total Time Asleep	0.069
T+2 Total Time In Bed	0.064
T-1 Withings Total Time Asleep	0.056
T+2 Time Spent in Light Stage	0.047
T-2 Time Spent in REM Stage	0.046
Total Time In Bed	0.045
User	0.045
T-2 Total Time In Bed	0.041
Weight	0.040
T-2 Withings Total Time Asleep	0.039
Healthy	
Feature	**Values**
Weight	0.080
Temp	0.049
MAP	0.047
sys	0.045
T-2 Time Spent in REM Stage	0.045
T+1 Withings Total Time Asleep	0.041
dia	0.040
Time Spent in Light Stage	0.039
Time Spent in Deep Stage	0.037
T+2 Withings Total Time Asleep	0.036

Table B19: Generalized Model - RF Feature Importance, SDS

All	
Feature	**Values**
T+2 AW Number of Wake-Ups	0.049
T+1 AW Consolidated Time Awake During Sleep	0.047
User	0.034
T+2 AW Consolidated Time Awake During Sleep	0.033
AW Total Time in Bed	0.033
T-2 AW Min HR	0.029
T+2 AW Min HR	0.028
T+1 AW Number of Wake-Ups	0.028
AW Min HR	0.026
T-2 Time Spent in REM Stage	0.024

Gender - Male	
Feature	**Values**
T+1 AW Max HR	0.039
T-2 AW Max HR	0.038
T+2 AW Consolidated Time Awake During Sleep	0.034
AW Max HR	0.033
T-2 AW Total Time in Bed	0.033
T-1 AW Max HR	0.028
T-2 AW Total Time Asleep	0.028
T+2 AW % of Time Asleep While In Bed	0.026
T-2 Time Spent in REM Stage	0.025
T+2 % of Time Asleep While In Bed	0.022
Gender - Female	
Feature	**Values**
AW Total Time in Bed	0.039
T+1 AW Consolidated Time Awake During Sleep	0.037
User	0.027
T+2 AW Total Time Asleep	0.026
AW Min HR	0.025
Time Spent in Light Stage	0.024
T+2 Time Spent in REM Stage	0.023
AW Consolidated Time Awake During Sleep	0.022
T-2 AW Min HR	0.022
T-2 Time Spent in Deep Stage	0.021
Income - Low	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.043
T+2 AW Consolidated Time Awake During Sleep	0.036
T+2 AW Number of Wake-Ups	0.034
Time Spent in Light Stage	0.028
T-2 W Consolidated Time Awake During Sleep	0.027
T+1 AW Number of Wake-Ups	0.027
T+2 AW % of Time Asleep While In Bed	0.026
T-2 AW Number of Wake-Ups	0.022
T-1 AW Mean HR	0.021
AW Consolidated Time Awake During Sleep	0.021
Income Medium High	
Feature	**Values**

AW Min HR	0.047
T-1 AW Total Time in Bed	0.032
AW Total Time in Bed	0.028
T+2 AW Min HR	0.028
T+1 AW Min HR	0.026
T-2 AW Min HR	0.024
T-2 AW Mean HR	0.024
AW Number of Wake-Ups	0.023
AW Consolidated Time Awake During Sleep	0.022
T+1 AW Consolidated Time Awake During Sleep	0.022
Employment Students	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.047
T+2 AW Number of Wake-Ups	0.041
T+2 Time Spent in REM Stage	0.040
T+2 AW Min HR	0.036
T+1 AW Number of Wake-Ups	0.035
T+2 AW Consolidated Time Awake During Sleep	0.030
Time Spent in REM Stage	0.029
AW Number of Wake-Ups	0.022
T-2 Time Spent in REM Stage	0.022
User	0.021
Employment Workers	
Feature	**Values**
AW Min HR	0.040
AW Total Time in Bed	0.040
T-1 AW Total Time in Bed	0.036
User	0.030
Time Spent in REM Stage	0.028
AW Max HR	0.027
T+2 AW Max HR	0.024
AW Consolidated Time Awake During Sleep	0.024
T+2 AW Total Time in Bed	0.024
AW Mean HR	0.022
Age 18-24	
Feature	**Values**
Time Spent in REM Stage	0.043
T+2 Time Spent in REM Stage	0.042

T+1 AW Min HR	0.040
T-2 AW Min HR	0.040
T+2 AW Min HR	0.038
T+1 AW Consolidated Time Awake During Sleep	0.038
T-1 AW Min HR	0.035
AW Min HR	0.032
T-2 AW Mean HR	0.029
T+2 AW Total Time in Bed	0.026
Age 25-34	
Feature	**Values**
T+2 AW Number of Wake-Ups	0.058
T+2 AW Consolidated Time Awake During Sleep	0.047
T+2 AW Mean HR	0.042
T+1 AW Min HR	0.034
T-2 % of Time Asleep While In Bed	0.033
AW Number of Wake-Ups	0.031
T+2 AW Min HR	0.030
T+1 AW Mean HR	0.027
T+1 AW Consolidated Time Awake During Sleep	0.027
AW Mean HR	0.024
Age 35-44	
Feature	**Values**
T-1 Withings Total Time Asleep	0.042
T+1 AW Mean HR	0.031
T-2 Time Spent in REM Stage	0.030
T+1 Withings Total Time Asleep	0.028
T+2 AW Mean HR	0.027
T+2 Time Spent in Light Stage	0.026
T+2 AW Max HR	0.025
AW Min HR	0.025
T-2 AW % of Time Asleep While In Bed	0.024
T-2 Time Spent in Deep Stage	0.023
Healthy	
Feature	**Values**
T+2 AW Consolidated Time Awake During Sleep	0.051
T-2 AW Total Time in Bed	0.044
T+1 AW Consolidated Time Awake During Sleep	0.043
T+2 AW Total Time in Bed	0.036
User	0.035

T+2 AW Number of Wake-Ups	0.032
T-2 AW Max HR	0.029
T-1 AW Total Time in Bed	0.025
T+2 Time Spent in REM Stage	0.024
T+1 AW Min HR	0.021

Table B20: Generalized_Imb Model - RF Feature Importance, D

All	
Feature	**Values**
MAP	0.032
ECG_DC	0.023
Weight	0.019
dia	0.019
Empatica_MSE2	0.019
Empatica_DC	0.019
ECG_SDNN	0.018
ECG_AR_AbsolutePower_HF	0.018
ECG_Stress Index	0.018
Empatica_AR_LFHF	0.018
Gender - Male	
Feature	**Values**
MAP	0.030
Weight	0.029
AW Min HR - Interval	0.029
Temp	0.020
Empatica_FFT_HF	0.020
sys	0.019
dia	0.019
Empatica_resp	0.019
Empatica_SampEn	0.018
Empatica_MSE2	0.018
Gender - Female	
Feature	**Values**
Weight	0.026
ECG_DC	0.022
ECG_SDNN	0.021
ECG_Stress Index	0.021
ECG_SNS_Index	0.019

ECG_Dcmod	0.019
ECG_AC	0.018
Empatica_AR_AbsolutePower_VLF	0.018
ECG_AR_AbsolutePower_HF	0.018
ECG_FFT_AbsolutePower_HF	0.018
Income - Low	
Feature	**Values**
MAP	0.023
Empatica_DC	0.021
Empatica_MSE2	0.020
Weight	0.018
ECG_AC	0.017
Empatica_MSE3	0.017
ECG_Stress Index	0.017
ECG_Dcmod	0.017
ECG_RMSSD	0.017
ECG_DC	0.017
Income Medium High	
Feature	**Values**
ECG_DC	0.030
Empatica_DC	0.025
dia	0.024
ECG_SD1SD2	0.020
HRV-1	0.020
Empatica_FFT_AbsolutePower_HF	0.019
Weight	0.019
Empatica_FFT_RelativePower_VLF	0.019
Short Term Min	0.019
ECG_AC	0.018
Employment Students	
Feature	**Values**
MAP	0.027
Weight	0.023
Empatica_DC	0.020
Empatica_MSE2	0.020
ECG_RMSSD	0.019
ECG_SD HR	0.019
ECG_Stress Index	0.019
ECG_SDNN	0.019

sys	0.019
Empatica_MSE3	0.018
Employment Workers	
Feature	**Values**
MAP	0.026
Weight	0.024
Empatica_FFT_RelativePower_VLF	0.019
Empatica_Mean RR	0.018
HRV-1	0.018
Short Term Min	0.017
Empatica_Max HR	0.016
ECG_DC	0.016
AW Mean Steps	0.015
Empatica_SD HR	0.015
Age 18-24	
Feature	**Values**
Weight	0.029
sys	0.019
AW Max Steps	0.019
AW Mean Steps	0.018
dia	0.018
Empatica_MSE4	0.017
Empatica_MSE2	0.016
ECG_FFT_RelativePower_LF	0.016
ECG_SD1SD2	0.016
Temp	0.016
Age 25-34	
Feature	**Values**
MAP	0.027
Weight	0.024
HRV-1	0.020
ECG_DC	0.016
Empatica_FFT_LFHF	0.016
Empatica_FFT_RelativePower_VLF	0.015
Empatica_FFT_RelativePower_LF	0.015
ECG_AR_AbsolutePower_HF	0.015
ECG_AR_AbsolutePower_LF	0.015
Empatica_AR_RelativePower_VLF	0.015
Age 35-44	

Feature	Values
Temp	0.019
Weight	0.019
Empatica_FFT_HF	0.018
Empatica_MSE11	0.018
ECG_AR_AbsolutePower_HF	0.017
ECG_DC	0.016
HRV-1	0.016
Empatica_FFT_LF	0.016
Empatica_MSE4	0.016
Empatica_Mean RR	0.015
Healthy	
Feature	**Values**
MAP	0.035
Empatica_resp	0.022
ECG_DC	0.021
Empatica_MSE2	0.020
dia	0.019
ECG_AR_AbsolutePower_HF	0.018
Empatica_MSE3	0.018
ECG_FFT_AbsolutePower_HF	0.017
sys	0.017
ECG_SDNN	0.017

Table B21: Generalized_Imb Model - RF Feature Importance, DECG

All	
Feature	**Values**
ECG_DC	0.050
ECG_Stress Index	0.047
ECG_AC	0.046
ECG_SDNN	0.046
ECG_AR_AbsolutePower_HF	0.046
ECG_SNS_Index	0.044
ECG_FFT_AbsolutePower_HF	0.044
ECG_Max HR	0.044
ECG_AR_AbsolutePower_LF	0.044
ECG_FFT_RelativePower_LF	0.044
Gender - Male	
Feature	**Values**

ECG_AR_HF	0.062
ECG_Mean RR	0.059
ECG_PNS Index	0.054
ECG_FFT_HF	0.050
ECG_Max HR	0.049
ECG_SD1SD2	0.047
ECG_RMSSD	0.047
ECG_DC	0.045
ECG_AC	0.045
ECG_SNS_Index	0.045
Gender - Female	
Feature	**Values**
ECG_Stress Index	0.050
ECG_AC	0.050
ECG_DC	0.049
ECG_SDNN	0.048
ECG_FFT_AbsolutePower_HF	0.047
ECG_AR_AbsolutePower_HF	0.047
ECG_SNS_Index	0.046
ECG_RMSSD	0.046
ECG_Max HR	0.044
ECG_SD HR	0.044
Income - Low	
Feature	**Values**
ECG_DC	0.048
ECG_SD HR	0.048
ECG_Stress Index	0.047
ECG_Mean RR	0.046
ECG_SNS_Index	0.046
ECG_AC	0.046
ECG_AR_AbsolutePower_HF	0.045
ECG_RMSSD	0.045
ECG_SDNN	0.045
ECG_AR_AbsolutePower_LF	0.044
Income Medium High	
Feature	**Values**
ECG_DC	0.055
ECG_AC	0.051
ECG_SD1SD2	0.050

