Signal Processing and Machine Learning Algorithms for Stress Monitoring using Wearable Sensor Technologies

A dissertation presented by:

Talha Iqbal, M.Sc., B.Eng.

to:

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

School of Medicine,

College of Medicine, Nursing and Health Sciences,

University of Galway.

in fulfilment of the requirements for the degree of

Doctor of Philosophy.

Supervised by:

Prof. William Wijns, Dr Atif Shahzad and Dr Adnan Elahi

March 2023

Table of Contents

Declaration of Originality

I, the Candidate **Talha Iqbal**, certify that this thesis entitled **"Signal Processing and Machine Learning Algorithms for Stress Monitoring using Wearable Sensor Technologies"**:

- is all my work.
- has not been previously submitted for any degree or qualification at this University or any other institution.
- and where any work in this thesis was conducted in collaboration, an appropriate reference to published work by my collaborators has been made and the nature and extent of my contribution have been clearly stated.

Name:

Talha Iqbal

Abstract

In recent years, there has been a notable increase in depression, anxiety, pathological stress, and other stress-related diseases. Stress is a known contributor to several life-threatening medical conditions and triggers acute cardiovascular events, as well as one of the root causes of several social problems. According to the statistics from the World Health Organization, stress is associated with several medical and social problems, and these problems are seriously affecting the health and well-being of not only adults but also children and youngsters. The recent development of miniaturized and flexible biosensors has enabled the development of connected wearable solutions to monitor stress and intervene in time to prevent the progression of stress-induced medical conditions. Therefore, a vast interest has been developed to investigate the underlying mechanisms of stress and monitor various biophysiological and biochemical responses of the body to stress. The review of the literature on different physiological and chemical indicators of stress, which are commonly used for quantitative assessment of stress, and the associated sensing technologies shows that prolonged exposure to stress triggers the adrenocorticotrophic hormonal (ACTH) system and causes the release of cortisol hormones from the adrenal cortex that boosts the alertness of the body. As a result, there is an increase in blood supply to muscles, heart rate, respiratory rate, and cognitive activity, along with several other responses. The variable and contradictory evidence in the literature on the use of either physiological or biochemical stress markers leads to the conclusion that neither of these biomarkers in isolation can provide sufficient means of monitoring stress. Therefore, a combination of physiological and chemical stress biomarkers, with contextual information, can be a more reliable solution for stress monitoring.

The current standard for stress evaluation is based on self-reported questionnaires and standardized stress scores. There is no gold standard to independently evaluate stress levels despite the availability of numerous biophysiological stress indicators. Moreover, there is no clear understanding of the relative sensitivity and specificity of these stress-related biophysiological indicators of stress in the literature. An extensive statistical analysis and classification modelling of biophysiological data gathered from healthy individuals, undergoing various induced emotional states was performed to assess the relative sensitivity and specificity of common biophysiological indicators of stress. The key indicators of stress, such as heart rate, respiratory rate, skin conductance, RR interval, heart rate variability in the

electrocardiogram, and muscle activation measured by electromyography, are evaluated as gathered from an already existing, publicly available WESAD (Wearable Stress and Affect Detection) dataset. Respiratory rate and heart rate were the two best features for distinguishing between stressed and unstressed states. Both parameters can be estimated using a single photoplethysmography (PPG) sensor. The heart rate is estimated by counting the number of peaks in the PPG signal. Most of the existing algorithms for the estimation of respiratory rate using photoplethysmography (PPG) are sensitive to external noise and may require the selection of certain algorithm-specific parameters, through the trial-and-error method. Thus, a new algorithm to estimate the respiratory rate using a photoplethysmography sensor signal for health monitoring is proposed. The algorithm is resistant to signal loss and can handle low-quality signals from the sensor. The results endorse that integration of the proposed algorithm into a commercially available pulse oximetry device would expand its functionality from the measurement of oxygen saturation level and heart rate to the continuous measurement of the respiratory rate with good efficiency at home and in a clinical setting.

Additionally, as the public availability of datasets for the development of stress monitoring devices is limited, a clinical study was performed. The dataset created is an open-access dataset named Stress-Predict dataset. The inclusion of an additional feature, i.e., respiratory rate data along with stress and baseline labels within the dataset, makes the dataset more desirable and unique from all the other publicly available Empatica E4-based datasets. The dataset and outcomes of this study contribute to understanding any accuracy gaps in current stress monitoring and help improve these technologies or develop new technologies for stress monitoring. Most wearable stress monitoring systems are built on a supervised learning classification algorithm trained on simple statistical features. For accurate stress monitoring, it is essential that these features are not only informative but also well-distinguishable and interpretable by the classification models. Thus, a correlation-based time-series feature selection algorithm is proposed and evaluated on the stress-predict dataset. The outcome of the study suggests that it is vital to have better analytical features rather than conventional statistical features for accurate stress classification.

One of the most challenging tasks in physiological or pathological stress monitoring is the labelling of the physiological signals collected during an experiment. Commonly, different types of self-reporting questionnaires are used to label the perceived stress instances. These questionnaires only capture stress levels at a specific point in time. Moreover, self-reporting

is subjective and prone to inaccuracies. Traditional supervised machine learning classifiers require hand-crafted features and labels while on the contrary, the unsupervised classifier does not require any labels of perceived stress levels and performs classification based on clustering algorithms. The analysis and results of this comparative study demonstrate the potential of unsupervised learning for the development of non-invasive, continuous, and robust detection and monitoring of physiological and pathological stress.

Acknowledgements

I sincerely thank all those whose help, support and guidance over the years have been instrumental in the completion of this thesis. With a profound sense of gratitude, I express my sincere thanks to my supervisors Prof. Wijns, Dr Shahzad, and Dr Elahi for their constant guidance, inspiration, valuable advice, education, and support during my research work. I would cordially thank Prof. Wijns for allowing me to work in Smart Sensors Lab (SSL) and generously providing needed resources for my doctoral project. Particularly, I wish to express my special appreciation to Dr Shahzad and Dr Elahi for their patience, motivation, forbearance, and tremendous knowledge, of both research work and the write-up, which they have imparted to me during the project. This work would not have been possible without their guidance and persistent help. Their continued mentoring and invaluable advice, on both research and career-building, over the years have inculcated in me the essential skills needed for research pursuits. I convey my thanks to the graduate research committee members, Prof. Stewart Redmond Walsh, Dr Karen Doyle, and Dr Martin Glavin, for their guidance and encouragement during this project.

I cordially acknowledge and express my sincere thanks to Dr Andrew Simpkin, Dr Davood Roshan, Prof. Jane Walsh, and Dr Gerard Molloy for sharing their valuable knowledge and expertise in statistical analysis, software, protocol designing and research in general. I would also such as to thank Dr Sandra Ganly, Nicola Glynn, John Killilea, and Eileen Coen for helping me conduct the clinical study. I am also indebted to friends and colleagues in Smart Sensors and Translational Medical Device (TMD) Lab Muhammad Farooq, Pau Redon, Bilal Ameen, Muhammad Saad Muneeb, Patricia Vazquez, Muhammad Riaz-ur-Rehman, Xinlei Wu, Daixin Ding, Jiayue Huang, Nadia Hussain, Sarita Gundeti, Hannah Ryman, Nadeem Soomro, Benazir Abbasi and Haroon Zafar for their support, encouragement, and making my working hours more pleasant during my PhD.

From my heart and soul, I am immensely grateful to my loving mother who incessantly prayed and waited for the completion of this PhD journey, and by supporting and admiring my father whose invaluable support and guidance gave me the strength to persistently endeavour during my academic career. I am thankful to my loving brother and sister who

continuously missed me as I missed them during my stay in Ireland. The love, support, prayers, and guidance from my family throughout these 4 years and my life immensely contributed to the accomplishment of my doctoral program. Indeed, I would such as to dedicate my PhD to my parents, siblings, and my country Pakistan. I would thank Dr Hazrat Ali, who inspired me to contribute to humanity through scientific knowledge and inspired in me the passion for research during my master's program back at COMSATS University Islamabad, Abbottabad Campus, Pakistan.

I also thankfully acknowledge the Science Foundation Ireland (SFI) for providing funding for this research project. Last but not the least, I would also such as to thank the people of Ireland, more specifically Galway, for being so kind, compassionate, and supportive. Being an international student from Pakistan, I always found Galway as my second home. The Irish people, their culture, and their hospitality were fabulous and unmatchable. The people of Ireland have imparted many good societal responsibilities to me that will surely shape me to become a productive and useful citizen of my country Pakistan.

Acronyms

TSST	trier social stress test
VASS	visual analogue scale
PSS	perceived stress scale
ECG	electrocardiogram
ACC	accelerometer
GYRO	gyroscope
GPS	global positioning system
PPG	photoplethysmogram
EMG	electromyography
BVP	blood volume pulse
ACTH	adrenocorticotrophic hormonal
HRV	heart rate variation
SVM	support vector machine
ANN	artificial neural network
DT	decision tree
CAD	computer-aided diagnostics
STAI	state-trait anxiety inventory
MIST	montreal imaging stress task
PSQ	perceived stress questionnaire
PCA	principal component analysis
PSD	power spectrum density
FFT	fast Fourier transform

kNN	k-nearest neighbour
IBI	inter-beat-interval
LDA	linear discriminant analysis
LC-MS/MS	liquid chromatography-mass spectrometry
RIA	radioimmunoassay
NIRS	near-infrared spectroscopy
ELISA	enzyme-linked immunosorbent assay
DAG	directed acyclic graph
LVQ	learning vector quantization
KSOM	korhonen self-organizing map
MLP	multilayer perceptron
HMM	hidden Markov mode
HPA	hypothalamic-pituitary-adrenal
CBG	corticosteroid-binding globulin
ISF	interstitial fluid
PSTD	posttraumatic stress disorder
PANAS	positive and negative affect schedule
LFI	lateral flow immunosensor
QCM	quartz crystal microbalance
TEB	thoracic electrical bioimpedance
LOOCV	leave-one-out-cross-validation
ESQI	entropy signal quality index
BIRCH	balanced iterative reducing and clustering using hierarchies
DBSCAN	density-based spatial clustering of applications with noise

OPTICS	ordering points to identify the clustering structure
SOM	self-organizing maps

List of Figures

List of Tables

Chapter 1
Introduction

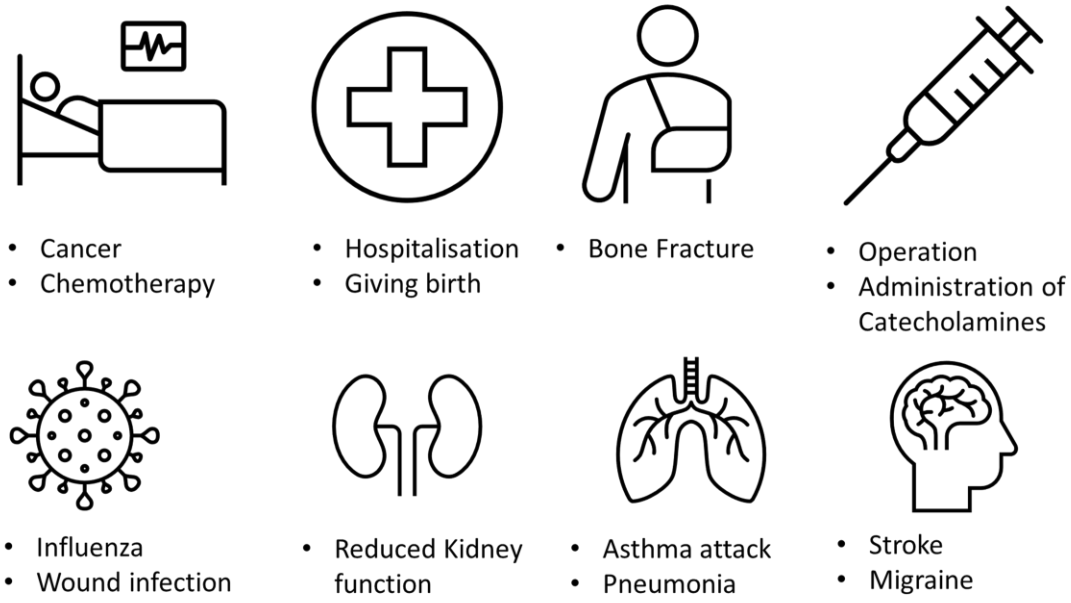
This chapter provides an overview of the proposed PhD thesis. It provides an introductory discussion about real-life stress, different stress triggers, a problem statement, a proposed solution, a brief literature survey, and discovered challenges in the development of a real-time stress monitoring system. The chapter also highlights the key contributions of this PhD to the stress monitoring field.

1.1 Stress and Stress Triggers/Stressors

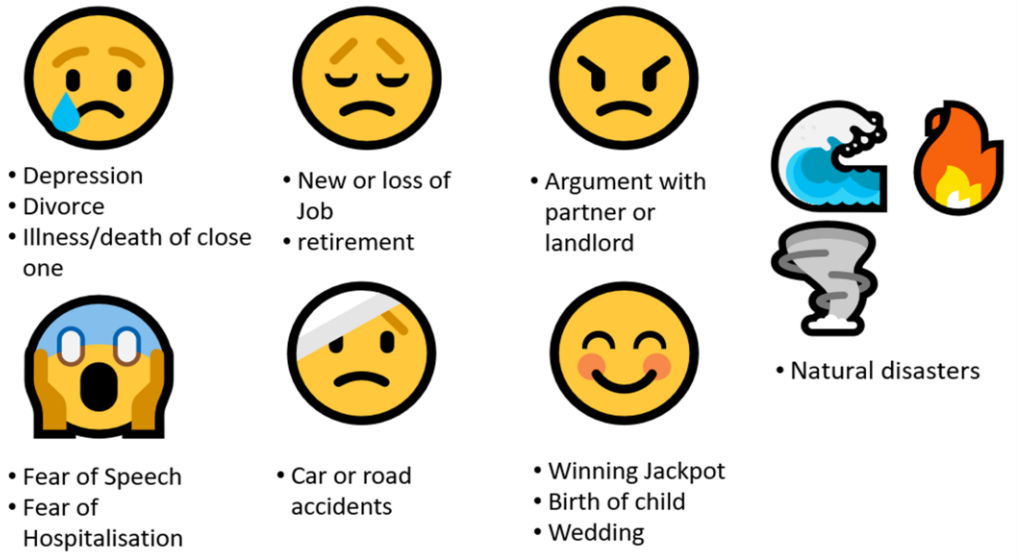
Any stimulus or threat that causes physical, emotional, and/or psychological strain can be called stress [1]. It is a feeling of being surprised or unable to deal with emotional or mental pressure. The stimuli that cause stress are called stress triggers or stressors. These triggers can be divided into two groups: one includes emotional triggers while the other includes physical triggers [2]. Emotional triggers include events such as depression, divorce, fear of speech, death of any family member, being involved in a car or road accident, arguing with a landlord or partner, and natural disasters. There are also some happy emotional triggers such as winning a jackpot, the birth of a child and a wedding. On the other hand, physical triggers of stress include stroke, migraine, asthma attack, giving birth, cancer, chemotherapy, wound infection, kidney failure or surgery, among others. Figure 1.1 summarizes different emotional and physical triggers. All these triggers or stressors affect the daily lifestyle and well-being of individuals and all of us as a community as well.

Generally, stress cannot be recognised easily. Several signs can indicate one might be expressing stress. Sometimes the stress is induced due to some obvious reason but sometimes even small daily-life stress from work, family, or friends could induce stress. Stress can be for a short time (called acute stress) or can last for a long time (resulting in chronic stress). Our body automatically responds to both types of stress and releases adrenaline and noradrenaline in the case of short-term/acute stress while releasing cortisol in the case of long-term/chronic stress. Stress hormones such as adrenaline/noradrenaline and cortisol are released into the body to regulate blood flow and blood pressure. The

release of these hormones due to sympathetic excitement causes an abrupt change in blood pressure, and high blood pressure can cause some plaque events. Figure 1.2 shows the response of our body to a stress condition.



(a) Physical triggers of stress



(b) Emotional triggers of stress

Figure 1.1 Physical and emotional triggers/stressors in real-life.

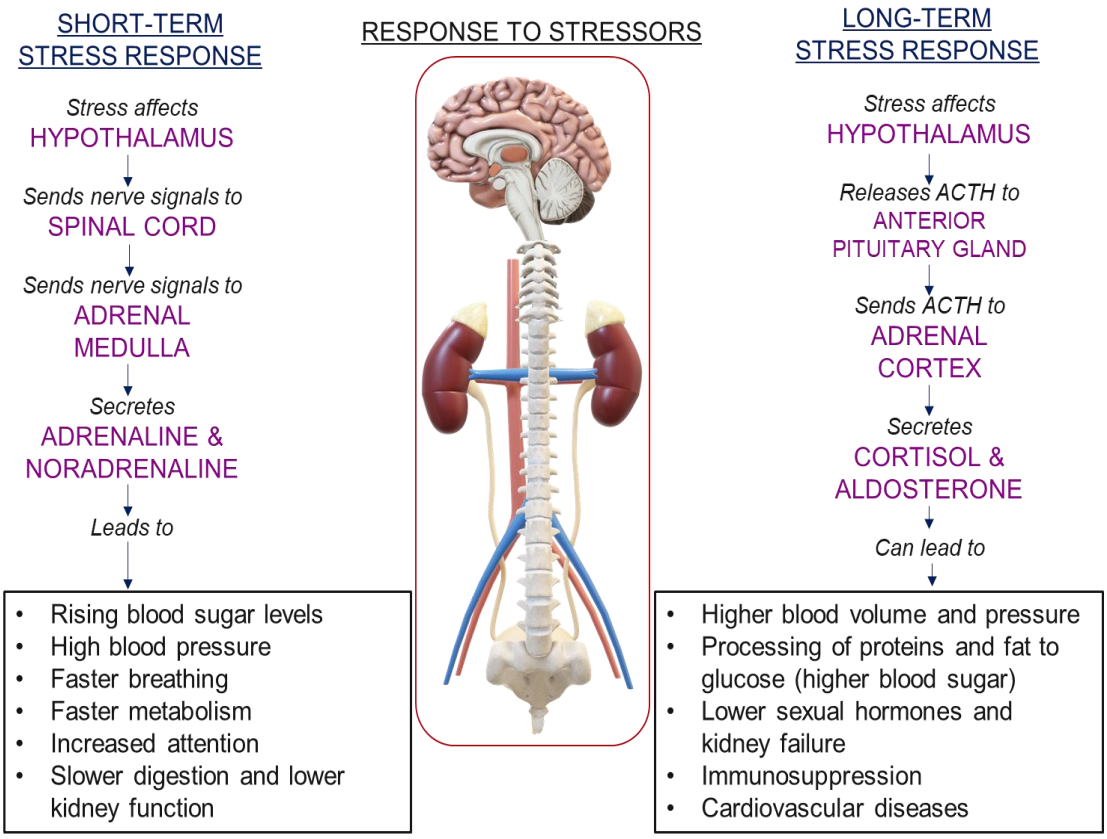


Figure 1.2 The response of our body to different stress conditions.

1.2 Clinical Need

In recent years, there has been a noticeable increase in pathological stress, anxiety, depression, and other stress-related diseases [3]–[5]. There is a vast literature present that concludes that physical and emotional stress is strongly related to heart failure and hypertension which may by themselves result in stroke or heart attack. According to the World Health Organization[1], the physical and mental health, as well as the overall well-being of a person, are compromised by stress [6]. According to the British Health and Safety Executive (HSE), 50% of all work-related illnesses in 2021 were due to stress [7]. Stress can be of two types: short-term stress also known as acute stress and long-term stress known as chronic stress. Acute stress may not cause severe health issues in young and healthy individuals, but if a stressor (a stimulus that causes stress) remains too persistent, then this

[1] *Available: https://www.hse.gov.uk/statistics/causdis/.*

chronic stress might lead to anxiety and depression, not only in young people but in children too [8]. Chronic stress is also a well-known contributor to several life-threatening diseases such as different heart diseases, diabetes, high blood pressure, and obesity and an acute episode of stress might cause stroke or heart attack due to arterial inflammation [9], [10].

In clinical settings, stress monitoring is performed using self-reporting questionnaires, surveys, or daily diaries of an individual [11]. These detection methods have several drawbacks including:

- Bias: The reported stress levels might not be truly indicative of stress as individuals can underreport their stress level due to social desirability or can over-report stress for gaining attention and sympathy.
- Accuracy: These reports cannot be considered an accurate reflection of an individual's stress levels as they might not be aware of certain stressors or might not be able to demonstrate their emotional state. The language used in the questionnaires might also be not fully understandable by the individual and thus results in inaccurate ratings.
- Generalizability: Most of these questionnaires are often based on a certain time frame and cannot be generalizable to other time frames. Moreover, the individual might not remember the event or feelings that occurred in past which results in recall bias.

Thus, new approaches are required to develop a stress monitoring system that can perform timely detection, continuous monitoring, and management of stress levels to provide a healthier lifestyle.

1.3 Literature Survey

Continuous stress monitoring can provide detailed information about the stress levels of an individual, helps in identifying stress patterns and lead to early detection as well as intervention for stress-related diseases. There has been a vast interest in investigating the underlying techniques of detection, monitoring and management of various biophysiological as well as biochemical responses of the body to stress [12]. Determination of a reliable and specific indicator or biomarker of stress is a critical step toward accurate monitoring of stress, which will potentially enable the mitigation of pathological conditions

at fairly early stages. In the last two decades, there has been a significant improvement in the development of physiological and biochemical sensing methods that can be used to detect and monitor stress.

To develop a reliable stress monitoring device, it is important to understand how our body responds physiologically and biochemically to stress triggers. Stress can be interpreted as a disturbance in the homeostatic balance of an individual, through which the body tries to cope with the stressor and is called a stress response [13]. The stress response might be different in some cases, but generally, under the influence of stressors, the stress response triggers the sympathetic nervous system and inhibits the parasympathetic nervous system. The sympathetic nervous system releases different hormones such as cortisol, alpha-amylase, and adrenaline [14], [15] that result in the change in overall respiratory rate, heart rate, blood pressure, skin temperature, muscle tension, and cognitive activity, along with several other biochemical markers. These hormonal changes and changes in biomarkers prepare an individual for a *fight-or-flight* reaction in response to stress [16].

Biochemical markers, such as cortisol and alpha-amylase [17] have high sensitivity to stress detection and offer ease of collection. However, several studies have demonstrated the potential of using different physiological signals such as skin temperature (TEMP), electromyography (EMG), blood volume pulse (BVP) and electrodermal activity (EDA) in response to stress [18]. The current standard clinical practice for stress detection is based on self-reporting visual scales (such as Visual Analogue Scale, VASS) or self-reporting questionnaires (such as the Perceived Stress Scale, PSS) [19]. Recent development in the field of wearable biosensor technology has made the measurement of biophysiological signals of the body for the detection and evaluation of stress more interesting, as these indicators can be continuously measured through non-invasive wireless systems.

Many research studies have explored the relationship between biophysiological signal variation and stress levels [20]–[23]. Most studies use machine learning methods to classify stressed physiological signals from nonstress signals. The commonly used physiological signals in these studies include respiratory rate (RESP) and electrocardiogram (ECG) signals combined with activity monitoring sensors (such as an accelerometer (ACC), gyroscope (GYRO), and global positioning system (GPS)). There are some less frequently used stress-indicating signals that can also be recorded and used for stress monitoring.

These indicators include photoplethysmogram (PPG), TEMP, EMG, BVP and EDA [24]–[27]. All these biophysiological indicators are not specific to stress and can be affected by other conditions such as when a person is happy, sad, in anger or excited [28]. Thus, varying stress detection accuracies can be observed in the published literature.

1.4 Challenges in the Development of Stress Monitoring Devices

Despite all advances, there is no single clinically validated stress monitoring device that can be used clinically to monitor continuous stress in daily life. Some commercially available devices claim to detect stress levels in users. Peake *et al.* [29] have reviewed these devices and have discussed their strengths and effectiveness in stress monitoring. It is important to note that these devices are not clinically tested and lack accuracy as they are trained on limited available data due to privacy concerns. Calibration for stress detection is done using self-reporting questions and does not consider self-report bias [30]. The detection of stress levels based on a single physiological measurement such as heart rate variation is also a problem as measured signals can be affected by other confounding factors (for example, happiness, exercise, or shock) [31].

This lack of a reliable stress monitoring device can be explained by several key challenges that need to be addressed to develop and clinically validate a stress monitoring model. Firstly, there is no universally accepted definition of stress. According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), stress is the physiological and psychological response towards a perceived change that threatens the well-being or survival of someone, resulting in a series of physiological responses and adaptations [32]. Philosophers define the concept of stress from various perspectives. According to Martin Heidegger, stress is not merely a physiological or psychological response, but a fundamental aspect of our existence [33]. Michel Foucault saw stress as a product of societal power structure [34]. Overall, there is no one definition of stress. The above-mentioned stress definitions could be revised and used clinically for understanding and assessing patient experiences.

Secondly, there is a lack of corresponding gold-standard reference/ground-truth values that can be somehow associated with stress definition. For example, biochemical researchers

report cortisol as a vital stress detection hormone. In most of these studies, questionnaire-based self-reporting has been the ground truth to assess the stress levels of the subjects [35], [36]. However, the correlation value between the self-reporting questionnaire and the cortisol level is reported to be in the range of 0.26 to 0.36 [37], [38]. This poor correlation could be associated with poor reporting in the questionnaires as quantifying stress levels is difficult and thus leads to difficulties in the prediction accuracy of cortisol levels.

The third challenge is recording the different biophysiological data, especially in the natural environment [39]. Data collection in the presence of different sources of noise and errors is a significantly challenging task. For example, to record a day-long ECG data, ECG electrodes must be connected to the body of the subject for the whole day. The adhesion of electrodes will degrade over time, thus producing a lot of noisy signals. The movement of the subject will alter the electrode connection and will also cause noisy spikes in the signal. Moreover, there is a probability of losing data during wireless transmission. The same problems (of connection with the body, noise due to movement, and chances of data loss) are also incurred while recording other signals such as respiratory rate, EMG, EDA, skin temperature and BVP. The fourth challenge is handling the confounding variables. The stimulus that is indicative of induced stress in an individual can easily be obfuscated by a movement of the limb, a change in posture, or any other physical movement that results in an altered low-quality signal. For example, respiratory-induced amplitude variation in the PPG signal is caused by inhale and exhale cycle. Separating a good/high-quality signal from a low-quality signal for stress analysis and diagnosis is, therefore, a challenging task.

The fifth challenge is to identify and calculate discriminative features that are specific to the stressed condition and can easily be characterized as a stress response from all other physiological stimuli. The final challenge is the development of a classification technique that computes all the stress-related features, trains, and validates the model for real-time stress monitoring. This is the most difficult challenge to solve as no reference/gold standard dataset is available that could be used for the training and validation of any stress classification models. Currently, self-reporting questionnaires are the only clinically accepted method to determine the labels for stressed signals. If self-reporting questionnaires are used to determine the labels of stress, then for consistency analysis the

threshold is set at 0.7 to declare the concordance [24]. This thresholding also reflects inherited variabilities and biases in the self-reported data.

Thus, it is essential to overcome the above-mentioned challenges and develop an accurate stress monitoring device to improve the overall quality of life by providing in-time interventions.

1.5 Objectives

The objectives of this thesis are designed to consider the gaps in the available literature and to mitigate the above-mentioned challenges. The challenge of having a universal definition of stress and a standard reference dataset still exists and can be solved by the collaborative efforts of researchers and clinicians.

The challenges of dealing with noisy, motion-affected, and low-quality signals can be solved by investigating a device that is less effected by noise, could be worn easily and can provide readings even during motion. One of the possible solutions is the use of PPG wrist-worn devices. The PPG device works on light reflections and does not need a firm connection with the body. However, an algorithm is also required to extract/filter the signal from noise and motion artefacts and accurately estimate the stress-related features.

As cofounding variables and features are also found in the recorded signals, a time-series feature selection algorithm is required for mitigating this problem. For accurate stress monitoring, it is essential that these features are not only informative but also well-distinguishable and interpretable by the classification models.

The accessibility of continuous signal from a wearable device, an accurate stress-features estimation and a feature selection algorithm will certainly improve the overall stress-predictive model's performance and partially, address the last challenge of having an accurate classification model.

The thesis objectives are as follows:

1. Literature reports different physiological and biochemical indicators of stress but lacks a quantitative measure of those indicators. Thus, review different biophysiological and biochemical indicators of stress to determine quantitatively measurable stress indicators.

2. Several machine learning classifiers for stress classification have been reported in the literature. As accurate stress classification is highly dependent on these classifiers, review the different machine learning algorithms used as predictive models for stress detection and identify the shortcomings resulting in different predictive accuracies.

The effectiveness of stress detection is dependent on the detection methodology and analysed parameter/indicator, as every method and parameter has its strengths and limitations. Thus,

3. Investigate the most promising techniques used for biochemical indicator (cortisol) detection. Discuss the challenges and the device's sensitivity, and specificity in the detection of cortisol presence.
4. Investigate and shortlist the biophysiological indicators as well as predictive algorithms to get the most sensitive and specific indicators of stress (that are less affected by other factors).

Studies report that most of the stress indicators (parameters) derived from physiological signals are noisy, of low quality, and have correlated features. These problems result in decreased classification accuracies as well as raise a question about the generalizability of the model. Thus,

5. Develop and evaluate an algorithm that can extract the stress-specific information from a physiological signal, which should be able to deal with low-quality signals and deals with other co-founding factors to detect and classify stress accurately.
6. Investigate and provide a solution for the identification and calculation of discriminative features of the stressed condition which should easily be characterized as a stress response from all other physiological stimuli.

1.6 Thesis Contributions

This PhD is a translational research and tech innovation project with software as a core part. The aim is to alleviate the above-mentioned challenges and propose a feasible solution for real-time stress management.

The specific novel contributions of this thesis are summarised and listed below:

1. The literature on different physiological and chemical indicators of stress used for quantitative measurement of stress and associated sensing technologies was reviewed to formulate the list of physiological and biochemical indicators of stress (objective 1). This contribution aimed to find approximated quantitative measures of a person's homeostatic imbalance and discuss any cofounders that might have caused variation in the reported results.

   Publication: Talha Iqbal, Adnan Elahi, Pau Redon, Patricia Vazquez, William Wijns, and Atif Shahzad. "A review of biophysiological and biochemical indicators of stress for connected and preventive healthcare." Diagnostics 11, no. 3 (2021): 556.

2. Different machine learning algorithms that are used for stress prediction were gathered and compared as accurate stress detection is vastly dependent on the predictive model (objective 2). This contribution gives a better in-look into state-of-the-art machine learning algorithms, their use as stress level classifiers and trade-offs (between computation time vs accuracy vs price of the device) that are to be made while using these algorithms.

   Publication: Talha Iqbal, Adnan Elahi, Atif Shahzad, and William Wijns. "Review on Classification Techniques used in Biophysiological Stress Monitoring." Arivx (2022).

3. Cortisol is considered the most vital and potentially clinically useful biomarker for stress monitoring and estimation. This contribution provides an overview of the most promising techniques currently used for cortisol detection and the challenges associated with them. This work helped in determining the best possible wearable method/technology to collect and analyse cortisol levels (objective 3).

   Publication: Talha Iqbal, Adnan Elahi, William Wijns, and Atif Shahzad. "A Comprehensive Review of Cortisol Detection Methods for Stress Monitoring in Connected Health." Health Sciences Review (2023): 100079.

4. A statistical and classification analysis of all the stress physiological and biochemical indicators reported in the literature was performed. This contribution identified the indicators that are more specific for stress monitoring and are less affected by any other physiological and emotional events. The respiratory rate and heart rate were the two best stress indicators for distinguishing between stressed and unstressed states (objective 4).

Publication: Talha Iqbal, Pau Redon-Lurbe, Andrew J. Simpkin, Adnan Elahi, Sandra Ganly, William Wijns, and Atif Shahzad. "A sensitivity analysis of biophysiological responses of stress for wearable sensors in connected health." IEEE Access 9 (2021): 93567-93579.

5. A comparative analysis of unsupervised versus supervised machine learning algorithm for stress state classification. This contribution explores the potential feasibility of unsupervised learning classifiers to be implemented in stress-monitoring wearable devices, as most wearable stress-monitoring systems are built on a supervised learning classification algorithm (objective 4).

   Publication: Talha Iqbal, Adnan Elahi, William Wijns, and Atif Shahzad. "Exploring Unsupervised Machine Learning Classification Methods for Physiological Stress Detection." Frontier in Medical Technology - Diagnostic and Therapeutic Devices 4 (2022).

6. Development of a novel algorithm for estimation of respiratory rate from raw Photoplethysmography (PPG) signal (objective 5). This contribution proposed a new algorithm that could accurately estimate the respiratory rate from raw PPG signals. Additionally, the impact of different window sizes on the estimation of respiratory rate was also determined in this study.

   Publication: Talha Iqbal, Adnan Elahi, Sandra Ganly, William Wijns, and Atif Shahzad. "Photoplethysmography-based respiratory rate estimation algorithm for health monitoring applications." Journal of Medical and Biological Engineering 42 (2022): 242–252.

7. Perform a clinical study to evaluate the proposed respiratory rate estimation algorithm and publish the developed dataset of healthy volunteers (objective 5). The dataset and outcomes contribute to understanding any accuracy gaps in current stress monitoring solutions and help improve these technologies or develop new technologies for stress monitoring.

   Publication: Talha Iqbal, Andrew Simpkin, Davood Roshan, Nicola Glynn, John Killilea, Jane Walsh, Gerard Molloy, Adnan Elahi, Sandra Ganly, Hannah Ryman, Eileen Coen, William Wijns, and Atif Shahzad. “Stress Monitoring Using Wearable Sensors: A Pilot Study and Stress-Predict Dataset.” Sensors 22, no. 21 (2022): 8135.

8. Proposed an optimised time-series feature extraction algorithm for accurate stress classification (objective 6). This contribution aims to explore the efficacy of a time-series feature extraction algorithm applied to the heart rate and respiratory rates of healthy volunteers (Stress-Predict dataset) and propose an optimised featured engineering algorithm for accurate stress classification.

   Publication: Talha Iqbal, Adnan Elahi, Atif Shahzad, and William Wijns, "Improved Stress Classification using Automatic Feature Selection from Heart Rate and Respiratory Rate Time Signals." Applied Sciences 13, no. 5 (2023): 2950.

1.6.1 Publications

There are 8 deliverables of this PhD (all published) excluding the thesis, listed:

[1] **Iqbal, Talha**, Adnan Elahi, Pau Redon, Patricia Vazquez, William Wijns, and Atif Shahzad. "A review of biophysiological and biochemical indicators of stress for connected and preventive healthcare." Diagnostics 11, no. 3 (2021): 556.

[2] **Iqbal, Talha**, Adnan Elahi, Atif Shahzad, and William Wijns. "Review on Classification Techniques used in Biophysiological Stress Monitoring", Arvix (2022).

[3] **Iqbal, Talha**, Adnan Elahi, William Wijns, and Atif Shahzad. "A Comprehensive Review of Cortisol Detection Methods for Stress Monitoring in Connected Health.", Health Sciences Review (2023): 100079.

[4] **Iqbal, Talha**, Pau Redon-Lurbe, Andrew J. Simpkin, Adnan Elahi, Sandra Ganly, William Wijns, and Atif Shahzad. "A sensitivity analysis of biophysiological responses of stress for wearable sensors in connected health." IEEE Access 9 (2021): 93567-93579.

[5] **Iqbal, Talha**, Adnan Elahi, William Wijns, and Atif Shahzad. "Exploring Unsupervised Machine Learning Classification Methods for Physiological Stress Detection." Frontier in Medical Technology - Diagnostic and Therapeutic Devices 4 (2022).

[6] **Iqbal, Talha**, Adnan Elahi, Sandra Ganly, William Wijns, and Atif Shahzad. "Photoplethysmography-based respiratory rate estimation algorithm for health monitoring applications." Journal of Medical and Biological Engineering 42 (2022): 242-252.

[7] **Iqbal, Talha**, Andrew Simpkin, Nicola Glynn, John Killilea, Jane Walsh, Gerard Molloy, Adnan Elahi, Sandra Ganly, Hannah Ryman, Eileen Coen, William Wijns, and Atif

Shahzad. "Stress Monitoring Using Wearable Sensors: A Pilot Study and Stress-Predict Dataset." Sensors 22, no. 21 (2022): 8135.

[8] **Iqbal, Talha,** Adnan Elahi, Bilal Amin, Atif Shahzad, and William Wijns, "Improved Stress Classification using Automatic Feature Selection from Heart Rate and Respiratory Rate Time Signals." Applied Sciences 13, no. 5 (2023): 2950.

1.7 Thesis Organisation

The rest of the thesis is organised as follows: Chapter 2 describes the background of stress monitoring and stress triggers, evaluation of stress monitoring methods in terms of biophysiological and biochemical indicators of stress, comparison of different stress prediction and classification techniques used in the field of biophysiological stress monitoring, and a comprehensive review of cortisol detection as a method of stress monitoring. The chapter concludes with a discussion on the grey areas identified after the literature survey toward accurate and reliable stress monitoring devices. Chapter 3 presents the statistical analysis of biophysiological and biochemical indicators of stress to determine the most specific indicator of stress. The chapter also explains the publicly available dataset used in the study and how it was composed. The shortlisted indicators are selected based on univariant and multivariant analysis results. Chapter 4 is about an optimised unsupervised machine learning algorithm for accurate stress classification and discusses the results on publicly available datasets. A novel algorithm and the reason for the development of a new algorithm to estimate the respiratory rate using photoplethysmography for health monitoring are discussed in Chapter 5. To test the developed algorithm, we planned and conducted a clinical trial. All the details of the clinical trial and developed dataset (named a stress-predict dataset) are presented in chapter 6. Chapter 7 explores the potential time-series features for an accurate stress classification and proposes an optimised feature engineering algorithm. Lastly, the overall conclusion of cardio-predict; stress monitoring solution and future directions are provided in chapter 8.

Chapter 2
Stress Monitoring Background

This chapter[2] presents:

- A review of the literature on different physiological and chemical indicators of stress, which are commonly used for quantitative assessment of stress, and the associated sensing technologies. It provides a list of different biophysiological along with biochemical signals/indicators associated with stress detection and is used in literature. (Section 2.1)
- A review of commonly used machine learning classification techniques and a discussion on choosing a classifier, which depends upon several factors other than accuracies, such as the number of subjects involved in an experiment, type of signal processing, and computational limitations. (Section 2.2)
- Finally, it gives a detailed overview of the most promising techniques currently used for biochemical indicator (cortisol) detection. It discusses the challenges such as the feasibility of the device used for cortisol collection, the correlation of cortisol levels with stress, invasive and non-invasive device's sensitivity, and specificity in the detection of cortisol presence and other issues associated with cortisol detection. (Section 2.3)

The work presented in this chapter achieved the first three objectives of the thesis (related to the review /investigation of existing technologies for stress monitoring in the literature) and helped in:

- Gaining an understanding of current research knowledge, key findings, and research gaps in the field of stress monitoring.

[2] *The following body of the chapter is copy of the papers published in parts in Diagnostics 2021 (Section 2.1), ArviX 2022 (Section 2.2) and in Health Science Review 2022 (Section 2.3). I am the first lead author in the papers, which is co-authored with my supervisors. The conceptualization, formal analysis, investigation and visualization were also done by me. Designed methodology and validation were led by my supervisors and me. I led all parts of the work with the support of my supervisors.*

- Identifying trends such as types of studies and methods used in the research field of stress monitoring.
- Identifying potential ethical issues that are associated with the data collection and processing, and steps to address them.
- Finally, enhancing the quality of research by designing future studies to avoid duplication of the already existent work which eventually led to saving time and effort.

2.1 Review of Physiological and Biochemical Indicators of Stress

Stress is a known contributor to several life-threatening medical conditions and a risk factor for triggering acute cardiovascular events, as well as a root cause of several social problems. The burden of stress is increasing globally and, with that, is the interest in developing effective stress-monitoring solutions for preventive and connected health, particularly with the help of wearable sensing technologies. The recent development of miniaturized and flexible biosensors has enabled the development of connected wearable solutions to monitor stress and intervene in time to prevent the progression of stress-induced medical conditions. This section presents a review of the literature on different physiological and chemical indicators of stress, which are commonly used for quantitative assessment of stress, and the associated sensing technologies.

2.1.1 Introduction

In recent years, we have seen a notable increase in anxiety, depression, pathological stress, and other stress-related diseases. Generally, stress harms the physical and mental health and well-being of a human [3]–[5]. In particular, chronic stress increases the chances of cardiovascular disease [40], diabetes, stroke, and obesity [9], [41]. According to the statistics from the World Health Organization, stress is associated with various medical and social problems, and these problems seriously affect the health and well-being of not only adults but also children and young people [42]. Therefore, a vast interest has been developed to investigate the underlying mechanisms of stress and monitor various biophysiological and biochemical responses of the body to stress [12]. A reliable biomarker or indicator of stress could provide accurate monitoring of stress, potentially enabling the prevention of pathological conditions at the early stages. In the past two decades, there has been significant development in physiological and biochemical sensing technologies. These sensors provide an excellent platform for connected health solutions and preventive care for various conditions caused by or associated with stress [8], [43].

Stress can be defined as a disturbance in an individual's homeostatic balance, with which the body attempts to cope, and this is known as the stress response [13]. Stress can be acute, i.e., an immediate response to a stressor, or chronic, i.e., a state caused by a constant stress

stimulus [44]. Chronic stress can lead to a stage where the body can no longer achieve homeostatic balance and the individual can no longer deal with the stressors. Activation of the stress response triggers a variety of changes in the body, caused by the stimulation of the sympathetic nervous system and the inhibition of the parasympathetic system. The stress response can vary but generally includes the release of stress hormones that increase the alertness of the body. As a result, there is an increase in heart rate, the blood supply to the muscles, respiratory rate, skin temperature (due to higher blood circulation), and cognitive activity, among several other responses. Stress-specific hormonal responses and other biomarkers affected by the stress response are commonly used to quantitively assess or monitor stress [5], [12], [13].

Most of the studies reported in the literature on stress monitoring follow a similar experimental approach, where sensors collect biophysiological data in the stress and non-stress states. First, stress is induced in a controlled environment (laboratory) [45] or real life [46] using mental arithmetic, the TSST, or the Stroop test. Then, various features are extracted from the sensors' data and machine learning (ML), or pattern recognition is used to differentiate the stress state from the non-stress state (or baseline). Machine learning algorithms can be divided into two basic types. The first is supervised learning, in which input is fed along with the classification labels to the model for prediction and classification. The second is unsupervised learning, in which no labels are given at the input and the model is designed to group the input data based on some inherent patterns or similarities. Usually, the data from the sensors are recorded on the device and then transferred to a computer or the cloud for processing and analysis. In some cases, especially in a simulated driving scenario, participants' wearable sensors are directly connected to a computer, and real-time analysis is performed during the experiment. Various ML techniques have been used for classification, for example, support vector machine (SVM) [27], Bayesian networks (BN), artificial neural network (ANN) [47], fuzzy logic, decision tree (DT) [48], and other computer-aided diagnostics (CAD) tools [49]. A detailed review of the most frequently used machine learning algorithms is provided in the next section.

The aforementioned ML methods are benchmarked against the reference obtained via the subject's self-reported assessment form or a psychometric questionnaire. The commonly used questionnaires are the Perceived Stress Scale (PSS), Stress Response Inventory (SRI)

[50], Holmes and Rahe Stress Inventory (Life Events), and COPE inventory. These techniques obtain emotional, behavioural, and cognitive stress responses, and they are used as ground truth. Ground truth is a reference or baseline value that helps in differentiating the stress state of the subject from a non-stress state. This is valuable for developing classification models because it makes it much easier to objectively compare two different states. The drawback of using questionnaires as ground truth is that they are designed for dedicated events and are highly subjective. Moreover, these conventional questionnaires rely on events that occurred previously; thus, they lack generalization.

There is vast literature available that shows the association of a higher heart rate with stress. This change in heart rate changes the blood flow within the body. Heart rate and heart rate variability can be monitored using an electrocardiogram (ECG) signal, while the change in blood flow can be measured through blood volume pulse (BVP), derived from a photoplethysmography (PPG) signal [51], [52]. Some studies have discussed sweat released during stress, which changes the skin conductance measured by the electrodermal activity (EDA) measurement device [48]. Muscle tension is also related to stress and is monitored using electromyography (EMG) [53], [54]. Sometimes, chronic stress can also cause mild fever (between 99 and 100 °F), as well as anxiety and restlessness. Thus, skin temperature (ST) and accelerometer (ACC) sensors can also help in detecting stress [55]–[57].

During the stress period, the body prepares itself for a 'fight or flight' response and catecholamines are released within the body to cope with stress. Thus, measuring plasma catecholamines can also help in the assessment of stress [58]. The role of arginine vasopressin (AVP) during the acute stress response has also been widely discussed in the literature. Copeptin is considered a stable biomarker of AVP release. Copeptin increases significantly with the increase in cortisol, prolactin, and adrenocorticotropic hormones, which are directly related to the stress response of the human body. Therefore, monitoring the level of copeptin and prolactin hormones in the blood can help detect stress [59]. Alpha-amylase is one of the major salivary enzymes and is secreted in saliva in response to psychological stressors [60]. Cortisol is a primary stress hormone released in the bloodstream during stress and causes an increase in glucose levels [61]. Thus, monitoring cortisol levels in the blood also helps to monitor stress levels. All the above-mentioned

hormones are measured using different available enzyme-linked immunosorbent assay (ELISA) kits.

There is a considerable body of literature available on stress monitoring using physiological or biochemical responses of the human body. However, there is no consensus on the sensitivity and specificity of these biophysiological and biochemical responses for stress identification. This sensitivity and specificity may be associated with the sensitivity of the response to stress, the sensitivity of sensors, the type of stimulators, the sample size in the study, the design of the experiment, and other variables [62]. Nevertheless, the sensitivity and specificity of measurable responses to stress are critical for long-term monitoring of stress in the context of preventive and personalized care. This study presents an up-to-date review of the literature on biophysiochemical indicators of stress with a focus on connected and preventive healthcare. In this study, we provide summaries of the available literature on stress indicators and a critical review addressing the sensitivity and specificity of the sensors, as well as indicators of stress, in the discussion section.

Table 2.1 shows the bio-signals that are mostly used for stress monitoring, which include biophysical and biochemical markers. Figure 2.1 shows the placement of different biosensing devices used for stress monitoring, while Figure 2.2 presents a list of biophysiological and biochemical indicators of stress.

Table 2.1 Most used bio-signals for stress monitoring.

S. No.	Bio-signals	Ref	Units*
1	Skin conductance (also known as electrodermal Activity, EDA)	[48], [63]–[66]	μS
2	Electrocardiography (ECG)	[27], [65], [67], [68]	mV
3	Electroencephalograph (EEG)	[69], [70]	μV
4	Respiration rate (Resp), blood pressure (BP), and blood volume pulse (BVP) using photoplethysmography (PPG)	[63], [71]–[73]	Breaths/min, mmHg and mV
5	Skin temperature (ST)	[63], [70], [74]	°C
7	Electromyography (EMG)	[65], [67], [71]	μV
8	Plasma catecholamines, copeptin and prolactin, steroids samples	[75]–[77]	mcg/24-h, ng/mol, ng,
9	α-amylase samples	[75], [77]	μL
10	Cortisol samples	[78]–[80]	nmol/L
* Here, μS *is* micro siemens, mV is millivolts, μV is microvolts, °C is degrees centigrade, mcg is micrograms, ng is nanograms, μL is microliters, and nmol/L is nanomoles per litre.			

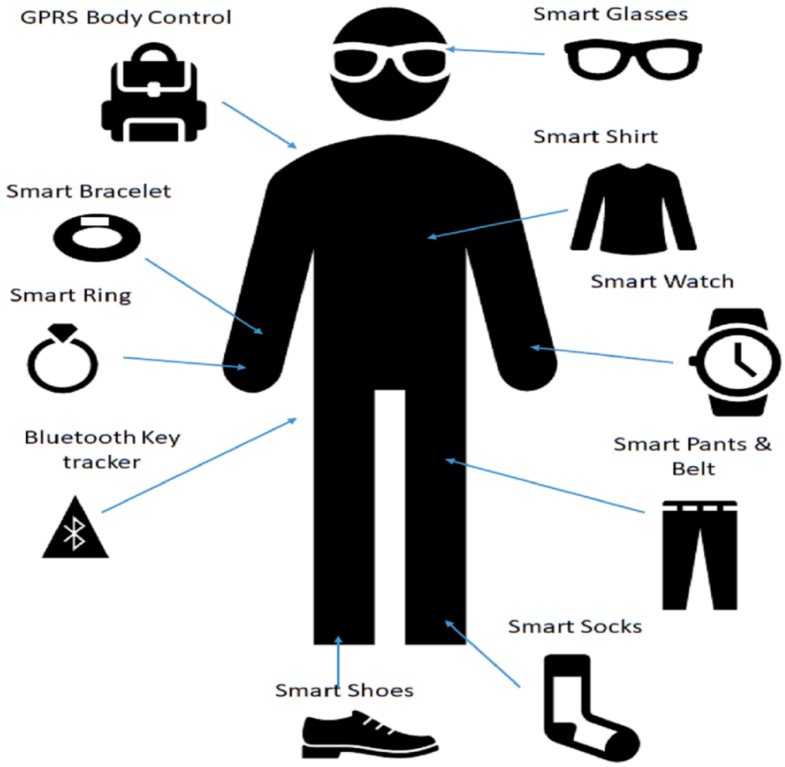


Figure 2.1 Possible sites for placement of smart sensors.

2.1.1.1 Stress Assessment Tests

Stress can be assessed subjectively via structured questionnaires (and self-reporting forms), which is also standard clinical practice, or objectively by measuring various responses of the body to stress [81]. The most used tools in clinical stress assessment are based on self-reported questionnaires (for example, Cohen's Perceived Stress Scale, PSS) or self-reported visual scales (for example, Visual Analogue Scale for Stress, VASS). Biomedical researchers are more interested in using biochemical markers for detecting stress, such as cortisol and α-amylase [17], and they trigger a stress state in the subjects under test via the Trier Social Stress Test (TSST) [82]. On the other hand, there are various studies available that assess stress by measuring physiological signals of the body in response to stress [18]. Details of some commonly used stress assessment tests and questionnaires are presented in Table 2.2.

Table 2.2 Stress assessment test and brief details.

Test Name	Stress Assessment Method
Mental Arithmetic Test	Participants are asked to solve arithmetic questions (subtraction, multiplication) within a time frame to induce stress.
Trier Social Stress Test (TSST)	Requires participants to deliver a speech on any given topic in a short time to prepare. After the speech, the participants are also asked to perform some verbal calculations. Both tasks are performed in the presence of an evaluating audience.
Stroop Test	Participants are shown the names of different colours written in various font colours and are asked to tell the font colour rather than reading the word.
Perceived Stress Scale (PSS)	Participants fill out the questionnaire by rating the questions about their feelings and thoughts. The total score varies from 0 (no stress) to 40 (highest stress).
Visual Analogue Scale for Stress (VASS)	In this test, participants are asked different questions during a given task or experiment to rate their stress on the scale as no stress, moderate stress, or high stress rather than a numerical value. Most of the time, a 5-point (smiley) scale is used for stress assessment.
Stress Response Inventory (SRI)	The Stress Response Inventory consists of 39 questions scored in the range of 0 to 156. These questions are categorized into 7 factors, i.e., tension, fatigue, depression, aggression, anger, somatization, and frustration. A high score means high perceived stress.
COPE Inventory	There are 28 items of self-reporting questions designed to measure the efficiency of how participants cope with a stressful event. A score is given to each question on a scale of 1 (low stress) to 4 (high stress). The total scoring determines the participant's stress coping style, i.e., approach coping or avoidant coping.

Holmes and Rahe Stress Inventory	Measures the amount of stress incurred within the past year. Participants select events that occurred in their life from the 43 life stress-related events. Each event has different scores. Participants accumulating a score greater than 300 are at a higher risk of illness, while a score lower than 150 suggests a slight risk of illness.
State-Trait Anxiety Inventory (STAI)	Participants validate 20 questions that measure the state and trait of anxiety. Participants respond to the questions on a scale of 1 to 4, where 1 denotes the least stress while 4 denotes a high-stress state.
Montreal Imaging Stress Task (MIST)	MIST consists of three stages, i.e., rest, control, and experiment. In the resting stage, the participant looks at the static screen of the computer. In the control stage, the participant is asked to solve a series of mathematical problems, while in the experiment stage, some difficult and time-constrained arithmetic tasks are given to induce high stress.
Perceived Stress Questionnaire (PSQ)	Participants fill out two types of questionnaires consisting of 30 questions; the first questionnaire has questions about stressful experiences and feelings over the last two years, while the second one has questions about stress during the last month. Participants must score each question from 1 (no stress) to 4 (stressed).

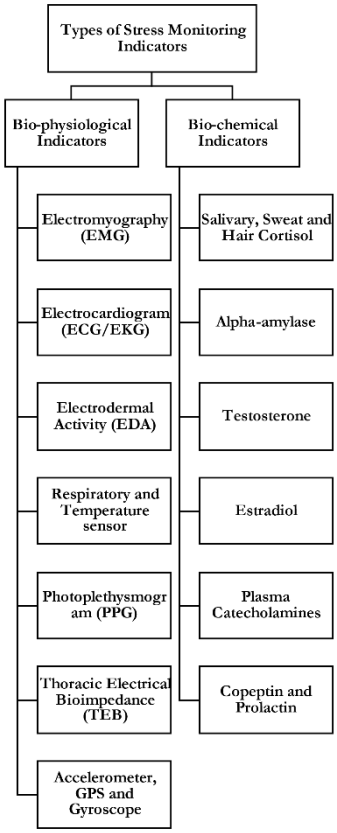


Figure 2.2 Types of stress indicators: physiological (left column) and biochemical (right column).

2.1.2 Biophysiological Indicators

In the literature, most of the studies related the increased heart rate and abrupt changes in the skin conductance, respiratory rate, and blood pressure with stress and used them as ground truth. It can be observed that some studies measured the same signal(s) and used the same stressor(s) and the same ground truth (method), but the reported results showed significant differences. Figure 2.3 shows the reported prediction accuracies versus the prediction algorithm using different stress indicators/markers. For prediction, most of the authors used SVM (with different kernels) as a predictor, while the highest prediction accuracy was achieved using an artificial neural network predictor. Table 2.1 summarizes the different biophysiological indicators-based studies and their conclusions.

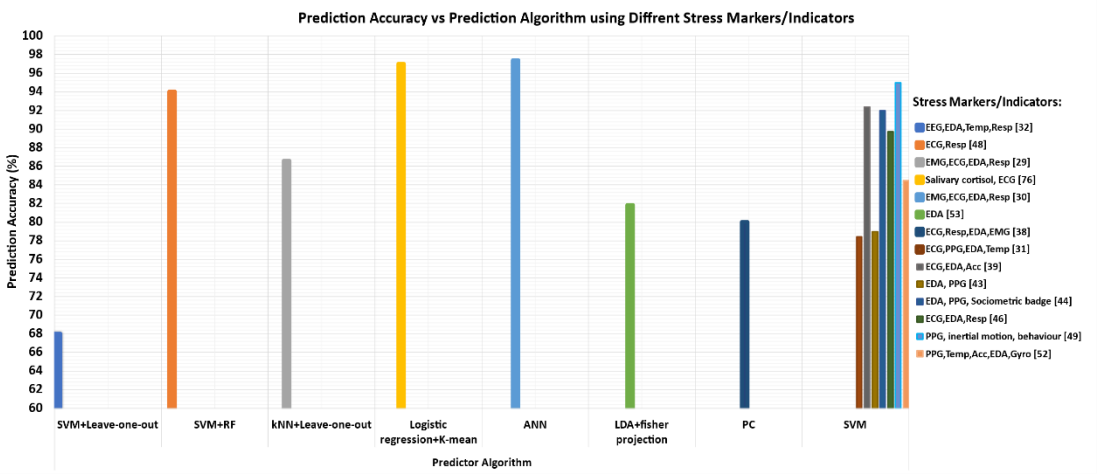


Figure 2.3 Reported prediction accuracies of various prediction algorithms using different stress indicators/markers. In the figure, SVM is a support vector machine, RF is random forest, kNN is a k-nearest neighbour, ANN is an artificial neural network, LDA is a linear discriminator analysis, and PC is a principal component analysis. *Reference number in legend are from paper [27].

Table 2.1 Biophysiological parameters (sorted year-wise).

Ref	Year	Signals	No. of subjects	Stressors	Conclusion	Features extracting Methodology/ Classifier
[66]	2020	Galvanic skin response (GSR)	10	Predefined PYSIONET dataset and driving on the highway, in the city	For the classification, the authors designed a binary logistic regression model and achieved an overall accuracy of 85.3% on data from PYSIONET, while they achieved 83.2% accuracy on the validation data, analysed through cross-validation. The authors also proposed that their developed model can be embedded in existing wearable GSR sensor devices	GSR from the Empatica E4 watch and binary logistic regression-based classifier along with cross-validation technique

					and, thus, can enable the detecting and monitoring the driving stress in real-time.	
[83]	2019	PPG, EDA, GSR, and ACC	21	Summer camp (training, the contest, and free day)	If individual data of each person are enough to design a person-based model, it should be developed; otherwise, each person should be clustered in accordance with their behaviour under stress and then a clustering model should be developed to increase the classification accuracy of the general model.	Clustering models such as kNN
[84]	2019	EDA	1	Driving on the highway, in the city	From the experiment, the authors concluded that their classification results indicate that road type and traffic conditions are important features related to driving stress. The authors reported an accuracy of 80.3% with a sensitivity of 85%, a specificity of 78%, and positive predictivity of 70% while using logistic regression as a classifier.	EDA from Empatica E4 watch and logistic regression-based classifier
[85]	2018	ECG	1	Daily life stress	The variation of the stress index shows high concordance with the work schedule of the subject and, thus, can provide an acceptable solution for the comparison of stress levels of different individuals.	By combining time- and frequency-domain nonlinear features
[73]	2018	PPG, Temp, ACC, and EDA	28	City Car Driving simulator	An accuracy of 68.31% for four states (i.e., normal, stressed, drowsiness, and fatigue) and an accuracy of 84.46% for three states (i.e., normal, stressed, and drowsiness or fatigue) classification.	Used pulse intervals and compared their values with standard pulse interval values. Winner-takes-all (WTA) and max-wins voting (MWV) methods were used along with an SVM classifier for classification
[86]	2018	EDA only (ECG, EMG, and respiratory)	11	Driving on the highway, in the city	After using Fisher projection and linear discriminant analysis (LDA) on the data collected from the dataset, the authors claimed to achieve a classification accuracy of 81.82%.	Fisher projection and linear discriminant analysis (LDA)

[65]	2017	ECG, EDA, and respiratory	14	Real driving environment	Using a full feature set, SVM with linear kernel gave the highest inter-drive classification precision. For the cross-drive level, SVM with RBF kernel gave a precision score of 89.7%.	SVM with linear kernel, SVM with RBF kernel
[48]	2017	EDA, ST, ACC, and PPG	5	Randomly generated equations (solved verbally)	Without contextual information, the stress detection was not in the range of acceptable accuracy, while, when they included the context information, the detection F-score jumped to 0.9 from 0.47 and precision increased to 95% from 7%.	Time-domain features such as mean inter-beat interval (IBI) of a sample, standard deviation, square root of the mean of the squares of the differences between adjacent IBI samples, and the percentage of the differences between adjacent IBI samples that are greater than x ms (x = 20, 50, 70) along with SVM classifier
[27]	2017	ECG and respiratory	39	Montreal Imaging Stress Task (MIST)	An accuracy of 84% using random forest features and SVM classifier in discriminating three stages of stress, while, for binary classification, i.e., rest and stress, they achieved an accuracy of 94%.	Random forest features and SVM classifier
[87]	2017	PPG and inertial motion and driver behaviour	28	Euro truck driving simulator version 2	Sequential feature selection with RBF kernel SVM classifier was able to achieve a classification accuracy of 95%, which shows the suitability of their glove as the driver's stress detection device.	RBF kernel-SVM classifier
[88]	2016	EDA, PPG, and sociometric badge for recording	18	STAI (State-Trait Anxiety Inventory) and TSST	Achieved higher accuracy, i.e., 92%, with SVM (RBF kernel) classifier, as compared to linear kernel SVM (80%), AdaBoost (67%) and KNN (62%), when using a selected set of features.	SVM (RBF kernel) classifier, linear kernel SVM, AdaBoost, and KNN
[72]	2015	EDA and PPG	5	Trier Scope Stress Test (TSST)	Detected stress of each participant with an average accuracy of 78.98%, i.e., combining all five participants' stress detection accuracies using SVM.	Support vector machine (SVM) classifier

[89]	2015	ECG and TEB measurements	40	Films, and games based on the addition	The measurement showed high potential in the use of ECG and skin activation (TEB) signals for the detection of long periods of stress or sudden increment in mental work overload or emotional responses of the people. With the MLP classifier, authors achieved an error rate of 21.23%, 4.77%, and 32.33% for activity identification, emotional state, and mental activity, respectively.	Low-pass filtering and decimated intermediate frequency-based algorithms along with multilayer perceptron classifier
[90]	2014	ECG, respiratory, body temp, GSR	10	Hajj pilgrimage	During sleep, the activity of the upper body and the duration of sleep contributed most to the detection of stress. The body temperature can be neglected as it did not contribute anything. The stress state was classified with an accuracy of 73% using SVM as a classifier.	Support vector machine (SVM) classifier
[91]	2013	ECG, EMG, GSR, and ST (only concern on ECG and HRV)	60	Stroop word–colour test	The optimum features of ECG were extracted through a fast Fourier transform (FFT). Accuracy of 91.66% and 94.66% was achieved using probabilistic neural network (PNN) and k-nearest neighbour (kNN) classifier, respectively.	Fourier transform (FFT) along with PNN and kNN classifier
[64]	2012	ECG, EDA, and ACC	20	Stroop color test and mental arithmetic problems based on Montreal Imaging Stress Task (MIST)	The inclusion of accelerometer data improved the stress detection process (92.4%) in a mobile environment.	Decision tree classifier, 10-fold validation, and least complex classifier
[71]	2011	ECG, respiratory, EDA, and EMG	18	Perceived stress scale (PSS) questionnaire	80% accuracy indicates the suitability of used features for the detection of stress in a subject.	Principal component analysis (PCA) technique
[70]	2010	ECG, PPG, EDA, and ST	22	Public speaking, mathematics, mental, social, and physical challenges	The SVM-based model detects stress with high precision and recall rate (68% accuracy), especially when they used personalized information with SVM.	Support vector machine (SVM) classifier-based model

[92]	2010	EEG, EDA, PPG, and respiratory	15	International Affective Picture System (IAPS)	Characteristics of the EEG signal were extracted by wavelet coefficients and Higuchi's algorithm, as well as correlation dimension. An accuracy of 82.7% using the Elman classifier was achieved.	Wavelet coefficients and Higuchi's algorithm, as well as correlation dimension with the Elman classifier
[93]	2009	ECG, ACC, GPS, EDA, and respiratory	3	Mental arithmetic test, Stroop colour-word test	The PDM algorithm shows lower inter-subject variance and interestingly showed some comparable performance between and within subjects, where PSD performance decreased when used between subjects.	Principle dynamic modes (PDM) algorithm
[68]	2004	ECG, EMG, EDA, and respiratory	24	Some audio, visual, and cognitive stimuli	Achieved a classification accuracy of 97.4% using data intervals of 5 min and found the highest correlation between heart rate and skin conductance metrics.	Artificial neural networks (ANN)
[63]	2004	ECG, PPG, EDA, and ST	50	Rest, highway drive, and city drive	Achieved a recognition rate of 78.4% for three emotional and 61.8% for four emotional states.	Support vector machine (SVM) classifier
[67]	2000	ECG, EMG, EDA, and respiratory	10	Garage exit, city road, toll booth, highway driving, ramp turnaround, two-lane merge, bridge crossing, and entering the garage	The combination of all features from the four types of sensors had an overall accuracy of 86.6%.	K-nearest neighbours (k-NN) classifier

2.1.3 Biochemical Indicators

The biochemical indicators of stress provide better detection and monitoring of stress, but most of these indicators can only be measured invasively. If measured noninvasively, as in some cases, the extraction of the required hormone from the sample collected takes time and, thus, cannot be used in real-time stress-monitoring devices. There is also scope for integrating some of the biochemical indicators, such as cortisol level measurement, with physiological indicators, such as heart rate, respiratory rate, and activity monitoring, to develop a more robust and accurate device for stress monitoring. Table 2.2 summarizes the biochemical indicators-based studies and their conclusions.

Table 2.2 Biochemical parameters (sorted year-wise).

Ref	Year	Signals	No. of subjects	Stressors	Conclusion
[77]	2019	ST, HR, pulse wave (EDA, ECG, PPG),	40	Trier Social Stress Test (TSST)	There was no clear correlation between physiological parameters and perceived stress

		copeptin, prolactin (blood), cortisol, and alpha-amylase (saliva)			levels. Moreover, alpha-amylase peak level time is 10 to 15 min after stress onset and, thus, should be measured within that time frame. Alpha-amylase and cortisol were measured in the morning (at that time, intra-individual variability is high).
[94]	2019	PPG and endocrine (salivary) cortisol	32	Childhood Trauma Questionnaire (CTQ)	No significant effect of early life stress on heart rate (autonomic indicator) and salivary cortisol (endocrine indicator), but the authors suggested that heart rate is a better indicator (of stress) than salivary cortisol as it is more sensitive to individual stress reactivity than salivary cortisol.
[95]	2019	Oxy-haemoglobin (oxy-Hb)	4	Decision-making and memory recall tests	Whenever high stress occurred, the average difference value of oxy-haemoglobin (oxy-Hb) increased.
[96]	2018	Biochemical (salivary cortisol) and physiological (HRV measures) domains	30	Academic final examination, Psychological Stress Response Inventory	The salivary cortisol levels were negatively correlated with the HRV indicator of parasympathetic activity, while they were positively related to the HRV indicator of sympathetic activity. The results also showed that the value of the mental stress index (MSI) was very sensitive to acute stress and could predict stress with an accuracy of 97%.
[76]	2016	Steroid hormones in hair	40	Perceived Stress Questionnaire (PSQ)	The concentrations of steroids in the hair were a decisive predictor of the increased long-term HPA axis. Furthermore, this biomarker could capture stress even after burdening events or any physical activity was finished.
[97]	2013	Sweat and saliva samples	17	Intense exercise	Intense exercise could increase the concentration of cortisol in hair, which was not decreased by hair washing.
[98]	2012	Hair cortisol	-	Daily life stress (3 months)	Identified some gaps in the currently available literature: firstly, to clarify the mechanism underlying cortisol incorporation into hair, and, secondly, to determine the factors that cause variation in hair cortisol such as the effect of hair washing.
[75]	2010	Salivary alpha-amylase, plasma catecholamines, BP, and HR	33	College academic final exam	The salivary alpha-amylase level changed significantly, but a partial correlation was found, statistically, between salivary alpha-amylase and blood pressure, heart rate, and plasma catecholamines.
[80]	2008	Cortisol level and EMG	16	Arithmetic problems, tasteless gum	Fast chewing had a greater effect on the stress release than a slow chewing rate, while the integration of EMG signals did not show any major difference in the 3 chewing rates.

[79]	2004	Salivary cortisol (in the blood)	12	The psychosocial stress test, Trier social test (TSST), and reading test	The level of salivary alpha-amylase was significantly lower in smoking females than non-smokers, while it was higher in smoking males than in non-smokers. Identified that the production of salivary cortisol affected the association of norepinephrine and amylases. Activation of the parasympathetic nervous system decreased the overall saliva production and volume. Therefore, the volume of saliva and amylase levels should be measured relative to the saliva produced.
[99]	2004	BP, pulse rate (PPG), and saliva	58	Vertical Visual Analogue Scale (V-VAS) and State-Trait Anxiety Inventory	There was a positive correlation between salivary cortisol and 24 h blood pressure.
[78]	2003	Total testosterone, free testosterone, oestradiol, androstenedione, and cortisol	30	Early follicular phase	Women with low levels of testosterone and androstenedione presented less competitive feelings. Moreover, oestradiol levels were unrelated to any competitive feeling.

2.1.4 Conclusion - Review of Physiological and Biochemical Indicators

Table 2.1 and Table 2.2 summarized the types of stressors used in each study reviewed in this study, along with the types of bio-signals collected by the authors to measure and monitor stress. It can be observed that some studies measured the same signals and used the same stressors, but the reported results showed significant differences. For example, the measured signals were the same in [67], [68], [71]; however, the classification accuracy achieved by each study is different (highest: 97.4%, lowest: 80%). Furthermore, the studies [63] and [70] also used the same signal for stress detection but reported a classification accuracy of 78.4% and 68%, respectively. In [48], the authors concluded that the physiological signals alone cannot provide acceptable accuracy for stress detection and that contextual information should also be included during the data collection. This was also evident in the results by [64], with improved accuracy of 92.4% when using contextual information. On the other hand, [86] achieved an accuracy of 81.82% by using only EDA (skin conductance), and, similarly, [65] reported an accuracy of 89.7% by using four physiological parameters and no context information.

The possible reasons for this variation in prediction accuracies can be due to the variations in the experimental setup (real and controlled environment), use of different features extracted from the raw data (time- and frequency-domain features), different lengths of

recorded data (varied from 5 min to 1.5 h), different placement of sensors (chest worn, wrist-worn, and foot-worn), and the different number of subjects recruited for the experiment.

Despite the abovementioned high classification accuracies using biophysiological parameters, some studies, for example, [77] and [94], reported that there is no clear correlation between perceived stress and biophysiological parameters. Additionally, these studies suggested that biochemical markers of stress should be considered when designing a stress monitoring system. Interestingly, in [94], the authors suggested that biophysiological markers of stress (heart rate) can be a better indicator than biochemical markers (salivary cortisol, a most frequently used biochemical indicator of stress). Contrary to this, [75] and [99] suggested a positive or partial correlation of salivary cortisol with physiological stress indicators (i.e., heart rate, heart rate variability, respiratory rate).

Reported accuracies are collected and plotted in Figure 2.3. It is important to note that there are different methods of feature extraction from a raw signal, as well as different ways of calculating accuracies. Commonly used tools, along with accuracy, to evaluate the performance of the different indicators and classifiers include confusion matrix, specificity, sensitivity, recall, f-score, the area under the curve, positive predictive value, negative predictive value, and likelihood ratio (positive and negative), as described in [100]–[102]. In the literature reviewed, authors might have used different matrices for stress-relevant feature extraction and classification; thus, reported accuracies may not be comparable.

The variable and contradictory evidence in the literature on the use of either physiological or biochemical stress markers leads to a conclusion that neither of these biomarkers in isolation can provide sufficient means of monitoring stress. Therefore, a combination of physiological and chemical stress biomarkers, with contextual information, can be a more reliable solution for stress monitoring. A multisensory platform with data-driven personal insights can help track and intervene in cases of stress in the high-risk population. There is still a need for a novel, more sensitive, and more specific stress monitoring system that should be easily implemented and adopted by medical professionals and home-based consumers.

2.2 Review of different machine Learning Algorithm used for Stress Classification

Cardiovascular activities are directly related to the response of a body in a stressed condition. Stress, based on its intensity, can be divided into two types i.e., Acute stress (short-term stress) and Chronic stress (long-term stress). Repeated acute stress and continuous chronic stress may play a vital role in inflammation in the circulatory system and thus leads to a heart attack or a stroke. In this study, we have reviewed commonly used machine learning classification techniques applied to different stress-indicating parameters used in stress monitoring devices. These parameters include Photoplethysmography (PPG), Electrocardiograph (ECG), Electromyograph (EMG), Galvanic Skin Response (GSR), Heart Rate Variation (HRV), skin temperature, respiratory rate, Electroencephalograph (EEG) and salivary cortisol, used in stress monitoring devices. This study also provides a discussion on choosing a classifier, which depends upon several factors other than accuracy, such as the number of subjects involved in an experiment, type of signals processing and computational limitations.

2.2.1 Introduction

Stress can be of two types, Physical Stress and Mental Stress. Physical stress is often caused by a poor diet, sleep deprivation, too much work or may be due to illness. Mental stress is triggered due to worrying about the illness of a loved one, the death of close ones, retirement or money or being fired from work [3]. Generally, most of the stress comes from our daily responsibilities. Work pressure and obligations, which are mental and physical, are not always noticeable to us. In response to stress, incurred in daily life, our body automatically alters our blood pressure, respiration, heart rate, blood flow to muscles and metabolism. The response tries to help our body to react fast, yet efficiently, to a high-pressure situation [4].

Stress situation can become a threat to our well-being and health. If no adjustments are made in time to cope with its effects. It is very important to realize all the external events. It does not matter how we are perceiving those events. These events can cause stress and may cause you to feel 'out of control'.

Fatigue, headaches, sleeping problems, digestive problems, and muscle tension are some most common effects of stress. A long term and unmanaged stress may cause heart disease, high blood pressure, diabetes, and obesity [103]. Stress may also cause anxiety, restlessness, and inability to focus. These mental and physical changes may affect our weight loss progress as eating junk food can be used to cope with stressful situations [5]. Several people get addicted to tobacco or illegal drugs as they use it as a stress management mechanism. If the stress state period goes on for a very long time, it does increase the chances of having a heart attack, hypertension, or stroke in a person [40].

According to American Psychological Association (APA), stress is linked to 6 leading causes of death including heart disease, depression, anxiety disorder, diabetes etc. Centre for Disease Control and Prevention reported that 110 million people die every year as a direct result of stress i.e., 7 people in every 2 seconds.

One needs to realize that without proper monitoring and management of stress, the situation will get more difficult to contain. People are getting depressed, are very easily angered, and have started to withdraw from themselves. As stress is becoming one of the movers, the performance of people has declined. People think this economic recession will stay such as this and there is no way to fight it. All these situations can be avoided if people knew how they can fight and win from the effects of stress.

Recognition of a high-stress state is very essential. For this purpose, one can monitor stress using physiological indicators of stress such as increased heart rate, blood pressure, respiratory rate, sweaty hands, and fast pulse rate. Unfortunately, most people, cannot recognize these physical symptoms. Such people can use stress monitoring devices that will inform them in time about an increase in their stress level and thus they will be able to control it beforehand by either doing meditation or exercise. Literature suggests that the following parameters, individually or with different combinations, can be considered for stress monitoring and are discussed below:

- **PPG:** A Photoplethysmography (PPG) is a light-based plethysmogram which is used to detect changes in the volume of the blood running in the micro-vascular system of the body. Usually, a pulse oximeter is used to obtain PPG. A pulse oximeter illuminates skin using a light-emitting diode (LED) and calculates the change in

absorption of light through a photodiode. As the flow of blood to the skin is modulated with different types of other physiological systems, thus, we can use PPG to monitor hypovolemia, breathing and conditions of other circulatory systems [104]. For monitoring of the stress, a PPG signal obtained from a pulse oximeter is used to calculate the oxygen level within the body of the human, often called oxygen saturation. For a normal person, this oxygen saturation level is between 95 to 100%. A saturation level below 90% is considered abnormal and can lead to a clinical emergency [105]. The waveform of the PPG signal differs from patient to patient and to the location as well as the way the pulse oximeter is attached.

- **ECG:** Electrocardiography (ECG, also known as EKG) is a method used to record the electrical activities of the heart with respect to time by placing electrodes on the skin. These electrodes detect small electrical changes under the skin that are generated by depolarization and re-polarization of heart muscles in electrophysiologic patterns for each heartbeat. ECG is commonly used to detect cardiac problems. In stress monitoring applications, an ECG signal is used to calculate Heart Rate Variation (HRV). In a stress state, the heart rate varies significantly and thus can be monitored by using HRV.
- **EMG:** Electromyography (EMG) is a technique for recording and evaluating the electrical activities produced by muscles of the skeleton. An instrument called an electromyograph is used to generate a record called electromyogram. EMG is the recording of the muscle cells whenever these cells are activated electrically or neurologically. Medical abnormalities such as Polymyositis and Muscular dystrophy, recruitment order, and activation levels can be analysed using EMG signals. Moreover, we can also analyse the biomechanics of animal or human movement.
- **GSR:** Galvanic Skin Response (GSR) is used to measure skin's electrical conductance. The sympathetic nervous system can be triggered through strong emotion, which results in more sweat being released by sweat glands. The signal is obtained using two electrodes attached to two fingers of the same hand and is often used to monitor the quality of sleep.
- **EEG:** Electrophysiological method for monitoring and recording electrical activities of the brain is called electroencephalography (EEG). It is, most of the time, a non-invasive method with electrodes located on the scalp but sometimes we can use

invasive electrodes. Voltage fluctuations are measured through EEG which causes the ionic current in the neurons of the brain. These readings are analysed with respect to a period and diagnosis are made on spectral content or event-related potentials.

- **Respiratory Rate:** Respiration rate is defined as the number of times someone takes a breath in 60 seconds, mostly calculated by observing the rise and fall of the subject's chest. Measuring the respiration rate while the person is sleeping is the most difficult task and cannot be done using general lab devices. Respiratory rate can also be measured from blood volume pulses. There are two ways to do so; by calculating the time change in-between two successive heartbeats and the change in the amplitude of the blood volume. The infrared sensor on the ring takes samples at 250 Hz from arteries and capillaries of the finger and thus does not disturbs the sleeping subject [106]. The collected samples show inter-beat-interval, yielding data on heart rate variability, heart rate and respiratory rate.
- **Salivary Cortisol:** Cortisol is a steroid hormone, which is produced in the adrenal gland by zona fasciculata in response to stress. For a normal person, the concentration of salivary cortisol ranges from 10.2 to 27.3 with +/ 0.8 nmol/L in the morning and ranges from 2.2 to 4.1 with +/ 0.2 nmol/L at night [107]. A high level of cortisol circulating in the human body shows the sustained stressed condition of the human and may create an allostatic load. This allostatic load can cause various physical changes in the body. Changed levels of cortisol can be observed in form of mood disorders, anxiety disorders, illness, stomach pain, fear, and other physiological and psychological disorders [108]. Collection protocols and approved collection methods of salivary cortisol are defined by Salimetrics USA [109].

Stress can be triggered by using a questionnaire or can be caused due to physio- and sociological factors. Different classification algorithms are used to recognize stress states. In this study, we will focus more on different machine learning techniques as these algorithms are more accurate and popular as well as state-of-the-art ways of monitoring and recognizing stress in humans.

2.2.2 Inducing Stress using Questionnaire Methods

A questionnaire method is mostly used to measure mental stress. In this method, stress is triggered in the subject using some questions and recording their response time along with

PPG and EEG signals. These response signals are then fed to a neural network to classify the stress states into low, normal, and highly stressed states. Such an assessment of stress is often called a subjective assessment. Most of the research community uses a 'Stroop test' to measure the mental health of the subject. The Stroop test is a colour-naming activity, mostly designed on a computer as a game. In this game, subjects are asked to call the name of the colour irrespective of what is written with that colour. For example, the word BLUE is written with PURPLE colour, so the subject has to say PURPLE as an answer to this question. Stroop test designed by Nagananda *et al.* in [110] uses five colours; BLUE, YELLOW, GREEN, RED and PURPLE and classifies stress into low, medium, and high-stress levels using a simple neural network.

Besides the Stroop test, several research fellows designed their test questionnaires. Kallus *et al.* [111] designed a RESTQ that measured the frequency of stress and activities related to stress recovery. The authors designed five different versions of RESTQ based on the types of subjects one wants to use. RESTQ-Basic for general usage and had seven stress scales. RESTQ-Sport for athletes with five recovery scales. RESTQ- Coach for coaches, RESTQ-CA for adolescents and RESTQ- Work for the subject's work context. Every version had its time frame of three days/nights or seven days and nights. The output is indicated on a scale of 0 to 6 i.e., never to always, respectively.

Boynton *et al.* [112] presented a very interesting study about the selection, design, and development of a self-defined questionnaire. The authors argued that anyone can design a list of questions and print it but designing a well effective and generalized questionnaire needs creative imagination and careful planning. The authors also discussed different perspectives on a questionnaire that should be considered while designing or developing it. A new questionnaire often fails to provide high-quality generalized data, thus, whenever possible one should use previously validated questionnaires and rephrase them appropriately for their targeted audiences and the information they require. The authors conclude that a nicely explained and carefully designed questionnaire will always lead to improved response rates.

2.2.3 Review of Stress Classification Machine Learning Algorithms

Recently, the development of different machine learning algorithms has greatly helped to develop tools that assist doctors to support patient care and predicting any mental disorders. Machine learning techniques are widely used as a decision boundary-making tool in complex data analysis of health. Supervised machine learning algorithms produce general hypotheses from externally supplied labelled features. These hypotheses are then used to make predictions about new incoming features [113]. The literature included in this review is measured on the following criteria:

- Should be frequently used classification methods.
- Should have achieved good classification accuracy (>50%).
- Should be about the classification of physical and mental stress states.