ECG_AR_AbsolutePower_LF	0.047
ECG_FFT_LF	0.047
ECG_Stress Index	0.047
ECG_SDNN	0.047
ECG_SD HR	0.046
ECG_FFT_RelativePower_LF	0.046
ECG_Mean RR	0.046
Employment Students	
Feature	**Values**
ECG_Stress Index	0.055
ECG_RMSSD	0.054
ECG_AR_AbsolutePower_HF	0.054
ECG_SD HR	0.052
ECG_SDNN	0.049
ECG_FFT_RelativePower_LF	0.048
ECG_DC	0.047
ECG_SNS_Index	0.047
ECG_FFT_AbsolutePower_HF	0.047
ECG_PNS Index	0.047
Employment Workers	
Feature	**Values**
ECG_DC	0.046
ECG_SD1SD2	0.046
ECG_AR_AbsolutePower_LF	0.045
ECG_FFT_RelativePower_LF	0.045
ECG_AC	0.045
ECG_PNS Index	0.045
ECG_Mean RR	0.045
ECG_AR_LFHF	0.045
ECG_SDNN	0.045
ECG_Max HR	0.044
Age 18-24	
Feature	**Values**
ECG_FFT_HF	0.049
ECG_SD1SD2	0.049
ECG_FFT_RelativePower_LF	0.048
ECG_Mean RR	0.047
ECG_Max HR	0.047
ECG_AR_HF	0.046

ECG_FFT_LFHF	0.046
ECG_AR_AbsolutePower_LF	0.046
ECG_SD HR	0.045
ECG_RMSSD	0.045
Age 25-34	
Feature	**Values**
ECG_AR_RelativePower_LF	0.050
ECG_Mean RR	0.049
ECG_AR_LFHF	0.048
ECG_AR_AbsolutePower_HF	0.048
ECG_DC	0.048
ECG_SD HR	0.048
ECG_SNS_Index	0.048
ECG_AR_AbsolutePower_LF	0.046
ECG_FFT_LF	0.045
ECG_FFT_LFHF	0.045
Age 35-44	
Feature	**Values**
ECG_AR_HF	0.061
ECG_SD1SD2	0.059
ECG_FFT_HF	0.052
ECG_AR_RelativePower_LF	0.049
ECG_AR_LFHF	0.047
ECG_AR_AbsolutePower_HF	0.047
ECG_FFT_RelativePower_LF	0.047
ECG_SD HR	0.046
ECG_AC	0.046
ECG_Mean RR	0.046
Healthy	
Feature	**Values**
ECG_DC	0.048
ECG_AC	0.048
ECG_AR_AbsolutePower_HF	0.046
ECG_Max HR	0.046
ECG_FFT_LFHF	0.046
ECG_SD1SD2	0.046
ECG_SDNN	0.045
ECG_Stress Index	0.045
ECG_RMSSD	0.045

ECG_FFT_AbsolutePower_HF	0.044

Table B22: Generalized_Imb Model - RF Feature Importance, DA

All	
Feature	**Values**
AW Mean HR - Interval	0.037
ECG_DC	0.036
HRV-1	0.035
AW Mean Steps	0.035
ECG_SDNN	0.035
ECG_FFT_AbsolutePower_HF	0.034
ECG_Stress Index	0.034
ECG_SNS_Index	0.034
AW Min HR - Interval	0.033
ECG_AC	0.033
Gender - Male	
Feature	**Values**
AW Min HR - Interval	0.056
ECG_AR_HF	0.055
AW Mean HR - Interval	0.044
ECG_Mean RR	0.043
ECG_FFT_HF	0.040
ECG_PNS Index	0.040
ECG_AR_LF	0.038
ECG_SD1SD2	0.037
Short Term Min	0.037
ECG_RMSSD	0.037
Gender - Female	
Feature	**Values**
ECG_SDNN	0.040
ECG_Stress Index	0.039
ECG_DC	0.038
ECG_AC	0.038
ECG_FFT_AbsolutePower_HF	0.037
ECG_RMSSD	0.037
ECG_AR_AbsolutePower_HF	0.036
ECG_SNS_Index	0.036
ECG_SD HR	0.034

AW Mean Steps	0.034
Income - Low	
Feature	**Values**
AW Mean Steps	0.037
HRV-1	0.035
ECG_DC	0.034
ECG_Stress Index	0.034
ECG_SD HR	0.034
AW Min HR - Interval	0.034
ECG_RMSSD	0.033
ECG_AC	0.033
ECG_SDNN	0.033
ECG_AR_AbsolutePower_HF	0.033
Income Medium High	
Feature	**Values**
AW Mean HR - Interval	0.046
ECG_DC	0.045
ECG_SD1SD2	0.040
ECG_AC	0.038
ECG_SDNN	0.036
AW Min HR - Interval	0.035
ECG_Stress Index	0.035
Short Term Min	0.035
HRV-1	0.035
ECG_FFT_AbsolutePower_HF	0.034
Employment Students	
Feature	**Values**
ECG_Stress Index	0.042
ECG_RMSSD	0.041
ECG_SDNN	0.039
ECG_AR_AbsolutePower_HF	0.038
ECG_FFT_AbsolutePower_HF	0.038
AW Min HR - Interval	0.037
ECG_SD HR	0.037
ECG_DC	0.036
ECG_PNS Index	0.036
HRV-1	0.036
Employment Workers	
Feature	**Values**

AW Mean HR - Interval	0.041
HRV-1	0.036
AW Mean Steps	0.034
ECG_AR_HF	0.034
ECG_PNS Index	0.033
AW Min HR - Interval	0.033
AW Max Steps	0.033
ECG_SD1SD2	0.032
ECG_RMSSD	0.032
ECG_DC	0.032
Age 18-24	
Feature	**Values**
AW Mean Steps	0.045
ECG_FFT_HF	0.038
ECG_FFT_LFHF	0.035
ECG_SD1SD2	0.035
ECG_FFT_RelativePower_LF	0.035
ECG_AR_HF	0.034
ECG_Max HR	0.034
AW Max HR - Interval	0.034
ECG_Mean RR	0.034
AW Mean HR - Interval	0.033
Age 25-34	
Feature	**Values**
AW Mean HR - Interval	0.040
HRV-1	0.040
ECG_AR_LFHF	0.036
ECG_DC	0.036
AW Min HR - Interval	0.035
ECG_AR_AbsolutePower_HF	0.035
AW Mean Steps	0.034
ECG_AR_RelativePower_LF	0.034
ECG_SD HR	0.033
ECG_AR_AbsolutePower_LF	0.033
Age 35-44	
Feature	**Values**
ECG_AR_HF	0.047
ECG_SD1SD2	0.046
ECG_FFT_HF	0.039

ECG_AR_RelativePower_LF	0.038
HRV-1	0.037
ECG_AR_LFHF	0.036
AW Min HR - Interval	0.036
AW Max HR - Interval	0.036
AW Mean Steps	0.035
Short Term Min	0.035
Healthy	
Feature	**Values**
AW Mean HR - Interval	0.036
AW Mean Steps	0.036
ECG_AC	0.036
ECG_SDNN	0.036
HRV-1	0.036
ECG_DC	0.036
ECG_FFT_AbsolutePower_HF	0.034
ECG_SD1SD2	0.033
ECG_Stress Index	0.033
AW Min HR - Interval	0.033

Table B23: Generalized_Imb Model - RF Feature Importance, DAW

All	
Feature	**Values**
Weight	0.065
Temp	0.037
MAP	0.035
ECG_SDNN	0.034
ECG_Stress Index	0.033
ECG_DC	0.030
AW Mean HR - Interval	0.030
ECG_AR_AbsolutePower_HF	0.030
AW Min HR - Interval	0.029
ECG_AC	0.029
Gender - Male	
Feature	**Values**
Weight	0.076
AW Min HR - Interval	0.047
Temp	0.044

MAP	0.044
dia	0.038
ECG_AR_HF	0.037
sys	0.036
ECG_SD1SD2	0.035
AW Mean HR - Interval	0.035
ECG_Mean RR	0.032
Gender - Female	
Feature	**Values**
Weight	0.054
ECG_SDNN	0.036
ECG_Stress Index	0.035
ECG_FFT_AbsolutePower_HF	0.035
ECG_AC	0.034
ECG_DC	0.034
ECG_AR_AbsolutePower_HF	0.033
Temp	0.033
ECG_RMSSD	0.032
ECG_SNS_Index	0.031
Income - Low	
Feature	**Values**
Weight	0.049
Temp	0.037
MAP	0.034
ECG_FFT_AbsolutePower_HF	0.032
ECG_SD HR	0.032
AW Min HR - Interval	0.031
ECG_RMSSD	0.031
ECG_Stress Index	0.031
ECG_DC	0.030
ECG_AR_AbsolutePower_HF	0.030
Income Medium High	
Feature	**Values**
Weight	0.086
Temp	0.048
AW Mean HR - Interval	0.037
dia	0.036
ECG_DC	0.033
MAP	0.033

ECG_Stress Index	0.033
ECG_SD1SD2	0.031
ECG_AC	0.031
ECG_SDNN	0.030
Employment Students	
Feature	**Values**
Weight	0.085
MAP	0.039
ECG_Stress Index	0.034
Temp	0.033
ECG_RMSSD	0.032
ECG_AR_AbsolutePower_HF	0.032
sys	0.032
ECG_SDNN	0.032
ECG_FFT_AbsolutePower_HF	0.031
ECG_SD HR	0.031
Employment Workers	
Feature	**Values**
Weight	0.073
Temp	0.046
AW Mean HR - Interval	0.036
MAP	0.036
dia	0.032
sys	0.030
ECG_SD1SD2	0.030
HRV-1	0.029
AW Min HR - Interval	0.029
AW Mean Steps	0.027
Age 18-24	
Feature	**Values**
Weight	0.083
AW Mean Steps	0.039
Temp	0.033
ECG_FFT_LFHF	0.031
ECG_FFT_RelativePower_LF	0.031
ECG_FFT_HF	0.031
ECG_FFT_LF	0.030
MAP	0.030
ECG_AR_HF	0.029

ECG_RMSSD	0.029
Age 25-34	
Feature	**Values**
Weight	0.083
AW Mean Steps	0.039
Temp	0.033
ECG_FFT_LFHF	0.031
ECG_FFT_RelativePower_LF	0.031
ECG_FFT_HF	0.031
ECG_FFT_LF	0.030
MAP	0.030
ECG_AR_HF	0.029
ECG_RMSSD	0.029
Age 35-44	
Feature	**Values**
Weight	0.077
ECG_SD1SD2	0.044
dia	0.042
MAP	0.038
ECG_AR_RelativePower_LF	0.037
ECG_AR_HF	0.035
ECG_AR_LFHF	0.035
sys	0.035
Temp	0.033
HRV-1	0.032
Healthy	
Feature	**Values**
Weight	0.075
MAP	0.039
Temp	0.037
ECG_SDNN	0.032
HRV-1	0.030
ECG_AC	0.030
ECG_Stress Index	0.029
sys	0.029
ECG_DC	0.029
ECG_AR_HF	0.029