The selection of a learning algorithm is a critical step. Usually, an algorithm is evaluated based on its number of correctly predicted outcomes over the total number of prediction attempts (which is called prediction accuracy). There are three ways to test a classifier. One is by splitting the training set into training and evaluation sets. The ratio of the split should be at least 70% and 30%, respectively. Cross-validation is the second option to test the performance of the classifier. Here training set is split into mutually exclusive and equal-sized subsets. A classifier is then, for every subset, trained on the union of all subsets. An error rate of every subset is calculated, and the average error rate determines the classifier's performance. There is a special type of cross-validation called Leave-one-out.

Computationally this method of cross-validation is expensive but is used whenever we require greater accuracy in terms of an error rate of a classifier. Error rate does depend upon the number of parameters such as the size of the training set, a dimension of the problem, hyper-parameters tuning and use of relative features of the problem. The third and final way to measure the performance of the classifier is a statistical comparison of the classifier's accuracies when trained on specific datasets [114].

Machine learning methods can be divided into four major types: Logic-based Algorithms, Perceptron-based Methods, Statistical learning Techniques and Support Vector Machines (SVM) [115]–[118]. Each type of algorithm has different sub-learning algorithms. Some of these learning algorithms are illustrated as follows:

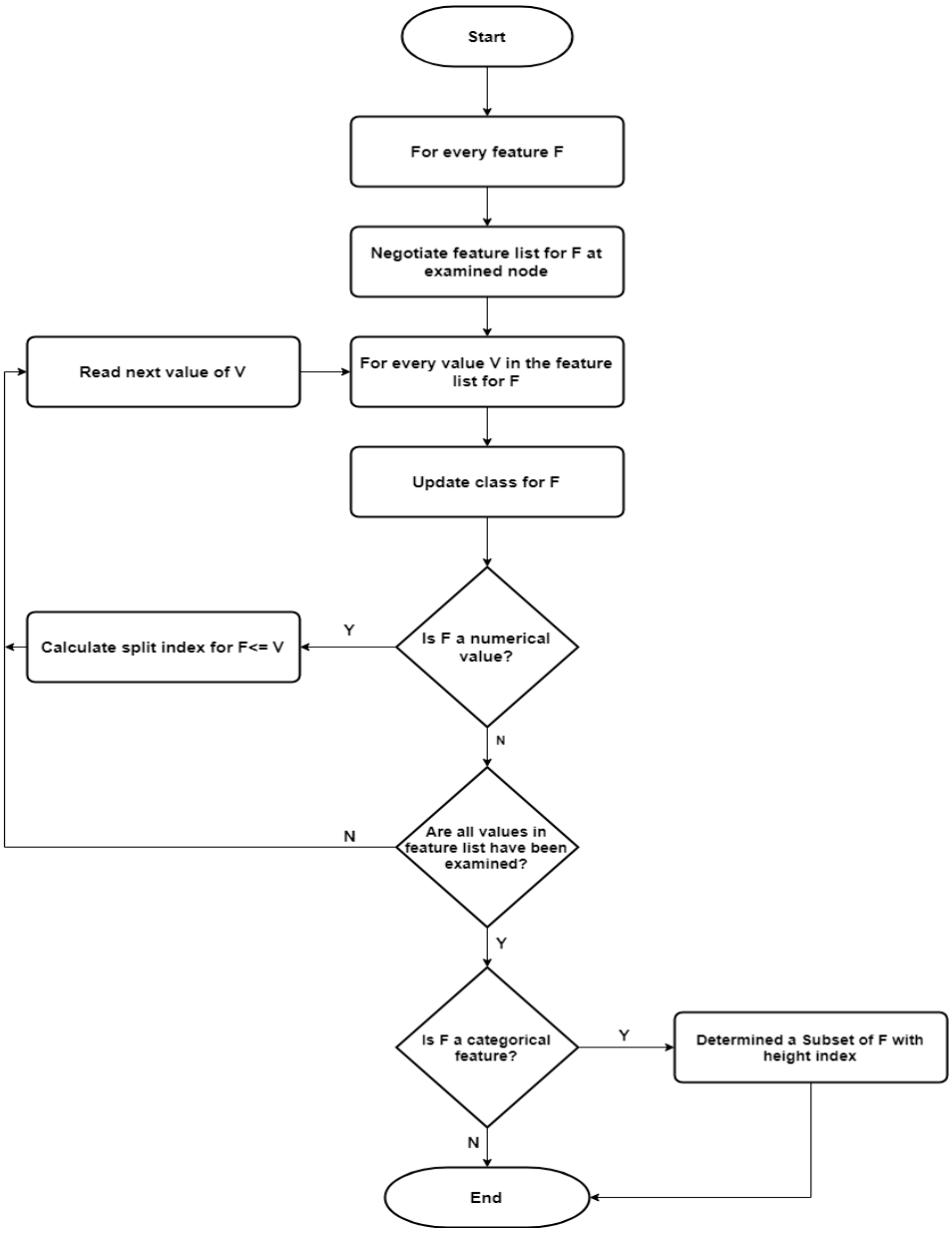


Figure 2.4 Flowchart of C.45 decision tree algorithm.

2.2.3.1 Decision Tree Classifier

The decision tree classifies by sorting input instances based on feature values [119]. Each node of the decision tree shows a classified feature from an input instance while each branch shows an assumed nodal value. Classification of instances starts from the root and is sorted depending on their feature values. The best divisor of input training data becomes the root node of the decision tree. Figure. 2.4 shows the steps involved in classification using a decision tree. This divide-and-conquer strategy is efficient and fast and can be an

important classifier if we have hundreds of thousands of input instances. A pseudo-code for designing a decision tree is presented in [120].

2.2.3.2 Artificial Neural Network Classifier

Artificial Neural Network (ANN) is used for classification whenever instances in the training dataset cannot be linearly separated [121], [122]. An overview of the Artificial Neural Network (ANN) is provided in [123]. ANN is created by the connection of many neurons (units) patterned as shown in Figure. 2.5.

Neurons of the network are divided into three layers: an input layer, which receives incoming information from the training dataset; an output layer, which gives us processed results (most of the time probabilities); and a hidden layer, which is in-between the input and output layer. If there is one-way communication between neurons of the network i.e., only from input to output, then the network is called a feed-forward network. The outcome of an ANN depends upon three factors: the architecture of the network, weights associated with each neuron in the network, and input along with activation functions used for each neuron.

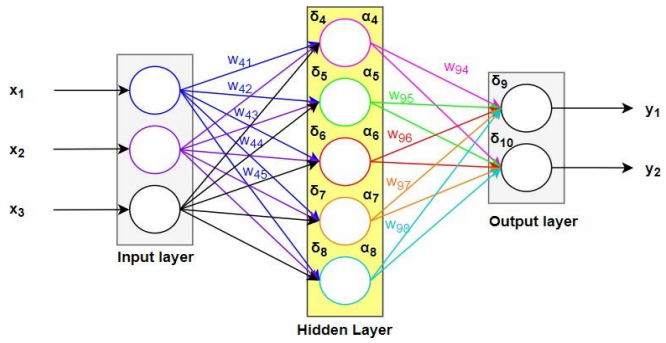


Figure 2.5 Feedforward Artificial Neural Network.

All weights are updated in such a way that it brings the outcome (a result of the classifier) nearer to the desired output. The most popular algorithm for updating weights is known as the Back Propagation (BP) algorithm and is defined as (see equation 2.1):

$$\Delta W_{ji} = \eta\, \delta_j\, O_i \quad (2.1)$$

Where W is the weight, δ_j is the next level neuron while O_i output of the previous node. η is the learning rate.

2.2.3.3 Bayesian Network Classifier

To represent probability relationships of input instances (features) in form of graphs, Bayesian Network is used. The structure of the Bayesian Network (BN) is a Directed Acyclic Graph (DAG) and there is a one-to-one correspondence between its nodes. Arcs in DAG show the influence of different features on each other. Conditional independence can be detected if there is no arc representing casual influences in-between features or are no descendants' nodes from this node (feature).

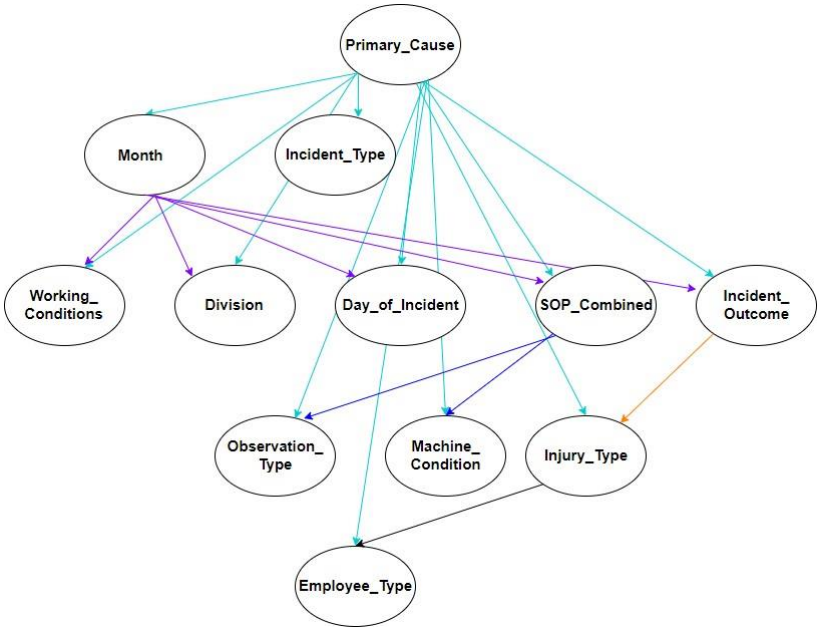


Figure 2.6 Structure of Bayesian Network.

Typically, learning a Bayesian network is a two-fold task: first, learn the DAG structure and how the BN structure is created using DAG, then determine BN parameters. Pseudo-code for training BN is presented in [120]. Figure. 2.6 shows the structure of a general Bayesian network.

2.2.3.4 Naive Bayesian Classifier

A naive Bayesian classifier is a very simpler form of a Bayesian network. Naive Bayesian (NB) has only one parent node in its DAGs, which is an unobserved node, and has many children nodes, representing observed nodes (Figure. 2.7). NB works with a strong assumption that all the child nodes are independent of their parent node and thus, one may say that Naïve Bayesian classifier is a type of estimator.

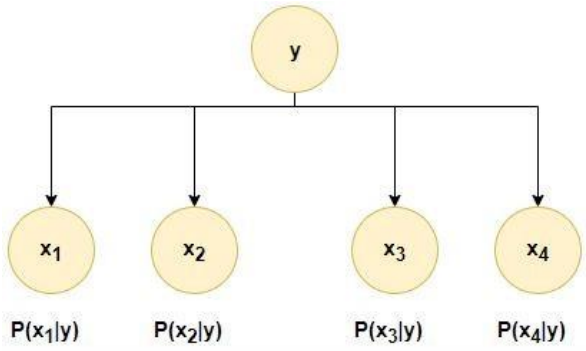


Figure 2.7 Structure of Naive Bayesian.

Mathematically,

$$R = \frac{P(i|X)}{P(j|X)} = \frac{P(i)P(X|i)}{P(i)P(X|j)} = \frac{P(i)\ \Pi\ P(X_r|i)}{P(j)\ \Pi\ P(X_r|j)} \qquad (2.2)$$

From equation 2.2, one can conclude that a larger probability value will indicate that the class label assigned to a feature (child node) is its actual label. The threshold for classification is as; if the value of R > 1 then predicts *i* otherwise predict *j*.

2.2.3.5 k-Nearest Neighbour Classifier

k-Nearest Neighbour (kNN) is one of the simplest instance-based learning algorithms. The working of kNN is as; it classifies all the proximity instances, in a database, into a single group and then when a new instance (feature) comes, the classifier observes the properties of the instance and places it into the closest matched group (nearest neighbour). Figure. 2.8 shows a flowchart for working the kNN classifier. For accurate classification, initializing a value to k is the most critical step in the kNN classifier.

2.2.3.6 Nearest Cluster Classifier

Nearest Cluster Classifier is a classification technique proposed to reduce the training set of k-Nearest Neighbour (kNN) and enhance its performance by using the clustering method, (proposed in [124]). The main goal of this method is to classify the given test samples according to their nearest neighbour tag. This algorithm first clusters the given (training) set into many different partitions. After getting these partitions, different clustering algorithms are executed to eliminate many clusters from those partitions. Then, the central label of previously produced clusters is calculated using the majority vote method between patterns of class labels in the cluster. A set of the most accurate clusters

are used as a training set for the final 1-NN classifier. In last, the class label of the upcoming test sample is calculated according to the class label of his nearest cluster centre [124].

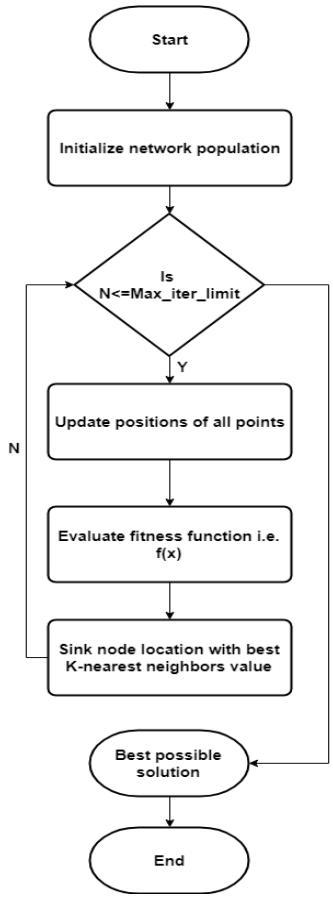


Figure 2.8 Flowchart of kNN classifier.

2.2.3.7 Learning Vector Quantization Classifier

The Learning Vector Quantization (LVQ) algorithm is an artificial neural network-based algorithm that enables us to choose the number of training features and learns what those features should look such as. LVQ is more likely a collection of codebook vectors. These vectors are consisting of a list of numbers that have the same input as well as output features as their training set. In this form of the neural network, every vector of the codebook is considered a neuron, each feature on the codebook vector is considered weight and the collection of vectors of the codebook makes a network [125]. The prediction procedure of LVQ is the same as that used in the kNN (k-Nearest Neighbour) algorithm. For prediction, all the vectors in the codebook are searched and the most similar K is found. Then the output is summarized for those selected K instances. By default, the value of K is

considered as 1 and the best-matched vectors in the codebook are called Best Matching Units (BMU). Figure. 2.9 shows the general architecture of the LVQ classifier.

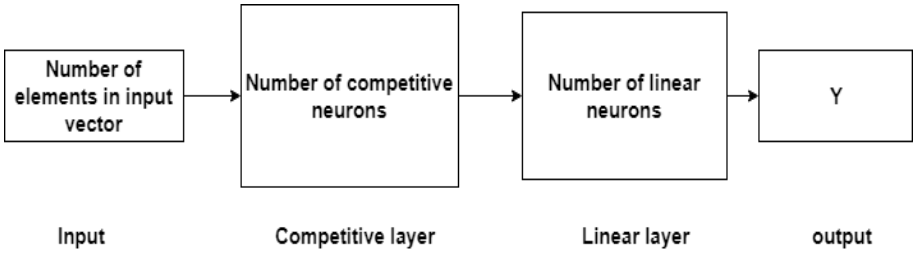


Figure 2.9 LVQ network architecture.

2.2.3.8 Kohonen Self-Organizing Map Classifier

The Kohonen Self-Organizing Map (KSOM) learning algorithm is originally presented in 1982 and uses vector quantization with similarities to patterns [126]. This algorithm is generally used for clustering problems having very complex datasets. Sizes of topology maps and learning are the two parameters that should be considered before designing a KSOM-based classifier.

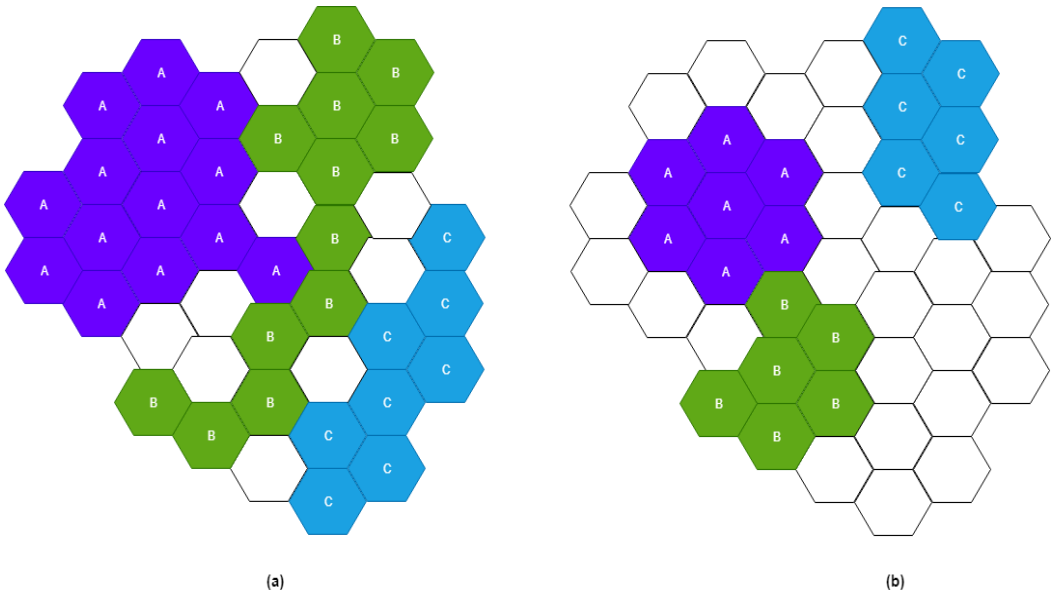


Figure 2.10 Result of (a) Original KSOM (b) Revised KSOM algorithm on Iris dataset.

The complete dataset is repeatedly trained using different-sized maps and a suitable size is found for accurate cluster classification. The Euclidean Distance model is used to calculate the distance between two nodes. Even grouping and clustering done by KSOM is quite

accurate but as every clustered group lies close to another group, it sometimes leads to overlapping of clusters and the problem of non-linear separation. To tackle this problem, [104] presents a variant of the original KSOM classification algorithm. They used a different approach for distance calculation and Figure. 2.10 shows a result comparison of the original and revised KSOM classifier. The steps used for designing the KSOM network for classification are shown in Figure. 2.11.

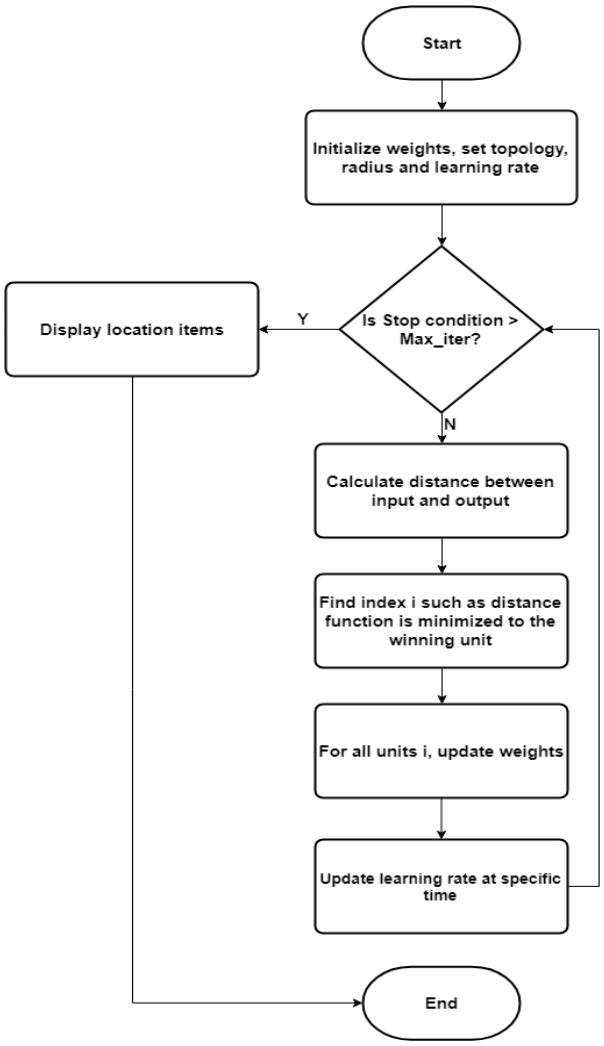


Figure 2.11 Steps in original KSOM learning algorithm.

2.2.3.9 Principal Component Analysis

Principal Component Analysis (PCA) is generally used to reduce the dimensions of a high-dimension dataset to a small subspace before using it to train any learning or classification algorithms. PCA transforms data into a new coordinate system having a low-dimension

subspace. In this coordinate system, the first axis represents the first principal component that represents the greatest aggregate of variance in a dataset [127]. From Figure. 2.12, we can see that by calculating two principal components, we can cover the variance of the whole dataset and all captured elements are independent of each other.

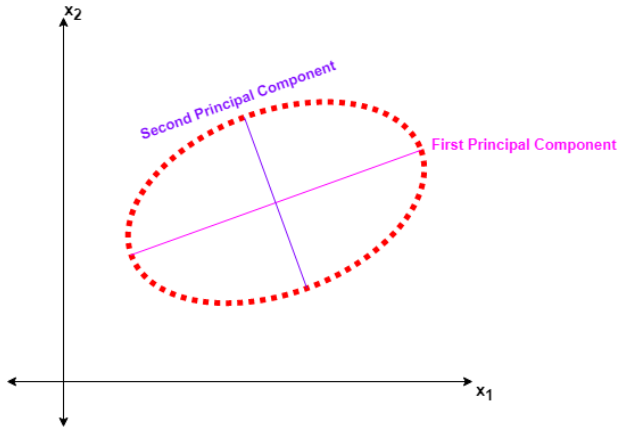


Figure 2.12 Two Principal Components of a dataset having two variables X_1 and X_2.

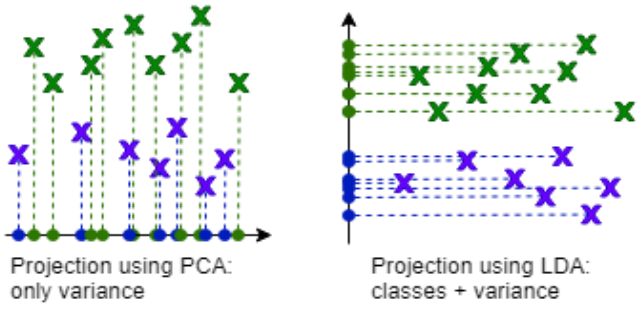


Figure 2.13 Comparison of PCA and LDA.

2.2.3.10 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is the most used dimensionality reduction algorithm. LDA is used during a pre-processing step in pattern classification and other machine-learning applications. LDA is calculated by following the steps in [128]; compute d-dimensional mean vectors for each class in a given dataset, compute the in-between-class scatter matrix and within-class scatter matrix, then compute Eigenvectors and their corresponding eigenvalues for both scatter matrices, select linear discriminants for the new feature subspace by sorting the eigenvectors in descending order using eigenvalues, and on the final step, transforming the samples onto new subspace by simply doing matrix

multiplication. Generally, LDA calculates an extra axis that shows a maximum separation between the two classes, as shown in Figure. 2.13.

2.2.3.11 Logistic Regression

Logistic Regression is one of the simplest machine learning algorithms mostly used for binary classification problems. This algorithm can be implemented easily and is used as a baseline algorithm for other two-class classifiers. There are three types of logistic regression algorithms [129]:

- If the targeted features have only two outcomes such as Spam or Not spam emails or Diabetic or Not diabetic, then this problem is solved by using binary logistic regression.
- If the targeted features have three or more nominal categories, then multinomial logistic regression. For example, the prediction of the type of clothing.
- If the targeted features have three or more ordinal categories, then ordinal logistic regression is such as rating any product between 1 to 5.

Logistic regression estimates and describes the relationship between independent and dependent binary features within a dataset. Figure. 2.14 shows a classification boundary calculated by a logistic regression algorithm.

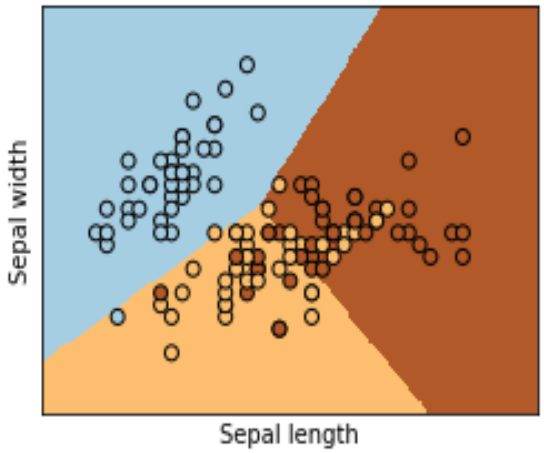


Figure 2.14 Classification using Logistic Regression on Iris dataset.

2.2.3.12 ZeroR and OneR classifier

ZeroR is the simplest classification technique, also known as the Zero Rule algorithm, that relies only on the target and ignores all predictors. This algorithm only predicts the majority

class of the given dataset. It is used for determining a baseline performance as a benchmark for other classification methods, although there is no predictability power in ZeroR. This algorithm constructs a frequency table for the target and selects its most frequent value [130].

OneR is an abbreviation of One Rule. It is a simple yet accurate classification technique that can produce one rule for every predictor present in the data. Then rule with the smallest error rate is selected and named 'one rule'. The rule for the predictor is created by the construction of a frequency table having two columns: target and its frequency. OneR has slightly less accurate as compared to state-of-the-art algorithms used for classification but is too easy for humans to interpret. Following is the pseudo-code for designing an oneR algorithm [131]:

1. For every predictor,
2. For every value of predictor, make the following rule.
    a. Count the appearance of each target value.
    b. Find the most frequently appearing class.
    c. Make the rule assignment of the selected class to the predictor value.
    d. Calculate the error rate for each predictor.
3. Choose the predictor with the smallest error rate.

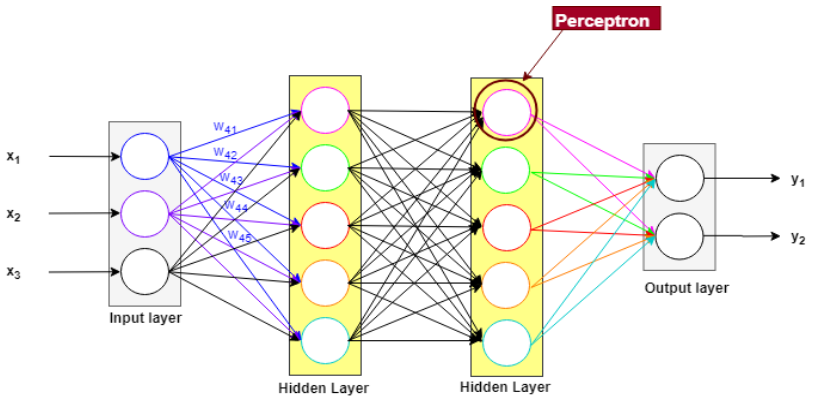


Figure 2.15 Structure of MLP Classifier.

2.2.3.13 Multi-Layer Perceptron Classifier

The multilayer perceptron is built using several neurons connected in different layers, mostly to solve a non-linear classification problem. Each perceptron is used to categorize

small linearly discrete regions of the input problem while the output of each perceptron is integrated to generate the final output. Neurons in multilayer perceptron are organised as the input layer, one or more hidden layers and the output layer, as shown in Figure. 2.15. The learning rule for this technique is called the backpropagation rule or generalised delta rule.

2.2.3.14 Genetic Algorithm

A Genetic Algorithm (GA) is an optimization technique that follows principles of natural selection and genetics. This technique is frequently used for finding an optimal or nearer optimal solution to complex problems that otherwise may take a long time to get solved. Genetic algorithms have a pool of possible solutions to a given problem. These solutions undergo mutation and recombination just as in natural genetics, making new children, and this process recurs again and again over different generations. Every individual or candidate solution is provided with a fitness value (label) depending upon their objective function. The fittest individual is given more chances to mate and produces fitter offspring (individual). Figure. 2.16 shows the general steps involved in the Genetic Algorithm.

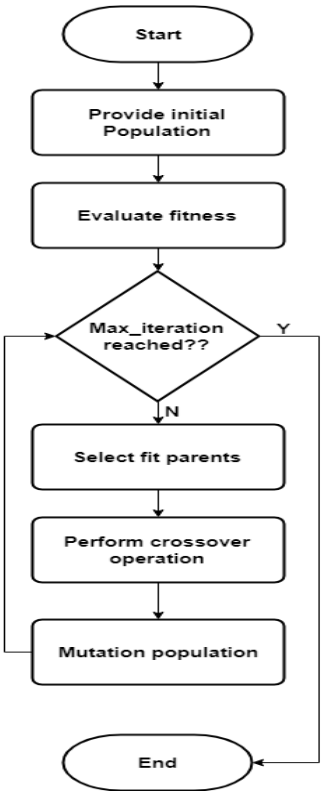


Figure 2.16 Steps involved in Genetic Algorithm.

2.2.3.15 Decision Forest

The decision Forest classification technique is a supervised learning technique. This classifier performs best if we must predict only two target classes at the output. Decision forest is typically implemented using the multiple decision trees model. Hyper-parameters of each decision tree model are tuned before training and testing the decision forest classifier. When combined, each decision tree votes for the popular output class, which is mostly used whenever we have an ensemble model.

Working off decision forest is as [132]; it creates many individual trees using the whole dataset, keeps starting point of each tree different (usually this selection is random), and every tree in the decision forest gives a non-normalized histogram of frequency labels at the output, these histograms are then summed up and normalized to get a probability of each label, and at last, the tree having highest confidence value is given higher weight in the last decision of the forest.

2.2.3.16 Decision Jungle

Decision Jungles is an extension of the decision forest. A jungle is consisting of an ensemble decision Directed Acyclic Graph (DAG). Hyper-parameters to be initialized are [133] resampling method; you should know how you want to create each tree (could either be begging or replicate technique), specify how to train your model; either you want to train your model using a single parameter or you want to train it via parameter range, the number of decision DAGs; define a maximum number of graphs which will be created, maximum depth of the decision DAGs; determine maximum depth allowed for each graph, a maximum width of the decision DAGs; determine maximum width allowed for each graph, the maximum number of iteration per decision DAGs; determine the number of iterations over given dataset to build each DAG.

Decision jungle allows branches of the trees to merge thus this algorithm can be generalized. Moreover, this algorithm can create non-linear boundaries during classification and is robust to noisy features.

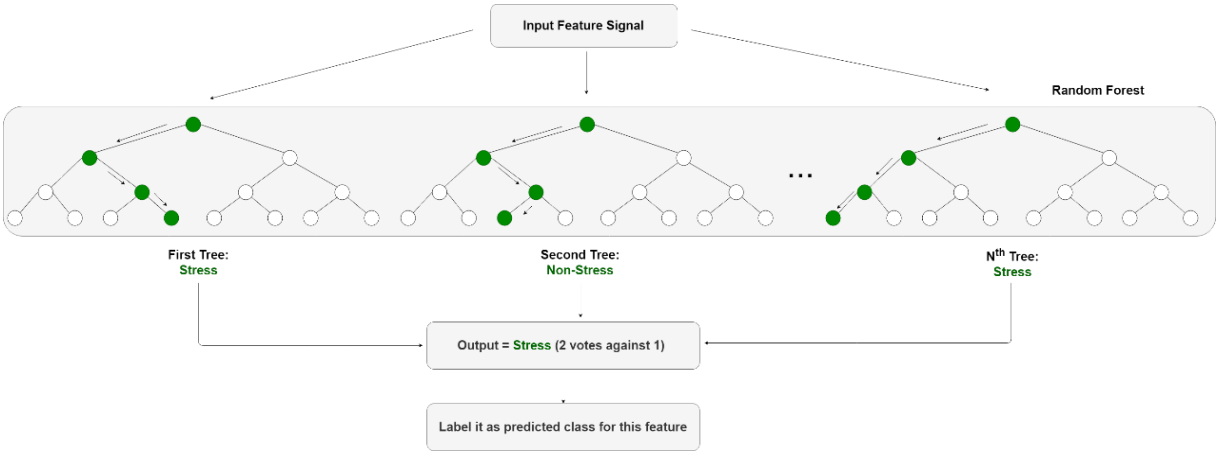


Figure 2.17 Simplified Random Forest Classification, classifying stress and non- stress.

2.2.3.17 Random Forest

Random Forest is a supervised machine learning algorithm. This algorithm creates random trees (forests) that are somewhat such as decision trees and the training method selected is always begging, as in begging we combine learning models linearly to increase the overall accuracy. While growing new trees, a random forest adds more randomness to the existing model. Instead of finding the most important target feature for node splitting, this algorithm searches for the best feature in the random subset of target features. In this way, we get wide diversity which in-return results in a better model. So, as a random forest only considers a random subset of features for splitting a node, we can make the trees of the model more random by using random thresholding of every feature rather than looking for the best threshold value [134]. Figure. 2.17 shows the classification of an input instance (feature signal) using a random forest classifier.

2.2.3.18 One vs All Multiclass Model

This model is used for the prediction of three or more three classes, specifically when the target outcome is dependent on categorical or continuous prediction variables. This algorithm can also be used for binary classification that needs multiple classes at the output [135]. The model is created using binary classifiers for each of the multiple class outputs. Every binary classifier is assigned to individual classes as a complement to all other classes present in the model. The final prediction is done by running all the binary classifiers and selecting the prediction having the highest probability score. A group of individual models is made, and results are then merged at the end, to create a single model which can predict

all classes. Therefore, we can choose any binary classifier as a basic classifier for the one-vs-all model. To configure such a model, one should know that there are no configurable parameters of this model. So, optimization should be done in the binary classifier which is to be provided as the input for building this model.

2.2.3.19 Ada-boost

Boosting refers to a group of techniques that creates a strong classifier using many weak classifiers. To find a weak classifier, a different machine learning-based algorithm having varied distribution is used. Each learning algorithm generates a new weak classification rule. This process is iterated many times and at the end, a boosting algorithm is formed by combining all newly generated weak classifiers rules to make a strong rule for prediction. A few steps should be followed for the selection of the right distribution [136]:

- Step 1: Give all the distributions to the base learner and assign equal weights to every observation.
- Step 2: If the first base learner gives any prediction error, then pay more attention to the observations causing this prediction error. Then, apply a new base learner.
- Step 3: Until the base learning limit is reached, or the desired accuracy is achieved, keep repeating Step 2.

Adaboost is usually used along with decision trees. The Adaboost model is created successively one after another, and the weight of every training instance (feature) is updated which affects the overall learning performed by the next tree in the line. After the generation of the first tree, for each training instance (feature), the performance of the tree is weighted i.e., how much consideration the next incoming instance (feature) will get from the newly generated tree up next. After creating all the trees, a prediction is made for new data. Figure. 2.18 shows classification done through the Ada-boost technique.

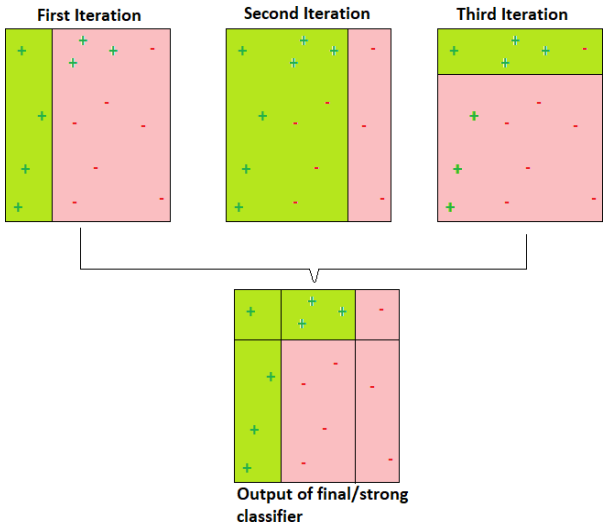


Figure 2.18 Working of AdaBoost Classifier.

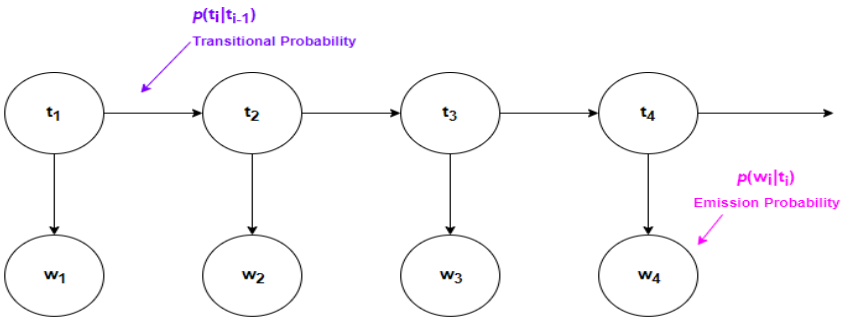


Figure 2.19 Transition and Emission probabilities in the Hidden Markov Model.

2.2.3.20 Hidden Markov Model

The hidden Markov Model (HMM) is a statistical version of the Markov model and is assumed to be a Markov method having unobserved (hidden) states. In a simple Markov model, an observer can see the states directly and that is why the Markov model has only parameters related to state transitional probabilities. On the other hand, in HMM the transitional states are not directly visible but output, which is dependent on the states, is visible. As every state has a specific probability distribution over possible output tokens, the sequence of the tokens produced by HMM gives information about the state arrangements. This phenomenon is known as pattern theory. The word hidden does not refer to the parameters of the model. It refers to the sequence of the states through which the model passes. The parameters of the Hidden Markov Model are exactly known. HMM, the framework contains the following components [137]:

- States, e.g., labels. Usually denoted by $T = t_1, t_2, ..., t_N$.
- Observations, e.g., words. Usually denoted by $W = w_1, w_2, ..., w_N$.
- Two Special States: t_{start} and t_{end}. These states are not directly associated with observations.

The states and observation-related probabilities are an initial probability distribution over states, a final probability distribution over states, a matrix with the probabilities going from one state to another state, called transition probability, and a matrix with the probabilities of an observation generated from a state called emission probability. Figure. 2.19 shows probabilities incurred in HMM.

2.2.3.21 Support Vector Machine Classifier

The Support Vector Machine (SVM) classification technique is the most precise method of solving a classification problem. These are built around a perception of margin i.e., data is separated into two classes, on each side of the hyperplane. SVM classifier is a binary classifier, so for the multi-class classification problem, multiple machines are trained [138]. SVM aims maximization of the margin between instances (features) of the two classes and to minimize generalization error, usually incurred in other classifiers. Figure. 2.20 shows how two different sets of features are classified using SVM.

Data points that lie on the margin of an optimized hyperplane are called support vector points and linear combinations of these points from solutions to the classification problem; all other points are ignored. Mathematically, SVM uses the QP (quadratic problem) with N dimensions, where N shows training samples.

Table 2.3 show 21 algorithms reviewed for the task of stress level monitoring.

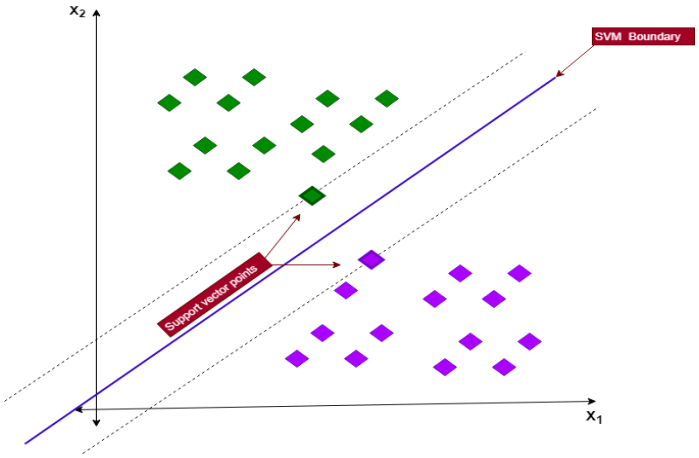


Figure 2.20 Maximum margin hyperplane for SVM trained model.

Table 2.3 Comparison Table of Machine Learning Approach for Stress Monitoring and Identification.

S.No.	Method/Technique	Ref	No. of subjects	Signals/Devices Used	Accuracy (%)
1	AdaBoost	[19]	15	Motion as well as physiological data using sensors	77.2
2		[88]	18	Physiological and social responses	94.3
3	ANN	[139]	10	ECG and PPG	82.5
4		[140]	20	ECG and PPG	98.8
5		[8]	55	Leaf node evidence, context, workload, and student health trait.	85.3
6	Decision Forest	[141]	30	ECG, Body Temperature, Pulse Oximeter, GSR, Blood Pressure, and Glucometer	96
8	Decision Tree	[19]	15	Motion as well as physiological data using sensors	74
9		[142]	34	Blood Volume Pulse, GSR, Skin Temperature and Pupil Diameter	88.02
10		[143]	42	Heart Rate Variability (HRV)	79
11	Genetic Algorithm	[144]	60	Pupil diameter (PD), ECG and PPG signals	N/A
12	HMM	[145]	60	Salivary cortisol, heart rate, and self-report ratings + moment of eyebrows and mouth	75-88
13	kNN	[19]	15	Motion as well as physiological data using sensors	70.5
14		[140]	20	ECG and PPG	97.5
15		[88]	18	Physiological and social responses	87.3
16		[146]	11	EEG signals + salivary cortisol	63.46
17	KSOM	[147]	20	GSR, PPG (Respiration rate), ECG and HRV.	81.6
18	LVQ	[148]	50	Skin conductance, electrocardiogram, and respiration rate	86.8
19	LDA	[19]	15	Motion as well as physiological data using sensors	78.05
20		[149]	21	Heart Rate (HR) and Pulse Transit Time (PTT)	60.8*
21		[150]	33	Electrodermal activity (EDA)	78.35*
22	Logistic Regression	[151]	42	Electroencephalogram (EEG) signals	89*
23	MLP	[89]	40	ECG and electrical bio-impedance (TEB) signals	67.7
24		[152]	3	ECG, EMG, and GSR	95.2*

25	Naïve Bayes	[142]	33	Blood Volume Pulse, GSR, Skin Temperature and Pupil Diameter	78.65
26		[152]	3	ECG, EMG, and GSR	64*
27	Nearest Class Centre	[150]	33	Electrodermal activity (EDA)	74.52*
28	OneR	[152]	3	ECG, EMG, and GSR	49.35*
29	One-v-all	[141]	30	ECG, Body Temperature, Pulse Oximeter, GSR, Blood Pressure and Glucometer	72
30	PCA	[153]	25	Heart rate monitor, respiration rate monitor and electrodermal activity sensors	60
31	Random Forest	[19]	15	Motion as well as physiological data using sensors	77.21
32		[154]	20	Heart Rate Variation (HRV) and simple heart rate signals.	97.2
33	SVM	[142]	32	Blood Volume Pulse, GSR, Skin Temperature and Pupil Diameter	90.1
34		[149]	21	Heart Rate (HR) and Pulse Transit Time (PTT)	62.3*
35		[144]	60	Pupil diameter (PD), ECG and PPG signals	90.1
36		[155]	50	Heart rate monitoring sensor (ECG, PPG), skin conductance (GSR) and skin temperature monitoring	91.26
37		[156]	15	ECG, EEG and EDA signals + Salivary samples	86
38		[8]	55	Leaf node evidence, context, workload, and student health trait.	90.8
39		[152]	3	ECG, EMG, and GSR	95.2*
40		[146]	11	EEG signals + salivary cortisol	71.72
41	ZeroR	[152]	3	ECG, EMG, and GSR	12.5*

*Note: * means it is an average of all the accuracies achieved by the author/s in different scenarios. OneR and ZeroR have accuracies of less than 50% (selection criteria) but are baseline algorithms which is why they are included here.*

2.2.4 Conclusion - Review of different Machine Learning Algorithms

This study focused on the classification of stress states. In total, 20 different machine learning algorithms are discussed in this review study that uses different parameters for training and prediction of stress, refer to Table 2.3. All the parameters considered, are correlated with the stress. Few of them are distinct parameters, for example, GSR and HR, while some are in conjunction with other parameters to monitor and recognize stress, for example, Skin temperature (ST) and EMG. But only using the parameters shown in Table 2.3, stress cannot be defined. We also require information about the context to interpret the data collected from the sensors and to understand what was going on at the time of reading collection. This context information can be gathered using mobile phones (IoT based) or from computers as both devices are been used frequently during our daily life work.

The study also gave a better look at state-of-the-art machine learning algorithms and their use as stress-level classifiers. We can see that only 6 out of 20, namely, SVM, Random

Forest, kNN, MLP, Ada-Boost and Decision Forest machine learning algorithms were able to recognize and classify the stress state of a subject with an accuracy of more than 90%. In terms of a consistently higher classification rate, the SVM classifier can be considered the best classifier for stress monitoring procedures, but one must see the number of subjects and type of sensors (signals) required for training and testing of the classifier. The question that which classification algorithm is best for utilization in stress monitoring devices remains in place as different people have different perceptions about the usage of machine learning algorithms. There is a trade-off that must be considered between computation time, accuracy, and the cost of the device. It is a matter of balancing the computational time versus the accuracy versus the price of the device while selecting an algorithm.

2.3 A Comprehensive Review of Cortisol Detection Methods for Stress Monitoring

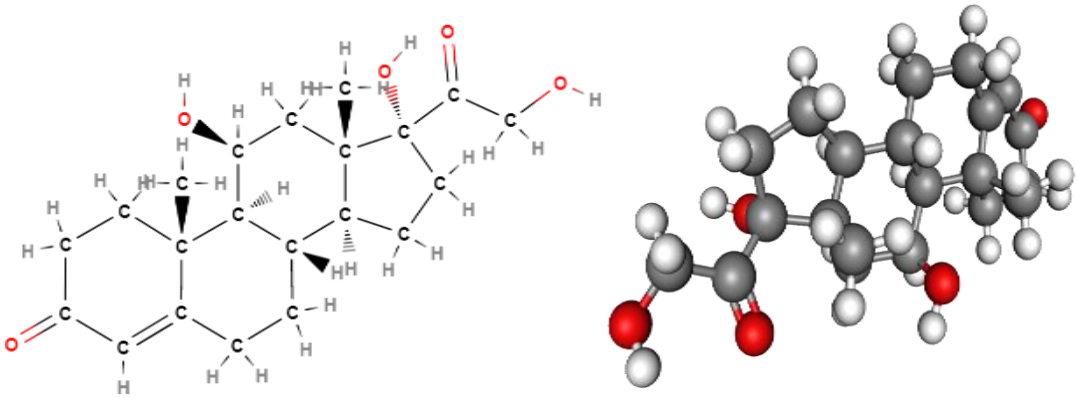
Everyday responsibilities and lifestyle issues are the main cause of physical and psychological stress, which deteriorates the individual's health. Prolonged exposure to stress triggers the adrenocorticotrophic hormonal (ACTH) system and causes the release of cortisol hormones from the adrenal cortex [157], [158]. Many other biomarkers are affected by stress, but cortisol is considered the most vital and potentially clinically useful biomarker for stress estimation and monitoring. Accurate and timely detection of increased cortisol levels might improve the diagnosis, treatment, and prevention of stress-related diseases such as anxiety disorders, metabolic dysregulation, and cardiovascular diseases. Unfortunately, most of the cortisol assessments are currently performed only in laboratories and there is no point-of-care solution for ambulatory/real-time cortisol assessment. This review aims to provide an overview of the most promising techniques, currently used for cortisol detection and the challenges associated with them. The review also provides a feasibility report about measuring cortisol levels in different bio-fluids (for example, urine), a correlation of perceived stress with cortisol levels, and methods/devices used in the laboratory as well as in the ambulatory environment for cortisol detection. The overall conclusion suggests that significant research efforts and investments are required for the development of an accurate, rapid, and repeatable cortisol measuring device that can be used for connected health applications.

2.3.1 Introduction

Stress is a physical, emotional, and physiological response of the body to an internal or external stimulus, categorizing it as one of the major threats to mental health [159]. The increasing psychological stress levels due to an altered lifestyle, globalization, and competition are of serious concern, causing life-threatening diseases such as depression, heart attacks and stroke [160]. Thus, the accurate and precise detection of physiological as well as psychological stress is gaining the attention of researchers and investigators for personalized health monitoring and diagnostics [161]. Current stress diagnostic approaches include the measurement of stress effects, stress exposure, self-reporting questionnaires and assessment of different biomarkers [162]. Among these approaches, an effective

interpretation of a biosensor in a biomarker-based stress assessment system is considered unequivocal for an effective diagnostic approach [163].

In recent years, wearable stress monitoring devices have been developed to relate stress with abnormalities in the environment and to gain vital information for real-time diagnosis and treatment. Since the development of the Trier Social Stress Task (TSST) [164] in the year 1993, our knowledge about the neuro-endocrine processes associated with physiological and/or psychological stress has increased significantly. Research with laboratory-based stress induction tests such as TSST aims to understand the impact of stress on the hormonal stress responses of the human body under controlled conditions [82]. The TSST emphasizes the critical role of cortisol blood levels in the detection of stress [165]. Several studies have performed experiments with students, police officers, nurses, and athletes to link the levels of cortisol with stress and have classified cortisol as the most prominent biomarker for stress detection [166]–[173].



(a) (b)

Figure 2.21 Molecular and 3D structure of Cortisol ($C_{21}H_{30}O_5$). In (a) C is for Carbon, H is for Hydrogen and O is for Oxygen molecule. In (b) Black shows Carbon, Grey shows Hydrogen and Red shows Oxygen molecule. (Generated using: https://molview.org/?cid).

Cortisol ($C_{21}H_{30}O_5$), is a steroid hormone with a molecular weight of 362.46 g/mol. It is a well-known biomarker of psychological and physiological stress [174], [175] (Figure 2.21). The level of cortisol plays an important part in regulating blood pressure, carbohydrate metabolism and glucose levels. It also contributes to the homeostasis of cardiovascular, renal, immune, endocrine and skeletal systems [166], [167], [176]. Abnormally increased levels of cortisol interfere with blood amino acid and fatty acid levels, resulting in

depression of the immune system and causing inflammation. Severely increased levels of cortisol contribute to the development of Cushing's disease along with symptoms of obesity, bone fragility and fatigue [177], while decreased levels of cortisol lead to Addison's disease manifested by arterial hypotension, weight loss and darkened scars/skinfolds [178]. The most dominating effects of cortisol are indicative of emotional or psychological stress and that is why cortisol is also called the 'stress hormone' [179].

Currently, in the clinic, total cortisol (i.e., the sum of protein-bound and free fractions) is measured. Free cortisol is the only biologically active fraction and is liable for all cortisol-related effects in the body and could be found in blood (serum and plasma), saliva, urine and other biological fluids [180]. Although the reference values of cortisol levels that could be translated to an individual's physical or psychological stress are yet to be determined, it would be of value to develop an accurate diagnostic system that allows repeated measurements of free cortisol [181]. The currently available methods for the determination of free cortisol levels are mostly laboratory-based. These strategies require large samples, are time-consuming, laborious as well as expensive and are not suitable for point-of-care diagnostics [182]–[184]. Furthermore, these current set-ups only provide a glimpse at cortisol levels within the sample submitted to the laboratory and do not provide a realistic representation of its variations. There is also an influence of the time of the day on the cortisol levels within the body. Thus, one tends to rely upon 24 h urine analysis of cortisol metabolites, instead of cortisol sampling. Nevertheless, nearly continuous, and real-time monitoring of cortisol levels is essential to provide a more reliable assessment for better diagnosis and treatment of stress-related conditions. Recent advances in technology have shown prominent results in the development of such systems.

To review the existing technologies developed to self-test cortisol levels, this section highlights the recent efforts made to develop strategies for the detection of cortisol in a laboratory and in point-of-care/ambulatory (where healthcare is provided close to the patient, for example at home) settings. Moreover, this review discusses the correlation of cortisol levels with induced stress, the comparison of different cortisol detection techniques and provides future directions regarding the development of a real-time cortisol measuring stress monitoring device. In the present literature, most of the review papers are specific to either cortisol detection in the laboratory or cortisol detection in point-of-care/ambulatory

settings. In comparison with previously published reviews, shown in Table 2.4, this review will answer the following questions (that no other single review covers):

- The correlation between induced stress and cortisol levels.
- The feasibility of cortisol detection using invasive and non-invasive methods.
- The devices/techniques used for detecting cortisol in laboratory settings.
- The devices/techniques used for detecting cortisol in point-of-care diagnostic settings.
- The current research status regarding developing a real-time cortisol measuring stress monitoring device and the future direction.

Table 2.4 Comparison of recent review papers with the proposed review

Ref	Correlation of cortisol with stress	Feasibility (invasive/ non-invasive)	Techniques Ambulatory settings	Techniques Laboratory settings	Future Stress management	Cortisol measuring methods reviewed
Kaushik *et al.* (2014) [161]	✓	×	✓	✓	×	5
Hogenelst *et al.* (2019) [181]	✓	×	✓	×	✓	3
Steckl *et al.* (2018) [185]	×	×	×	✓	✓	4
Parlak *et al.* (2021) [186]	✓	×	✓	×	✓	6
Zhang *et al.* (2022) [187]	×	×	✓	×	×	5
Zea *et al.* (2022) [188]	×	×	×	✓	✓	6
Proposed	✓	✓	✓	✓	✓	**6**

2.3.2 Feasibility of different Sources for Cortisol Sampling

The cortisol hormones are secreted by adrenal glands located above our kidneys. It is the end-product of an important component of the human's body adaptive system called the hypothalamic-pituitary-adrenal (HPA) axis. HPA regulates the physiological processes of the body under different environmental factors [189]. In response to triggers, the hypothalamus in the brain releases a corticotrophin-releasing hormone (CRH) acting at the pituitary glands. The pituitary glands release adrenocorticotrophic hormones (ACTH) into the blood that travels to the adrenal cortex. The adrenal cortex responds to ACTH by increasing the production of cortisol, which then participates in modulating several

physiological processes. The secreted cortisol steers its way to the circulatory system and can be found in several biological fluids in detectable quantities. Figure 2.22 illustrates the process of the body's response to a stressor (a stimulus that induces stress). This section assesses the feasibility, advantages, and disadvantages of sampling different sources of cortisol based on literature data.

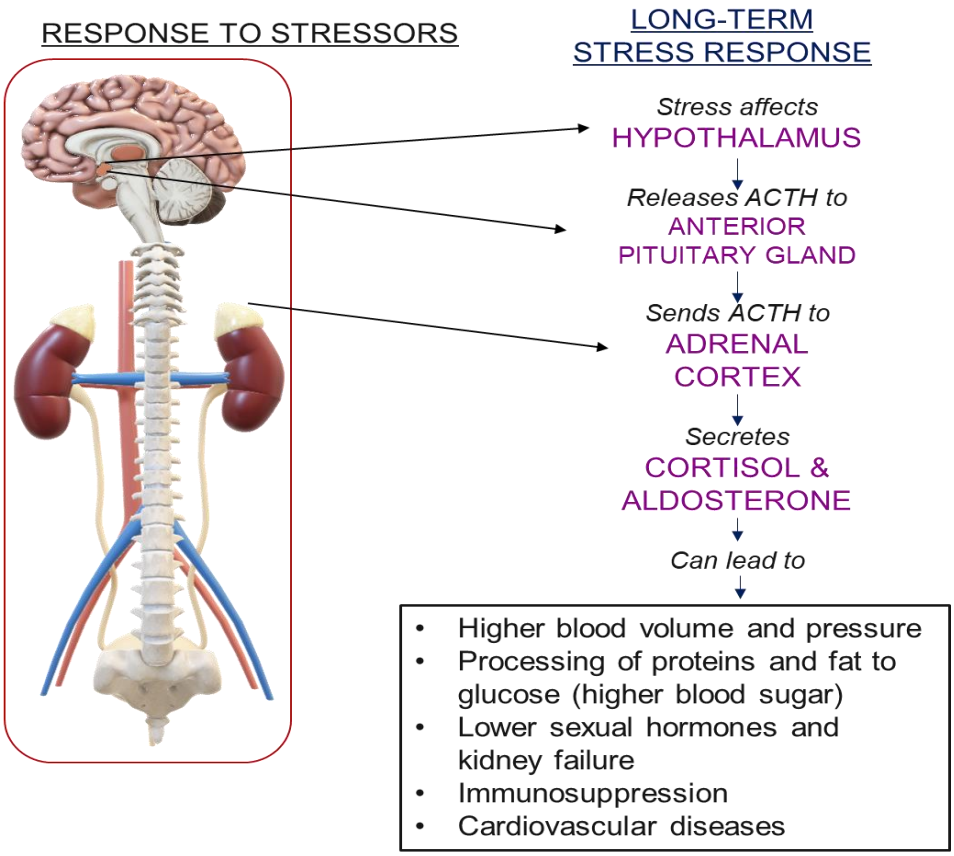


Figure 2.22 Illustration of cortisol production as a response to long-term stress.

2.3.2.1 Salivary cortisol

Over recent years, salivary cortisol detection has gained considerable attention for the development of stress monitoring systems. This is mainly due to the advantages associated with saliva as a convenient bio-fluid. The first and most important fact is the well-documented literature that shows a strong correlation between blood and salivary cortisol concentrations [190]. The other important advantage of salivary cortisol is that it is entirely in a free state. Harvesting salivary cortisol samples is also a completely non-invasive and painless process. A standard operating procedure has been established for saliva collection, which leads to reduced variability of measurements. Along with the above-mentioned

advantages, there are some drawbacks as well. The nominal values of salivary cortisol during the diurnal cycle vary from 0.5μg/dL to 0.05 μg/dL [191]. Thus, a highly sensitive assay with the ability to detect low ranges of cortisol is required. Also, the salivary cortisol is highly unstable at room temperature causing problems with storage throughout the on-site sampling and processing period. In the mornings, the mean salivary cortisol concentration is in the range of 3.6 nmol/L to 8.3 nmol/L while at late nights the concentration drops to 2.95 nmol/L to 2.1 nmol/L [192].

2.3.2.2 Hair cortisol

Human hair grows at a predictable rate of almost 1cm per month. The cortisol hormone is known to be found in the shaft of the hair. Also, the proximal 1cm segment of hair (closer to the scalp) approximates the previous month's cortisol production [193]. The hair cortisol is hypothesized to reflect the free cortisol fraction rather than the total cortisol concentration in the blood plasma and serum. The hair cortisol reference values vary from 1.7 to 153.2 pg/mL (1.7x 10-6 μg/mL to 0.0001532 μg/mL) [194]. The hair cortisol measurement is a non-invasive way of obtaining a biological sample. Hair cortisol levels are representative of long-term exposure to stress. Because the data are "collected" over several months, determining the association of hair cortisol with stress will require careful correlation with frequently obtained clinical data, not available so far in the literature.

2.3.2.3 Urine cortisol

Cortisol levels measured in urine are referred to as 24h urinary-free cortisol (UFC). Only free and active cortisol is present in the human urine, qualifying urine samples as a relevant bio-fluid for cortisol detection. The normal range of UFC levels is 36 μg/24h to 137 μg/24h [195]. Although the collection of 24h urine for the UFC test is a non-invasive and painless method, it also poses some issues regarding convenience and reliability. Since the sample is collected over 24 hours, the patient must carry a special urine collecting container all day and must be relatively confined to a given location for 24 hours. Also, the container needs to be stored in a refrigerator from the time of collection to its delivery to the laboratory. Moreover, with the entire volume (24h urine), creatinine is also needed to be measured to verify that the collection is complete. Factors such as pregnancy and medication (such as ketoconazole, adrenalux and metyrapone) can alter the concentration of cortisol in the urine

sample. The requirement for 24h collection of urine renders the UFC measurement unfit for real-time cortisol detection in point-of-care settings.

2.3.2.4 Blood (serum and plasma) cortisol

The assays used to measure cortisol levels in the blood measure the sum of bound and free cortisol, the latter being the active form. The Cortisol Free Index (CFI) is calculated using Coolen's equations [180]. Under non-stress conditions, 10 to 15% of the blood cortisol is bound to serum albumin while 80 to 90% is bound to corticosteroid-binding globulin (CBG). The remaining 5-10% is in a biologically active state and participates in cortisol-initiated effects [196]. The normal range of blood cortisol varies from morning to evening. The nominal value of total blood cortisol level is 0.05 μg/mL to 0.25 μg/mL (25 μg/dL at 9 am versus 2 μg/dL at midnight) [161]. There are several drawbacks to sampling cortisol from blood, making it a suboptimal sampling site. The main drawbacks are listed below:

- Collecting blood samples requires trained staff and sterilized equipment with the potential effect of stress, fear, and pain on cortisol levels.
- Cortisol molecules are unstable at room temperature and the plasma requires special handling and storage environments.
- Sampling blood needs vein puncturing which can be painful and is often perceived as stressful.
- Although the typical time of cortisol spiking after a stressful event is 10 to 15 minutes in humans, venepuncture and the apprehension that is associated with the need for blood sampling can initiate stress-induced cortisol spiking.
- The time of the day needs to be considered for the interpretation of the result and a standardized sampling technique in a quiet environment is desirable.

The cost of equipment, staff, handling and storage makes blood cortisol sampling a shunned option.

2.3.2.5 Interstitial fluid cortisol

Interstitial fluid (ISF) is the extracellular fluid that envelops cells in the tissues. ISF is similar to blood plasma in composition. Generally, small to moderate-sized molecules (0.5-5nm), such as ethanol, glucose, and cortisol (362.46 g/mol) are found in ISF in the same ratio as in blood plasma. The use of microneedles to obtain ISF in a painless, minimally invasive

manner, has been successfully developed [197]–[200]. For cortisol detection, ISF needs to be harvested at a very slow rate of 10 μL/h which limits its applicability in the point-of-care/ambulatory setting. The concern regarding biodegradation and biocompatibility of microneedles, risk of infection, continuous extraction of body fluid and other sterilization issues need to be addressed carefully for the successful implementation of the ISF cortisol detection approach.

2.3.2.6 Sweat cortisol

Sweat cortisol measurements have found reference values of cortisol concentration in the sweat, ranging from 8.16 ng/mL to 141.7 ng/mL (0.00816 μg/mL to 0.1417 μg/mL) [97]. For sweat collection, sweat patches are frequently used which is an effective non-invasive method to do so. However, due to very limited knowledge of cortisol correlation in sweat and several other factors such as humidity, temperature, physical activity, geographic condition, and individual genetic factors, there are inherent limitations to the development of a reliable and repeatable sweat sampling device for cortisol detection.

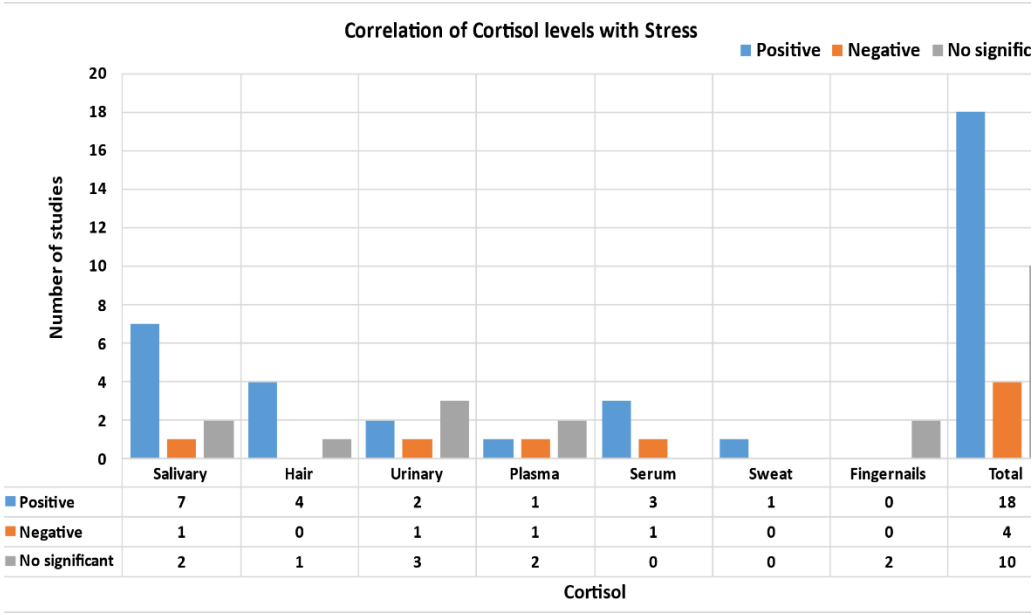


Figure 2.23 The number of reviewed literatures reporting positive, negative and no significant correlation of cortisol level, measured via different technique, with induced stress.

2.3.3 Perceived Stress and Cortisol levels: Correlation Analysis

To determine the correlation between stress and cortisol levels there is vast literature available on non-invasive methods of cortisol detection, such as cortisol in saliva, hair,

urine, sweat and fingernails while limited literature is available on invasive methods of cortisol detection (such as, in plasma and serum). The following literature has been reviewed to determine the true relationship between stress-induced cortisol levels in different biospecimens. Figure 2.23 illustrates the number of reviewed literature reporting positive, negative and no significant correlation of cortisol level, measured via different techniques, with induced stress.

Overall, the most frequently used cortisol for monitoring and detection of physiological as well as psychological stress is salivary cortisol. The literature review also revealed that, in most cases, cortisol levels within the body are either positively or negatively related to stress. This correlation highly depends on the source that is used to measure cortisol levels. Stress is mostly positively related to cortisol levels determined in hair, saliva, and serum. Alternatively, there is also some contrary relationship reported in the previously reported literature. For example, Schmalbach *et al.* [201] reported a negative correlation between stress with cortisol levels in saliva. When determining correlation using plasma and urinary cortisol, different studies report contradictory results i.e., 19 reported positive, 4 negative and 9 reported no significant relationship of stress with cortisol levels. Table 2.5 summarizes the studies reviewed to determine the correlation between cortisol levels and stress in this section.

Table 2.5 Summary of studies determining the correlation between cortisol levels and stress.

Cortisol	Ref	Year	Sample size	Measurement Technique	Correlation
Salivary	[202]	2019	250	Radioimmunoassay	Positive
	[203]	2005	64	Radioimmunoassay	Positive
	[204]	2002	40	Radioimmunoassay with a time-resolved fluorometric detection	Positive
	[205]	2016	31	Cotton Salivettes (Salivette kit)	Positive
	[201]	2020	52	Salivette kit	No significant
	[206]	2020	12	Salivette Code Blue	Positive
	[207]	2020	112	Salivette kit	negative
	[208]	2020	79	ELISA kit	Positive
	[209]	2020	42	ELISA kit	No significant
	[210]	2020	53	Salivette kit	Positive
Hair	[211]	2019	57	Salimetrics high-sensitivity cortisol EIA kit	No significant (less or low stress) / Positive (high stress)
	[212]	2016	270	Liquid chromatography-tandem mass spectrometry (LCTMS)	Positive
	[213]	2015	174	Liquid chromatography-tandem mass spectrometry (LCTMS)	Positive

	[214]	2013	55	Enzyme-linked immunoassay (ELISA) kit	Positive
	[215]	2019	75	ELISA kit	Positive
Urinary	[216]	2003	32	Radioimmunoassay	Negative
	[217]	2016	110	Polypropylene container	Positive
	[218]	2006	28	Polypropylene container	No significant
	[219]	2014	80	ELISA kit	No significant
	[220]	2011	104	Radioimmunoassay	No significant
	[221]	2021	6878	Liquid-Chromatography Mass Spectrometry (LC-MS/MS)	Positive
Plasma	[222]	2001	102	From a cannula inserted in the forearm vein	No significant
	[223]	2006	68	Radioimmunoassay	Positive
	[224]	2017	91	Intravenous catheter	Negative
	[225]	2016	85	Radioimmunoassay	No significant
Serum	[226]	2017	21	Commercial enzyme-linked immunosorbent assay (ELISA) kits	Positive
	[227]	2016	44	Chemiluminescence assay	Positive
	[228]	2020	34	Chemiluminescence assay	Positive
	[229]	2021	106	Electrochemiluminescence immunoassay	Negative
Sweat	[230]	2020	48	ELISA kit	Positive
Fingernails	[231]	2018	47	ELISA kit	No significant
	[232]	2018	51	ELISA kit	No significant (15 days) / Positive (45 days)

2.3.4 Application of cortisol detection in clinical research and practice

Cortisol concentration in the body is associated with many clinical outcomes such as hypertension, dyslipidemia, depression, and anxiety. There are several somatic health factors, chronic and acute stressors, and psychopathological factors that affect the concentration level of cortisol. A detailed list of these factors has been reviewed by Wester *et al.* [233]. In terms of stress monitoring, the different concentrations of cortisol can be indicative of cardiometabolic status, chronic stress, and/or psychopathology factor. Each factor is discussed in detail as follows.

2.3.4.1 Cardiometabolic status

Cardiometabolic factors (such as obesity, hypertension, diabetes, and dyslipidemia) are highly associated with increased levels of cortisol [234]. Manenschijn *et al.* [235] and Feller *et al.* [236] found a positive correlation between cortisol levels with diabetes type 2 disease in an elderly population. Increased level of cortisol is also found in patients with myocardial

infarction as determined by Pereg *et al.* [237]. Stalder *et al.* [238] showed the relationship between higher cortisol levels and adverse metabolic syndrome. Several studies [239]–[242] have shown a positive correlation between cortisol levels and obesity.

Altogether, these findings strongly suggest that long-term cortisol exposure increases the risk of worsening cardiometabolic status and needs to be monitored thoroughly.

2.3.4.2 Chronic stress

Both psychological and physical stress can result in hyperactivation of the HPA axis which results in the release of cortisol hormones in the body. The relationship between HPA activation and perceived stress is complex as conflicting results are reported in the literature. Kalra *et al.* [243] reported a negative relationship while Karlen *et al.* [244] reported a positive association of cortisol concentration with perceived stress. On the contrary, [245] and [246] reported no significant correlation between cortisol levels with stress.

Thus, further research is required to determine the true relation between perceived stress and cortisol concentration. The effect of factors such as sex, region, age, and sample size on cortisol concentration should also be investigated.

2.3.4.3 Psychopathology factors

Anxiety and mood disorders are linked to cortisol levels in the body. Burke *et al.* [247] determined that cortisol levels take a long time to return to baseline levels in people with depression as compared to the non-depressed population. Veldhorst *et al.* [248] conducted a small study with patients with major depression disorder and matched the cortisol levels with the control group. The authors found increased cortisol concentrations in the depressed patient's group compared to the control group. Traumatic event and post-traumatic stress disorder (PTSD) is also associated with the triggering of the HPA axis. In the literature, mixed results are found about the association of cortisol levels with PTSD condition. Studies such as [249] and [250] showed no significant relationship between cortisol and PTSD while [251] and [252] reported significant changes in cortisol concentration in PTSD and non-PTSD groups.

The literature review shows that the association between cortisol levels and PTSD depends on the type of traumatic event, sample characteristics examined and the timespan between trauma and cortisol assessment.

2.3.5 Cortisol assessment in laboratory settings

Many immuno-sensing applications of cortisol using anti-cortisol antibodies have been developed for the diagnosis of stress in laboratory settings. In this section, we have reviewed some state-of-art methods of cortisol measurement for stress detection in laboratory settings. Table 2.6 summarizes the different methods of anti-bodies-based cortisol mechanisms along with their minimum limit-of-detection, advantages, and limitations. Figure 2.24 illustrates the different laboratory-based sensors and their functionality.

Table 2.6 List of anti-bodies-based cortisol mechanisms along with their minimum detection limit, advantages, and drawbacks.

Method	References	Lowest Detection Limit	Advantages	Drawbacks
Enzyme-Linked Immunosorbent Assay (ELISA)	[253]–[255]	0.25 pg/ml	Highly sensitive, stable reagents and robust	Requires enzyme labelling
Polyaniline-protected gold nanoparticles (PPAuNPs)	[256], [257]	1 pM	Oxidation-reduction stability and sensitive	pH value of the PBS solution might affect the detection accuracy
Lateral Flow Immunosensor (LFI)	[258]–[260]	3.5 μg/L	Highly sensitive and selective, easy to use/interpret, kit available, no special training requires	Semi-quantitative, Uncertain sample size reduces accuracy and precision
Quartz Crystal Microbalance (QCM)	[261], [262]	11 pg/ml	High sensitivity and specificity	Environmental noises affect measurements
Chemiresistor Immunosensor	[263]–[265]	1 pg/ml	Excellent binding label-free selectivity of cortisol	Non-specific adsorption of high molecular weighted hormone from a saliva sample
Surface Plasmon Resonance (SPR)	[182], [266]	1 μg/L	Highly sensitive, robust, simple, reproducible	A pre-treatment procedure such as partial purification is required for accuracy

2.3.5.1 Summary

All these systems are portable, highly sensitive, smaller in size than conventional sensors, low cost and provide comparable results with standard electrochemical analyzers. Among the mentioned cortisol measurement methods, the ELISA has the highest sensitivity (can detect the presence of cortisol as low as 0.25 pg/ml), great reagent stability and is easily reproducible. Along with all these advantages, these systems require multiple steps that should be performed carefully for accurate detection thus making them less reliable, non-user-friendly and not suitable for home base care systems, thus far.

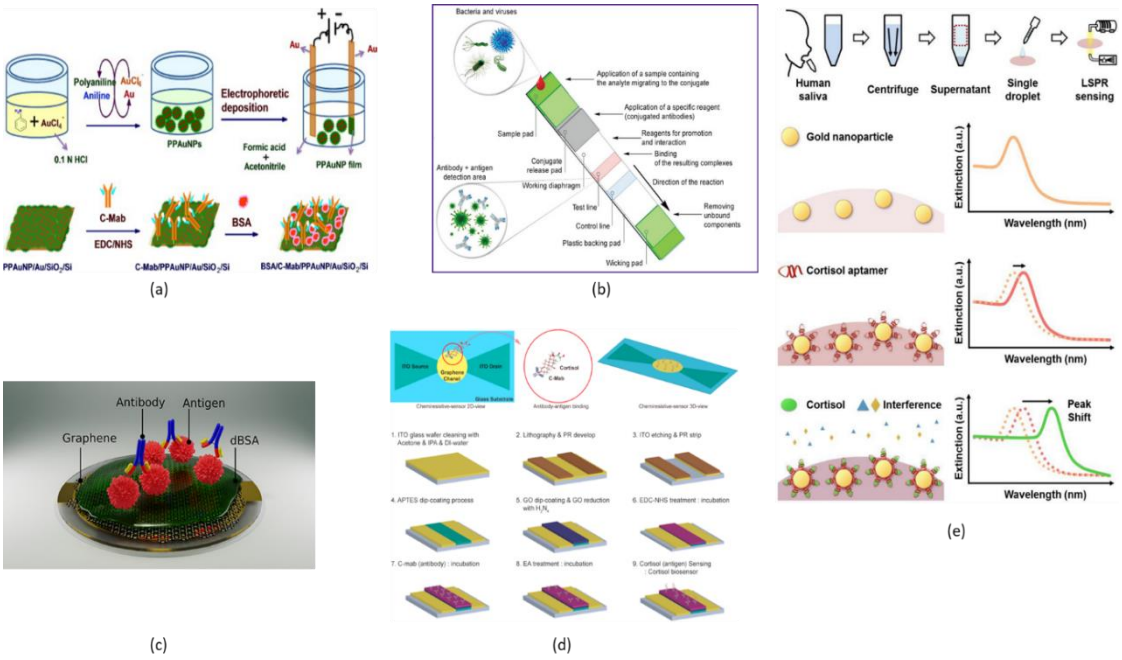


Figure 2.24 (a) Shows the fabrication schematics of PPAuNP electrode; reproduced from [256] (b) Represents the schematics of LFI mechanism; reproduced from [259] (c) illustrates the graphical abstract of working of QCM; reproduced from [262] (d) is schematics of chemiresistor sensor fabrication; reproduced from [264] and (e) shows the steps involve in the sampling and detection of salivary cortisol using SPR; reproduced from [266].

2.3.6 Cortisol assessment in point-of-care/ambulatory settings

In this section, we reviewed the literature investigating ambulatory cortisol measurement methods and devices. Most of the proposed devices are only a proof-of-concept and have either ability or potential to assess the stress on the go in a real-life environment. Thus, there is a lot to improve before these methods could be translated to point-of-care or home-based devices. Following are some cortisol assessment technologies categories based on the

cortisol immobilizing method used for point-of-care analysis along with their properties, advantages, and limitations. Table 2.7 summarizes all the literature reviewed in this section. Figure 2.25 shows the most popular techniques in the literature for real-time cortisol detection.

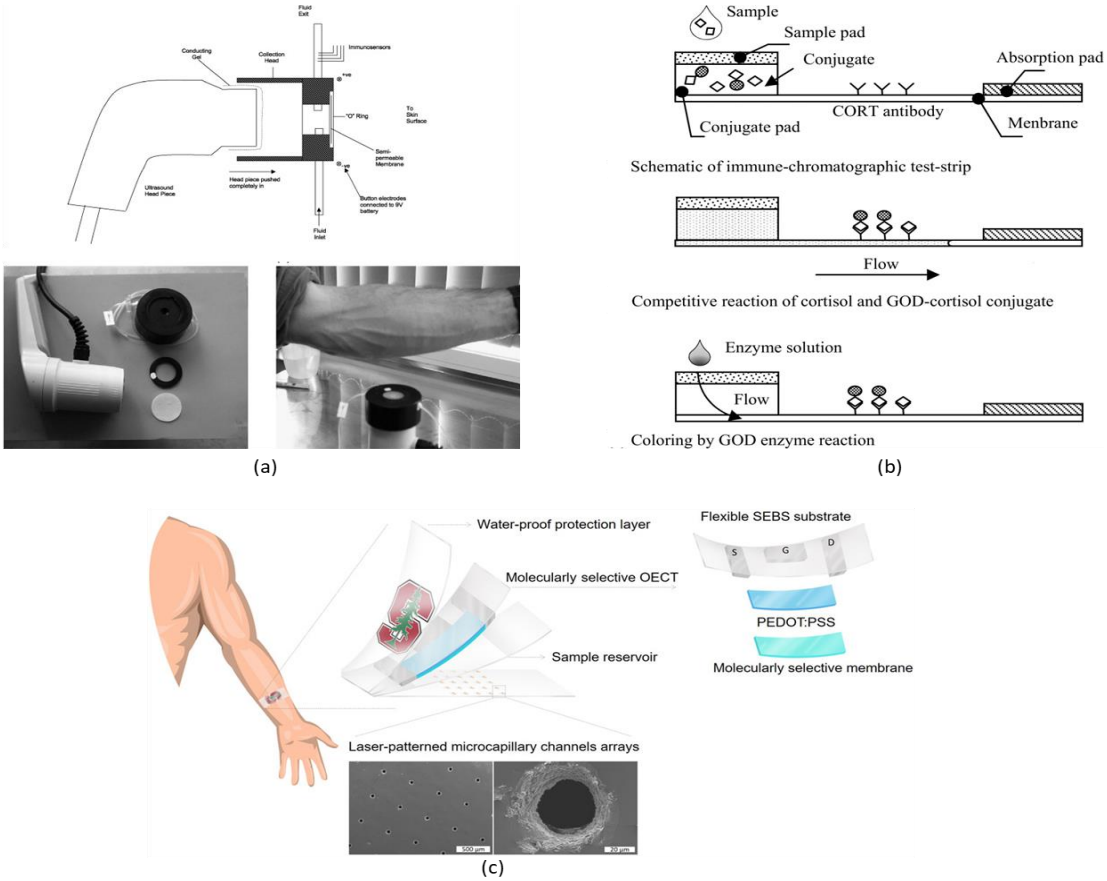


Figure 2.25 A platinum based transdermal device with an O ring, membrane, collection chamber and ultrasound delivery device and use of the device in human is shown in (a); reproduced from [276]. (b) demonstrates the working principle of an immune-chromatographic chip; reproduced from [277]. (c) shows the proposed schematics of patch type wearable sweat cortisol sensor; reproduced from [282].

Table 2.7 List of Immobilizing Matrix-based cortisol detecting mechanisms along with their properties, advantages, and drawbacks for ambulatory assessment.

Immobilizing Matrix	Ref	Detection technique	Range	Response time	Advantages	limitations
Self-Assembled Monolayer (SAM) membrane and variants	[267]	Electrochemical immunosensor	0.1 to 10 ng/ml	35 min	Portable, highly sensitive, smaller in size, low	Require multiple steps for accurate detection, expert needed,
	[268]		0.036 to 36 ng/ml	30 min		

	[269], [270]		10 pM to 500 nM	30-35 min	cost and accurate	not user-friendly
Nitrocellulose membrane and variants	[271]	Lateral flow immunoassay (LFIA)	3.5 ng/ml to 1280 ng/ml	5 min	Portable, handheld devices with highly correlated results	For quantification, bulky reading equipment is required
	[272]		0.9 ng.ml to 25 ng/ml	20 min		
	[273], [274]	Colorimetric LFIA	0.3 ng/ml to 60 ng/ml	30 min	Simple, portable, easy to use and best for ambulatory settings	Require 3D printer for a printer the adapter and cartridge holder
Platinum Electrode	[275], [276]	Electrochemical immunosensor	0.1 ng/ml to 100 ng/ml	5 min	Portable, handheld devices with highly correlated results	Require the application of 9V onto the skin for 2 minutes
Immune-chromatographic chip	[277]	Calorimetric chromatography	1 ng/ml to 10 ng/ml	25 min	Portable, handheld, and reusable with a monitor and disposable test strips	Indirect measurement of the cortisol
	[278], [279]	Colorimetric LFIA	1 ng/mL to 70 ng/mL	15 min		
Polystyrene pad	[280]	Electrochemical immunosensor	0.4 ng/ml to 11.3 ng/ml	15 min	A reusable sensor as having a disposable disk	Bulky reading equipment is required
Without any immobilizing antibody or enzyme	[281]	Functionalized nanoparticles	30 pg/ml to 10 μg/ml	2.5 min	Highly sensitive against cortisol hormone and provides a strong correlation with the ELISA kit	Only proof-of-concept devices, not available to end-users
	[282]	MIP-based electrochemical transistor	0.01 μM to 10 μM	<1 min		

2.3.6.1 Summary

Among the mentioned cortisol detection techniques, the functionalized nanoparticles and MIP-based electrochemical transistors are the fastest results-producing devices that can provide results of cortisol levels within 1 minute. These techniques are highly sensitive to cortisol and provide strongly correlated results with an ELISA kit. The only drawback is no such device (using these detection methods) is commercially available and is a proof-of-concept. Among the commercially available devices, the platinum electrode-based

electrochemical immunosensor has the highest sensitivity (can measure cortisol levels between 0.1 ng/ml to 100 ng/ml) and is an easily portable handheld device that provides highly accurate results.