Table B24: Generalized_Imb Model - RF Feature Importance, DW

All	
Feature	**Values**
Weight	0.374
Temp	0.199
MAP	0.168
dia	0.130
sys	0.130
Gender - Male	
Feature	**Values**
Weight	0.335
Temp	0.207
MAP	0.170
dia	0.145
sys	0.144
Gender - Female	
Feature	**Values**
Weight	0.32
Temp	0.22
MAP	0.18
dia	0.14
sys	0.14
Income - Low	
Feature	**Values**
Weight	0.278
Temp	0.221
MAP	0.197
dia	0.155
sys	0.148
Income Medium High	
Feature	**Values**
Weight	0.353
Temp	0.213
MAP	0.157
dia	0.147
sys	0.131
Employment Students	
Feature	**Values**

Weight	0.369
Temp	0.188
MAP	0.185
dia	0.132
sys	0.127
Employment Workers	
Feature	**Values**
Weight	0.338
Temp	0.216
MAP	0.169
dia	0.140
sys	0.137
Age 18-24	
Feature	**Values**
Weight	0.329
Temp	0.225
MAP	0.170
dia	0.140
sys	0.136
Age 25-34	
Feature	**Values**
Weight	0.366
Temp	0.191
MAP	0.170
dia	0.137
sys	0.136
Age 35-44	
Feature	**Values**
Weight	0.294
Temp	0.200
MAP	0.183
dia	0.164
sys	0.160
Healthy	
Feature	**Values**
Weight	0.429
Temp	0.185
MAP	0.162

sys	0.128
dia	0.097

Table B25: Generalized_Imb Model - RF Feature Importance, DEmpatica

All	
Feature	**Values**
Empatica_resp	0.027
Empatica_MSE3	0.023
Empatica_FFT_HF	0.022
Empatica_MSE2	0.021
Empatica_FFT_AbsolutePower_HF	0.021
Empatica_MSE15	0.021
Empatica_DC	0.020
Empatica_ApEn	0.020
Empatica_MSE13	0.020
Empatica_FFT_RelativePower_VLF	0.020
Gender - Male	
Feature	**Values**
Empatica_MSE13	0.037
Empatica_MSE3	0.035
Empatica_resp	0.032
Empatica_MSE9	0.027
Empatica_FFT_LF	0.026
Empatica_MSE2	0.026
Empatica_MSE7	0.026
Empatica_PNS Index	0.024
Empatica_MSE15	0.024
Empatica_FFT_HF	0.024
Gender - Female	
Feature	**Values**
Empatica_NN50	0.025
Empatica_FFT_AbsolutePower_HF_log	0.022
Empatica_FFT_AbsolutePower_HF	0.022
Empatica_MSE18	0.022
Empatica_TINN	0.021
Empatica_SampEn	0.021
Empatica_MSE20	0.021

Empatica_DC	0.021
Empatica_FFT_HF	0.021
Empatica_SD1SD2	0.020
Income - Low	
Feature	**Values**
Empatica_DC	0.026
Empatica_MSE3	0.025
Empatica_MSE18	0.024
Empatica_MSE2	0.023
Empatica_MSE15	0.023
Empatica_resp	0.022
Empatica_FFT_AbsolutePower_HF	0.022
Empatica_SD HR	0.022
Empatica_MSE14	0.022
Empatica_DET	0.022
Income Medium High	
Feature	**Values**
Empatica_MSE3	0.0260
Empatica_DC	0.0257
Empatica_resp	0.0255
Empatica_FFT_HF	0.0246
Empatica_AR_RelativePower_VLF	0.0242
Empatica_Mean RR	0.0237
Empatica_FFT_RelativePower_VLF	0.0235
Empatica_FFT_AbsolutePower_HF_log	0.0229
Empatica_MSE2	0.0226
Empatica_FFT_AbsolutePower_HF	0.0220
Employment Students	
Feature	**Values**
Empatica_resp	0.028
Empatica_MSE3	0.028
Empatica_DC	0.027
Empatica_MSE2	0.027
Empatica_MSE12	0.025
Empatica_FFT_RelativePower_LF	0.024
Empatica_FFT_AbsolutePower_HF	0.024
Empatica_FFT_HF	0.023
Empatica_ApEn	0.023

Empatica_Shannon	0.023
Employment Workers	
Feature	**Values**
Empatica_D2	0.026
Empatica_MSE13	0.026
Empatica_MSE3	0.025
Empatica_NN50	0.025
Empatica_Min HR	0.024
Empatica_MSE15	0.024
Empatica_FFT_AbsolutePower_HF	0.024
Empatica_SNS_Index	0.023
Empatica_MSE9	0.023
Empatica_PNS Index	0.023
Age 18-24	
Feature	**Values**
Empatica_MSE4	0.028
Empatica_FFT_AbsolutePower_VLF_log	0.028
Empatica_AR_LF	0.026
Empatica_FFT_AbsolutePower_VLF	0.026
Empatica_MSE13	0.025
Empatica_Max line length	0.024
Empatica_DC	0.024
Empatica_MSE2	0.023
Empatica_SD1SD2	0.023
Empatica_ApEn	0.023
Age 25-34	
Feature	**Values**
Empatica_resp	0.027
Empatica_FFT_HF	0.024
Empatica_AR_RelativePower_VLF	0.023
Empatica_FFT_LFHF	0.022
Empatica_Shannon	0.022
Empatica_MSE6	0.022
Empatica_MSE14	0.022
Empatica_REC	0.022
Empatica_MSE19	0.022
Empatica_AR_LF	0.022
Age 35-44	

Feature	Values
Empatica_MSE3	0.036
Empatica_FFT_HF	0.034
Empatica_resp	0.029
Empatica_alpha2	0.028
Empatica_Mean RR	0.026
Empatica_Max line length	0.025
Empatica_MSE18	0.024
Empatica_MSE17	0.024
Empatica_SNS_Index	0.024
Empatica_DC	0.023
Healthy	
Feature	**Values**
Empatica_resp	0.034
Empatica_MSE3	0.031
Empatica_FFT_HF	0.029
Empatica_MSE2	0.024
Empatica_MSE15	0.021
Empatica_MSE13	0.021
Empatica_FFT_AbsolutePower_HF	0.020
Empatica_MSE9	0.020
Empatica_MSE8	0.020
Empatica_DC	0.020

Table B26: Generalized_Imb Model - RF Feature Importance, SDA

All	
Feature	**Values**
T+2 AW Number of Wake-Ups	0.043
T+2 AW Consolidated Time Awake During Sleep	0.039
T+1 AW Consolidated Time Awake During Sleep	0.038
AW Consolidated Time Awake During Sleep	0.028
T-1 AW Total Time in Bed	0.026
T+1 AW Number of Wake-Ups	0.026
T-2 AW Number of Wake-Ups	0.026
AW Number of Wake-Ups	0.025
T-2 AW Mean HR	0.025
AW Min HR	0.024
Gender - Male	
Feature	**Values**

AW Min HR	0.045
T-2 AW Total Time Asleep	0.033
AW Total Time in Bed	0.031
T+1 AW Consolidated Time Awake During Sleep	0.030
T-2 AW Min HR	0.028
T+2 AW Max HR	0.027
T-2 Total Time in Bed	0.026
ECG_AR_HF	0.025
T+1 AW Max HR	0.025
T+1 AW Mean HR	0.024
Gender - Female	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.052
T+2 AW Consolidated Time Awake During Sleep	0.036
T+2 AW Number of Wake-Ups	0.035
AW Total Time in Bed	0.028
T+1 AW Number of Wake-Ups	0.026
AW Consolidated Time Awake During Sleep	0.025
T+1 AW Mean HR	0.024
ECG_RMSSD	0.024
AW Min HR	0.023
T-1 AW Total Time in Bed	0.022
Income - Low	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.045
T+2 AW Consolidated Time Awake During Sleep	0.044
T+2 AW Number of Wake-Ups	0.036
T+1 AW Number of Wake-Ups	0.029
T+2 AW Total Time in Bed	0.028
AW Total Time in Bed	0.027
AW Consolidated Time Awake During Sleep	0.027
T+2 % of Time Asleep While In Bed	0.026
T-2 AW Number of Wake-Ups	0.024
T-1 AW Total Time in Bed	0.024
Income Medium High	
Feature	**Values**
AW Min HR	0.034
T-2 AW Mean HR	0.032

T+1 AW Consolidated Time Awake During Sleep	0.031
T-2 AW Min HR	0.030
T+2 AW Min HR	0.029
T+1 AW Min HR	0.029
Short Term Min	0.027
T+2 AW Mean HR	0.026
AW Min HR - Interval	0.026
T-1 AW Mean HR	0.026
Employment Students	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.058
T+2 AW Consolidated Time Awake During Sleep	0.053
T+2 AW Number of Wake-Ups	0.050
T+1 AW Number of Wake-Ups	0.037
AW Consolidated Time Awake During Sleep	0.028
AW Number of Wake-Ups	0.025
T-2 AW Number of Wake-Ups	0.024
T+2 AW Min HR	0.023
T-2 AW Min HR	0.023
ECG_Stress Index	0.022
Employment Workers	
Feature	**Values**
AW Min HR	0.031
AW Total Time in Bed	0.028
AW Min HR - Interval	0.025
AW Mean HR - Interval	0.024
T+1 AW Max HR	0.024
T-2 Consolidated Time Awake During Sleep	0.024
T-2 AW Mean HR	0.024
T+2 AW Max HR	0.024
T+1 AW Consolidated Time Awake During Sleep	0.024
T+2 AW Total Time in Bed	0.023
Age 18-24	
Feature	**Values**
T+2 AW Consolidated Time Awake During Sleep	0.035
T+2 AW % of Time Asleep While In Bed	0.034
T+1 AW Consolidated Time Awake During Sleep	0.031
T+2 AW Number of Wake-Ups	0.027

AW Total Time in Bed	0.027
T-2 AW Min HR	0.025
T-2 AW Max HR	0.024
T+2 AW Total Time in Bed	0.024
T+2 AW Min HR	0.024
ECG_FFT_RelativePower_LF	0.024
Age 25-34	
Feature	**Values**
T-2 AW Total Time in Bed	0.035
T+1 AW Consolidated Time Awake During Sleep	0.033
T-2 AW Total Time Asleep	0.031
T+2 AW Number of Wake-Ups	0.031
AW Total Time in Bed	0.031
T+2 AW Consolidated Time Awake During Sleep	0.030
T+2 AW Total Time in Bed	0.028
AW Consolidated Time Awake During Sleep	0.025
T-1 Total Time in Bed	0.024
HRV-1	0.023
Age 35-44	
Feature	**Values**
AW Min HR	0.034
T-2 % of Time Asleep While In Bed	0.028
T+1 AW Mean HR	0.027
T+2 AW Mean HR	0.026
Short Term Min	0.025
AW Min HR.1	0.024
T+2 AW Min HR	0.023
T-2 AW Total Time in Bed	0.023
T-2 AW Min HR	0.023
T-1 AW Min HR	0.023
Healthy	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.029
T-2 AW Max HR	0.027
T+2 AW Number of Wake-Ups	0.026
T+2 AW Consolidated Time Awake During Sleep	0.026
T-2 AW Min HR	0.026
T-2 AW Total Time Asleep	0.024
AW Number of Wake-Ups	0.024