2.3.7 Conclusion - Review of Cortisol Detection Methods

This review highlights the current efforts made in recent years to develop cortisol detection technologies for stress monitoring in real-life. This section (section 2.3) presents the feasibility report of cortisol extraction using different bio-fluids, the correlation of cortisol levels in different body fluids, the status of laboratory-based reliable cortisol assessment methods and ambulant cortisol measurement techniques along with their analytics. The study showed that the non-invasive methods (such as saliva, urine and sweat) of sampling and quantifying the cortisol levels are more feasible than invasive methods (plasma or serum) as invasive interventions induce extra stress and thus affect the true value/state of stress. Furthermore, cortisol secretion is highly correlated to physical and psychological stress, with scarce contrary results. To determine the true relationship between stress and cortisol, further high-quality analysis is required. The effect of sex, ethnicity, and treatments on the association between stress and cortisol levels must also need to be investigated.

The literature search also resulted in several promising strategies, technologies, and devices for cortisol assessment in daily life. Many studies reported a good sensitivity and specificity of cortisol sampling and showed comparable accuracies in cortisol detection when compared to gold-standard or conventional methods. Some of the devices (such as [273], [274], [282]) are portable, rapid, and easy to use. However, it is be noted that all the devices or technologies discussed in section 2.3 are not commercially available and are only proof-of-concept.

At present, there is no single best analytical method for the development of ambulatory cortisol technology and thus, this field appears to be relatively immature and novel. With significant research efforts and additional investments in the future years, we are hopeful to overcome the remaining procedural and technological hurdles and make a true ambulatory corticosteroid diagnostic wearable stress monitoring device for a home-based care system.

2.4 Chapter Conclusion

In mental health research, psychological and sociological stress is considered one of the most important as well as most complex areas of the current century. High stress does become a threat to the health of a person. The reasons behind this extensive stress are complex personal, social, and ecological environments, diversity in expressing stress as well as multiple transactions of humans due to his/her surroundings. Even though stress is a routine characteristic of life, nowadays, if it becomes continuous and increasing, an individual might show problematic symptoms which threaten their health as well as people in their surroundings.

Stress arises from events that threaten the homeostatic stability of a person. The human biological system is very complex, and stress evokes different physiological and cognitive reactions in the human body [87]. For this reason, stress markers established until now do not provide any reliable assessment of the quantitative stress response. To the best of our knowledge, there is no easily applicable and repeatable method available that can compare the stress response levels of one person in different situations. Moreover, the stress response of two different persons is also different. The goal of this review study was to find approximated quantitative measures of a person's homeostatic imbalance, determine the feasible sensory technology and understand the trade-off required while using a machine learning classifier for stress monitoring.

Generally, in a stress induction experiment, questionnaires are used as a standard stress state reference. These psychometric questionnaires are not designed to be used in general applications. Moreover, these questionnaires are subject to individual variability and depend upon the person's self-perception about their condition. Among the research community and professionals in the medical field, there is no agreement on the reference standards for monitoring stress levels and measurement methods. This lack of a standard for stress evaluation can be due to the variability in stimuli of stress, to which each human reacts differently. Furthermore, the available literature aimed to address one or few stress responses in an individual study rather than comprehensively describing the physiological stress response. The use of biochemical markers may result in better and more promising results in the detection of stress, but one of the biggest drawbacks of biochemical markers of stress levels is their relationship with the intensity of perceived stress. The reason is that

this relationship between biochemical hormones and stress is both complex and understudied.

The literature presents a variable and conflicting conclusion regarding the use of physiological or biochemical stress markers. A multisensory platform with data-driven personal insights can help track and intervene in cases of stress in the high-risk population. Integrating all the available sensors on to one platform is unrealistic. Thus, further analysis is required to determine the most sensitive and specific indicator/s of stress, which are susceptible to other physical or emotional stimuli.

Moreover, the review also showed that accurate stress monitoring is highly dependent on the selection of a stress classifier. It is important to consider the classifier's training performance and the trade-off between computational time, accuracy, and price of the device while selecting a predictive classification model. The use of questionnaires score for stress labelling is a difficult and inaccurate process. Thus, investigating the use of unsupervised machine learning algorithms for stress monitoring is deemed useful.

The review also revealed two major reasons behind the degraded classification performance of the stress classifiers. These reasons are; the lack of features that could translate/show the well-distinguishable patterns between stress and baseline readings and the presence of highly correlated features within the dataset results in compromised generalizability and overfitting. For accurate classification, the stress-related features should not only be informative but should also be well-distinguishable and interpretable by the classification models. Thus, further studies are required to develop feature extraction and feature selection models that can accurately estimate stress-related indicators/features and are well-understandable by the classifiers.

Chapter 3
Statistical Analysis of Stress Indicators

The literature review revealed a long list of different signals/indicators of stress with different conclusions for each as well as a combination of these indicators. Most stress detection and monitoring studies report only the classification results and lack a statistical analysis of the extracted features. Also, there is no clear understanding of the relative sensitivity and specificity of these stress-related indicators of stress in the literature. Thus further statistical analysis and classification modelling of these signals/indicators were required to assess the relative sensitivity and specificity of common physiological as well as biochemical indicators of stress.

This chapter[3] presents a statistical and classification analysis, performed on the list of biophysiological indicators of stress to determine the most specific and sensitive stress indicator/s using the WESAD dataset [283]. This work covers the fourth objective of the thesis which was investigating and shortlisting the stress indicators that are more sensitive and specific for stress monitoring and are less affected by other environmental factors.

The analyzed physiological indicators included heart rate, respiratory rate, skin conductance, RR interval, heart rate variability in the electrocardiogram, and muscle activation measured by electromyography. A comparative analysis has been performed by applying a deviance analysis and t-test to validate the hypothesis that the physiological data for each variable for the stress and non-stress (baseline) states is statistically differentiable, and logistic regression was applied to identify the strongest predictor of stress. The results of the study suggest that respiratory rate is the strongest (stand-alone) predictor while heart rate emerged as the second-best predictor of stress.

---

[3] *The following body of the chapter is exact copy of the paper published in IEEE Access (2021). I am the first lead author in the paper, which is co-authored with my supervisors. The conceptualization, formal analysis, investigation and visualization are done by me. Methodology design and validation were led by my supervisors. I led all parts of the work with the support of my supervisors.*

An analysis of different biochemical indicators is also presented (Section 3.6). The analysed biochemical indicators included adrenaline, noradrenaline, copeptin, prolactin, plasma catecholamines, alpha-amylase, estradiol, testosterone, and cortisol. The investigation of different biochemical parameters revealed that there is a gap in determining an accurate way of stress detection. However, the detection of cortisol levels in the sweat showed the potential of being the most significant stress indicator due to its characteristics of the non-invasive collection, risk-free and no pain while collecting the sample. The sweat colourimetric sensors are frequently developing and have shown great potential of becoming the only method of detecting stress in real-time using the biochemical parameter.

3.1 A Sensitivity Analysis of Biophysiological Responses of Stress for Wearable Sensors in Connected Health

ABSTRACT: Stress is known as a silent killer that contributes to several life-threatening health conditions such as high blood pressure, heart disease, and diabetes. The current standard for stress evaluation is based on self-reported questionnaires and standardized stress scores. There is no gold standard to independently evaluate stress levels despite the availability of numerous biophysiological stress indicators. With an increasing interest in wearable health monitoring in recent years, several studies have explored the potential of various biophysiological indicators of stress for this purpose. However, there is no clear understanding of the relative sensitivity and specificity of these stress-related biophysiological indicators of stress in the literature. Hence this study aims to perform statistical analysis and classification modelling of biophysiological data gathered from healthy individuals, undergoing various induced emotional states, and to assess the relative sensitivity and specificity of common biophysiological indicators of stress. In this chapter, several frequently used key indicators of stress, such as heart rate, respiratory rate, skin conductance, RR interval, heart rate variability in the electrocardiogram, and muscle activation measured by electromyography, are evaluated based on a detailed statistical analysis of the data gathered from an already existing, publicly available WESAD (Wearable Stress and Affect Detection) dataset. Respiratory rate and heart rate were the two best features for distinguishing between stressed and unstressed states.

INDEX TERMS: Stress monitoring, biophysiological stress response, sensitivity analysis, heart rate, respiratory rate, skin conduction, electrocardiogram, electromyograph.

3.2 Introduction

It is well understood that every human being is exposed to some level of stress more than once in their lifetime. Stress can be defined as a non-specific response of our body to meet a certain demand in extreme conditions [19]. It has been seen that stress generally has negative effects on the mental health and well-being of a person [284]. Acute stressors

(stimuli that cause stress) may not impose any health burden on young and healthy people having an adaptive and good coping response, but if the stressors are too persistent or too strong, these stressors may lead to depression and anxiety [77]. Chronic stress is known to contribute to life-threatening conditions such as heart disease, high blood pressure, diabetes, and obesity, and an acute episode of stress can trigger a heart attack or stroke by causing arterial inflammation [285].

The current standard for clinical evaluation of stress is based on self-reported questionnaires or standardized stress scores, such as the Perceived Stress Scale (PSS) [19]. However, with the recent development in wearable biosensor technologies, a huge interest has been seen in measuring biophysiological responses to stress for the evaluation and monitoring of stress. To develop a reliable device for stress monitoring, it is important to understand how stress affects the human body from a physiological and biochemical point of view. Under the influence of a stressor, the stress triggers the sympathetic nervous system, causing the release of various hormones such as adrenaline or cortisol [14], [15]. The release of these hormones leads to changes in heart rate, and respiratory rate, and causes muscle tension among other physiological responses. These changes in the body prepare the individual for a physical fight or flight reaction. The changes caused in both the biochemical and physiological state of the human body in response to stress can be observed and used as an indicator of stress. Physiological indicators are of particular interest due to the possibility of measuring these indicators non-invasively.

The wearable sensor technology has progressed to the level that several physiological parameters can be measured continuously as well as wirelessly. Some real-time stress-detecting models have been described [37], [38], [286]. Many sensor-based stress monitoring devices and research studies exploited the relationship between stress and resulting physiological variations [20]–[23], [287]. These include machine learning techniques to detect stress from physiological and activity data collected from respiration (RESP), electrocardiogram (ECG) and accelerometer (ACC) sensors [24], [25]. Other less frequently used indicators are blood volume pulse (BVP), skin temperature (TEMP), electromyography (EMG), photoplethysmogram (PPG) and electrodermal activity (EDA), which have also been recorded and used for stress monitoring [26], [27]. These physiological indicators are not specific to stress response, therefore, the stress prediction

based on the physiological indicators may have varying accuracy, be it for any individual indicator or a combination of these indicators.

Besides the physiological indicators, several biochemical indicators are also used for stress detection. In humans, these indicators include the level of cortisol, adrenaline, alpha-amylase, copeptin and prolactin [76], [77], [288]. Among these indicators, cortisol is considered the primary stress hormone [60]. Several techniques have been proposed to measure cortisol levels in saliva, sweat and hair [289]–[291]. This chapter is intended to focus on and provide a sensitivity analysis of biophysiological indicators of stress rather than biochemical ones. A comprehensive review of biochemical stress indicators can be found elsewhere [292].

3.3 Related Work

Han *et al.* [27] proposed a stress detection technique that detects three levels of stress i.e. no stress, moderate stress, and high perceived stress using ECG and PPG signals. The authors collected data from 39 subjects and reported a classification accuracy of 84% using a random forest and support vector machine (SVM) classifier for the three-stage classification of stress. For binary classification i.e. rest and stress, an accuracy of 94% was achieved. Choi *et al.* [73] proposed a wearable device to measure the stress, drowsiness, and fatigue of vehicle drivers. Stress indicators they measured were (Galvanic Skin Response) GSR, activity data from the accelerometer, skin temperature, and PPG signals of 28 drivers. The authors reported an accuracy of 68.3% for four classes i.e. normal, stressed, drowsiness, and fatigue and an accuracy of 84.5% for three classes i.e. normal, stressed, drowsiness or fatigue classification. Mohino-Herranz *et al.* [89] assessed the mental fitness of different subjects. They used ECG and Thoracic Electrical Bioimpedance (TEB) signals to monitor the stress of 40 subjects. The authors achieved error rates of 21.2%, 32.3%, and 4.8% for activity identification, mental activity, and emotional state, respectively. Liu *et al.* [86] determined the feasibility of the EDA signal parameter and developed a stress-monitoring device. The authors used only EDA signals for the detection of the stress of 11 drivers. After computing Fisher projection and Linear Discriminant Analysis (LDA), the authors reported a classification accuracy of 81.8% by only using EDA.

Lee *et al.* [87] and Healey *et al.* [67] wanted to develop a wearable glove that could detect the stress of drivers and collect data from 28 and 10 drivers, respectively. The authors of both studies recorded PPG signals for analysis. Lee *et al.* achieved a classification accuracy of 95% while Healey *et al.* reported an accuracy of 62.2% using only PPG signal (respiratory rate) recorded through sensors on the driver's gloves. Similarly, Wijsman *et al.* [71] and Healey *et al.* [67] developed an algorithm for detecting mental stress using physiological signals. These signals included ECG, PPG, EMG, and EDA from 18 and 10 subjects, respectively. Wijsman *et al.* [71] claimed to reach 80% accuracy for two-class (i.e. rest and stress) classification and concluded that the accuracy indicates that these features are suitable to be used for an individual's stress detection while Healey *et al.* reported a classification accuracy of 86.6% using the same physiological signals.

Chen *et al.* [65], Shi *et al.* [70] and Kim *et al.* [63] developed a stress detection system based on multimodal features and kernel-based classifiers using ECG, EDA and PPG signals. The studies collected data from 14, 22, and 175 subjects, respectively. Chen *et al.* analysed the data in terms of precision, sensitivity and specificity. While using a full feature set, SVM with a linear kernel gave the highest inter-drive classification precision. For the cross-driver analysis, the SVM with radial basis function (RBF) kernel gave a precision score of 89.7%. Shi *et al.* concluded that the SVM-based model detected stress with high precision and recall rate and classification accuracy of 68%. Kim *et al.* reported that they achieved a classification accuracy of 78.4% for three emotional states classification problems and an accuracy of 61.8% for four-state classification problems. Sun *et al.* [293] determined the mental as well as the physical stress of 20 subjects during different physical activities. The authors used ECG, EDA, and accelerometer signals. They reported a classification accuracy of 92.4% using accelerometer data along with ECG and EDA physiological signals. The inter-subject classification accuracy was reported to be 80.9%.

Mozos *et al.* [88] and Sandulescu *et al.* [72] presented a stress detection methodology for people who suffer from stress in social situations. Both studies used EDA and PPG signals for stress detection collected from 5 and 18 subjects, respectively. After experimentation, Mozos *et al.* reported an accuracy of 92 % with the SVM (RBF kernel) classifier, in comparison to Linear kernel SVM (80%), AdaBoost (67%) and k-nearest neighbours (KNN) (62%), when using a selected set of features. Sandulescu *et al.* were successful in

classifying the stress of each participant with an average accuracy of 79%. The authors concluded that their approach is a good starting point for the detection of a subject's stress state in real-time. Such detection alongside some intervention in real-time may improve quality of life.

Muaremi *et al.* [294] presented a stress detection system using a smartphone and wearable chest belt. The authors evaluated their system in a real-world environment with 35 test subjects studied for 4 months. The prediction accuracy was calculated using the leave-one-out-cross-validation (LOOCV) method. The system achieved a 55% accuracy using mobile phone features only (accelerometer) while a 59% prediction accuracy was obtained using the heart rate variability (HRV) feature. The combination of both features gave a prediction accuracy of 61%. Lai et al [295] described an intelligent stress monitoring assistant (SMA) prototype and used a deep learning-based method for stress detection using the WESAD dataset. The authors used Residual-Temporal Convolutional Network (Res-TCN) to recognise and detect stress states with an accuracy of 86% and 96%, respectively. Smets et al [296] used a data-driven approach for stress detection, using real-life data obtained from 1002 subjects in five consecutive, free-living days. The authors found a significant difference between the ECG, skin conductance and skin temperature for different stress levels. They compared their self-reported data with the standard digital phenotypes-based wearable device and achieved the F1-score (a measure of test accuracy using precision and recall) of 0.43, which suggests that the physiological stress response varies greatly between individuals. Thus, stress detection systems should apply personalised models for accurate stress detection.

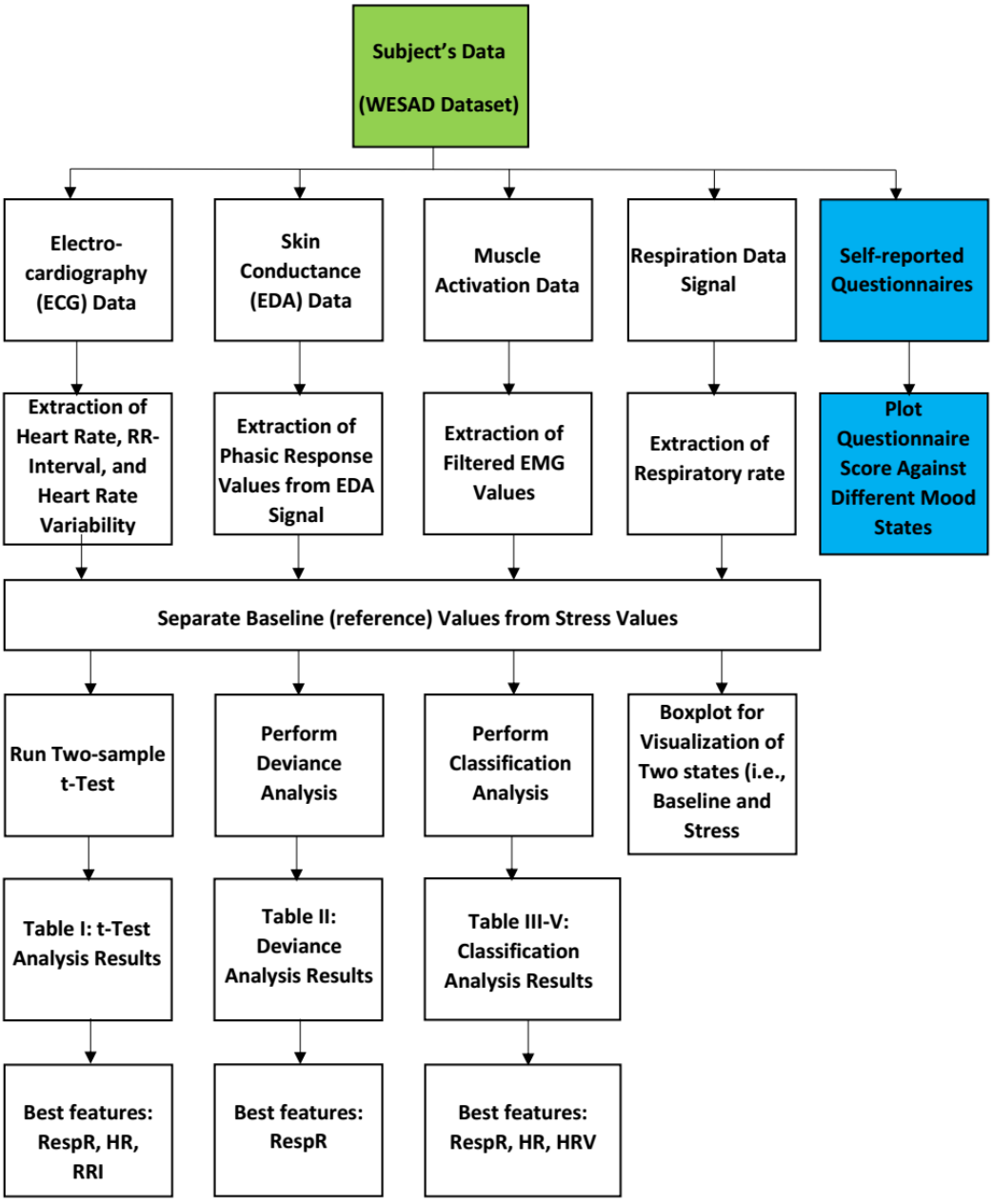


Figure 3.1 Block diagram of the proposed Study.

From a review of the related literature, it can be noticed that the best features for stress detection and monitoring are still unclear. Several studies have used the same physiological parameters and have implemented the same classifier, yet reported different accuracies. It is important to note that no study has previously tried to find what parameter is the best predictor for stress. The first step should be to establish a statistically significant difference between the baseline and stress state before developing any machine learning model. Most

studies have reported machine learning models for stress identification using various features available from the sensor that was used to collect the data.

The results reported in these studies may only apply to those particular features or sensors. The main objective of this study is to analyze the relative importance of the most common and clinically relevant biophysiological stress indicators and identify the most useful specific indicators for a wearable sensor-based stress monitoring solution. Most previous studies have focused on the determination of sensor/signal ratios. Our study ranks biophysiological stress indicators in order of diagnostic performance using single and multivariable (deviance) analysis. This is a commonly used approach to assess the predictive model, in this case, the stress state response. Figure 3.1 summarizes the pipeline of the proposed work.

The only other study closest to this work is by Zhen et al. [297]. According to the authors, the improper imposition of workload on pilots is the most critical cause of human error. Thus, the authors studied different physiological responses of pilots during flight. These parameters included eye blinking, saccade, pupil diameter, fixation, respiratory rate, and heart rate. They performed statistical analysis to check the sensitivity and diagnostic ability of the aforementioned physiological parameters. They collected data from 12 healthy student pilots and applied a one-way ANOVA test to the collected data. After the experiment, they concluded that from all the physiological parameters, pupil diameter and respiratory rate turned out to be the most sensitive parameters in distinguishing different stages. The diagnostic capability of the parameters was different. Respiratory rate and eye blinking were directly related to the difficulty of the task (stress) while other parameters were affected by external factors, for example, fatigue and attention.

The advantages of our study over the Zhen *et al.* study are three-fold. First, the set of stress-measuring features analysed in their study is different from ours. Secondly, we have performed descriptive and regression analyses. Thirdly, we developed a classification task to evaluate the sensitivity and specificity of selected biophysiological parameters for stress detection. The analysis was performed on a publicly available dataset collected in the Wearable Stress and Affect Detection (WESAD) project [19]. The main objectives of this study are:

- Descriptive analysis of the commonly used stress monitoring features.
- Regression analysis for the selection of the most important features that can be used for stress monitoring devices in the future.
- Implementation of a uni-variable and multi-variable classification model (using logistic regression) to classify the stress state from the non-stress state of an individual.

3.4 Methodology

3.4.1 Study Participants

The data were collected using two multimodal devices: a chest-worn device (BioSignalPlux RespiBAN Professional); and a wrist-worn device (Empatica E4). Some recent studies that have used WESAD datasets are Reiss *et al.* [298], Jiang *et al.* [299], Aridas *et al.* [300] and Taufeeq *et al.* [301]. The data included a high-resolution measurement of BVP, EMG, EDA, ECG, RESP, TEMP, and movement from ACC. All the participants were healthy graduate students of the University of Siegen, Germany [19]. Study participants with mental disorders, heavy smoking, pregnancy, or those suffering from any cardiovascular and other chronic diseases were excluded from the study. A total of 17 individuals participated in the study but the data of two participants were incomplete due to the malfunctioning of sensors and were therefore removed from the dataset. There were 12 males and 3 females in the remaining 15 subjects with a mean age of 27.5 ± 2.4 (SD) years. Some of the variables were missing in the data from subject no. 11. Thus, all analyses for this study were completed using 14 subjects. In the dataset, there are 11,500,000 baseline (non-stress) samples and 6,400,000 stress samples.

3.4.2 Features Related to Stress

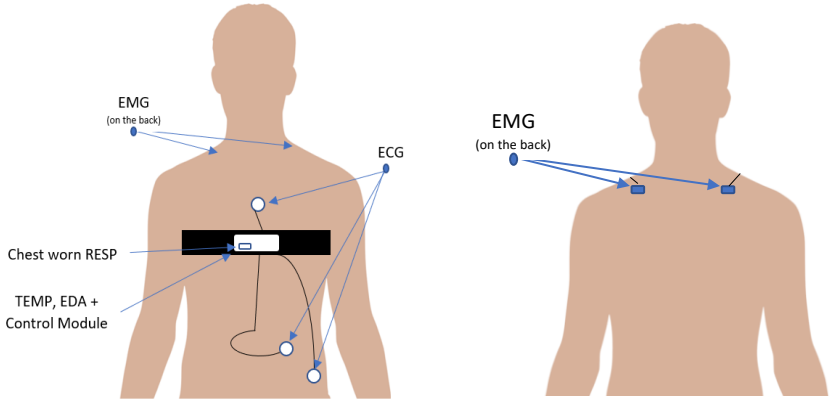
During stress, the heart rate usually increased, thus causing more blood to flow within the body. This change in blood flow can be measured through BVP, which is derived from a PPG signal. Changes in heart rate and heart rate variability can also be monitored using ECG signals [51], [52]. Stress also causes the release of sweat, thus changing skin conductance properties. This change is measured by the EDA device. There is vast literature available that demonstrates the association of muscle tension with stress. Muscle tension changes are measured using EMG signals [53], [54]. In some people, chronic stress causes a low-grade fever (between 99^o to 100^oF) and may also cause anxiety as well as

restlessness. Thus, Temperature (TEMP) sensors and accelerometer (ACC) readings can also be used to monitor stress [55]–[57].

3.4.3 Setup and Placement of Sensors

The chest-worn device, RespiBAN Professional, was used to record ECG, EMG, EDA, TEMP, and RESP along with additional ACC data. The placement of the device control unit and sensors is shown in Figure 3.2. The data from RespiBAN Professional was sampled at 700Hz. The ECG signal was recorded using a standard 3-lead approach (as shown in Figure 3.2) and an inductive respiration sensor was used to record the RESP signals.

The EDA signals were recorded from the abdomen and EMG was recorded from the muscles of the upper trapezius on both sides of the spine. In addition, Empatica E4 was worn on the dominant hand by all subjects, and BVP, EDA, TEMP, and ACC signals were recorded at the sampling rate of 64Hz, 4Hz, 4Hz, and 32Hz, respectively. All the participant data were recorded on the devices and then transferred to a computer through a wired connection. On the day of the study, upon arrival, participants were equipped with chest and wrist-worn sensors. A functionality test was performed to test the working of the sensors. After that, both devices were synchronised using a double-tap gesture, manually.



(a) Front View (b) Back View

Figure 3.3 Placement of RespiBAN Professional Device (a) shows the placement of different sensors on the front of the human body (b) shows the placement of EMG sensors on the back of the body.

3.4.4 Study Protocol

The study protocol was designed to record readings of 3 different states of the participants, i.e. baseline, amusement, and stress. Participants were also asked to complete a self-reporting questionnaire after each session and undergo a guided meditation session to get de-excited after amusement and stress conditions. Participants could not intake tobacco or caffeine one hour before the study commenced. Moreover, the participants were asked to avoid strenuous exercise on the day of the study. All study participants signed informed consent before commencing. A short sensor test was conducted while equipping the participants. Finally, both the devices (RespiBAN Professional and Empatica E4) were manually synchronized.

For baseline readings, participants were asked to stand or sit at a table and read a magazine. Baseline readings were recorded for 20 minutes and were labelled as a baseline state. An amusement state was induced by showing eleven different funny clips with a gap of 5 seconds between them. The total length of the amusement state was 392 seconds for each participant.

The stress condition was induced using the Trier Social Stress Test (TSST) [164]. TSST consists of mental arithmetic and a public speaking task. Both tasks are considered reliable to evoke stress [25] as they inflict a high mental load and are categorized as a social-evaluative threat to subjects. The participants had to deliver a speech for five minutes on their strengths and weaknesses in front of a panel. Participants were told that the judging panel is from the human resource department and that impressing them will increase their hiring chances. After the speech, the panel asked each participant to count backwards, with a gap of 17, from 2023 to 0. If the participant makes any mistake while counting, they had to start over. This exercise following the speech also lasted for five minutes. So, TSST was conducted for a total of 10 minutes. After TSST, participants were given a rest period of 10 minutes. After the amusement and stress period, participants were asked to perform some predefined meditation steps to de-excite and bring them back to a neutral state. Meditation included controlled breathing instructed through an audio track. After removing the sensors, participants were told that the panel was of normal researchers so that they can recover from the test-induced stress.

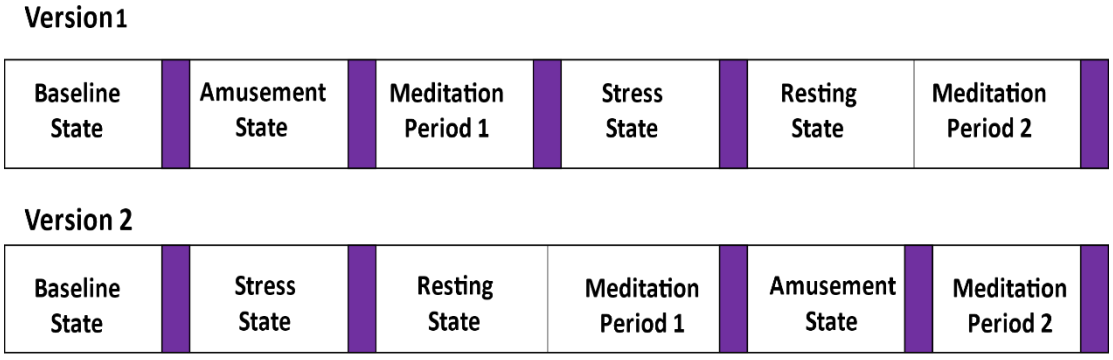


Figure 3.4 The two distinct protocol versions for the proposed study. The dark boxes indicate filling out self-reporting questionnaires.

As humans are naturally good at adapting to different situations quickly, two study protocols were designed for this study to keep the randomness and collect the true feelings of the subjects. The two protocols are shown in Figure 3.3. Half of the subjects followed the version 1 protocol while the other half followed the version 2 protocol.

3.4.5 Signal Processing and Feature Extraction

Raw data from all the sensors (ECG, EDA, EMG, and RESP) were collected using a 0.2-second non-overlapping sliding window, and all physiological features, except EMG, were computed using a 60-second non-overlapping sliding window. The window sizes were chosen following the recommendations of Koelstra *et al.* [302].

From raw signals of ECG, the heart rate was calculated using the Hamilton peak detection algorithm [303]. Moreover, heart rate variability (HRV) was derived from the locations of the peaks in ECG. Figure 3.4 shows the block diagram of the Hamilton peak detection algorithm. The algorithm works on the detection of the QRS complex in the ECG signal. The preprocessing steps involve rectification of the signal rather than squaring the signal as in [304], averaging sliding windows, and low and high pass filtering followed by some QRS detection rules. Rectification of the signal gives us better sensitivity to the detection algorithm, which is also indicated in [305]. The QRS complex detection rules are as follows:

- Ignore all the detected peaks preceding or following larger peaks by less than 200 milliseconds.
- If the peak is detected, check whether the signal contains both positive and negative peaks. If not, the detected peak represents a baseline shift.

- If a peak is detected within 360ms of the previously detected peak and had a maximum slope less than 50% of the maximum slope of the previous peak then assume it is a T-wave.
- If the detected peak is larger than the detection threshold then consider it a QRS complex otherwise consider it noise.

The detection threshold is calculated using estimates of QRS peaks and noise peaks heights and is mathematically represented as:

$$\text{Detection\_threshold} = \text{average\_noise\_peak} + TH * (\text{average\_QRS\_peak} - \text{average\_noise\_peak}) \quad (3.1)$$

In equation 3.1, TH denotes the threshold coefficient between 0.3124 and 0.475. Each time the QRS complex is detected, it is stored in a buffer with the previous eight most recent peaks while every non-QRS complex is stored in a buffer that contains the previous eight non-QRS peaks also called noise peaks. Through equation 1, we set the detection threshold between the mean or median of QRS and noise peaks. The noise detection is done similarly to [306]. The algorithm characterizes low-frequency noise by the interval between the end T-wave and the start of the P-wave while high-frequency noise by bandpass filtered beats outside the QRS complex. In this study, we have used the heart rate and RR interval extracted from the ECG signal using the above-mentioned algorithm.

The sympathetic nervous system controls the EDA response that provides high arousal states with high sensitivity. EDA signals were first passed through a low-pass filter with a critical frequency of 5 Hz, similar to work reported in [307], [308] and phasic (skin conductance response) and tonic (skin conductance level) components were extracted. The phasic component is a short-term response due to some stimulus while the tonic component shows a slow variation in baseline conductance. EDA features can be found in [309], [310]. In this study, we used phasic components.

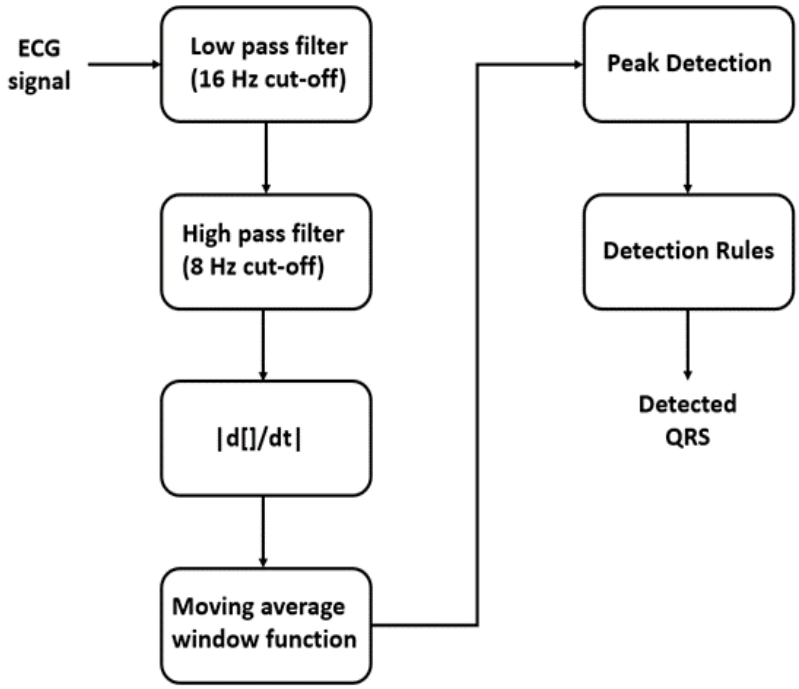


Figure 3.5 Steps to detect QRS complex using Hamilton peak detection algorithm.

The raw EMG signal was processed in two steps. In the first step, the DC component was removed using a high-pass filter and the peak frequency was calculated from the filtered signal by applying a 5-second window. In the second step, a raw EMG signal was passed through a low-pass filter with a cut-off frequency of 50 Hz, to suppress the power line noise, and features were extracted using the method described in [311]. A normalized root means squared (RMS) value of EMG voltage amplitude is used as a feature in this study.

The RESP signal was used to extract the respiratory rate (RspR). Before computing the features of respiration, the raw signal was filtered using a band-pass filter with critical frequencies of 0.1 and 0.35 Hz. A peak detection algorithm was used to identify minima and maxima in the signal and inspiration volume, respiration duration, respiration rate, and inhalation and exhalation ratio were derived as in [25].

3.4.6 Statistical Features

We normalized the data from the phasic component of skin conductance, muscle activation, heart rate (HR), RR-interval (RRI), heart rate variability (HRV) and respiratory rate (RspR) using min-max normalization to eliminate initial variation in the readings. Data for each are summarized using the mean and standard deviation separately for stress and

baseline scenarios. All features along with their mathematical representation are listed in the next subsections.

Let us suppose the above physiological signals are x and x_i is an i-th sample of the signal within the sliding window, where i= 1,…, n. Then:

1) *MEAN:* Mean is denoted by $\bar{x}$ and represents the mean value of a raw signal within a sliding window. The mean is calculated by the following equation:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (3.2)$$

2) *STANDARD DEVIATION:* Standard deviation is denoted by *S* and represents the deviation of raw signal around the mean of the signal within the sliding window. Standard deviation is calculated using the following equation:

$$\boldsymbol{S} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (3.3)$$

3) *MEDIAN:* Median corresponds to the cumulative percentage of 50% i.e. middle reading in a dataset. It is calculated using the equation:

$$\boldsymbol{median} = \left(\frac{n+1}{2}\right)^{th} \boldsymbol{value} \quad (3.4)$$

Here *n* is the total number of entries in a dataset.

3.4.7 Stress Evaluation Methodology: Questionnaire

To validate the protocol, four different self-reports were filled out by each participant after every session. First, participants filled out a Positive and Negative Affect Schedule, also known as PANAS. In the second place, six items were picked from the State-Trait Anxiety Inventory (STAI) to measure the anxiety level of each participant. Thirdly, a Self-Assessment Manikins questionnaire (SAM) was used to generate labels in the valence arousal space. Finally, nine items were included in a questionnaire from the Short Stress State Questionnaire (SSSQ) to identify the type of stress that prevailed [19]. The outcome of these questionnaires can be considered as subjective reports showing how the participants felt during the test and can be used to train any personalized model. However, for the defined dataset, the study protocol was used to differentiate between the three states

and therefore contribute to labelling different readings. The options of answers were given for each questionnaire. The PANAS questionnaire was answered using 5 points scale (1 = not at all and 5 = extremely). The questionnaire asked the subjects about their emotional state i.e. stressed, happy, sad, or frustrated. STAI questionnaire was answered on 4 points scale (1= not at all, 4= very much so) and included questions about the subject's feelings i.e. were they feeling nervous, relaxed, worried, pleasant, jittery, or at ease. Valance and arousal were scored on a scale from 1 = low to 9 = high. The SSSQ questionnaire included questions about what the subject's mindset was while answering the questionnaire. Subjects answered on 5 points scale where 1=not at all and 5=extremely.

The self-reports were also analysed to make sure that the designed experiment was suitable for inducing stress and manipulating the subject's affective states. Authors in [19] calculated the mean and standard deviation of the anticipated self-reports of three states i.e. baseline, amusement, and stress states along with their subscales. The result of the analysis is shown in Figure 3.5. After baseline and amusement states, the comparison of self-reports revealed that the amusement state had the desired effect on the subject i.e. the subject reported score was high in valence and arousal (dimensional approach, DIM) and less in STAI (anxiety).

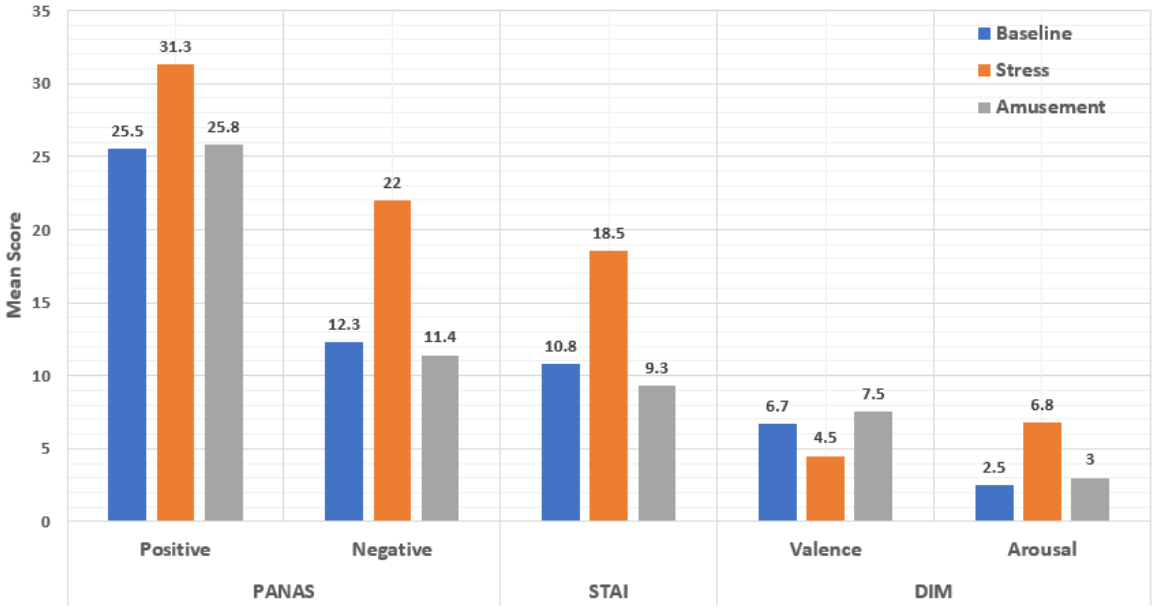


Figure 3.6 Analysis result of self-reported questionnaires.

The impact of induced stress was noticeably pronounced across all the questionnaires. Analysis of the SSSQ score revealed that the subjects felt more worried and engaged as compared to distressed during the Trier Social Stress Test (TSST) tasks. The score calculated is worrying = 10.6, engaged = 11.7 and stressed = 6. The higher value of the positive affect (PA) score shows that the subject felt energetic and concentrated during the TSST tasks which also resulted in a higher engagement score in SSSQ. The elevated score of negative affect (NA) indicated an increased level of the subject's stress. The dimensional approach (DIM) score also supports these observations by indicating an increase in arousal score and a decrease in valence score. We had a higher STAI score after TSST, as expected for a subject in a stressful state. Overall, the analysis of self-reported questionnaires revealed that the designed experimental protocol was suitable to induce desired effective stress in the subjects, especially in stress conditions.

3.4.8 Statistical Analysis

For the statistical analysis, only two-state data (Baseline and Stressed states) were used to evaluate the relative importance of each physiological indicator of stress for prediction. Two types of analysis were performed: 1) an independent analysis for each biophysiological indicator via a two-sample t-test under the null hypothesis that the mean biophysiological indicator is equal during the Baseline and the Stressed States; 2) a multivariable (deviance) analysis to rank the contribution of each biophysiological indicator in a logistic regression model, defined as follows:

$$log\left(\frac{p(Stress)}{p(Baseline)}\right) = c_0 + c_1 EDA + c_2 EMG + c_3 RRI + c_4 HR + c_5 RspR \quad + c_6 HRV \quad (5)$$

The logit link function $log\left(\frac{p(Stress)}{p(Baseline)}\right)$ is used (p is the probability) to relate the log odds of being stressed to the linear predictor where (c_0,c_1,c_2,c_3,c_4,c_5 and c_6 are the coefficients showing the direction of the relationship); 3) logistic regression classification analysis with 30-70 split, 4-fold cross-validation and leave-one-out cross-validation to determine the mean absolute error, root mean square error, classification accuracy, sensitivity and specificity of the model.

3.4.8.1 A two-sample t-test

The data statistics and the results of the t-test are provided in Table 3.1. The units of each feature are RspR (breaths per min), HR (beats per min), RRI (milli-sec), Phasic EDA (micro-siemens), and EMG (micro-volts) and HRV (milli-sec). The p-value > 0.05 shows the relevant feature has a non-significant mean difference between stress and baseline state values while the p-value <0.05 shows a significant difference in the mean values of stress and baseline condition.

3.4.8.2 Deviance analysis

In logistic regression, deviance can be used to assess how good the model is to predict the response (which in this case is stress state) – the lower the deviance, the better the fit to the sample data. To analyze the independent effect of the variables in determining stress, separate regression models were constructed for combinations of indicators, and the deviance is then used to measure the strength of the relationship between the response and independent variables. Deviance analysis using logistic regression was performed using MATLAB's statistics toolbox while the classification model was developed using Python code.

Table 3.1 Statistical analysis result of physiological parameters.

Features	Mean		Standard Deviation		Median		t-test (p-values)
	Baseline	Stress	Baseline	Stress	Baseline	Stress	
RspR	15.51	12.35	2.48	2.43	15.65	12.22	6.02e-08
HR	66.9	95.35	20.01	20.45	68.8	90.66	0.0003
RRI	871	710	157	132	863	702	0.003
Phasic EDA	3.75	5.24	2.92	4.27	2.58	3.89	0.310
EMG	9.5e-5	1.1e-4	5.3e-5	10.8e-5	8.1e-5	8.8e-5	0.660
HRV	51.13	46.97	25.79	22.92	44.33	42.95	0.668

*Units: RspR (breaths/min), HR (beats/min), RRI (milli-seconds), Phasic EDA (micro-siemens), EMG (micro-volts) and HRV (milliseconds)

3.4.8.3 Classification Methodology

For the regression model, logistic regression was selected instead of linear regression since our dependent variable in this study is binary i.e. Stress vs No-stress/Baseline. Logistic regression uses the maximum likelihood method to arrive at the solution. Also, the logistic

loss function causes large errors to be penalized to an asymptotically constant. The dataset has baseline (11,500,000 per subject) and stress (6,400,000 per subject) samples of the 14 subjects. Given that the number of subjects in the dataset is small, we used stratified k-fold cross-validation with k=14, to ensure that the results achieved are generalizable. Stratified k-fold cross-validation ensures the selection of the same proportion of samples of each class in each fold. The time complexity of the k-fold is measured by $O(Kn)$, where n is the number of samples. The $O(Kn)$ means that the experiment is repeated K time. When K approaches n, the time complexity becomes $O(n^2)$. So, it can be concluded that as the value of k increases, the systems become complex and computationally expensive. We evaluated the classification model using leave-one-out cross-validation (LOOCV) to have an unbiased estimate of the model performance. Stratified k-fold cross-validation differs from simple k-fold cross-validation by splitting the dataset in such a way that the mean values of all the splits are almost equal.

3.5 Results and Discussion

The data from 14 participants were used in the analysis, as some of the variables were missing in the data from subject no. 11. Figure 3.6 shows the distribution of the data in two states for each variable on boxplots. The data statistics and the results of the t-test are provided in Table 3.1. Logistic regression is used whenever the outcomes of the analysis are limited, which in this case is stress and baseline (unstress). Thus, logistic regression is used to perform a deviance analysis. Similarly, for the classification task, the response variables (classes) are categorical (yes/no or true/false), so the logistic regression classifier fits best for such type of classification problem and is used as a stress versus non-stress classifier.

3.5.1 A Two-Sample t-Test

Based on the analysis of the p-values, the magnitudes of the coefficients in the logistic regression, and the effect of each variable on deviance, it can be concluded that respiratory rate (RspR) is the best predictor of stress among these six variables. This result re-enforces the outcome of Zhen *et al.* [297] that respiratory rate is the most specific and sensitive parameter out of all the other physiological parameters and could be used as a stand-alone parameter to detect stress in the lab as well as in the natural environment. Heart rate (HR)

combined with respiratory rate can provide a slight improvement in the evaluation or monitoring of stress using wearable sensors. On the other hand, electrodermal activity and electromyograms are poor predictors of stress and may not add value to the wearable stress monitoring system.

3.5.2 Deviance Analysis

Table 3.2 shows the results of the deviance analysis of the fit for single and multi-variant logistic regression models. The values are sorted in decreasing deviance order. The lowest value model is the best. The deviance decreases when the model includes RspR compared to those without RspR. Interestingly, a single variable model, comprising only the RspR, fits better than the multivariable model using EDA, EMG, HR and RRI together. Using any other feature in combination with RspR achieves deviance of close to 0, suggesting a perfect fit for these 14 individuals. Without further samples, it is unclear which combination is optimal.

The box plot of six variables (EMG, EDA, RspR, RRI, HRV and HR) shows that there was evidence of a difference in mean values of RspR, RRI, HRV and HR between baseline and stress states. On the other hand, there was a little difference in mean EMG and EDA for stress and baseline states (see Figure 3.6), which is also evident in the results of the t-test (see Table 3.1).

From the above-mentioned results, it can be concluded that the data of EDA and EMG cannot be separated easily, which reinforces our p-value (i.e., the p-value is greater than 0.05) and deviance analysis (Table 3.2) results. The data of respiratory rate can easily be separated using any logistic fitting curve and thus, qualifies as a most distinctive feature to distinguish baseline state from stress state.

Table 3.2 Logistic regression model fitting (largest to smallest).

Features	Deviance
HRV	39.00
EMG	38.60
EDA	37.65
EDA + EMG	37.62
RRI	25.00
HR	23.73
EDA+EMG+HR	21.49

EDA+EMG+HR+RRI	21.06
RspR	4.70
All	0.00001

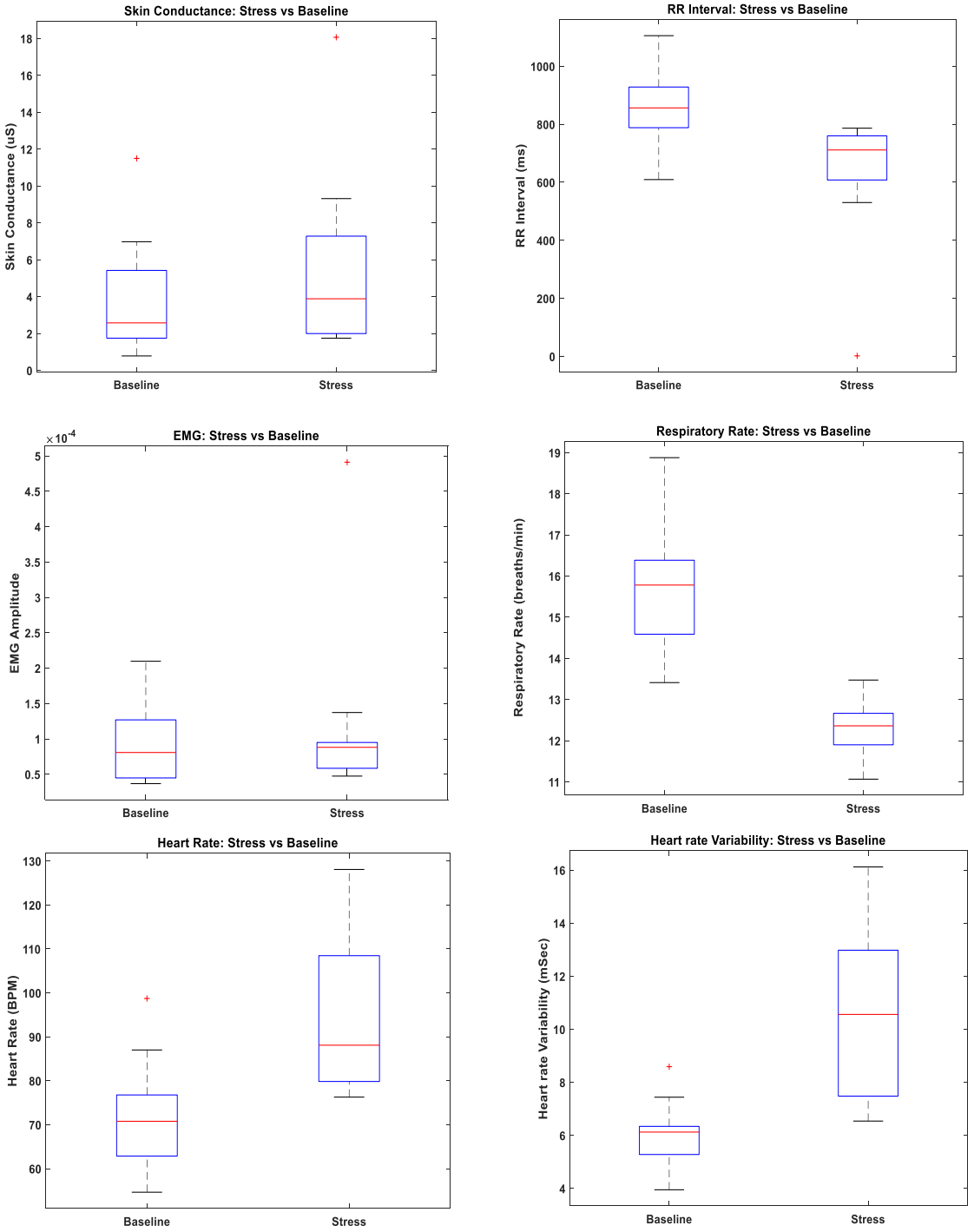


Figure 3.7 Boxplots of the data from two states (Baseline and Stress) for EDA, EMG, HR, RRI, RspR, and HRV, respectively.

3.5.3 Classification Methodology

The result of the logistic regression classification model is shown in Table 3.3, Table 3.4 and Table 3.5. The table shows the test-train split for classification, classification accuracy, sensitivity, specificity, 95% confidence interval of sensitivity and specificity along with likelihood ratios of the developed model. Table 3.3 shows the results of a single Test-Train run. As the results achieved may not be generalizable due to a small number of training and testing samples, therefore cross-validation is performed. Firstly, k-fold cross-validation is performed using k=4. However, since the number of subjects in the dataset is very small (n=14), 4-fold cross-validation may still be not generalizable. Therefore, leave-one-out-cross-validation is also performed to obtain a more robust estimate of the model performance. From the tables, we can see that among the analysed physiological parameters, respiration rate and heart rate give better accuracy than RR interval, skin conductance, muscle activation and heart rate variation. The combination of respiratory rate, heart rate and heart rate variability gives us almost the same accuracy as the combination of all six parameters. So, we can conclude that the combination of respiratory rate, heart rate and heart rate variability, which can be calculated using a single PPG sensor, is the best predictor of stress.

3.5.4 General Discussion

While these classification results indicate the potential of the logistic regression (machine learning) technique to predict stress using the above features, there is still a question of generalizability due to the very small size of the dataset, despite rigorous cross-validation. Therefore, these results need to be validated using a larger dataset. The main conclusion from all three analysis results is that RspR is the best singular feature for detecting stress (from the tables) while the combination of RspR and HR (RRI) are key multi-features of stress, with HRV emerging as the next best.

Table 3.3 Logistic regression classifier results (using 4-fold): Accuracy, Sensitivity, Specificity and Confidence Intervals, Likelihood ratio.

Features	Test- Train Split	Accuracy (%)	Sensitivity	Specificity	LR+	LR-
EMG	30-70 %	64.91	1.00	0.00	1.00	inf
EDA		63.82	0.93	0.09	1.03	0.73
RspR		77.21	0.87	0.59	2.11	0.22
RRI		76.40	0.91	0.49	1.78	0.18
HR		77.10	0.89	0.56	2.12	0.21

HRV		66.73	1.00	0.51	2.12	0.00
HR +RspR		83.82	0.90	0.72	3.25	0.14
HR+ RspR +HRV		88.88	1.00	0.83	6.00	0.00
All combined		88.90	1.00	0.86	7.00	0.00

Table 3.4 Logistic regression classifier results (using 4-fold): Accuracy, Sensitivity, Specificity and Confidence Intervals, Likelihood ratio, Variance, and Standard Deviance.

Features	**Test-Train Split**	**Accuracy (%)**	**Sensitivity**	**Specificity**	**95% Confidence Intervals of sensitivity and specificity**		**LR +**	**LR -**	**Variance**	**Standard deviance**
					Lower	**Upper**				
EMG	4-fold Cross Validation	64.96	1.00	0.00	[1.00, 0.00]	[1.00, 0.00]	1.0	Inf	0.00	0.00
EDA		59.23	0.87	0.88	[0.86, 0.08]	[0.86, 0.088]	0.95	1.55	0.01	0.10
RspR		77.00	0.87	0.59	[0.87, 0.59]	[0.87, 0.59]	2.11	0.22	0.00	0.01
RRI		71.37	0.85	0.47	[0.75, 0.49]	[0.75, 0.48]	1.59	0.33	0.01	0.10
HR		73.55	0.85	0.53	[0.85, 0.53]	[0.85, 0.53]	1.80	0.29	0.01	0.09
HR +RspR		79.89	0.86	0.69	[0.86, 0.69]	[0.86, 0.69]	2.79	0.21	0.00	0.06
HR + RspR +HRV		82.14	0.79	0.86	[0.57, 0.67]	[1.00, 1.04]	5.51	0.25	0.03	0.18
All combined		82.14	0.79	0.86	[0.57, 0.67]	[1.00, 1.04]	5.51	0.25	0.03	0.18

Table 3.5 a. Logistic regression classifier results (using 14-fold/LOOC validation): Accuracy, Sensitivity, Specificity and Confidence Intervals.

Features	**Test- Train Split**	**Accuracy (%)**	**Sensitivity**	**Specificity**	**95% Confidence Intervals of sensitivity and specificity**	
					Lower	**Upper**
EMG	Leave One Out Cross Validation	64.96	1.00	0.00	[1.00, 0.00]	[1.00, 0.00]
EDA		63.07	0.93	0.07	[0.93. 0.072]	[0.93, 0.07]
RspR		76.96	0.87	0.59	[0.87, 0.57]	[0.87, 0.57]
RRI		75.35	0.89	0.50	[0.89, 0.50]	[0.89, 0.50]

HR		76.38	0.88	0.55	[0.88, 0.55]	[0.88, 0.55]
HR +RspR		82.81	0.89	0.72	[0.89, 0.71]	[0.89, 0.72]
HR + RspR + HRV		85.70	0.86	0.86	[0.67, 0.67]	[1.04, 1.04]
All combined		85.71	0.86	0.86	[0.67, 0.67]	[1.04, 1.04]

Table 3.5 b. Logistic regression classifier results (using 14-fold/LOOC validation): Likelihood ratio, Variance, Standard Deviance and ROC AUC score.

Features	Test- Train Split	LR+	LR-	Variance	Standard deviance	ROC AUC
EMG	Leave One Out Cross Validation	1.00	inf	0.00	0.00	0.50
EDA		1.00	0.94	0.01	0.11	0.71
RspR		2.12	0.22	0.00	0.06	0.83
RRI		1.78	0.22	0.01	0.11	0.86
HR		1.95	0.22	0.01	0.11	0.88
HR + RspR		3.12	0.16	0.00	0.06	0.93
HR + RspR + HRV		5.97	0.16	0.51	0.22	0.93
All combined		6.00	0.16	0.52	0.23	0.93

3.6 Analysis of Biochemical Indicators of Stress

Different biochemical parameters were analysed based on their useability in a real-time environment. The levels of chemical hormones such as adrenaline, nor-adrenaline, copeptin, prolactin, and plasma catecholamines could be used for stress monitoring but the collection of all these hormones require invasive procedures or 24-hour urine collection. The levels of alpha-amylase in the saliva are also indicative of stress but require careful handling of the sample and an incubation time of 3 to 5 minutes. The levels of Estradiol and testosterone also decrease when stressed. The collection of these hormones is also non-feasible for real-time stress monitoring.

Recently, there have been ground-breaking advancements made in the field of cortisol detection for stress monitoring. Cortisol hormone can be collected non-invasively, risk-free, and almost with no pain through saliva, hair and sweat. A special Enzyme-Linked Immunosorbent Assay (ELISA) kit is commercially available that can be used for screening the levels of cortisol in saliva or hair and determining stress levels, but it is a long process

and cannot be repeatable (within 24 hours). The detection of sweat cortisol is still an active research topic. Sweat colourimetric sensors are frequently developing and have shown great potential of becoming the only method of detecting stress in real-time using the biochemical parameter. Table 3.6 summarizes the feasibility and usability of different biochemical parameters.

Table 3.6 Analysis of biochemical hormones/parameters, remarks, and feasibility.

Type	Hormone/Parameter	Remarks	Reason
Bio-chemical	Adrenaline & noradrenaline	not significant	Blood/urine (invasive and not feasible)
	Copeptin & prolactin	not significant	Blood/urine (invasive and not feasible)
	Plasma catecholamines	not significant	Blood/urine (invasive and not feasible)
	Alpha-amylase	not significant	Saliva (3 to 5 min incubation)
	Estradiol & testosterone	not significant	Sex hormones (invasive and not feasible)
	Saliva and hair cortisol	semi-significant	non-invasively, risk-free, and almost with no pain but lengthy procedure (ELISA)*
	Sweat cortisol	significant	non-invasively, risk-free, and almost with no pain (colourimetry techniques)
*Enzyme-Linked Immunosorbent Assay (mostly used for saliva testing but with modification can be used for hairs) • Immersed in a 50-ng/mL hydrocortisone solution for periods lasting 15 minutes to 24 hours			

3.7 Conclusion

Several human biophysiological variables have been explored to evaluate and monitor both physical and mental stress levels in recent literature. Many of these variables have been independently used in wearable sensor-based devices. This study is particularly focused on a comparative analysis of these variables in terms of sensitivity and prediction specificity for stress monitoring. The comparative analysis has been performed by applying a t-test to validate the hypothesis that the physiological data for each variable for the stress and non-stress (baseline) states is statistically differentiable, and logistic regression was applied to identify the strongest predictor of stress.

A logistic regression-based classifier was also trained and validated during this study to determine the classification accuracy of the model. The results of two types of statistical analysis and classification model suggest that respiratory rate is the strongest (stand-alone)

predictor of stress compared to other commonly used physiological variables that include heart rate, RR interval, heart rate variability in the ECG/PPG, skin conductance (electrodermal activity) and muscle activation (electromyogram). Heart rate (RRI) emerged as the second-best predictor of stress. The prediction model, consisting of the combination of respiratory rate, heart rate and heart rate variation, derived from a single sensor, gives accurate classification results as a combination of EDA, EMG, RspR, HR (RRI), and HRV. The latter is a more complex sensory system, prone to motion artefacts. The investigation of different biochemical parameters revealed that there is a gap in determining an accurate way of stress detection. Recently, cortisol levels have emerged as a stress hormone but research is needed to develop an accurate, real-time and non-invasive method of cortisol collection.

Furthermore, in the case of biophysiological parameters analysis, it is important to note that all efforts were focused to provide a fair comparison by using data from the same device and participants. However, there may be other excitation sources (of similar responses) that these experiments failed to capture. Therefore, including context to the data, for example, physical activity, will be key to effective monitoring of stress daily.

Chapter 4
Machine Learning Classification Methods for Stress Detection

From the literature review, it could be concluded that most of the studies in the literature use supervised machine learning methods to build a predictive model for stress classification while very few studies proposed the use of unsupervised machine learning as a predictive model. The supervised learning algorithms heavily rely on the reference data labelled during the recording phase. Commonly, different types of self-reporting questionnaires or diaries of an individual are used to label the perceived stress instances. These questionnaires/diaries only capture stress levels at a specific point in time, are subjective and prone to inaccuracies as an individual might under-report stress levels due to social pressure or might over-report stress to gain sympathy.

Conversely, unsupervised learning algorithms perform clustering and do not need labels for training, which makes them more suitable for the development of a robust and continuous stress monitoring device. Thus, this chapter[4] provides a comparative study of unsupervised machine learning (clustering) algorithms with supervised machine learning (linear, ensembles, trees, and neighbouring models) algorithms. The finding enhances the understanding of implementing the unsupervised learning classifiers in wearable devices and confirms that respiratory rate and heart rate are the potential indicators for accurate stress detection and backs the results found in chapter 3.

- The study used two publicly available datasets i.e., Stress Recognition in Automobile Drivers Dataset and SWELL-KW Dataset.
- The unsupervised machine learning algorithms used for comparison included (1) Affinity Propagation, (2) Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), (3) K-mean, (4) Mini-Batch K-mean, (5) Mean Shift, (6)

[4] *The following body of the chapter is exact copy of the paper published in Frontiers in Medical Technology (2022). I am the first lead author in the paper, which is co-authored with my supervisors. The conceptualization, formal analysis, investigation and visualization were also done by me. Designed methodology and validation were led by my supervisors and me. I led all parts of the work with the support of my supervisors.*

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and (7) Ordering Points to Identify the Clustering Structure (OPTICS).

- While the supervised machine learning algorithms included (1) Logistic Regression, (2) Gaussian Naïve Bayes, (3) Decision Tree, (4) Random Forest, (5) AdaBoost and (6) K-Nearest Neighbours classifiers.

The study concluded that although the classification accuracy achieved by supervised machine learning algorithms is better than unsupervised algorithms, the use of an unsupervised classifier is important for the development of a non-invasive, robust, and continuous stress monitoring device since labelling the physiological signal in the ambulatory environment is a difficult and inaccurate task. Furthermore, these unsupervised machine learning algorithms require further investigation and modification to discover the hidden patterns and structures within the data. This will help these (unsupervised) algorithms to surpass the accuracies provided by supervised learning algorithms. This is an open-research topic and needs big data (non-existent, yet) to formulate a high-performance unsupervised learning model. For this thesis, supervised classification techniques were used in further analysis for stress detection and monitoring. This work covers thesis objective 4 of investigating the machine learning predictive models for accurate stress detection.

4.1 Exploring Unsupervised Machine Learning Classification Methods for Physiological Stress Detection

Abstract: Over the past decade, there has been a significant development in wearable health technologies for diagnosis and monitoring, including application to stress monitoring. Most wearable stress monitoring systems are built on a supervised learning classification algorithm. These systems rely on the collection of sensor and reference data during the development phase. One of the most challenging tasks in physiological or pathological stress monitoring is the labelling of the physiological signals collected during an experiment. Commonly, different types of self-reporting questionnaires are used to label the perceived stress instances. These questionnaires only capture stress levels at a specific point in time. Moreover, self-reporting is subjective and prone to inaccuracies. This chapter explores the potential feasibility of unsupervised learning clustering classifiers such as Affinity Propagation, Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), K-mean, Mini-Batch K-mean, Mean Shift, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Ordering Points To Identify the Clustering Structure (OPTICS) for implementation in stress monitoring wearable devices. Traditional supervised machine learning (linear, ensembles, trees, and neighbouring models) classifiers require hand-crafted features and labels while on the other hand, the unsupervised classifier does not require any labels of perceived stress levels and performs classification based on clustering algorithms. The classification results of unsupervised machine learning classifiers are found comparable to supervised machine learning classifiers on two publicly available datasets. The analysis and results of this comparative study demonstrate the potential of unsupervised learning for the development of non-invasive, continuous, and robust detection and monitoring of physiological and pathological stress.

Keywords: Machine learning, unsupervised learning, supervised learning, stress monitoring, physiological signals, heart rate, respiratory rate.

4.2 Introduction

There has been a notable increase in depression, anxiety, stress and other stress-related diseases, worldwide [3]–[5]. Stress deteriorates the physical and mental well-being of a human. Particularly, chronic stress leads to a weakened immune system, substance

addiction, diabetes, cancer, stroke and cardiovascular disease [292]. Thus, it is of utmost importance to develop robust techniques that can detect and monitor stress continuously, in real-time. The concept of detecting stress is quite complex, as stress has physiological as well as psychological aspects to it. Furthermore, both these aspects are triggered by multiple factors and are difficult to capture [312]. The recent development of wearable sensor technology has made it easier to collect different physiological parameters of stress in daily life.

The use of psychological assessment questionnaires, filled out on different instances in a day, is the most common technique to determine human stress. These questionnaires are limited to capturing stress at a particular time and do not allow continuous as well as real-time stress monitoring [313]. The time-bound nature of these questionnaire-based assessments unveils a major problem for the validation of new stress monitoring systems as there is no precise recording of which task or activity caused the participants' stress. To develop an acceptable standard for continuous stress monitoring, Hovsepian *et al.* [24] used wearable devices and proposed a data-driven stress assessment model, called the *cstress* model. To collect the data in this study, the participants were asked to fill out an Ecological Momentary Assessment (EMA) questionnaire 15 times a day, at random hours. The collected EMA self-report acted as the reference value for stress validation. The *cstress* model compensated for the unpredictable lag that occurred between the stressor and its logging in the EMA self-report.

In the literature, several supervised learning algorithms have been utilised for the detection and classification of stress [19], [314], [315]. These machine learning algorithms include logistic regression, Gaussian Naive Bayes, Decision Tree, Random Forest, AdaBoost, K-Nearest Neighbours and many others [292]. Dalmeida and Masala [316] investigated the role of electrocardiograph (ECG) features derived from heart rate variation (HRV) for the assessment of the stress of drivers. A set of different supervised machine learning algorithms were implemented, and the best recall score achieved was 80%. Similarly, Wang *et al.* [317] combined the supervised ensemble classifier with an unsupervised learning classifier and used the driver's galvanic skin response (GSR) data to detect stress. Their proposed model was able to detect stress with an accuracy of 90.1%.

The physiological parameters that are frequently used for stress analysis are respiratory rate, heart rate, skin conductance, skin temperature, and galvanic skin response [283]. As supervised learning requires training labels for training the classifier, in most cases, either the labels are unavailable or inaccurate, in the real-time data collection [318]. Several studies have reported the challenges of labelling the stress states and the importance of addressing these issues for the further development of sensor-based stress monitoring systems [319]–[321]. The challenges of poor-quality reference data and human bias encourage the exploration of unsupervised machine learning algorithms for stress detection and monitoring, as the unsupervised algorithms do not require reference data.

4.3 Related Work; Unsupervised Learning Classification

Throughout the literature, most authors are dedicated to the use of techniques based on supervised learning classification while the use of unsupervised learning methods is relatively new in the stress monitoring field. Rescio *et al.* [322] implemented the k-means clustering algorithm for stress classification using heart rate (HR), galvanic skin response (EDA) and electrooculogram (EOG) signals of 11 volunteers. To induce stress, the participants were asked to perform a mental arithmetic task and complex LEGO assembly without instruction. Authors have reported a classification accuracy of 70.6% with heart rate, 74.6% with EDA and 63.7% with EOG used as a single variable unsupervised classification model. Huysmans *et al.* [312] proposed a Self-Organizing Maps (SOM) based mental stress detection model that uses skin conductance (SC) and the electrocardiogram (ECG) of the test subjects. The authors recruited a group of 12 subjects and asked them to complete three stress-related tasks (each of 2 minutes). The first task was the Stroop Word Color test, in which subjects had to select the colour of the word rather than the written word. The second task was the mental arithmetic task, in which the subjects had to count backwards from 1081 with a difference of 7. The final task was to talk about a common stressful event that ever happened to them. The authors reported the average test accuracy of 79.0% using the proposed SOM based classifier.

Ramos *et al.* [323] used Naïve Bayes and logistic regression models to classify the stress outside the laboratory settings. They collected the heart rate, breathing rate, skin temperature and acceleration data from 20 volunteers while they were performing physical activity (such as walking, cycling, or sitting). To induce stress, the authors used random

noises, verbal mathematical questions, and a cold-water test. The activity data was ignored and an accuracy of 65% was achieved by the authors. Maaoui *et al.* [321] investigated the use of three unsupervised learning classification methods (K-mean, Gaussian Mixture Model (GMM) and SOM) to determine the stress levels using a low-cost webcam. Along with the webcam, the authors also collected the heart rate (extracted seven attributes) of 12 student volunteers. The authors reported the classification error rate of the three algorithms as 13.05% (K-means), 44.04% (GMM) and 36.57% (SOM) classifier. Similarly, Fiorini *et al.* [324] compared the performance of three unsupervised classification techniques (K-means, K-medoids and SOM) with three supervised learning techniques (Support Vector Machine (SVM), Decision Tree (DT), and K-nearest neighbours (K-NN)). They collected ECG, EDA, and electric brain activity signals from 15 healthy individuals. The authors designed the study to induce three different emotional states (i.e., relaxed, positive and negative) by the means of social interaction. The reported classification accuracy for the best-performing unsupervised classifier (K-means) was 77% while for the same model the best-performing supervised classifier (K-NN) was 85%.

This chapter explores the possible use of unsupervised classification methods for physiological stress detection. To perform a comparative analysis of the performance of unsupervised learning algorithms against supervised learning algorithms, two publicly available datasets were used. A total of seven most common supervised and seven unsupervised learning algorithms were implemented in Python Programming Language. The implemented unsupervised algorithms are Affinity Propagation [325], Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) [326], K-Mean, Mini-Batch K-Mean [327], Mean Shift, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [328] and Ordering Points To Identify the Clustering Structure (OPTICS) [329]. For comparison, supervised learning algorithms such as logistic regression, Gaussian naïve Bayes, decision tree, random forest, AdaBoost and K-nearest neighbours, are implemented.

4.4 Material and Methods

To address the challenge of manual annotation and labelling of the physiological signal as stress or non-stress in a supervised learning setup, we investigated the efficiency of the commonly used unsupervised machine learning algorithms, illustrated in the literature. For

assessment of the efficiency of these methods and comparative analysis, two publicly available datasets were downloaded. The first dataset is provided by the Massachusetts Institute of Technology (MIT), named Stress Recognition in Automobile Drivers by Healey [330], and is available on Physionet, while the second dataset is called the SWELL-KW dataset, available on Kaggle [331]. Both datasets contain heart rate variation features and provide labelled heart rate and respiratory rate parameters. The efficiencies of the supervised and unsupervised learning algorithms were benchmarked and are provided using standard measures of accuracy, precision, recall, and F1-score matrices of each classifier.

4.4.1 Performance Assessment Matrices

The performance of the classifier is assessed using the following metrics:

- The accuracy of a classifier is defined as the percentage of total correctly predicted labels in the test dataset, given mathematically as (equation 4.1):

$$\text{Accuracy} = \frac{\text{true positive labels+true negative labels}}{\text{total readings}} \tag{4.1}$$

- The precision and recall are calculated using equations 4.2 and 4.3:

$$\text{Precision} = \frac{\text{true positive labels}}{\text{true positive labels+false positive labels}} \tag{4.2}$$

$$\text{Recall} = \frac{\text{true positive labels}}{\text{true positive labels+false negative labels}} \tag{4.3}$$

- The F1-score of a classifier is the harmonic mean of its precision and recall. Equation 4.4 shows how the F1 score is calculated, mathematically:

$$\text{F1} - \text{Score} = 2 * \frac{\text{Precision}-\text{Recall}}{\text{Precision+Recall}} \tag{4.4}$$

4.4.2 Data Collection

To explore the usability of unsupervised machine learning classifiers in stress monitoring and comparison with supervised learning methods, two publicly available datasets were downloaded. Details of both datasets are described below.

4.4.2.1 Stress Recognition in Automobile Drivers Dataset

The dataset is developed by Healey [68], [332] during her PhD program at MIT. The dataset consists of the electrocardiogram (ECG), galvanic skin response (GSR), electromyogram (EMG), respiratory rate and heart rate measured using wearable sensors along with stress/non-stress labels generated from a combination of questionnaires and captured videos of the drivers. A total of 18 young drivers were asked to drive in different stress-inducing scenarios, such as on highways, during rush hours and at red lights, as well as a non-stress scenario (marked as non-stress or baseline readings). To rate the driver's stress levels, three different methods were used. These methods included self-reporting questionnaires, experimental design and metrics defined by independent annotators based on the video recording of the drivers. The dataset has baseline readings along with three different stress level readings (low, medium, and high stress).

4.4.2.2 SWELL-KW Dataset

The SWELL-Knowledge Work (SWELL-KW) dataset [331] provides heart rate variability (HRV) indices from sensor data for stress monitoring in an office work environment. The experiment was conducted on 25 subjects, performing typical office work such as preparing presentations, reading emails, and preparing work reports. Three different working conditions were defined by the authors:

- Neutral/no-stress: the subjects were allowed to complete the given task with no time boundary.
- Time pressure (a stress condition): the time to complete the given task was reduced to 2/3 of the time the subject took in the neutral condition.
- Interruption (a stress condition): during this time, subjects received 8 different emails. Some of the emails were related to their task and were asked to take specific action while some emails were not related to their task.

The experiment recorded data on facial expression, computer logging, skin conductance and ECG signal. For labelling, the Rating Scale Mental Effort (RSME) [333] and Self-Assessment-Manikin Scale (SAMS) [334] were used. Moreover, all subjects were also asked to report their perceived stress on a 10-point scale (from not-stressed to very stressed) using a visual analogue scale.

4.5 Unsupervised Classification Algorithms

Most of the unsupervised classification algorithms are based on clustering algorithms. Clustering algorithms find best-suited natural groups within the given feature space. In this study, the sensor data for stress and non-stress states of the participants are considered as the feature vector. The most widely used unsupervised classifiers implemented in this study are introduced in the following subsections.

4.5.1 Affinity Propagation

Affinity propagation takes the input data points as a measure of similarity between two data points. Each data point within the dataset sends a message to all other data points about the target's relative attractiveness. Once the sender is associated with its target (stress/no stress), the target becomes an exemplar. All the points with similar exemplars are combined to form one cluster. The classifier finds a set of different exemplars (representative points of each cluster) that best summarises the data points within the dataset [325].

4.5.2 BIRCH Classifier

Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) classifier constructs tree structure from which classification cluster centroids are obtained. The BIRCH classification algorithm utilises the tree structure to cluster the input data. The tree structure is called a clustering feature tree (CF Tree). Each node of the tree is made of a clustering feature (CF). The BIRCH clusters multi-dimensional input data entities to produce the best number of clusters with the available memory and time constraints. The algorithm typically finds good clusters within a single scan but can improve the quality with some additional scans [326].

4.5.3 K-Mean Classifier

The K-mean classifier is one of the most frequently used, unsupervised learning classifiers. The algorithm assigns the group label to each data point to minimise the overall variance of each cluster [327]. The algorithm starts with a random group of centroids, considering each centroid as a cluster, and performs repetitive calculations to adjust the position of centroids. The algorithm stops the optimization of clusters when the centroids are stable (no change in their values) or a defined number of iterations is achieved.

4.5.4 Mini-Batch K-Mean Classifier

Mini-Batch K-mean classifier is a modified version of the K-mean classifier. The classifier clusters the dataset using mini-batches of the data points rather than using whole data. This classifier is also robust to statistical noise and performs the classification of a large dataset more quickly [327].

4.5.5 Mean Shift Classifier

The mean shift classifier finds the underlying density function and classifies the data based on the density distribution of the data points in feature space [335]. The mean shift classification algorithm tries to discover different blobs within a smooth density of the given dataset. The algorithm updates the candidates for centroids that are then considered as the mean of the points with the given region. These candidates are filtered to eliminate near-duplicate centroids to form the final set of centroids, that form the clusters.

4.5.6 DBSCAN Classifier

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) finds the highest density areas in the given feature domain and expands those areas, forming clusters of feature space (stress/non-stress) [328]. The DBSCAN finds neighbourhoods of a data point exceeding a specified density threshold. This threshold is defined by the minimum number of data points required within a radius of the neighbourhood (minPts) and the radius of the neighbourhood (eps). Both the parameters are initialised manually at the start of the algorithm.

4.5.7 OPTICS Classifier

Ordering Points To Identify the Clustering Structure (OPTICS) is derived from the DBSCAN classifier, where a minimum of samples are required as a hyper-parameter to classify the data as a cluster (feature) [329].

4.6 Supervised Classification Algorithms

This study also implemented supervised classifiers, logistic regression, Gaussian naïve Bayes, decision tree, random forest, AdaBoost and K-nearest neighbours for comparison of results with the unsupervised classifiers. All these algorithms are briefly defined below. Interested readers are referred to Chaitra and Kumar *et al.* [336] for details.

4.6.1 Logistic Regression Classifier

Logistic Regression is one of the simplest machine learning algorithms mostly used for binary classification problems. Logistic regression estimates and classifies based on the relationship between independent and dependent binary features within a dataset.

4.6.2 Gaussian Naïve Bayes Classifier

The Naive Bayesian classifier is a probabilistic classifier. Naive Bayesian (NB) has only one parent node in its Directed acyclic graphs (DAGs), which is an unobserved node and has many children nodes, representing observed nodes. NB works with a strong assumption that all the child nodes are independent of their parent node and thus, one may say that the Naïve Bayesian classifier is a type of estimator.

4.6.3 Decision Tree Classifier

The Decision tree classifies by sorting input instances based on feature values. Each node of the decision tree shows a classified feature from an input instance while each branch shows an assumed nodal value. Classification of instances starts from the root and is sorted depending on their feature values.

4.6.4 Random Forest Classifier

The Random Forest is a supervised machine-learning algorithm. This algorithm creates random trees (forests) that are somewhat such as decision trees and the training method selected is always begging, as in begging learning models are linearly combined to increase the overall accuracy. While growing new trees, a random forest adds more randomness to the existing model. Instead of finding the most important target feature for node splitting, this algorithm searches for the best feature in the random subset of target features. In this way, we get wide diversity which in-return results in a better model. So, as a random forest only considers a random subset of features for splitting a node, we can make the trees of the model more random by using random thresholding of every feature rather than looking for the best threshold value.

4.6.5 AdaBoost Classifier

Boosting refers to a group of techniques that creates a strong classifier using many weak classifiers. To find a weak classifier, a different machine learning-based algorithm having

varied distribution is used. Each learning algorithm generates a new weak classification rule. This process is iterated many times and at the end, a boosting algorithm is formed by combining all newly generated weak classifiers rules to make a strong rule for prediction. A few steps should be followed for the selection of the right distribution:

- Step 1: Give all the distributions to the base learner and assign equal weights to every observation.
- Step 2: If the first base learner gives any prediction error, then pay more attention to the observations causing this prediction error. Then, apply a new base learner.
- Step 3: Until the base learning limit is reached, or the desired accuracy is achieved, keep repeating Step 2.