T-1 AW Max HR	0.024
ECG_AR_HF	0.024
T+1 AW Max HR	0.024

Table B27: Generalized_Imb Model - RF Feature Importance, SDAW

All	
Feature	**Values**
T+2 AW Number of Wake-Ups	0.037
T+2 AW Consolidated Time Awake During Sleep	0.031
T+1 AW Consolidated Time Awake During Sleep	0.029
AW Total Time in Bed	0.026
Weight	0.025
AW Min HR	0.025
T+2 AW Min HR	0.024
T+2 Time Spent in REM Stage	0.024
T+1 AW Number of Wake-Ups	0.024
T-2 AW Min HR	0.023
Gender - Male	
Feature	**Values**
T+1 AW Max HR	0.025
Weight	0.024
T-2 AW Number of Wake-Ups	0.024
MAP	0.022
T+2 AW Consolidated Time Awake During Sleep	0.022
sys	0.021
Temp	0.020
T-2 AW Total Time Asleep	0.020
dia	0.019
ECG_FFT_RelativePower_HF	0.019
Gender - Female	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.033
AW Total Time in Bed	0.029
T+2 AW Number of Wake-Ups	0.028
T+2 AW Consolidated Time Awake During Sleep	0.027
T+1 AW Number of Wake-Ups	0.021
AW Min HR	0.020

Weight	0.020
T-2 Time Spent in Deep Stage	0.020
ECG_Stress Index	0.019
T-2 AW Min HR	0.019
Income - Low	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.042
T+2 AW Consolidated Time Awake During Sleep	0.037
T+2 AW Number of Wake-Ups	0.031
T+1 AW Number of Wake-Ups	0.028
T+2 Time Spent in REM Stage	0.027
T+2 % of Time Asleep While In Bed	0.021
T-2 AW Number of Wake-Ups	0.020
AW Consolidated Time Awake During Sleep	0.020
AW Total Time in Bed	0.019
T-2 AW Consolidated Time Awake During Sleep	0.019
Income Medium High	
Feature	**Values**
AW Min HR	0.037
T+2 AW Min HR	0.027
AW Total Time in Bed	0.025
T+1 AW Min HR	0.025
T+2 AW Number of Wake-Ups	0.024
T-2 AW Min HR	0.024
T-2 AW Mean HR	0.024
T+1 AW Consolidated Time Awake During Sleep	0.023
Temp	0.023
T+2 AW Consolidated Time Awake During Sleep	0.023
Employment Students	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.039
T+2 Time Spent in REM Stage	0.036
T+2 AW Consolidated Time Awake During Sleep	0.036
T+2 AW Number of Wake-Ups	0.033
T+2 AW Min HR	0.030
Time Spent in REM Stage	0.029
T+1 AW Number of Wake-Ups	0.026
T+1 AW Min HR	0.025

T-2 Time Spent in REM Stage	0.025
T-2 AW Min HR	0.025
Employment Workers	
Feature	**Values**
AW Min HR	0.027
T-2 AW Min HR	0.024
AW Total Time in Bed	0.024
T+2 AW Max HR	0.023
T+1 AW Max HR	0.023
Time Spent in Deep Stage	0.022
Weight	0.021
T+2 AW Min HR	0.020
AW Mean HR - Interval	0.018
Time Spent in REM Stage	0.018
Age 18-24	
Feature	**Values**
T+2 % of Time Asleep While In Bed	0.034
T+2 Time Spent in REM Stage	0.033
Weight	0.030
T-2 AW Min HR	0.029
AW Total Time in Bed	0.026
T-2 Time Spent in REM Stage	0.025
T-1 AW Mean HR	0.025
Time Spent in REM Stage	0.025
T+1 AW Min HR	0.023
T+2 AW Consolidated Time Awake During Sleep	0.023
Age 25-34	
Feature	**Values**
T+2 AW Number of Wake-Ups	0.038
Weight	0.031
T+2 AW Consolidated Time Awake During Sleep	0.030
T+2 AW Mean HR	0.029
T+1 AW Consolidated Time Awake During Sleep	0.027
T+2 AW Min HR	0.026
T-2 AW Mean HR	0.024
T+1 AW Min HR	0.022
T-2 Time Spent in Deep Stage	0.022
T-2 AW Number of Wake-Ups	0.021
Age 35-44	

Feature	Values
T+2 Time Spent in Light Stage	0.024
T+1 AW Mean HR	0.023
T-2 % of Time Asleep While In Bed	0.023
T-2 Time Spent in Deep Stage	0.022
T+1 Withings Total Time Asleep	0.022
T-2 Time Spent in REM Stage	0.021
T+1 AW Min HR	0.021
Withings Total Time Asleep	0.021
AW Min HR	0.020
T-2 AW Total Time Asleep	0.020
Healthy	
Feature	**Values**
Weight	0.026
AW Total Time in Bed	0.026
T-2 AW Min HR	0.023
T+2 AW Number of Wake-Ups	0.022
T+2 Time Spent in REM Stage	0.022
T-2 Time Spent in REM Stage	0.021
T-2 AW Total Time in Bed	0.020
T+2 AW Consolidated Time Awake During Sleep	0.020
T+1 AW Consolidated Time Awake During Sleep	0.020
AW Min HR	0.019

Table B28: Generalized_Imb Model - RF Feature Importance, SDW

All	
Feature	**Values**
Weight	0.089
Temp	0.059
dia	0.058
sys	0.058
MAP	0.056
T-2 Time Spent in REM Stage	0.053
T+2 Time Spent in REM Stage	0.048
Withings Total Time Asleep	0.046
T+2 Withings Total Time Asleep	0.044
T+1 Withings Total Time Asleep	0.041
Gender - Male	

Feature	**Values**
T-2 Withings Total Time Asleep	0.075
T+2 Withings Total Time Asleep	0.071
T-1 Withings Total Time Asleep	0.069
Temp	0.065
T-2 Time Spent in Light Stage	0.064
MAP	0.058
Weight	0.055
Withings Total Time Asleep	0.054
sys	0.052
T+2 Total Time In Bed	0.050
Gender - Female	
Feature	**Values**
Weight	0.077
Temp	0.051
MAP	0.049
sys	0.048
T-2 Time Spent in REM Stage	0.047
T+1 Withings Total Time Asleep	0.045
T+2 Time Spent in Deep Stage	0.042
T+2 Time Spent in REM Stage	0.042
Withings Total Time Asleep	0.042
T+2 % of Time Asleep While In Bed	0.041
Income - Low	
Feature	**Values**
Weight	0.059
T+2 Time Spent in REM Stage	0.055
Total Time In Bed	0.052
T-2 Time Spent in REM Stage	0.051
Time Spent in Light Stage	0.048
Time Spent in REM Stage	0.047
Temp	0.046
Withings Total Time Asleep	0.044
% of Time Asleep While In Bed	0.044
MAP	0.043
Income Medium High	
Feature	**Values**
Weight	0.075

Temp	0.066
MAP	0.054
sys	0.054
dia	0.053
T-2 Time Spent in REM Stage	0.047
T+2 Withings Total Time Asleep	0.045
T+1 Withings Total Time Asleep	0.045
Time Spent in REM Stage	0.045
% of Time Asleep While In Bed	0.043
Employment Students	
Feature	**Values**
Time Spent in REM Stage	0.072
Weight	0.072
T+2 Time Spent in REM Stage	0.071
T-2 Time Spent in REM Stage	0.061
Time Spent in Light Stage	0.045
T+2 % of Time Asleep While In Bed	0.044
MAP	0.042
% of Time Asleep While In Bed	0.040
T-2 Time Spent in Light Stage	0.039
T-2 Time Spent in Deep Stage	0.038
Employment Workers	
Feature	**Values**
Temp	0.063
Weight	0.061
dia	0.058
sys	0.055
MAP	0.054
T+2 Withings Total Time Asleep	0.051
Time Spent in Deep Stage	0.048
T+2 Total Time In Bed	0.046
T+1 Withings Total Time Asleep	0.045
T-1 Withings Total Time Asleep	0.045
Age 18-24	
Feature	**Values**
Weight	0.097
Time Spent in REM Stage	0.066
T+2 Time Spent in REM Stage	0.066

Temp	0.049
T-2 Time Spent in REM Stage	0.048
% of Time Asleep While In Bed	0.041
MAP	0.041
sys	0.041
dia	0.040
T+2 Withings Total Time Asleep	0.040
Age 25-34	
Feature	**Values**
Weight	0.106
% of Time Asleep While In Bed	0.060
T-2 % of Time Asleep While In Bed	0.057
Temp	0.049
T+2 % of Time Asleep While In Bed	0.048
T-2 Time Spent in Deep Stage	0.046
MAP	0.045
Time Spent in REM Stage	0.044
dia	0.041
sys	0.040
Age 35-44	
Feature	**Values**
T+2 Total Time In Bed	0.072
T+2 Withings Total Time Asleep	0.063
T-1 Withings Total Time Asleep	0.063
T+2 Time Spent in Light Stage	0.055
Weight	0.054
T-2 Time Spent in REM Stage	0.053
Temp	0.052
T-2 Withings Total Time Asleep	0.049
Total Time In Bed	0.049
MAP	0.049
Healthy	
Feature	**Values**
Weight	0.076
T-2 Time Spent in REM Stage	0.060
T+2 Withings Total Time Asleep	0.054
T+1 Withings Total Time Asleep	0.047
Temp	0.047
Time Spent in Light Stage	0.047

Withings Total Time Asleep	0.046
T+2 Time Spent in REM Stage	0.044
Time Spent in REM Stage	0.042
Time Spent in Deep Stage	0.042

Table B29: Generalized_Imb Model - RF Feature Importance, SDS

All	
Feature	**Values**
T+2 AW Number of Wake-Ups	0.042
T+1 AW Consolidated Time Awake During Sleep	0.041
T+2 AW Consolidated Time Awake During Sleep	0.039
AW Total Time in Bed	0.039
T-2 AW Min HR	0.034
T+2 AW Min HR	0.031
T-1 AW Min HR	0.031
T+1 AW Number of Wake-Ups	0.030
AW Min HR	0.030
T-2 Time Spent in REM Stage	0.030
Gender - Male	
Feature	**Values**
T+2 AW Consolidated Time Awake During Sleep	0.048
T+1 AW Max HR	0.044
T-2 AW Total Time in Bed	0.038
T-1 AW Max HR	0.035
T+2 Time Spent in Deep Stage	0.035
T-2 AW Max HR	0.034
T-1 AW Mean HR	0.033
% of Time Asleep While In Bed	0.033
AW Mean HR	0.032
T-2 Time Spent in REM Stage	0.032
Gender - Female	
Feature	**Values**
AW Total Time in Bed	0.045
T+1 AW Consolidated Time Awake During Sleep	0.044
T+2 AW Consolidated Time Awake During Sleep	0.032
T+2 AW Number of Wake-Ups	0.032
T+2 Time Spent in Deep Stage	0.031
AW Min HR	0.028

T-2 Time Spent in Deep Stage	0.027
T+2 Time Spent in REM Stage	0.027
T+2 AW Total Time Asleep	0.027
AW Consolidated Time Awake During Sleep	0.026
Income - Low	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.051
T+2 AW Number of Wake-Ups	0.042
T+2 AW Consolidated Time Awake During Sleep	0.037
T+1 AW Number of Wake-Ups	0.032
Time Spent in Light Stage	0.032
T-2 W Consolidated Time Awake During Sleep	0.031
Time Spent in REM Stage	0.030
T+2 AW % of Time Asleep While In Bed	0.030
T+2 Time Spent in REM Stage	0.029
T-2 Time Spent in Light Stage	0.029
Income Medium High	
Feature	**Values**
AW Min HR	0.048
T+2 AW Min HR	0.039
AW Total Time in Bed	0.038
T-2 AW Min HR	0.036
T+1 AW Min HR	0.035
T+1 AW Consolidated Time Awake During Sleep	0.035
T+2 AW Number of Wake-Ups	0.035
AW Total Time Asleep	0.034
T-1 AW Total Time in Bed	0.034
T+2 AW Consolidated Time Awake During Sleep	0.033
Employment Students	
Feature	**Values**
T+1 AW Consolidated Time Awake During Sleep	0.058
Time Spent in REM Stage	0.042
T+2 Time Spent in REM Stage	0.041
T+2 AW Min HR	0.037
T-2 Time Spent in REM Stage	0.037
T-2 AW Min HR	0.036
T+2 AW Number of Wake-Ups	0.036
Time Spent in Light Stage	0.035