4.6.6 K-Nearest Neighbours Classifier

The k-Nearest Neighbour (kNN) is one of the simplest instance-based learning algorithms. The working of kNN is as follows. It classifies all the proximity instances, in a database, into a single group and then when a new instance (feature) comes, the classifier observes the properties of the instance and places it into the closest matched group (nearest neighbour). For accurate classification, initializing a value to k is the most critical step in the kNN classifier.

4.7 Results and Discussions

All the algorithms are implemented in python using the scikit learn library. Table 4.1 shows the hyper-parameter settings of all the classifiers discussed above. Figure. 4.1 demonstrates the overall steps involved in the implementation of the supervised and unsupervised classifiers.

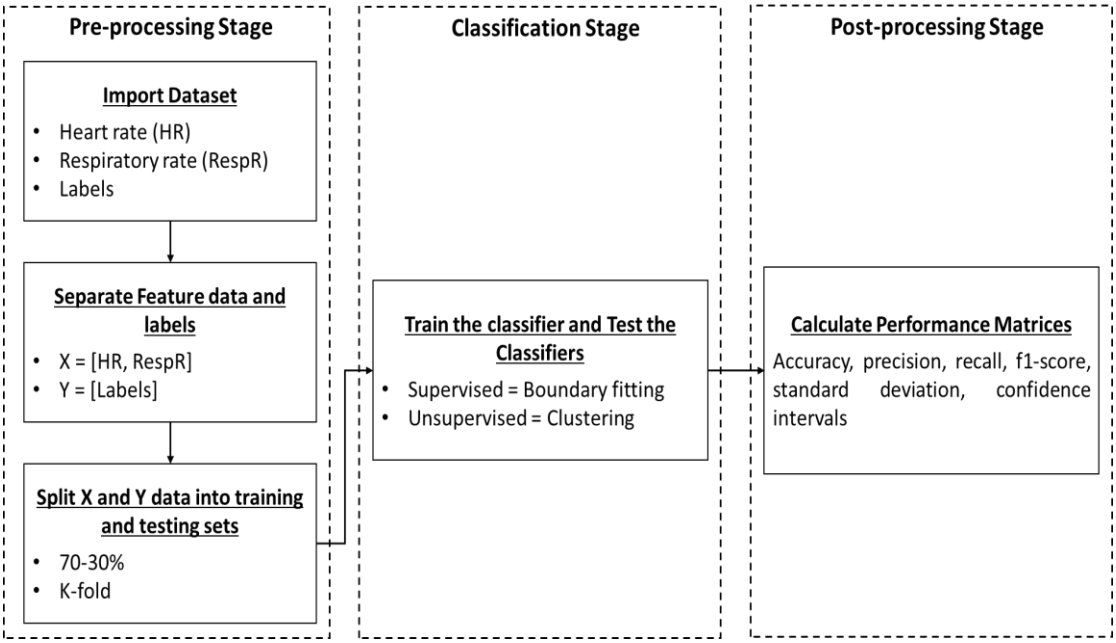


Figure 4.1 Block diagram of the implemented classification methods illustrating pre-processing, classification, and post-processing stages.

In the pre-processing stage, the heart rate, respiratory rate, and stress/non-stress label data are accumulated from the dataset. In the second step, the collected data is split into 70-30% or k-folds to have the training and testing sets. In the classification stage, the supervised learning classifiers are trained and tested to classify the input data into stress/non-stress using boundary fitting while in the case of unsupervised learning classifiers, clustering is performed on the input data and two clusters are formed corresponding to stress and non-stress data. In the final stage (post-processing), different performance evaluation metrics (accuracy, recall, precision, f1-score, standard deviation and 95% confidence intervals) are calculated and reported.

The performance of unsupervised and supervised learning algorithms was tested on the two datasets. The Stress Recognition in Automobile Drivers Dataset was a smaller dataset with 4,129 data points for each feature, i.e., heart rate and respiratory rate, along with stress/non-stress labels. The SWELL-KW dataset was a relatively larger dataset with a total of 204,885 data points for the heart rate feature along with stress/non-stress conditions. Each data point is considered a separate sample and is selected randomly for test and train sets, for supervised learning classifiers.

Table 4.1 Hyper-parameters settings and python library used for implementation.

Algorithm Type	Classifiers	Train-test Split	Hyper-parameters	Python Library
Supervised machine learning algorithm	Logistics Regression	70-30% and 10-fold cross-validation	• Solver = 'lbfgs' • Penalty = 'l2'	sklearn.linear_model
	Gaussian Naïve Bayes		• Variance smoothing = 1e-09	sklearn.naive_bayes
	Decision Tree		• Quality of split criterion = 'gini' • The value of max_depth was varied between range (1 to 11 with an increment of 1) • Maximum number of features to consider = 'auto'	sklearn.tree
	Random Forest		• Quality of split criterion = 'gini' • Maximum depth of trees = 11 • Maximum number of features to consider = 'auto' • Number of trees in the forest = 10	sklearn.ensemble
	AdaBoost		• The learning rate varied in range (0.01 to 1.1 with an increment of 0.01) • The maximum number of estimators at which boosting is terminated was varied between the range (of 50 to 200 with an increment of 10) • Algorithm = 'SAMME.R'	sklearn.ensemble
	K-Nearest Neighbours		• The number of neighbours required was set to 2	sklearn.neighbours
	K-Nearest Neighbours		• The number of neighbours required to be set at 5	sklearn.neighbours
The unsupervised machine learning algorithm	Affinity Propagation	70-30% and 10-fold cross-validation	• The damping factor was set at 0.8 to maintain the current value relative to the incoming value (weight 1 – damping) • Maximum iteration = 200 • Maximum number of iterations with no change in the number of estimated clusters = 15	sklearn.cluster
	BIRCH		• The threshold from which the radius of the subcluster should be lesser = 0.5 • Number of clusters = length of unique ids in the training set (default =2)	sklearn.cluster
	DBSCAN		• Maximum distance between two samples for consideration as neighbours (eps) = 0.50	sklearn.cluster

			• Minimum samples in the neighbourhood of a point to consider it as a core point = 9 • Distance calculation method = 'euclidean'	
	K-Mean		• The number of neighbours required was set to 2	sklearn.cluster
	Mini-Batch K-Mean		• The number of neighbours required was set to 2	sklearn.cluster
	Mean Shift		• Number of clusters = length of unique ids in the training set (default =2)	sklearn.cluster
	OPTICS		• Maximum distance between two samples for consideration as neighbours (eps) = 0.80 • Minimum samples in the neighbourhood of a point to consider it as a core point = 10	sklearn.cluster

In real-time, the unsupervised classifier is fed with control data and asked to classify the data into stress and non-stress condition. Then new data point is passed to the classifier and based on the centroids calculated using the control data, the new data point is placed in a specific cluster. For the comparison, a set of different supervised learning classifiers were also implemented, and the performance of the classifiers was evaluated using classification accuracies, precision, recall, and F1-scoring matrices. The results of the classifiers are discussed below.

4.7.1 Stress Recognition in Automobile Drivers Dataset

It is a well-known fact that all the traditional machine learning classifiers are data-hungry. As the Stress Recognition in Automobile Drivers dataset is a smaller dataset, the highest classification accuracy achieved (with 70-30% train-test split) using combined heart rate and respiratory rate along with supervised learning algorithm is 66.8% (AdaBoost classifier) while for single feature model i.e., heart rate and respiratory rate separately, the highest classification accuracy is 61.9% (Decision Tree classifier) and 66.8% (AdaBoost classifier), respectively. These results are better than previously reported accuracy values (52.6% and 62.2% for heart rate and respiratory rate models) [330]. Similarly, when combined heart rate and respiratory rate are used along with unsupervised learning classification, the highest classification accuracy achieved is 63.8% (Affinity Propagation classifier). If a single feature model is used, the highest accuracy for the heart rate feature model becomes 59.7% while for the respiratory rate feature model, it is 65% using the Affinity Propagation classifier. K-

fold cross-validation (with k = 10) was also performed using supervised learning models. The highest achieved accuracies for a single feature model are 59.9% for heart rate and 63.9% for respiratory rate while two feature models (heart rate and respiratory rate combined) gave an accuracy of 65.6%. Detailed analyses of different supervised and unsupervised learning algorithms are illustrated in Tables 4.2(a), 4.2(b) and 4.3.

4.7.2 SWELL-KW Dataset

The results of different supervised and unsupervised learning algorithms using the SWELL-KW dataset are illustrated in Table 4.4 (a), Table 4.4 (b) and Table 4.5. The highest classification accuracy achieved (with 70-30% train-test split) using a supervised learning algorithm is 74.8% (Decision Tree/Random Forest classifier), which is better than previously reported results for one physiological modality (accuracy = 64.1%) in [337] while for unsupervised learning is 68.3% (Mean shift classifier). The overall classification accuracies of the supervised classifiers do not change significantly with k-fold cross-validation applied to the data. The highest classification accuracy achieved using 10-fold validation is 75.0%.

The other performance matrices, precision, recall, and F1-score, of both datasets, follow similar performance trends as the accuracy for comparison of algorithms.

Table 4.2a. Results of supervised learning algorithms on Stress Recognition in Automobile Drivers Dataset.

Datasets	Classifiers	Feature	Test-Train Split	Classification Accuracy	Precision	Recall	F1-score
Stress Recognition in Automobile Drivers Dataset	Logistic Regression	Heart rate and Respiratory rate	70-30 %	59.3%	0.59	0.59	0.59
	Gaussian Naive Bayes			56.5%	0.60	0.59	0.59
	Decision Tree			63.4%	0.64	0.64	0.63
	Random Forest			65.0%	0.65	0.66	0.65
	AdaBoost			66.8%	0.67	0.66	0.65
	KNN=5			63.7%	0.63	0.63	0.63
	KNN=2			58.1%	0.60	0.57	0.56
Stress Recognition in Automobile Drivers Dataset	Logistic Regression	Heart rate		58.4%	0.59	0.58	0.58
	Gaussian Naive Bayes			56.0%	0.59	0.56	0.55

	Decision Tree			61.9%	0.66	.062	0.57
	Random Forest			56.2%	0.56	0.56	0.56
	AdaBoost			61.5%	0.61	0.61	0.60
	KNN=5			54.4%	0.54	0.54	0.54
	KNN=2			51.7%	0.55	0.52	0.50
Stress Recognition in Automobile Drivers Dataset	Logistic Regression	Respiratory rate		63.2%	0.70	0.63	0.55
	Gaussian Naive Bayes			63.4%	0.72	0.63	0.55
	Decision Tree			62.4%	0.64	0.62	0.63
	Random Forest			56.9%	0.57	0.57	0.57
	AdaBoost			66.8%	0.66	0.67	0.67
	KNN=5			59.5%	0.59	0.60	0.59
	KNN=2			54.0%	0.58	0.54	0.53

Table 4.2b. Results of supervised learning algorithms on Stress Recognition in Automobile Drivers Dataset (K-fold Cross Validation).

Datasets	**Classifiers**	**Feature**	**Test-Train Split**	**Classification Accuracy**	**Standard Deviation**	**Confidence Limits**	
						Lower	**Upper**
Stress Recognition in Automobile Drivers Dataset	Logistic Regression	Heart rate and Respiratory rate	10-fold cross-validation	61.5%	0.038	58.8%	64.2%
	Gaussian Naive Bayes			61.6%	0.022	58.9%	64.3%
	Decision Tree			64.1%	0.047	61.5%	66.8%
	Random Forest			64.0%	0.029	61.3%	66.6%
	AdaBoost			65.6%	0.036	62.9%	68.2%
	KNN=2			54.9%	0.051	52.2%	57.6%
	KNN=5			58.6%	0.034	55.9%	61.3%
Stress Recognition in Automobile Drivers Dataset	Logistic Regression	Heart rate		58.7%	0.20	57.2%	60.2%
	Gaussian Naive Bayes			56.4%	0.024	54.9%	57.9%
	Decision Tree			59.9%	0.019	58.4%	61.4%

	Random Forest			57.5%	0.027	56.0%	59.0%
	AdaBoost			59.9%	0.016	58.4%	61.4%
	KNN=5			52.0%	0.023	50.4%	53.5%
	KNN=5			56.1%	0.024	54.6%	57.6%
Stress Recognition in Automobile Drivers Dataset	Logistic Regression	Respiratory rate		58.3%	0.037	55.6%	61.0%
	Gaussian Naive Bayes			58.7%	0.038	56.0%	61.4%
	Decision Tree			61.4%	0.053	58.7%	64.0%
	Random Forest			59.4%	0.50	56.7%	62.1%
	AdaBoost			63.9%	0.036	61.2%	66.5%
	KNN=2			54.6%	0.039	51.9%	57.4%
	KNN=5			59.0%	0.052	56.3%	61.7%

Table 4.3 Results of unsupervised learning algorithms on Stress Recognition in Automobile Drivers Dataset.

Datasets	Classifiers	Feature	Test-Train Split	Classification Accuracy	Precision	Recall	F1-Score
Stress Recognition in Automobile Drivers Dataset	Affinity Propagation	Heart rate and Respiratory rate	70-30 %	63.8%	0.65	0.64	0.62
	BIRCH			54.9%	0.62	0.57	0.50
	DBSCAN			53.8%	0.56	0.54	0.41
	K-Mean			55.7%	0.62	0.56	0.52
	Mini-Batch K-Mean			53.0%	0.28	0.53	0.37
	Mean Shift			53.0%	0.28	0.53	0.37
	OPTICS			54.1%	0.54	0.54	0.53
Stress Recognition in Automobile Drivers Dataset	Affinity Propagation	Heart rate		59.7%	0.60	0.82	0.69
	BIRCH			49.1%	0.66	0.49	0.38
	DBSCAN			54.7%	0.30	0.55	0.39

	K-Mean			55.5%	0.61	0.55	0.53
	Mini-Batch K-Mean			54.8%	0.61	0.55	0.52
	Mean Shift			54.7%	0.30	0.55	0.39
	OPTICS			51.6%	0.51	0.52	0.51
Stress Recognition in Automobile Drivers Dataset	Affinity Propagation	Respiratory rate		65.0%	0.77	0.65	0.57
	BIRCH			57.4%	0.33	0.57	0.42
	DBSCAN			60.6%	0.62	0.61	0.53
	K-Mean			59.8%	0.63	0.60	0.60
	Mini-Batch K-Mean			60.3%	0.6	0.60	0.60
	Mean Shift			57.4%	0.33	0.57	0.42
	OPTICS			54.6%	0.49	0.55	0.46

Table 4.4a. Results of supervised learning algorithms on SWELL-KW Dataset.

Datasets	**Classifiers**	**Feature**	**Test-Train Split**	**Classification Accuracy**	**Precision**	**Recall**	**F1-score**
SWELL-KW Dataset	Logistic Regression	Heart rate	70-30 %	70.2%	0.70	0.70	0.64
	Gaussian Naive Bayes			70.3%	0.70	0.70	0.64
	Decision Tree			74.8%	0.74	0.75	0.73
	Random Forest			74.8%	0.74	0.75	0.73
	AdaBoost			74.6%	0.75	0.75	0.71
	KNN=5			71.8%	0.71	0.72	0.71
	KNN=2			62.7%	0.68	0.63	0.64

Table 4.4b. Results of supervised learning algorithms on SWELL-KW Dataset (K-fold Cross Validation).

Datasets	Classifiers	Feature	Test-Train Split	Classification Accuracy	Standard Deviation	Confidence Limits	
						Lower	Upper
SWELL-KW Dataset	Logistic Regression	Heart rate	10-fold cross-validation	70.2%	0.002	70.0%	70.4%
	Gaussian Naive Bayes			70.3%	0.002	70.4%	70.5%
	Decision Tree			74.8%	0.002	74.6%	75.0%
	Random Forest			75.0%	0.003	74.8%	75.2%
	AdaBoost			74.6%	0.003	74.4%	74.8%
	KNN=2			62.8%	0.002	62.6%	63.0%
	KNN=5			72.0%	0.003	71.8%	72.2%

Table 4.5 Results of unsupervised learning algorithms on SWELL-KW Dataset.

Datasets	Classifiers	Feature	Test-Train Split	Classification Accuracy	Precision	Recall	F1-Score
SWELL-KW Dataset	Affinity Propagation	Heart rate	70-30 %	66.5%	0.44	0.67	0.53
	BIRCH			68.1%	0.66	0.68	0.60
	K-Mean			66.7%	0.45	0.67	0.53
	Mini-Batch K-Mean			66.7%	0.45	0.67	0.53
	Mean Shift			68.3%	0.69	0.68	0.60
	DBSCAN			66.7%	0.45	0.67	0.53
	OPTICS			66.7%	0.45	0.67	0.53

Table 4.6 Results Comparison of supervised learning algorithms on datasets with previously reported work.

Datasets	Classifier Type	Ref	Feature	Highest Reported Classification Accuracy	Highest Achieved Classification Accuracy [This Study] with 70-30% SPLIT	Highest Achieved Classification Accuracy [This Study] with K-fold Validation
Stress Recognition in Automobile Drivers Dataset	Supervised learning algorithms	Table 5.8 of [23]	Respiratory Rate	62.2%	66.8%	63.9%
			Heart rate	52.6%	61.9%	59.9%
SWELL-KW Dataset		Table 4 of [31]	Heart rate	64.1%	74.8%	75.0%

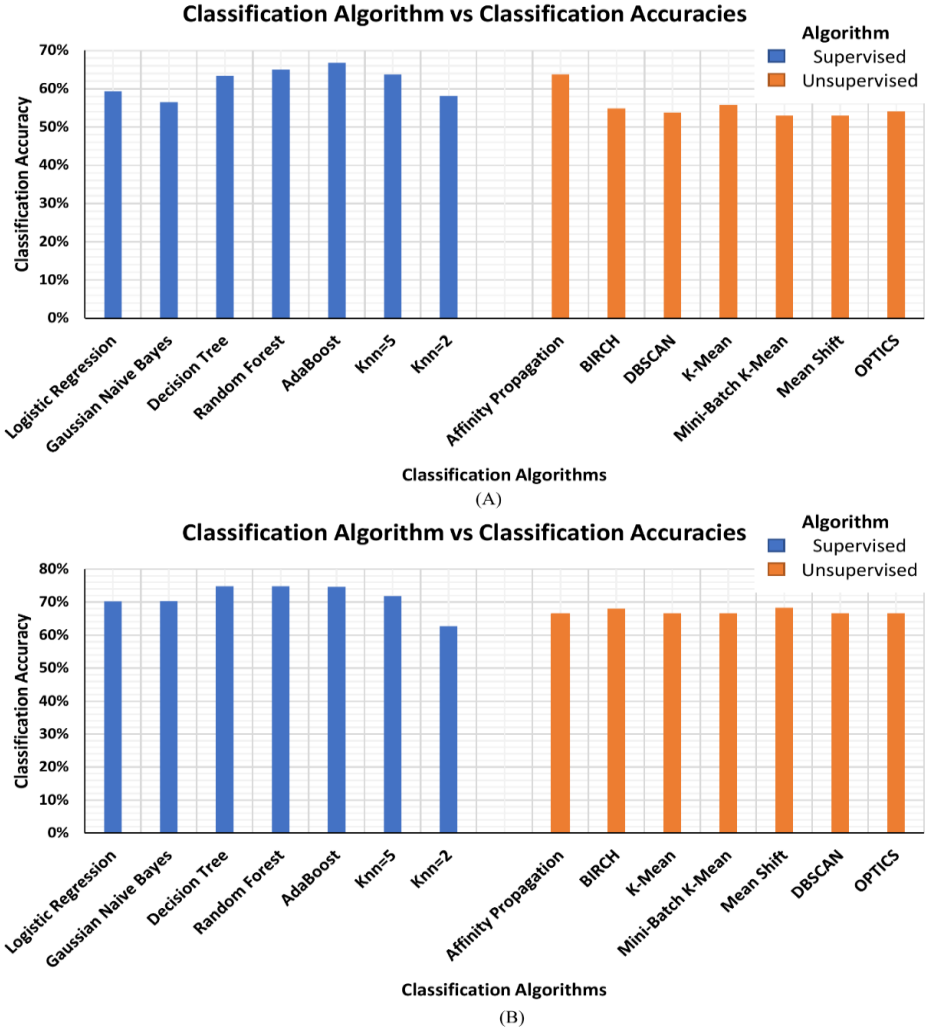


Figure 4.2 Bar-plot of classification accuracies of supervised and unsupervised classification algorithms using (A) Stress Recognition in Automobile Drivers Dataset and (B) SWELL-KW Dataset.

4.7.3 Summary

The results of the supervised learning classification algorithm are better than the previously reported results [330], [337] using the same datasets, see Table 4.6. As both datasets have real-time physiological signals, there are some outliers and noisy signals within the signal. Thus, intense pre-processing and outlier detection was performed to cleanse the dataset for better training of the classification algorithm. The achievement of better results than the previously published results reflects that our performed pre-processing step (thresholding and filtering) does help in developing a better classification model.

The authors acknowledge that these accuracies are not indicative of good performance but motivate the researchers to propose better supervised as well as unsupervised learning classification models for improved stress monitoring. Figure. 4.2 shows the bar plot of classification accuracies of supervised and unsupervised classification algorithms using Stress Recognition in Automobile Drivers Dataset (Figure. 4.2A) and SWELL-KW Dataset (Figure. 4.2B). The use of an unsupervised classifier is important for the development of a non-invasive, robust, and continuous stress monitoring device since labelling the physiological signal in the ambulatory environment is a difficult and inaccurate task. The results in Tables 4.2 to 4.5 show the comparison of classification efficiencies of supervised and unsupervised classification algorithms. The difference in the highest classification accuracies is comparable, i.e., for the Stress Recognition in Automobile Drivers dataset is 1% for the respiratory rate-based model and 3% for two feature-based models. While for the SWELL-KW dataset, the difference is 6.5%. The overall accuracies of the supervised classifiers are better than the unsupervised classifier but as an unsupervised machine learning classifier does not require any intense pre-training as well as stress/non-stress labels, these results are encouraging the researchers to use the unsupervised models in stress monitoring wearable devices. Further improvements in unsupervised algorithms to optimise use in stress monitoring can potentially provide even better detection accuracies.

4.8 Conclusion

Stress detection in a real-world environment is a complex task as labelling of the collected physiological signals is often inaccurate or non-existing. The questionnaires and self-reports are considered the only established way of getting the reference state of the participant's emotions. The supervised machine learning classifiers have been able to accurately classify the stress state from the non-stress state. The problem of stress level labelling has already been reported in many studies but has rarely been addressed.

One possible solution is the use of an unsupervised machine learning classifier as such algorithms do not require labelled data. In this study, we have implemented different unsupervised classification algorithms to explore the feasibility of unsupervised stress detection and monitoring in different stress-monitoring scenarios. For comparison, a set of different supervised learning algorithms was also implemented.

We have also performed an analysis to investigate the significant difference in the model performance using the standard deviations and confidence intervals. The performance of some models differs significantly from others. For instance, the performances of decision tree classifiers compared to k-nearest neighbours (k=2) on Stress Recognition in the Automobile Drivers dataset and random forest classifier compared to the logistic regression classifier on the SWELL-KW dataset are quite different. This leads us to the conclusion that a careful selection of classification models is required when aiming to develop an accurate stress detection system. The selection of the classifier is dependent on the type and shape of the data. It also depends upon the number of data points within the dataset.

The classification results indicate that unsupervised machine learning classifiers can show good performance in terms of classification accuracy, precision, recall and F1-score, without any training phase which is usually time-consuming and inaccurate. The findings enhance our understanding of the feasibility of unsupervised learning classifiers in wearable devices. Furthermore, these findings also may inform further approaches for the detection and monitoring of stress in an ambulatory environment.

Chapter 5
Photoplethysmography (PPG)-Based Respiratory Rate Estimation Algorithm

The finding of the previous two chapters (Chapter 3 and Chapter 4) suggest that respiratory rate along with heart rate is indeed the best predictor of stress conditions. In the literature, most of the studies estimate (extract) the information on respiratory and heart rates using ECG signals. The ECG devices come with a lot of drawbacks including the adhesiveness of electrodes, motion artefacts and a controlled environment (cannot be used 24/7). The use of a PPG signal (optical sensor) can be an alternative and has shown a good correlation with ECG signals (0.7). A PPG sensor comes in most smart wrist-worn watches and is feasible for long-time continuous data recording (as they have longer battery life).

The heart rate estimation is relatively easier (counting the number of peaks in a minute) but extraction of respiratory rate information from the PPG signal is a complex task. Most of the existing algorithms for the estimation of respiratory rate using PPG signal are sensitive to external noises and may require the selection of certain algorithm-specific parameters, through the trial and error method. Thus, this chapter[5] presents the newly developed algorithm to estimate the respiratory rate using a PPG sensor signal for health monitoring. The algorithm is resistant to signal loss and can handle low-quality signals from the sensor, which completes the fifth objective (first half) of the thesis.

It combines selective windowing, pre-processing and signal conditioning, modified Welch filtering and postprocessing to achieve high accuracy and robustness to noise. The proposed algorithm was developed and evaluated using the BIDMC dataset (containing 53 subjects' data, each recorded for 8 minutes). The results endorse that integration of the

[5] *The following body of the chapter is exact copy of the paper published in the Journal of Medical and Biological Engineering, JMBE (2022). I am the first lead author in the paper. The conceptualization, formal analysis, investigation and visualization were also done by me. Designed methodology and validation were led by my supervisors and me. I led all parts of the work with the support of my supervisors.*

proposed algorithm into a commercially available PPG device would expand their functionality from the measurement of oxygen saturation level and heart rate to the continuous measurement of the respiratory rate.

5.1 Photoplethysmography-Based Respiratory Rate Estimation Algorithm for Health Monitoring Applications

Abstract: *Purpose:* Respiratory rate can provide auxiliary information on the physiological changes within the human body, such as physical and emotional stress. In a clinical setup, the abnormal respiratory rate can be indicative of the deterioration of the patient's condition. Most of the existing algorithms for the estimation of respiratory rate using photoplethysmography (PPG) are sensitive to external noise and may require the selection of certain algorithm-specific parameters, through the trial-and-error method. *Methods:* This chapter proposes a new algorithm to estimate the respiratory rate using a photoplethysmography sensor signal for health monitoring. The algorithm is resistant to signal loss and can handle low-quality signals from the sensor. It combines selective windowing, pre-processing and signal conditioning, modified Welch filtering and postprocessing to achieve high accuracy and robustness to noise. *Results:* The Mean Absolute Error and the Root Mean Square Error of the proposed algorithm, with the optimal signal window size, are determined to be 2.05 breaths count per minute and 2.47 breaths count per minute, respectively, when tested on a publicly available dataset. These results present a significant improvement in accuracy over previously reported methods. The proposed algorithm achieved comparable results to the existing algorithms in the literature on the BIDMC dataset (containing data from 53 subjects, each recorded for 8 minutes) for other signal window sizes. Conclusion: The results endorse that integration of the proposed algorithm into a commercially available pulse oximetry device would expand its functionality from the measurement of oxygen saturation level and heart rate to the continuous measurement of the respiratory rate with good efficiency at home and in a clinical setting.

Keywords: photoplethysmography, respiratory rate, adaptive estimation, wearable sensors, health monitoring, algorithms.

5.2 Introduction

In recent years, respiratory rate (RespR), blood pressure (BP) and heart rate (HR) monitoring are considered essential for continuous and primary assessment of the patient's well-being [338]. The inhalation and exhalation processes can increase or decrease the blood flow within the body. Therefore, respiratory rate can be determined by measuring the changes in the heartbeats or blood flow [339]–[342]. There is significant evidence in the literature to suggest that irregular respiration is an imperative indicator of some serious illnesses [343]–[345]. The normal range of respiratory rate for children (1-5 years) is above 24 and less than 40 breaths per min while above 5 years, the normal range is between 12 to 25 breaths per minute. Any deviation from the normal range is an indicator of respiratory distress and requires instant clinical intervention [346], [347]. According to the World Health Organisation (WHO), elevated respiratory rate is observed in cases of chronic obstructive pulmonary disease, asthma, hypoxia, and pneumonia [348], [349].

In hospitals, respiratory rate is monitored using thoracic/abdominal plethysmography belts, oral/nasal pressure transducers, capnography, and transthoracic impedance pneumography [350], [351]. However, these devices are not as user-friendly as mobile wearable devices. Most smartwatches use a photoplethysmography signal to extract only the heart rate even though the PPG signal can also be used to extract RespR [352], [353]. While algorithms have been proposed in the literature to extract respiratory rates from PPG signals, each algorithm has certain limitations. Estimation of respiratory rate from a PPG signal can be done by using digital signal processing (DSP) techniques. Among these techniques, digital filters are commonly used to remove noise and extract variables of interest from the raw signals. However, the performance of DSP filters is highly dependent on the cut-off frequency of the filter. Other techniques include analytical methods, which are very sensitive to noise and result in very poor respiratory rate detection in presence of motion artefacts. Time-frequency analysis-based methods such as Wavelet transform addresses most of the common problems of filtering and analytical methods. It is less sensitive to noise and motion artefacts but requires the selection of more than one parameter such as the mother wavelet function and the total number of decomposition levels, which in practice are unknown [354], [355]. A list of various categories of respiratory rate estimation methods and their limitations is provided in Table 5.1.

One of the major challenges in respiratory rate estimation from PPG signal is respiratory-induced amplitude variation [356]. During the inhale cycle, the intra-thoracic pressure changes cause decreased stroke volume of the left ventricle, which leads to a smaller PPG amplitude. Similarly, during expiration, the left ventricle stroke volume increases, which results in increased pulse amplitude. In the literature, a variety of methods have been proposed for the estimation of RespR from a PPG signal. Liu *et al.* [352] have highlighted the merits and demerits of different respiratory rate estimation algorithms. Another key challenge in developing a respiratory rate extraction algorithm is the estimation of optimal window size for the segmentation of the signal. A shorter time window provides high resolution, low computational cost, and better real-time performance. In contrast, a longer window size provides better estimation accuracy [357]. The proposed algorithm is developed considering all the major limitations including noise and poor signal quality, the effect of window size, and the cut-off frequencies of the filters.

For evaluation of the proposed algorithm, the estimated respiratory rate is compared with the reference respiration data provided in the publicly available dataset called BIDMC dataset. The dataset is available at PhysioNet [332] while for filtering and signal processing, an open-source toolkit Heartpy [358] with some modifications was used. The performance of the proposed algorithm was benchmarked using accuracy assessment metrics against published results of existing algorithms.

Table 5.1 Respiratory rate estimation methods and their limitations.

Methods		**Limitations**
Digital Method	The digital technique (FFT, Welch, Notch) [359]	Highly dependent on the selection of cut-off frequencies.
Wavelet methods	Wavelet transforms [360], [361]	Requires the selection of more than one parameter such as the mother wavelet function and the total number of decomposition levels.
	Smart fusion [357]	
Adaptive estimations	Adaptive respiratory rate estimators [359], [362]	Very sensitive to noise and results in very poor respiratory rate estimation if there are any motion artefacts in the signal.
	Empirical mode decomposition (EMD) [363], [364]	
	Autoregression [365], [366]	

Analytical Methods	Artificial neural networks [367]	Often requires a relatively long time to converge and give an accurate estimation of respiratory rate.
	Principal component analysis (PCA) [368]	
	Complex demodulation [369]	
	Independent component analysis (ICA) [370]	

5.3 Proposed Algorithm

Figure 5.1 shows the pre-processing, signal analysis and post-processing steps of the proposed respiratory rate estimation algorithm.

5.3.1 Pre-processing steps

For pre-recorded datasets, there is one additional step of signal interpolation is included in the pre-processing stage and is explained below. All the other steps shown in Figure 5.1 are the same for pre-recorded as well as real-time PPG signals.

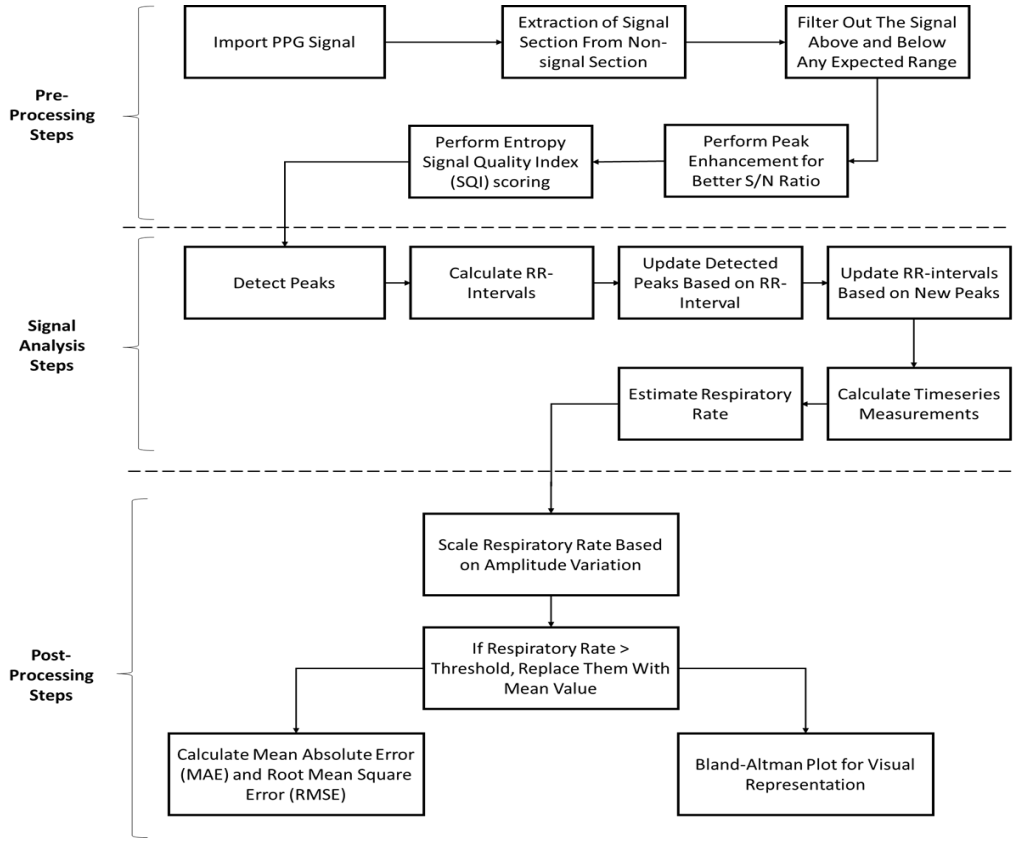


Figure 5.1 Pre-processing, signal analysis and post-processing steps of the respiratory rate estimation algorithm.

5.3.1.1 Signal interpolation

The first step in the pre-processing stage for a pre-recorded dataset is to extract the raw PPG signal values from the dataset and reference respiratory rate values. In the literature, several techniques have been proposed to address the problem of a missing signal [371], [372]. The simplest and easiest way to tackle this problem is to remove the signal points with missing values. However, eliminating the signal points from the dataset is not encouraged by statisticians as it comes under signal manipulation and leads to signal loss. It is recommended to interpolate for missing data as eliminating or inserting zeros causes a complete loss of the data (information or signal). This loss of data might play an important role in deriving conclusions or in determining any statistical outcome. Thus, replacing the missing signal, generally marked as NaN (not a number), with the mean value of two neighbouring signal samples, might not affect the overall signal behaviour and derived conclusions/results can be considered valid.

5.3.1.2 Digital filtering

The raw PPG signal has information on heart rate and respiratory rate as well as noise. The digital filtering method is used to remove the noise and extract the relevant information. The raw PPG signal is passed through a bandpass Butterworth filter to allow only the frequencies within the range of minimum and maximum respiratory rate (i.e., 0.1-0.4Hz or 6-24 breaths per minute), as shown in Figure 5.2.

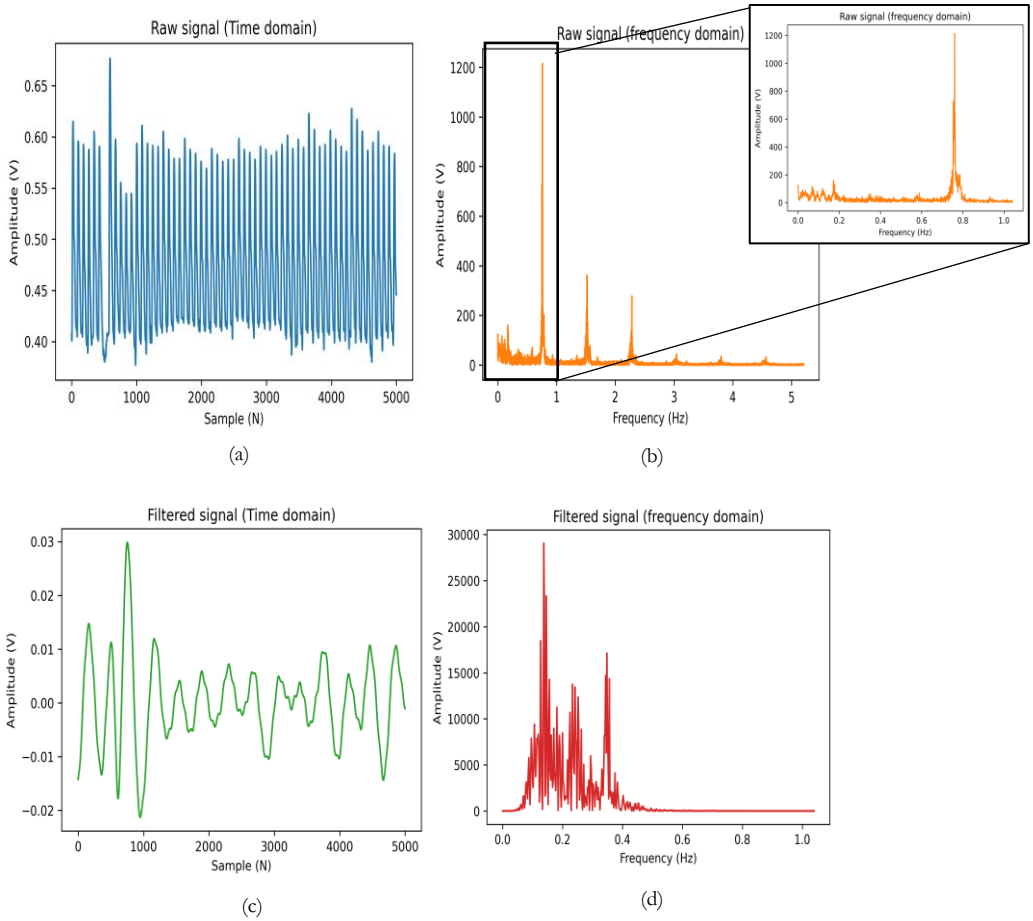


Figure 5.2 Extraction of respiratory rate signal from raw PPG signal. (a) shows the raw PPG signal imported from the dataset (b) is the frequency domain signal of the same raw PPG signal. (c) illustrates the filtered signal passed through band pass Butterworth filter with a cut-off frequency of 0.1-0.4 Hz while (d) is the frequency domain representation of the filtered signal. Note that only the frequencies between 0.1 and 0.4 are passed and all others are blocked (showing flat line).

5.3.1.3 Peak enhancement

To enhance the signal-to-noise ratio and to get better detection of the peaks, the algorithm performs peak enhancement using the peak enhancement algorithm. This function makes the higher peaks more dominant while suppressing the smaller peaks, for better detection of the peaks. This function scales the signal to the specified lower and upper range. The formula is as in equation (5.1):

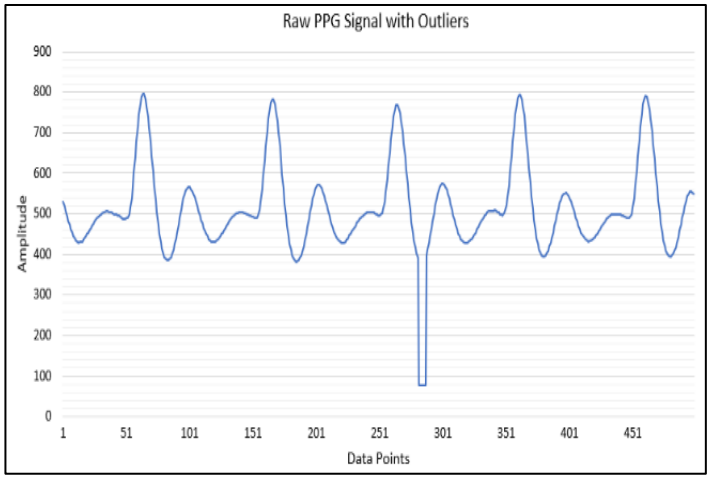
$$P_{enh} = rb * ((x - min(x))/range(x)) + l\_lt \quad (5.1)$$

Where P_{enh} is peak enhancement, x is the raw signal, rb is the range of upper limit and l_lt (lower limits), given by the user (by default it is set at 1024 and 0, respectively). The $range(x)$ is a valued range calculated by subtracting the maximum value of the analysed signal from the minimum value.