T+1 AW Min HR	0.034
T+1 AW Number of Wake-Ups	0.034
Employment Workers	
Feature	**Values**
AW Total Time in Bed	0.040
AW Min HR	0.034
T-2 AW Min HR	0.033
Time Spent in Deep Stage	0.032
T+2 AW Max HR	0.029
Time Spent in REM Stage	0.029
T-2 AW % of Time Asleep While In Bed	0.028
T-2 AW Total Time Asleep	0.028
T+1 Withings Total Time Asleep	0.028
T+2 AW Min HR	0.028
Age 18-24	
Feature	**Values**
T-1 AW Mean HR	0.046
T+2 AW % of Time Asleep While In Bed	0.044
Time Spent in REM Stage	0.043
T-2 AW Min HR	0.039
T+1 AW Consolidated Time Awake During Sleep	0.038
T-2 Time Spent in REM Stage	0.038
T-1 AW Min HR	0.037
T+2 Time Spent in REM Stage	0.036
AW Total Time in Bed	0.035
T-2 AW Mean HR	0.034
Age 25-34	
Feature	**Values**
T+2 AW Number of Wake-Ups	0.055
T+2 AW Mean HR	0.043
T+2 AW Consolidated Time Awake During Sleep	0.043
T+1 AW Consolidated Time Awake During Sleep	0.036
T+2 AW Min HR	0.034
T-2 AW Number of Wake-Ups	0.033
T+2 AW Max HR	0.033
T-2 % of Time Asleep While In Bed	0.033
T-2 Time Spent in Deep Stage	0.031
T+1 AW Min HR	0.030
Age 35-44	

Feature	Values
AW Mean HR	0.040
T-2 AW % of Time Asleep While In Bed	0.038
T+1 AW Mean HR	0.037
T+2 Withings Total Time Asleep	0.036
T+2 Time Spent in Light Stage	0.035
T+1 Withings Total Time Asleep	0.035
T-1 Withings Total Time Asleep	0.035
T-2 AW Total Time Asleep	0.035
T-2 Time Spent in REM Stage	0.034
T+1 AW Min HR	0.033
Healthy	
Feature	**Values**
T-2 AW Total Time in Bed	0.045
T+2 AW Consolidated Time Awake During Sleep	0.044
T+1 AW Consolidated Time Awake During Sleep	0.043
T-1 AW Total Time in Bed	0.038
T-2 % of Time Asleep While In Bed	0.036
T+2 AW Number of Wake-Ups	0.035
AW Total Time in Bed	0.035
T-2 AW Mean HR	0.032
T-2 AW Number of Wake-Ups	0.032
T-2 AW % of Time Asleep While In Bed	0.032

Appendix C – User Manual

Get Started

Welcome and thank you for participating in our study and contributing to our research on the use of wearable devices to improve population health and stress detection. We ask that you please take a moment to read these instructions to ensure you know how to properly install the devices and apps.

Study Package

You should have received a package with the devices below and 3 documents: **User Manual – Get Started, User Manual – Data Collection Schedule,** and **User Manual – Data Collection Protocol.**

The contents of the documents are as follows:

User Manual – Get Started: Please read this document first in order to properly set up the devices.

User Manual – Data Collection Schedule: Please read this document to understand the schedule for data collection.

User Manual – Data Collection Protocol: Please read this document to understand the data collection protocol for each device.

Devices of the Study

For this study, you should have received:

1. **iPhone** (iOS 14.1 or higher) and **charging cable**

2. **Apple Watch** (Series 4 or higher, watchOS 5.1 or higher), **small size band** and **charging cable**

3. **Withings Sleep** and **charging cable**

4. **Withings BPM Connect** and **charging cable**

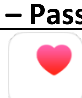
5. **Withings Wireless Scale**

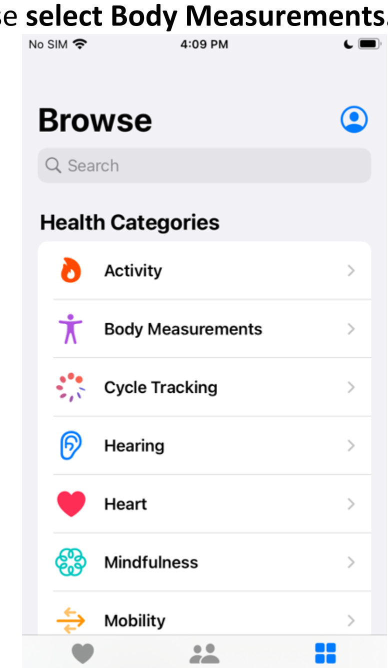
6. **Withings Termos**

7. **Empatica E4** and **charging cable**

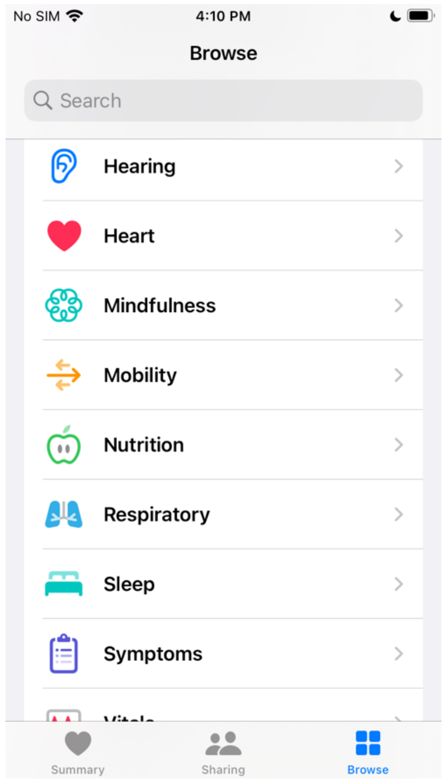
<u>Setting up the devices</u>

1. **<u>iPhone – Get Started – Password of iPhone: 000000</u>**

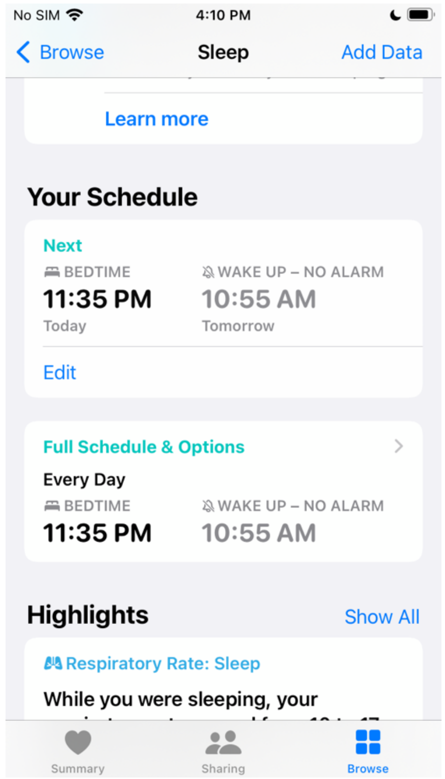
- Open the Health app .
- On the Browse tap, please **select Body Measurements.**



- Click on **Height.**
- On the top right corner, select **Add Data.**
- Please include your Height.
- On the Browse tap, please **select Sleep.**



- In **Your Schedule,** click **Edit** and insert your sleep schedule. Please include the most approximate estimate of the times you generally go to sleep and wake up.



2. Apple Watch – Get Started

Placement The Apple Watch band should fit closely but comfortably on the top of your wrist. Please adjust the band accordingly, not too tight or too loose and with room for the skin to breathe. You may tighten Apple Watch for workouts if necessary, and loosen the band when the workout is done. Please use the Apple Watch in your dominant wrist and adjust it so that the Digital Crown on the side of the Watch is nearest to the top of your wrist.

Too loose

If your Apple Watch doesn’t stay in place, or the sensors aren’t reading your heart rate, tighten the band a bit.

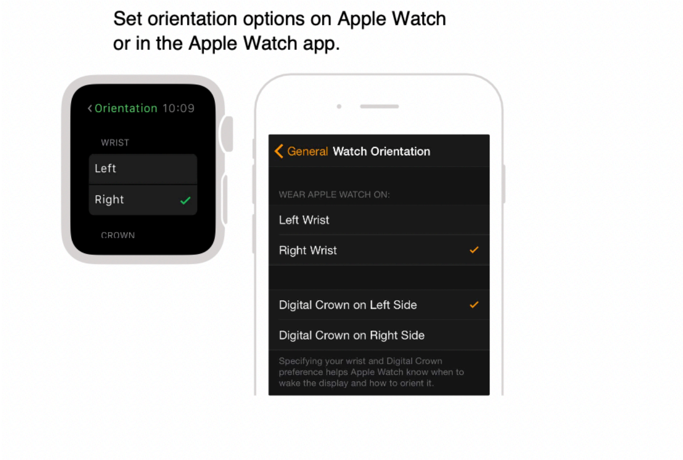
Just right

Your Apple Watch should be snug but comfortable.

In case band is too big The package also contains a smaller size band for the bottom of the Apple Watch. To switch bands, please hold down the band release button and slide the band across to remove it.

Adapted from: https://support.apple.com/en-us/HT204818

Select Crown Orientation and Dominant Hand The Digital Crown should be nearest to the top of your wrist. If necessary**,** please adjust your orientation as follows: open the Settings app, then go to General > Orientation. To change the settings in the Apple Watch app on iPhone, tap My Watch, then go to General > Watch Orientation.



Charging: Please follow the instructions below to charge your Apple Watch:

1.Plug in the charging cable into the USB port on your computer or the USB wall charger provided.
2.Place the concave end of the charging cable on the back of your Apple Watch. The concave end of the charging cable magnetically snaps to the back of your Apple Watch and aligns it properly.
3.You will hear a chime when charging begins (unless your Apple Watch is in silent mode) and see a green charging symbol on the watch face.
4.Charging fully takes about two to three hours. While the watch charges, you can tap it to check the battery level. A fully charged watch shows the green charging symbol encircled by a green circle on the watch face.

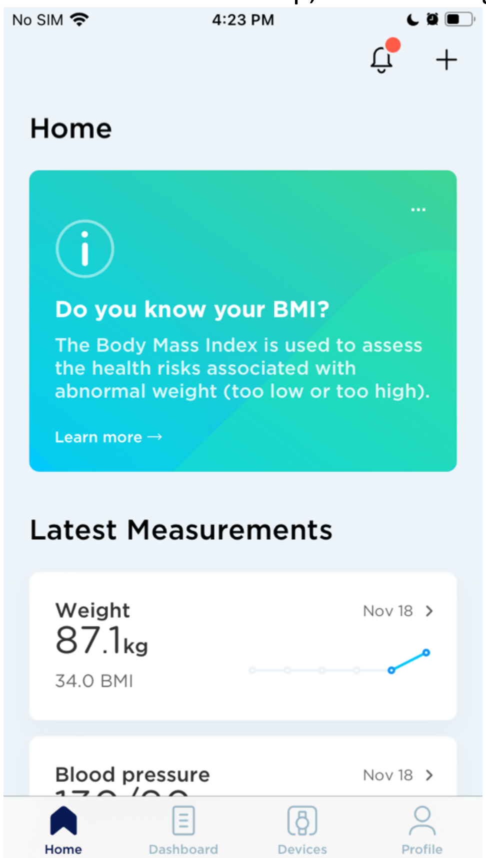
Your fully charged Apple Watch has battery life of up to 18 hours.

Adapted from: https://support.apple.com/en-ca/guide/watch/welcome/watchos

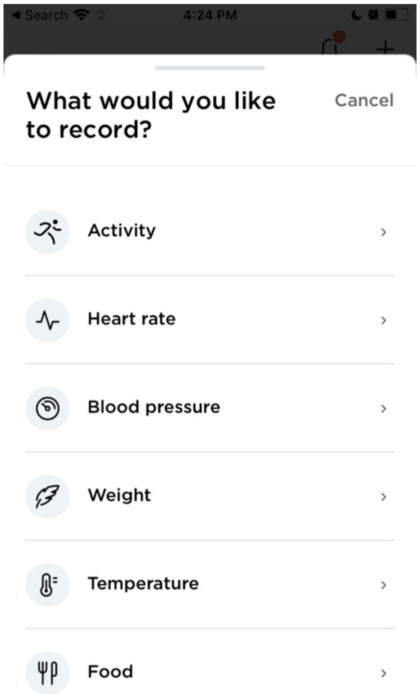
3. Health Mate – Get Started

1.Open the Health Mate app.

2. On the tab at the top, click the **+** sign at the top.

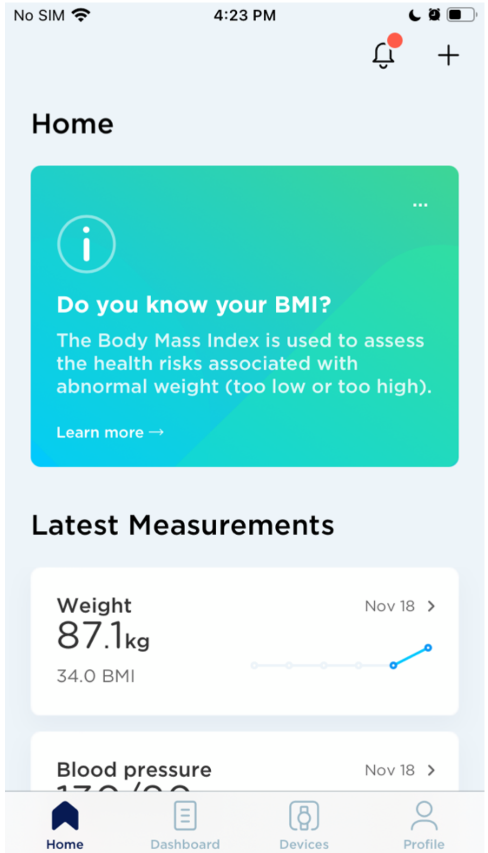


3. Select **Weight** and include an estimate of weight. This information is necessary for the scale to identify you as the user.

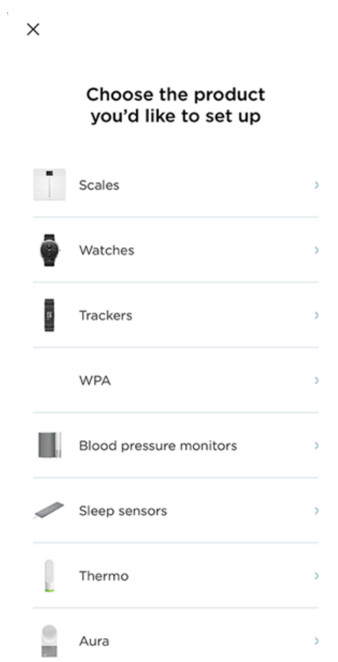


4. <u>Withings Sleep – Get Started</u>

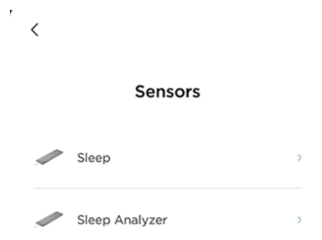
1. Open the Health Mate app.
2. Tap the Devices tab, scroll to the bottom and select Install a Device.



3. Tap Sleep Sensors

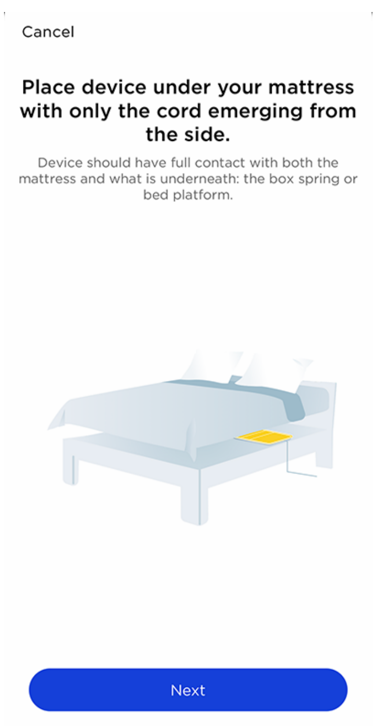


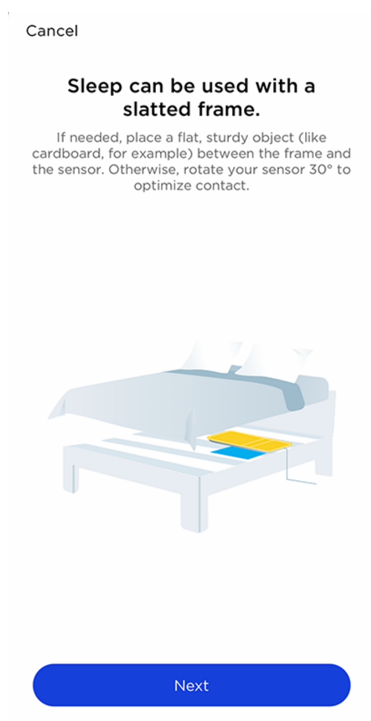
4. Tap Sleep

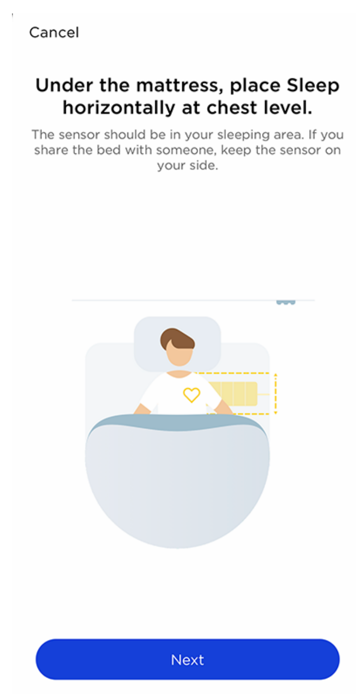


5. Tap Install

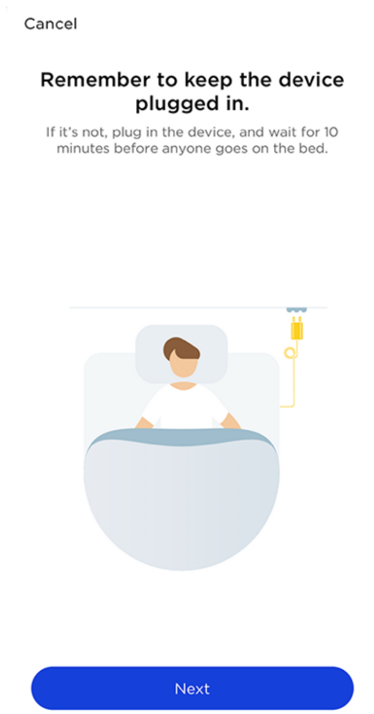
6. Place Sleep entirely under your mattress according to the provided instructions. You can also place it between the mattress and mattress topper/pad.







7. Plug Sleep using the provided adapter. Please keep the device plugged in throughout the duration of the study.



8. Pair the sensor to the iPhone by tapping Pair as requested by the app.

9. Tap the Wi-Fi network you want to use or tap Choose a different network. Please select the network that is most stable and with better connectivity.

10. Tap Next for the calibration of your Sleep to start. This step can last up to 10 minutes during which a buzzing sound can be heard. Please do not sit on the bed during the process. You will receive a notification in the Timeline of the Health Mate app once the calibration process is over.

11. If the calibration process went well, you can start using the Sleep Sensor! Thank you for following the instructions.
12. Please return the device inflated to the researcher at the end of the study.

Adapted from: https://support.withings.com/hc/en-us/articles/360020911714-Sleep-Sleep-Analyzer-User-Guide

5. **Withings BPM Connect – Get Started**

Charging BPM Connect lasts about 6 months per charge, so you shouldn't need to charge it during the study. However, if the battery of the device is low, you can charge your BPM Connect using the charging cable. To do so, connect the USB end of the charging cable to a power source. Please charge it for approximately 3 hours.

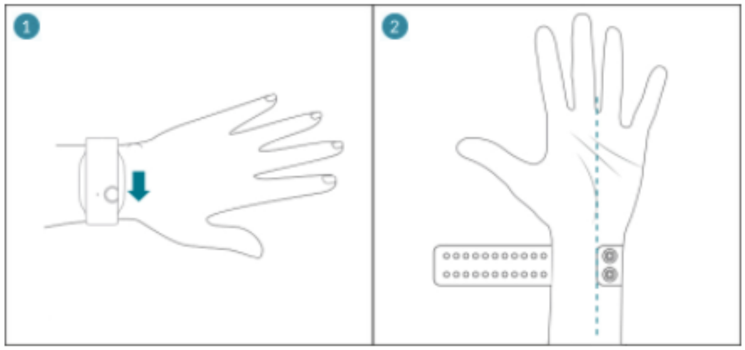
Adapted from: https://support.withings.com/hc/en-us/articles/360024569473-BPM-Connect-Charging-the-device

6. Empatica E4 – Get Started

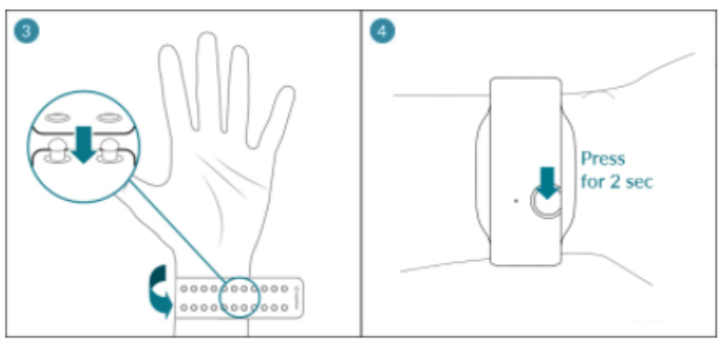
The Empatica E4 comes with pre-installed silver-plated electrodes (1), a USB dock placed under the device (2), and a USB MICRO-B Cable (4).

The E4 is easy to wear and adjust. To ensure proper fit and quality data, please follow the steps below:

1. Slide the loop towards the case and place the E4 wristband top-down on a surface.
2. Wear the E4 wristband on the non-dominant hand with the case on the top of the wrist. The EDA electrodes (under the snap-fastener) should line up on the bottom of the wrist. Line them up under the middle and ring fingers.



3. Wrap the band over snaps and tighten. To secure, connect one snap at a time. If too tight, loosen by one snap. Tighten the E4 wristband band enough to ensure the EDA electrodes do not change position on the skin during normal movement but not so much as to constrict blood flow or cause discomfort. Adjust the band by sliding up the wrist towards the elbow until it is snug. Reposition the band if it becomes loose during use.
4. The E4 wristband should fit snugly above the wrist joint. When the E4 wristband is properly secured, you should not be able to see any light escaping from the PPG sensor on the back of the wrist under the E4 wristband without lifting the housing from the wrist.
5. Press the button for 2 seconds to power on the E4 wristband.

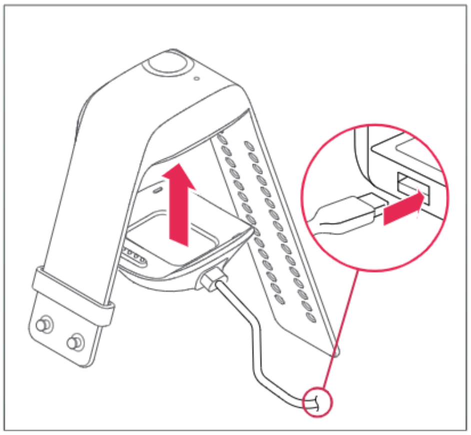


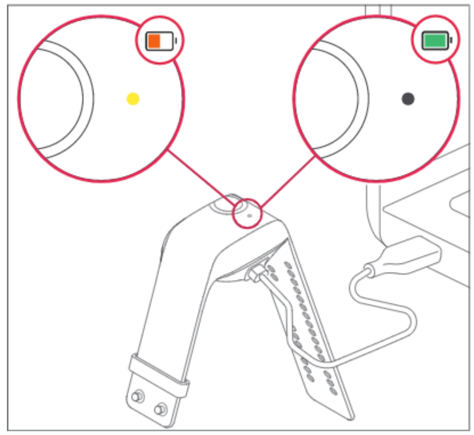
Adapted from: https://www.empatica.com/get-started-e4

Charging In order to charge the device please follow the instructions below.

- Snap the E4 into the dock and affix the dock via USB to a power source.
- The LED will turn YELLOW indicating it has received power and is charging.
- When E4 is fully charged, the LED will turn a solid green.

The charging dock is only used while charging the E4; **remove it before wearing the device**. It is a passive component. Charging typically takes between 1 and 2 hours.



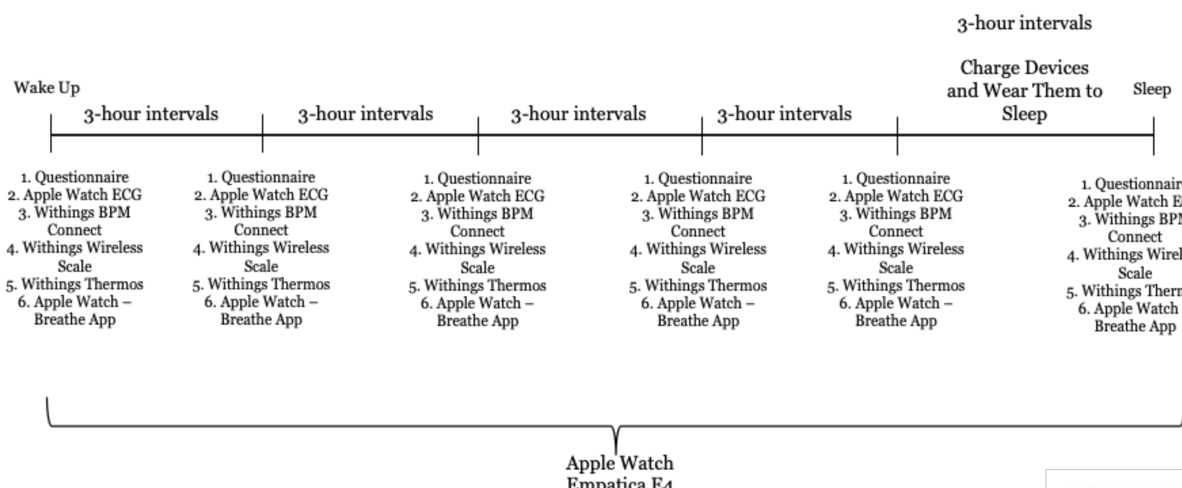


Data Collection Schedule

Welcome and thank you for participating in our study and contributing to our research on the use of wearable devices to improve population health and stress detection. If you haven't done so, we ask that you first read the **User Manual – Get Started** document.

If you already read the previous document, we ask that you please take a moment to read the following instructions to ensure you properly follow the schedule for data collection.

Please take a moment to look at the schedule below.



Throughout the study, we kindly ask that you **wear the Empatica E4 and the Apple Watch throughout the day and night**, taking the devices at the end of day for charging as detailed in the diagram and **wearing them again to bed.** Further, we ask that you leave the **Withings Sleep device plugged** for the duration of the study.

Obs: if the Empatica device is too far away from the phone, the session disconnects. We ask that you please check throughout the day in the Empatica E4 Realtime App to ensure data collection is still ongoing. If not, please start recording a new session as soon as possible.

We also ask that you perform the data collection protocols, detailed in the document **User Manual – Data Collection Protocol**, 6 times during the day. Please try taking the measurements approximately every 3 hours. Taking the measurement before or after the 3 hours shouldn't be a problem – the important is to take the measurements 6 times during the day according to intervals as regular as possible. If you miss one data collection, please take the readings as soon as possible.

The order for data collection is as follows:

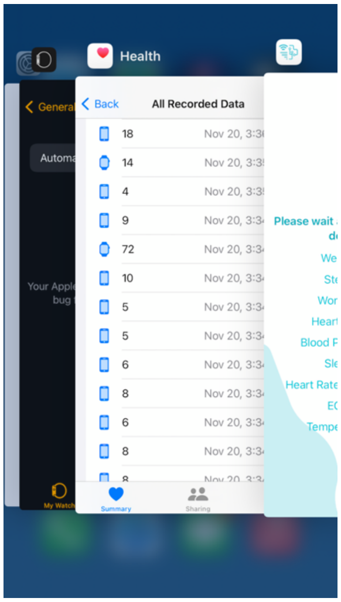
1. Fill the **Stress Questionnaire**.
2. Take an **Apple Watch ECG** Reading.
3. Take a Blood Pressure reading with the **Withings BPM Connect**.
4. Take a Weight reading with the **Withings Wireless Scale** (obs: For the scale, please take the readings every moment that you are at home; it is not expected that you will carry the scale with you throughout the day).
5. Take a Temperature reading with the **Withings Thermos.**
6. Take an Apple Watch Mindfulness app 5-minute reading.

Items 2, 3,4 and 5 can be done in any order. However, please fill out the **Stress Questionnaire** at the beginning of the cycle and use the **Apple Watch Breathe App** as the last measure in the cycle.

After data collection:

Once you are done with all the readings, please fully close the Apple Health and Health Mate app, open it again and wait until it syncs the new data.

Obs: To fully close an app in the iPhone, double click on the Home Button (the circle beneath the screen). You will see all open apps. Swipe up to close them.



<u>Data Collection Protocol</u>

Welcome and thank you for participating in our study and contributing to our research on the use of wearable devices to improve population health and stress detection. If you haven't done so, we ask that you first read the **User Manual – Get Started** document.

If you already read the previous document, we ask that you please take a moment to read the following instructions to ensure you know how to properly use the devices and apps and collect data for the study.

Before taking the readings, please ensure that you are wearing your Apple Watch device on your dominant hand and the Empatica E4 on your non-dominant hand throughout the day.

1. <u>Empatica E4</u>

At the start of the day, please wear your Empatica E4 as described in the **User Manual – Get Started Document.** In order to begin taking the reading:

- Launch the E4 realtime App
- Power ON the E4 - when the device is powered off a 2-sec button press will power it on. The LED indicator will blink light blue.

- Start streaming - tap “CONNECT E4 AND START STREAMING” and select your E4 from the list. The LED light turns a steady blue to indicate that streaming has started. In a few second the real-time streaming starts.

- **For the Empatica E4 device, you only need to start recording once throughout the day.**

- End session - press the "STOP RECORDING" button on the home screen. The E4 powers off and the session uploads automatically to Empatica secure cloud storage. The E4 will also power off if the Bluetooth connection is lost. e.g. is out-of-range. Please only end the session at the end of day or and charge the device.

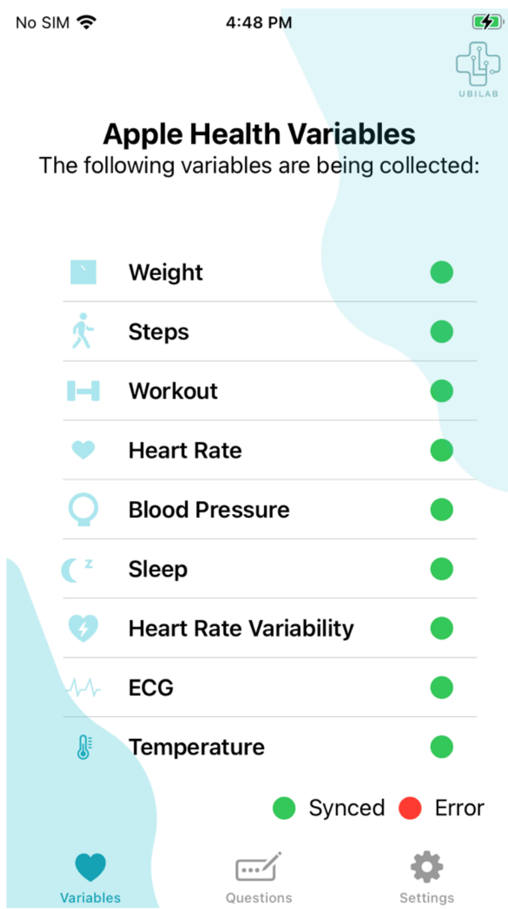
- **In occasions where the device should be removed (e.g., shower), please end the session and start it again once you wear the device.**

2. Stress Questionnaire

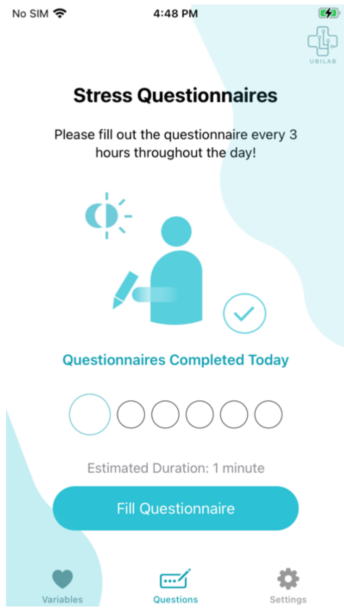
Important: <u>please ensure that the phone is connected to Wi-Fi before taking the stress questionnaire.</u>

1. Please access the MHP app.

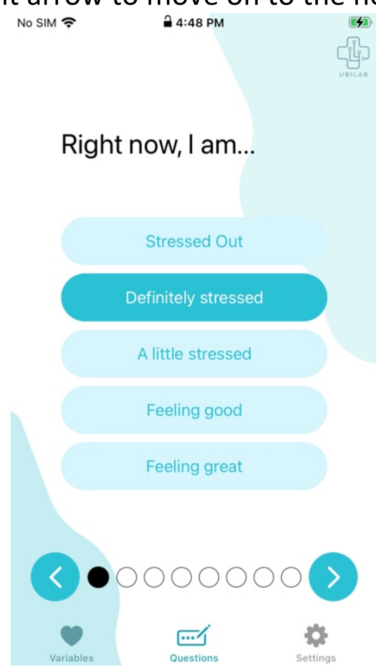
2. Please click on the tab **Questions (the second tab in the bottom)**



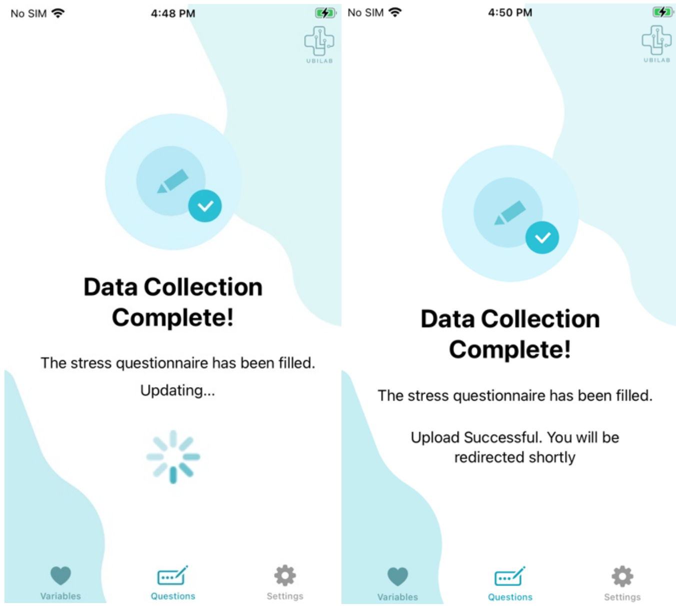
3. Please click on the **Fill Questionnaire** button



4. Please complete the form to the best of your abilities. After you respond to a question, click rhe right arrow to move on to the next one.



5. Once all questions are responded, the screen below will appear. Please wait until the **Updating**…message disappears. Please wait on the tab for a few seconds until you are redirected to the first questionnaire screen.



3. Apple Watch - ECG

Preparing for taking the reading:

- Rest your arms on a table or in your lap while you take a recording. Try to relax and not move too much.
- Make sure that your Apple Watch isn't loose on your wrist. The band should be snug, and the back of your Apple Watch needs to be touching your wrist.
- Make sure that your wrist and your Apple Watch are clean and dry.
- Make sure that your Apple Watch is on the wrist that you selected in the Apple Watch app. To check, open the Apple Watch app, tap the My Watch tab, then go to General > Watch Orientation.
- Move away from any electronics that are plugged into an outlet to avoid electrical interference.

Please follow the instructions below when taking an ECG reading:

1. Make sure that your Apple Watch is snug and on the wrist that you selected in the Apple Watch app. To check, open the Apple Watch app, tap the My Watch tab, then go to General > Watch Orientation.
2. Open the ECG app on your Apple Watch.
3. Rest your arms on a table or in your lap.
4. With the hand opposite your watch, hold your finger on the Digital Crown. You don't need to press the Digital Crown during the session.
5. Wait. The recording takes 30 seconds.

Please keep in mind that the Apple Watch cannot: detect a heart attack; blood clots or stroke; other heart-related conditions. If you are not feeling well, please contact emergency services.

Adapted from: https://support.apple.com/en-ca/HT208955

4. Withings BPM Connect

Preparing for taking the reading:

- Use BPM Connect on the left upper arm
- Rest 5 minutes before the measurement.
- Sit down in a comfortable position, legs uncrossed, feet flat or on the floor, arm and back supported.
- Do not speak or move during the measurement.
- You can wear one layer of clothes but it should not cover your left arm.
- The electrodes should be in contact with the skin.
- Take the measurement in a calm and quiet area.

Please follow the instructions below when taking a Blood Pressure reading:

- Unroll cuff and place your arm inside it.

- Tighten the cuff around your arm. The tube should be positioned against your inner arm.

- Place your arm on a table and level with your heart.

- Press the button to start BPM Connect. Press the button again to start the measurement.

- At the end of the measurement, results are displayed on the screen of BPM Connect. Press the button to validate the measurement. Press the button again to attribute the measurement. Results are sent via WiFi or Bluetooth in the HealthMate app. Please ensure the results were updated.

Adapted from: https://support.withings.com/hc/en-us/articles/360026536033-BPM-Connect-User-Guide

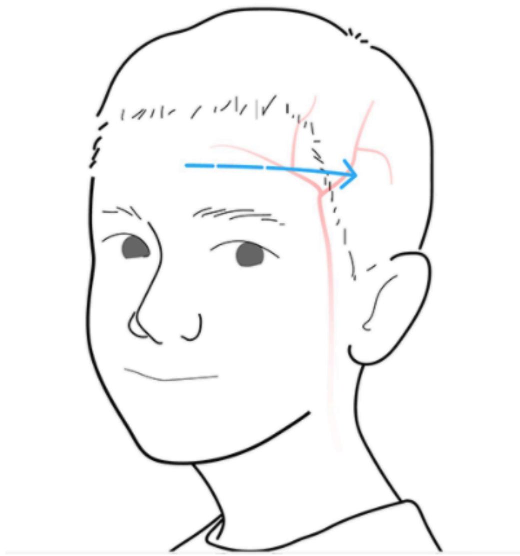
5. **Withings Wireless Scale**

Taking the reading:

- Step on the scale. Reading should start automatically.
- Adjust your body position according to the arrows that appear on the scale if necessary until the number shown as your weight starts blinking.

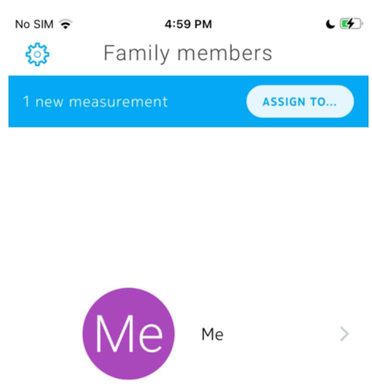
6. **Withings Thermos**

- Remove the protective cap (green cap at the bottom of the thermometer).
- Put the device in front of the forehead. The device does not need to touch the skin but it must be close.
- Starting from the center of the forehead, press and release the button of your Thermo and scan across the forehead in a straight line to the temple.

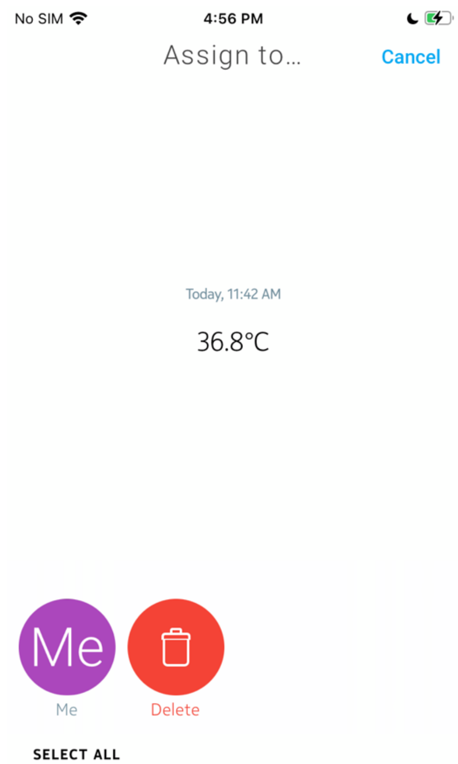


Direct contact with the skin is not necessary. Thermo can be up to 1 cm (0.5 in.) away from your skin.

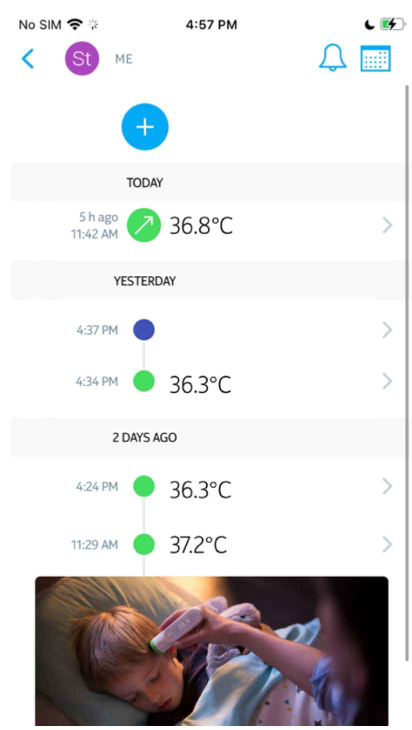
- Thermos vibrates at the end of the measurement and the result comes up on the display. If the result is RETRY, please take another measurement with the device closer to the forehead and moving more slowly, until the result comes up as your current temperature.
- Once the data is read, please assign the correct user. You can do this by sliding your finger up or down on the touch sensitive area of the display to select the correct user, and pressing the button to confirm your choice.
- Pleae open the **Thermos** app.
- If there is a blue banner at the top of the screen:



- Click on the Assign to button, and click **Me.**



- The new measurement should appear on top of the timeline.



7. **Apple Watch – Breathe/Mindfulness App**

Please remember to take this reading as your last reading of the cycle, after blood pressure, ECG, and stress questionnaire.

Taking a reading:

- On the Apple Watch, access the Mindfulness app
- Press the Digital Crown to go to the Home screen, then open the Mindfulness app.
- The session length should be set for 5 minutes to ensure data accuracy. In case the session length is different, please turn the Digital Crown to set the session's length to 5 minutes.
- Tap Start when you're ready. Remember to stay still while you breathe.
- Inhale as the animation grows and your watch taps your wrist. Then exhale as the animation shrinks and the taps stop.
- Breathe until the session ends and your watch taps you twice and chimes. When you're done, you can see your heart rate.
- When you use the Mindfulness app, your watch mutes some notifications, so you can focus. If you answer a call or move too much during a session, the session ends automatically, and you won't get credit.

Adapted from: https://support.apple.com/en-ca/HT206999

8. **Apple Watch – Workout App**

When you begin a workout, please use the **workout** app on the Apple Watch.

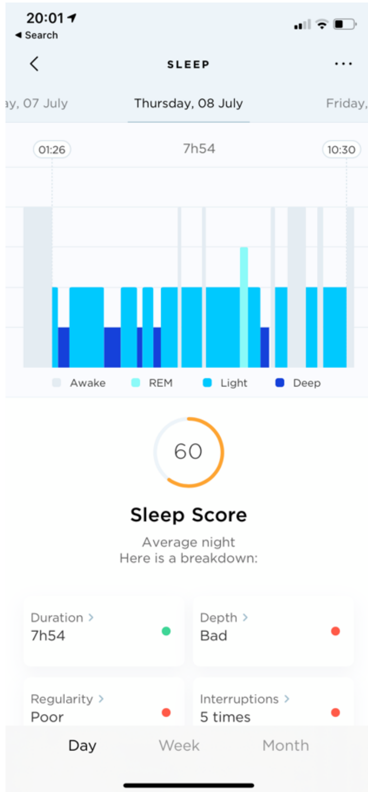
1. Open the Workout app.

2. Find the workout that best matches your activity and select it.

3. To end your workout, swipe right, then tap the End Button. .

9. <u>Withings Sleep</u>

- After setup as described in **User Manual – Get Started**, Withings Sleep will collect data automatically when plugged. Please keep the device plugged for the duration of the study.
- In the morning, please verify in the Health Mate app that sleep data was collected correctly.


20:01
Search
SLEEP
y, 07 July
Thursday, 08 July
Friday,
01:26
7h54
10:30
Awake
REM
Light
Deep
60
Sleep Score
Average night
Here is a breakdown:
Duration
7h54
Depth
Bad
Regularity
Poor
Interruptions
5 times
Day
Week
Month