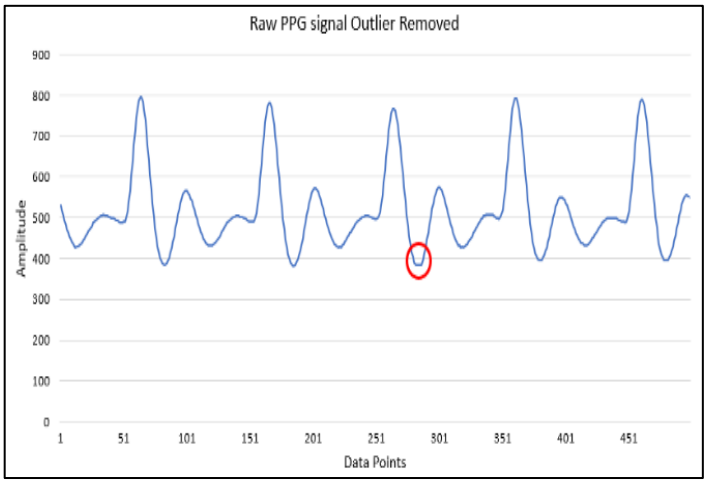
5.3.1.4 Outlier detection

Most of the time, PPG devices have loose contact with the body causing abrupt changes, possibly due to sudden moves of the finger or for other unknown reasons, in the raw signal. An outlier detection function is required to eliminate these baseline abrupt changes. These changes could not be completely removed by general digital filters as they contain wide-band frequency components. In the developed algorithm, the outlier can be removed using the Hampel filtering technique [373], [374].

The goal of the Hampel filter is to identify and replace outliers in each window analysed. It uses a sliding window of configurable width to go over the signal. The median and the standard deviation are calculated for each window, of *x* seconds, and expressed as the median absolute deviation (MAD). For each sample of *x*, the algorithm computes the median of a window composed of the sample and its six surrounding samples (if window size = 6), three per side. If a sample differs more than the median + three standard deviations, it is replaced with the median. As the algorithm uses 6 neighbouring samples (data point +3 per side) to cover most of the signal, only the last two sample/data points are left at the end. To include them for outlier detection, padding is performed. Figure 5.3 shows a sample example of how outlier removal works. Figure 5.3(a) exemplifies an outlier present in a PPG signal while Figure 5.3(b) shows the output of the Hampel filter and interpolated PPG signal after the removal of the outlier.



(a)



(b)

Figure 5.4 Outlier removal using Hampel filter with window size = 6.

5.3.1.5 Entropy-based signal quality index (ESQI)

One of the objectives of the proposed algorithm is to accurately extract the respiratory rate even from a low-quality signal. Entropy Signal Quality Index (ESQI) scoring was proposed by Selvaraj *et al.* [375]. It quantifies the difference in the probability distribution (PDF) of raw signal from the uniform (normal) distribution and provides a measure of uncertainty present in the analysed signal [376]. The equation to calculate ESQI is as:

$$E_{SQI} = -\sum_{i=1}^{N} x[i]^2 \; log_e(x[i]^2) \qquad (5.2)$$

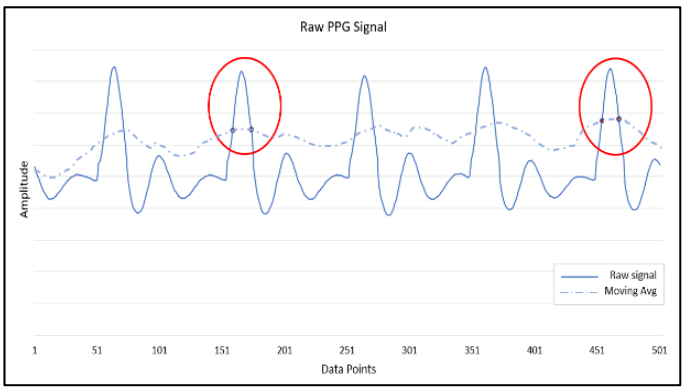
In equation 5.2, x is the raw PPG signal and N is the number of points within the analysed raw PPG signal. The signal quality assessment revealed that for some portion of the signal, $x[i]^2 \rightarrow 0$ (no or fewer fluctuations) then $log_e(x[i]^2) \rightarrow nan$. Thus, the ESQI value

returns as undefined. The algorithm skips the further computation for respiratory rate estimation if the ESQI is undefined (nan).

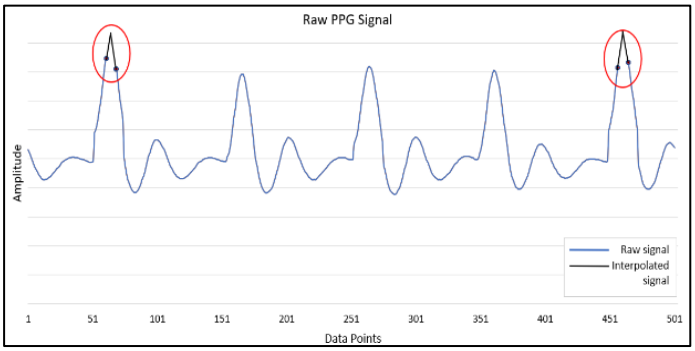
5.3.2 Signal analysis and respiratory rate estimation

5.3.2.1 Peak detection

The next step is to analyse the signal. The algorithm detects peaks within the peak-enhanced signal. This is a crucial step as all the further calculations will be dependent on the detected peak. Once peaks are detected, the algorithm calculates the systolic peak interval or RR interval. Peak detection can be done by using a moving average. An intersection threshold is made, and a Region of Interest (ROI) is selected between two intersection points where the signal amplitude is the highest, as shown in Figure 5.4(a). If the raw PGG signal had a clipped signal, the algorithm uses cubic spline interpolation to interpolate the missing signal before peak detection as shown in Figure 5.4(b). The red circles in Figure 5.4(a) mark ROI and two intersecting points are used to determine the peak while in Figure 5.4(b) they show the clipped signal (in blue) and cubic spline interpolated signal (in black).



(a) An example of peak detection using moving average from raw PPG signal



(b) An example of interpolation of clipped signal using cubic spline interpolation from raw PPG signal

Figure 5.5 Peak detection and interpolation of the clipped signal.

5.3.2.2 Respiratory rate estimation

Sometime after the detection of R-R peaks, there is a possibility of the detection of false peaks [377], [378]. The proposed algorithm looks for any anomalies in the detected peak array using RR intervals. The R-R peaks were not considered if the interval between two adjacent peaks is less than 30% of the mean interval of the analysed signal, as mentioned in [358]. Based on updated R-R peaks, new RR intervals are calculated. The new values of the RR interval are then used to calculate different time-series measurements. These measurements include heart rate in beats per minute (BPM), RR intervals or inter-beat intervals (IBI) and estimated respiratory rate in breaths per min. The respiratory rate is calculated by estimating power spectrum density (PSD) and respiratory frequency band (breaths/min) using Welch's method [379], [380]. Welch's method divides the inter-beat intervals (signal) into overlapping segments and computes a modified periodogram for each segment. Then an average of all periodograms along with an array of respiratory frequency bands are returned at the output. Table 5.2 describes the parameter initialisation values used in the proposed algorithm. The respiratory rate is the maximum frequency within the respiratory frequency band (breaths/min) that is, where PSD is maximum.

Table 5.2 Welch filter parameters for respiratory rate estimation.

S. No.	Parameter	Value/Method
1	Sampling frequency	125
2	Window	Hann Window
3	Number of overlapping points	50%
4	Length of FFT	Length of data
5	Scaling	Density
6	Averaging periodogram	Mean

5.3.3 Post-processing steps

Usually, the PPG waveform varies in synchrony with the respiratory cycle [381]. During the inhale cycle, the intra-thoracic pressure changes causing decreased stroke volume of the left ventricle, which leads to a smaller amplitude (PPG) pulse. Similarly, during expiration, the left ventricle stroke volume increases, which increases the pulse amplitude. This phenomenon is known as amplitude modulation of cardiac synchronous pulsatile waveform or respiratory-induced amplitude variation [356].

As in the prior stages, the proposed algorithm does not account for amplitude variation, due to active filtering and peak enhancement steps, which is essential for accurate peak detection; In the post-processing stage, the estimated respiratory rate is scaled based on the variation (changes) in the amplitude of the PPG signal. The scaling is essential as the proposed (and most of the existing algorithms) does not account for amplitude variation in the pre-processing stage. Thus, a separate scaling is applied to the estimated respiratory rate. This scaling factor depends on the range (difference between maximum and minimum value) of the signal and defined window size (of x seconds).

The proposed scaling method is generalisable and can work for a variety of PPG datasets. Nevertheless, there might be scenarios where some fine-tuning may be required for better estimation. As the last step, the algorithm makes sure the estimated values of the respiratory rate are within a specific threshold and do not exceed the maximum physiologically possible breathing rate.

5.4 Validation of Proposed Algorithm

To validate the proposed algorithm and assess its performance, a publicly available dataset, known as Berth Israel Deaconess Medical Centre (BIDMC) dataset, was used. The proposed algorithm was applied to the PPG data in the BIDMC dataset for the estimation of respiratory rate and the performance was benchmarked against various existing algorithms using the most common performance evaluation metrics.

5.4.1 BIDMC Dataset Overview

The dataset consists of electrocardiogram (ECG), photoplethysmogram (PPG) and impedance pneumogram respiratory signals of patients in the intensive care unit at Berth Israel Deaconess Medical Centre (BIDMC), Boston, USA [332], [382]. The presented dataset was proposed to evaluate the performance of any newly developed respiratory rate algorithm and reflect its potential usability in a real-world critical care environment. Table 5.3 shows the key statistical features of the BIDMC dataset. The dataset is comprised of 53 patients' recordings, each of 8-minute duration, containing:

Table 5.3 Key statistical features of the respiratory rate in the BIDMC dataset (unit = breaths per minute).

N	Validated	53	
	Outlier	2 (Subject 13 and 33)	
		With Outlier	Without Outlier
Mean		17.42	17.63
Median		17.89	17.89
Standard Deviation		3.22	2.62
Variance		10.39	6.86
Minimum		3.71	10.47
Maximum		24.67	24.67

- Physiological signals; were sampled at 125 Hz.
- Physiological parameters: such as respiratory rate, blood oxygen saturation levels and heart rate sampled at 1 Hz.
- Age and gender; are fixed parameters.
- Manually annotated individual breaths, annotated independently by two researchers.

5.4.2 Performance evaluation metrics

For the evaluation of the developed algorithm, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics were used, and the performance of the proposed algorithm was compared with existing algorithms.

- The mean absolute error (MAE) is an average measure of difference (error) between the reference value and the algorithm's estimated value of that observation. Mathematically, MAE is calculated using equation 5.3 and is as follows:

$$\text{MAE} = \sum_{i=1}^{N} \frac{|reference\ \ value_i - estimated\ value_i|}{total\ number\ of\ data\ points} \qquad (5.3)$$

- The root means square error (RMSE) is a square root of the mean of the square of estimation error. The RMSE shows the standard deviation of the estimation error and is considered a good measure of accuracy. Equation 5 illustrates RMSE mathematically.

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N} \frac{|((reference\ \ value_i - estimated\ value_i)|^2}{total\ number\ of\ data\ points}} \quad (5.4)$$

In Equations 5.3 and 5.4, the *reference value* denotes the real respiratory rate reported in the BIDMC dataset while the *estimated value* denotes the calculated respiratory rate.

5.5 Results and Discussion

To estimate the respiratory rate and perform estimation error analysis, data from 51 patients in the BIDMC dataset were used, discarding the two outliers mentioned in Table 5.3. The two patients that are excluded from this study are patient 13 and patient 33. Reference respiratory rate values of patient 13 are missing in the dataset while the raw data of patient 33 are too distorted to extract any meaningful information. The respiratory rate was calculated using window sizes of 10, 20, 30, 45, 60, 90, and 120 seconds, defined at pre-processing step. For comparison of the developed algorithm with other state-of-the-art algorithms [357], [382]–[384], a window size of 32 and 64 seconds was also used to estimate the respiratory rate to match the window sizes with the previously published algorithms.

The smaller window size yields less processing and computation time, but it can give inaccurate readings of respiratory rate. While using a larger window size, the overall accuracy of the estimation can be improved but it is difficult to estimate the lowest detectable respiratory rate [357]. Thus, a careful trade-off is needed while selecting any specific window size for the analysis.

5.5.1 Performance evaluation

The proposed algorithm was able to estimate the respiratory rate accurately for all the subjects within the BIDMC dataset excluding subject 18, see supplementary tables Chapter 5 S1 and Chapter 5 S2, in appendix chapter 5 (Appendix 9.2). Figure 5.4 shows the error in the estimation of respiratory rate using different window sizes with and without Entropy-based signal quality (ESQI) assessment. ESQI is generally used in offline data processing to select good-quality signals from raw data, and it improves estimation accuracy, as the poor-quality data is rejected. However, ESQI will result in a loss of data in online/real-time applications and may not be applicable in many cases. We wanted to assess whether the proposed algorithm can produce acceptable results without ESQI. The MAE and RMSE

values (in terms of breaths per minute) are close to each other endorsing the claim of accurate estimation of respiratory rate even using a low-quality signal.

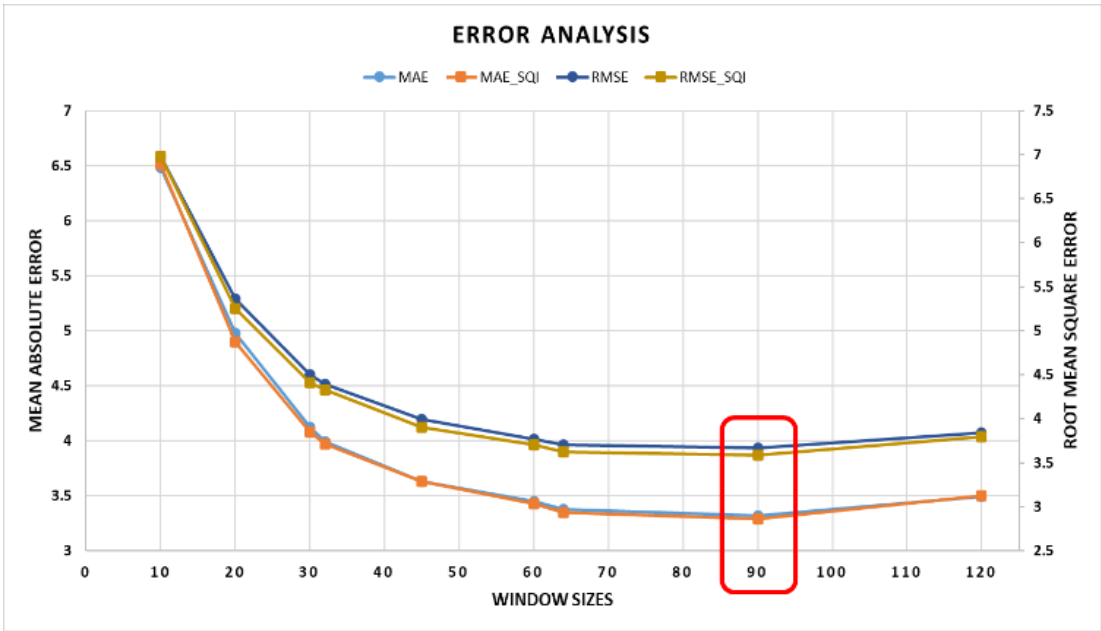


Figure 5.6 Error analysis of estimated respiratory rate (breaths per minute) using different window sizes, with and without ESQI.

From Figure 5.5, it can also be noticed that the minimum mean absolute error (MAE) and root mean square error (RMSE) are achieved by using a window size of 90 seconds for all the subjects. The error rate continues to decrease till a window size of 90 seconds (highlighted by the red box) and increases for a window size of 120 seconds.

5.5.2 Selection of best window size

The best window size for respiratory rate estimation varies from subject to subject. In the real-world scenario, the user can calibrate the respiratory rate monitoring device beforehand by taking regular breaths and manually entering them into the device. The device will then compare the estimated respiratory rate with the manually entered value and determine the suitable window size for respiratory rate estimation.

When MAE and RMSE are calculated using the best window size for each subject, the overall error decreases from 3.32 (breaths per minute) and 3.67 (breaths per minute) to 2.15 (breaths per minute) and 2.56 (breaths per minute) (without any signal quality assessment) while from 3.29 (breaths per minute) and 3.59 (breaths per minute) to 2.05 (breaths per

minute) and 2.47 (breaths per minute) (with ESQI assessment), respectively, as shown in Table 5.4. This improvement results in an over 35% reduction in mean estimation error.

Table 5.4 Error in respiratory rate estimation using 90 sec and best-suited window sizes (unit for MAE and RMSE = breath per minute).

Metrics (without E_{SQI} criteria)			**Metrics (with E_{SQI} criteria)**		
Window	**90 sec**	**Best suited**	**Window**	**90 sec**	**Best suited**
MAE	3.32	2.15	**MAE**	3.29	2.05
RMSE	3.67	2.56	**RMSE**	3.59	2.47

Table 5.5 Comparison of proposed respiratory rate estimation algorithm: Mean Absolute Error (MAE) and Window Sizes.

Algorithm	**MAE (breaths per minute)**	**Window Size**
Karlen *et al.* [357]	5.80	32
Pimentel *et al.* [382]	4.00	
Nilsson *et al.* [383]	5.40	
Fleming *et al.* [384]	5.20	
Proposed	**3.97**	
Karlen *et al.* [357]	5.70	64
Pimentel *et al.* [382]	2.70	
Nilsson *et al.* [383]	4.60	
Fleming *et al.* [384]	5.50	
Proposed	**3.35**	
Proposed	**2.05**	Best Window Size*
*Calculation is done using the best window size for each subject; see Table S2 (Appendix Chapter 4, appendix 9.2)		

5.5.3 Comparison with state-of-the-art respiratory rate estimation algorithms

The performance of the proposed algorithm was compared with Karlen *et al.* [357], Pimentel *et al.* [382], Nilsson *et al.* [383] and Fleming *et al.* [384]. These algorithms are representative of key studies that are either state-of-the-art or considered benchmark investigations. Table 5.5 shows the values of MAE for each algorithm using a window size of 32 and 64 seconds. The proposed algorithm gives compatible accuracy to the existing algorithms with the MAE of 3.97 (breaths per minute) and 3.35 (breaths per minute) for each window, respectively. The algorithm also gives the lowest MAE value i.e., 2.05 if the best window size for each subject is used.

5.6 Conclusion

In this study, we proposed an algorithm to extract the respiratory rate from a PPG signal. The algorithm is based on three steps that are pre-processing, signal analysis, and post-processing. In the pre-processing stage, the signal is analysed for required signal extraction, filtration of the signal above and below the expected range, and peak enhancement for increasing the signal-to-noise ratio. In the signal analysis stage, peak detection, peak-to-peak interval, error in peak detection and correction, updated peak-to-peak interval, calculation of different time-series measurements and estimation of respiratory rate is done. As the amplitude of the PPG signal is affected by respiratory rate, in the final stage, scaling is performed based on amplitude variation.

For the validation of the proposed algorithm, we used the BIDMC signal set and calculated the mean absolute error and root mean squared error. One of the aims of this study was to determine the impact of different window sizes on the calculation of respiratory rate in real-time. The results in Figure 5.5 suggest that a window size of 90 seconds is best for the estimation of respiratory rate using the BIDMC signal set, as it gives minimum MAE and RMSE values. The best window size to estimate the respiratory rate differs from person to person. If the best window size for each subject is used for the error analysis, then the maximum MAE and RMSE of our algorithm decrease to 2.05 and 2.47, respectively (see Table 5.3).

In the future, the scaling technique will be improved which will eventually improve the estimation accuracy furthermore. The scaling value is the only hyperparameter that might need to be determined empirically. The default method of scaling does work for most of the PPG data but may require improvement in some cases. To solve this problem, the algorithm needs to be evaluated on different datasets to determine a more generalizable scaling value to estimate the respiratory rate accurately.

The developed algorithm can estimate the respiratory rate from the PPG signal collected through a pulse oximeter to provide a simple, cheap, and signal-sensor solution. Integration of the proposed algorithm to a commercially available pulse oximetry device would expand its functionality from the measurement of oxygen saturation level (SpO2) and heart rate to

the continuous measurement of the respiratory rate with great efficiency in the clinical setting as well as in the ambulatory home-based environment.

Supplementary Files

In the appendix Chapter 5, Fig. Chapter 5 S1 to Chapter 5 Fig. S9 shows the Bland-Altman plot of all the subjects using a window size of 10, 20, 30, 32, 45, 60, 64, 120 and the best window in seconds. According to the United States Food and Drug Administration (FDA), repeated measurements through a device must lie within the allowed 36 (± 3* standard deviation) range to be classified as a Class II medical device [385]. In the proposed case, the bias values of all the Bland-Altman plots in the supplementary figures are near zero while 95% of the data lies within the limits of agreement (±1.96 * standard deviation), Tabulated in Table Chapter 5 A1 (Appendix Chapter 5). These results endorse that the developed algorithm can estimate respiratory rate close to the reference respiratory rate and thus can be implemented in the commercially available devices that collect PPG signals for long-term continuous respiratory rate measurements.

Chapter 6

A Pilot Study using Non-Invasive Wearable Device and Stress-Predict Dataset

While performing studies presented in Chapter 3, Chapter 4 and Chapter 5, a gap of limited, well-annotate, publicly available stress-related datasets was identified. Most of the publicly available datasets did not provide annotation of stress and baseline data readings, which made them difficult to use for the research. Moreover, if they had annotations, they lack the respiratory rate reading (which is determined to be the most sensitive and specific indicator of stress). Thus, a clinical study was designed to fill this gap by developing a publicly available stress-predict dataset and validating the developed algorithm (Chapter 5) on the healthy volunteer's data (to complete the fifth objective of the thesis).

This study[6] monitored the stress levels of healthy subjects using wrist-worn watches i.e, Emaptica E4. The optical sensor (PPG sensor) detected the signals corresponding to blood volume changes under the human skin and records them on the watch. While wearing these watches, the 35 healthy volunteers underwent a series of tasks and answered questionnaires designed to induce a level of everyday stress for 60 minutes. The changes in the blood flow during each task were recorded by the watch and were labelled as occurring during a stress-induced activity or a rest period (no stress). The information gathered from the watches

[6] *The following body of the chapter is exact copy of the paper published in Sensors (2022). I am the first lead author in these papers, which is co-authored with my supervisors. The clinical study is been approved by Clinical Research Ethics Committee, Merlin Park Hospital, Galway, Ireland on 19th January 2022 as: "Ref: C.A. 2731 - Stress levels monitoring using sensor-derived signals from non-invasive wearable device and dataset development" (approval letter attached in Chapter 5 Appendix). All the authors have a substantial contribution to preparing ethics applications. Nicola Glynn and John Killilea helped me in volunteer recruitment and data collection. Formal analysis and visualization were done with the help of Andrew Simphkin and Davood Roshan. Designed protocol and outcome validation was led by my supervisors. I led all parts of the work with the support of my supervisors.*

and the questionnaires is compiled together as Stress-Predict Dataset. The dataset is consisting of heart rate and estimated respiratory rate readings, performing three stress-inducing tasks (i.e., Stroop colour test, interview session and hyperventilation) along with a rest period in between each task.

Most stress detection and monitoring studies report only the classification results and lack a statistical analysis of the extracted features. Moreover, the studies that report statistical analysis results perform a group analysis (considering all participants as a single group). The limitation of the group analysis is that, within a group (each participant), the variability of stress-related parameters is quite high. For example, the normal heart-rate values of one participant could overlap with the stressed heart-rate value of another participant and vice versa. The group analysis exploits this variability and thus results in biased outcomes. In this study, an individual analysis was also performed, along with a group analysis to get better insightful information.

To validate that the difference in the stress versus baseline readings is significant, the linear mixed model was implemented for group analysis while the development of a personalized adaptive reference range allowed for individual-level monitoring of heart rates and respiratory rates. Both models validated the hypothesis that the physiological data collected during stress and non-stress/baseline task are statistically differentiable.

6.1 Stress Monitoring Using Wearable Sensors: A Pilot Study and Stress-Predict Dataset

Abstract: With the recent advancement in the field of wearable technologies, the opportunity to monitor stress continuously using different physiological variables has gained significant interest. The early detection of stress can help improve healthcare and minimizes the negative impact of long-term stress. This chapter reports the outcomes of a pilot study and associated stress monitoring dataset, named "Stress-Predict Dataset", created by collecting physiological signals from healthy subjects using wrist-worn watches with a photoplethysmogram (PPG) sensor. While wearing these watches, 35 healthy volunteers underwent a series of tasks (i.e., Stroop colour test, Trier Social Stress Test and Hyperventilation Provocation Test) along with a rest period in between each task. They also answered questionnaires designed to induce stress levels compatible with daily life. The changes in the blood volume pulse (BVP) and heart rate were recorded by the watch and were labelled as occurring during stress-inducing tasks or a rest period (no stress). Additionally, the respiratory rate was estimated using the BVP signal. Statistical models and personalised adaptive reference ranges were used to determine the utility of the proposed stressors and extracted variables (heart rate and respiratory rate). The analysis showed that the interview session was the most significant stress stimulus causing a significant variation in heart rate of 27 (77%) participants and respiratory rate of 28 (80%) participants out of 35. The outcomes of this study contribute to understanding the role of stressors and their association with physiological response and provide a dataset to help develop new wearable solutions for more reliable, valid, and sensitive physiological stress monitoring.

Keywords: Stress-predict dataset; photoplethysmogram (PPG); biomedical signal processing; adaptive reference ranges; non-invasive devices; health monitoring; heart rate; respiratory rate.

6.2 Introduction

Stress is known as a silent killer that contributes to several life-threatening health conditions such as high blood pressure, heart disease, and diabetes. According to the British Health and Safety Executive, 50% of all work-related illnesses in 2021-22 are due to stress [7]. Stress has negative effects on mental health as well as the overall well-being of a person

[284]. Short-term stress may not impose any threat to young and healthy people who have an adaptive coping response, but if the stressful experience is too persistent or too strong, it may increase the risk of developing chronic conditions associated with depression and anxiety [77]. Long-term stress is also known to increase the risk of life-threatening illnesses such as heart disease, high blood pressure, diabetes, and obesity while an acute episode of stress can potentially trigger a heart attack or stroke [285]. In clinical settings, the subjective experience of stress is evaluated from psychometric methods such as self-reported questionnaires e.g. the Perceived Stress Scale (PSS) [386] and/or State-Trait Anxiety Inventory (STAI) [88].

To develop a reliable objective stress monitoring device, it is essential to understand the effects of stress from the perspective of changes in the relevant physiological and biochemical variables. During standardized stress inducting procedures, the sympathetic nervous system of the body is triggered causing the release of different hormones (such as cortisol or adrenaline) [14], [15]. These hormones lead to changes in respiratory rate, and heart rate and trigger muscle tension among other physiological responses that prepare the body for fight or flight reactions. Both physical and biochemical changes can be used as indicators of stress and measured using different wearables. Some real-time stress monitoring devices/models are described in [308]–[310], [387]–[389]. There are multiple reasons behind the lack of a reliable objective stress monitoring device/model. Foremost of which is the absence of a universally acceptable definition of stress. Moreover, lack of gold standard ground-truth/reference values or data, collection of stress data in the natural environment, different confounding variables, identification of discriminative/specific stress features, and development of an accurate classifier model to classify stress data from baseline/normal are also contributing reasons for lack of unswerving stress monitoring device. Further details are explained in [283].

6.2.1 Related Work

The proposed study is inspired by several existing works in the field of wearable devices for stress detection and monitoring. The WESAD (Wearable Stress and Affect Detection) dataset [19] was created by using RespiBAN and Empatica E4 as wearables. The authors monitored the stress levels of 15 students while they were watching movies and taking a trier social stress test (TSST). The random forest classifier achieved an accuracy of 75.2%

using blood volume pulse (BVP), electrodermal activity (EDA) and temperature readings while distinguishing between three classes (baseline vs stress vs amusement). The SWELL-KW dataset [331] used video (for facial expression), computer logging, and Kinect (3D sensor for body posture) to monitor the stress levels of 25 people while they were performing typical knowledge work (making a presentation, reading, writing reports, email) under time pressure. The author reported averaged subjective experience scores using task load, mental effort, emotion, and perceived stress questionnaires for all the subjects. The study concluded that based on subjective scores, there was no significant effect of work conditions on perceived stress levels. The Affective-Road dataset [390] used Zephyr Bioharness 3.0 and Empatica E4 to study the stress levels under different driving conditions for 10 drives. The data were collected during driving for 1 hour 26 minutes on different types of roads and traffic conditions and no statistical or classification analysis was performed on the dataset. The author suggested that their prototype provides an accurate collection of different signals. Thus, vehicle manufacturing companies can embed the system into their vehicle and provide a real-world experimental dataset for studying the effect of road type on drivers' stress levels. Healey et al. [391] developed a wearable glove with a photoplethysmogram (PPG) sensor embedded, to monitor the stress levels of 10 drivers while they drove on different routes. Stress vs normal state classification was performed using electrocardiogram (EKG), electromyography (EMG), Respiratory rate and galvanic skin response (GSR) signals. An accuracy of 62.2% was reported by the authors using a sequential forward floating selection (SFFS) k-NN classifier. Shi et al. [70] developed a multi-node stress monitoring system based on ECG, EDA and PPG signals. They collected data from 22 subjects and reported that a support vector machine (SVM) model gave the highest accuracy of 68% in distinguishing stressed conditions from normal. Similarly, Muaremi et al. [294] were able to detect different stress levels using a smartphone and Wahoo wearable chest belt. They experimented on 35 subjects, collecting heart rate variability and smartphone application (questionnaires) data for 4 months. The combination of both information resulted in the highest three stress levels (low, moderate and high) classification accuracy of 61% using logistic regression-based leave-one-outcross-validation. Hosseini et al. [392] created a multi-sensor dataset of nurses working in the hospital during the COVID-19 outbreak. They used Empatica E4 watches to collect information about the electrodermal activities, heart rate, and skin temperature of the

subjects. The authors concluded that the device was unable to detect physiological differences across the various stress exposures.

From the literature review, it can be concluded that the optimal measurement approach for physiological stress monitoring is still unclear. Different studies have used the same physiological variables and classifiers but have reported significantly different classification accuracies. Furthermore, there is no clear understanding of the relative sensitivity and specificity of stress-related biophysiological indicators of stress (such as heart rate and respiratory rate) in the literature [283], [292]. All the above-mentioned datasets have a relatively small sample size and are more focused on performing classification analysis with reported accuracies in the range of 60 to 70% rather than performing statistical analysis of the dataset. These analyses are critical in understanding the relative importance of the most common and clinically relevant physiological stress indicators as well as in identifying the most specific indicators of stress for the development of a reliable stress monitoring device.

6.2.2 Study Objectives

Accurate monitoring of physiological stress levels has the potential to assist physicians in guiding their patients to adapt their lifestyle decisions e.g., individual occupational contexts, inform personalised treatment plans and ultimately improve their overall health. Therefore, this study aims to develop a *stress-predict dataset* and perform statistical analysis of biophysiological data collected from healthy individuals, who underwent induced psychological stress, to assess the relative sensitivity and specificity of common biophysiological indicators of stress and provide a stepping-stone towards the development of an accurate stress monitoring device. In this study, 35 healthy volunteers performed three different stress-inducing tasks (i.e., Stroop colour word test, Trier Social Stress Test and Hyperventilation Provocation Test session) with a baseline/relax period in between each task. Blood volume pulse (BVP), inter-beat-intervals (in milliseconds), and heart rate (in beats per min) were continuously recorded using Empatica watches. The key objectives of the study are as follows:

- Collect physiological data for wearable stress monitoring (Stress-Predict dataset).
- Perform statistical analysis and analyse the dataset to study the association of various physiological variables and stress levels.
- Assess the effectiveness of stress-inducing activities for experimental studies.

6.2.3 Key Contributions

The key contributions of this study are as follows:

- Collected PPG signals using an Empatica E4 watch (a wrist-worn device) and developed an open-access dataset.
- Estimated respiratory rate readings from the raw signal using a novel PPG-based respiratory rate estimation algorithm [393] and included them in the dataset.
- Performed individual-level statistical analysis using a novel method based on the Bayesian framework and time-efficient approximate Expectation-Maximisation (EM) algorithm [394].

The rest of the chapter is organised as follows: Section 6.3 provides an overview of the proposed protocol and data analysis metrics; Section 6.4 presents details of data features included in the dataset; Detailed analysis and results are provided in Section 6.5 while Section 6.6 concludes the chapter and discusses protocol limitations and provides future directions towards the development of an accurate stress monitoring device.

6.3 Material and Methods

6.3.1 Study design

This was a research study aimed at providing useful information and facts on stress in healthy individuals from recorded data using a wrist-worn watch. The study was a quasi-experimental repeated measures design where participants were assessed across a set of standardised psychological stress induction protocols over a 60-minute laboratory-based testing session. There was no longer-term follow-up on the participants involved. It was an opportunity to sample from a healthy individual population.

6.3.2 Selection and Recruitment of Participants

All study participants were selected and consecutively enrolled in the study based on the inclusion/exclusion criteria specified in Table 6.1. If the participant is eligible for inclusion and informed consent is obtained, the participants were entered into the study enrolment log and assigned a unique subject ID number. This healthy volunteer study was advertised via brochures and posters at the University Hospital Galway (UHG) and the University of Galway. The clinical research team also helped recruit volunteers. The study protocol and

patient informed consent forms were approved by the local Ethics Committee (on 19 January 2022 Ref: C.A. 2731).

Table 6.1 Selection Criteria.

Inclusion Criteria	Exclusion Criteria
Healthy (no underlying health condition)	No consent
Age between 18 and 75 years	Unhealthy
English-speaking (all ethnicities)	Breastfeeding mothers, pregnant women
Give consent	Colour-blind

6.3.3 Study Methodology and Protocol

The study, adapted from [395], was completely non-invasive and took approximately 60 minutes to complete for each participant. The consent form was given to the interested participant who was given sufficient time (2 days) to read, understand, and ask any question to the Lead Researcher/Investigator. During this period, the participants had to decide whether they wanted to participate in a research study. All participants were asked to read and sign the consent form before the start of the study. Moreover, at the beginning of the experiment, each participant was reminded of the order of phases, the duration of each phase, and what they were required to do in each phase.

It is well established that social evaluative acute stressors such as the colour word (Stroop-CW) test and the Trier Social Stress Test elicit the strongest physiological responses in laboratory settings when compared to cognitive challenges [82], [396]. There is also the argument that conducting interviews is a more ecologically valid analogue of real-world social stressors that we are interested in [397]. The two questionnaires (The PSS and STAI) are the most popular ways of assessing stress [387], [398]–[400]. These questionnaires help us to understand how different situations affect participants' feelings and anxiety. Furthermore, to estimate the respiratory rate from the PPG signal, a 2-minute Hyperventilation Provocation Test task was also performed to obtain a reference reading. After each task, the participant was asked to relax for 5 minutes. The following protocol, illustrated in Figure 1, was followed in the proposed study.

Stress-inducing tasks might induce some degree of lasting stress. If participants felt stressed during or after the study, the research team including clinical nurses made sure that they get enough time to relax before starting a new task or going home. Furthermore, the

participants were instructed to contact the Clinical Research Facility Galway, University Hospital Galway or the Student Health Unit, the University of Galway in the event of persistent stress.

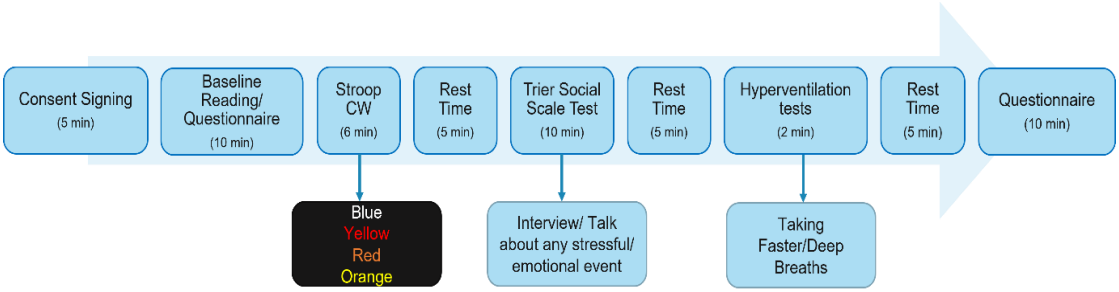


Figure 6.1 Study Protocol of the stress monitoring study including 3 stress-inducing tasks/sessions, 2 self-reporting questionnaires sessions and in-between rest sessions.

6.3.4 Study Sample Size Calculation

Previously, in a detailed literature survey and statistical analysis to determine the most sensitive and specific parameters for stress monitoring, we concluded that the respiratory rate (RR) is the most important parameter for the detection of stress conditions [19], [283], [297]. The results of these statistical analyses were published in [283]. For this study, an easier and quicker option would be to have power for a paired sample comparison, i.e., comparing RR within individuals when they are stressed vs not stressed. We have a within-person or paired design, as each person will undergo periods of stress and no stress.

A clinically significant difference in stressed vs. unstressed respiratory rate is a 10% difference [401]. The control respiratory rate and variability reported in [283] was 12.35, so a 10% increase is 13.58. The variability in Respiratory Rate from the same paper was 2.5. Using these summary statistics, a sample size of n=34 participants is required to achieve 80% power to detect a 10% change in RR, at the alpha 0.05 significance level. Thus, a total of 35 healthy volunteers (females and males) aged between 18 and 75 years old were recruited for the study, to allow for attrition.

6.3.5 Data Acquisition

In this study, an Empatica E4 watch (Figure 6.2) was used to measure individual physiological changes based on PPG, which has been previously used in several similar studies [19], [390], [392], [402]–[405]. The watch is a medical-grade device that is classified

as Class IIA Medical Device according to the 93/42/EEC Directive. Empatica E4 is a wireless multi-sensory platform designed to acquire real-time physiological data with ease.

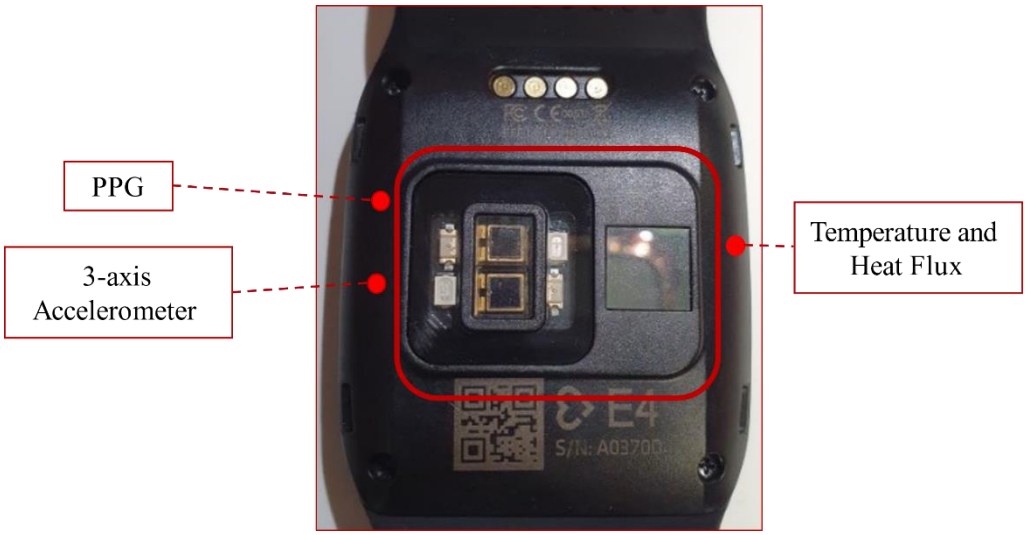


Figure 6.2 Empatica E4 watch.

6.3.5.1 Empatica E4 photoplethysmogram (PPG) sensor

The PPG sensor embedded in the watch has a sampling rate of 64 Hz. The raw PPG signal is filtered to get a clean blood volume pulse (BVP) signal, which is then passed to the heart rate (HR) and inter-beat intervals (IBI) estimation algorithm. There are 2 green and 2 red light-emitting diodes (LEDs) that transmit light onto the skin. To receive the reflected light, 2 photodiodes in place have an area of sensitivity of 14 mm^2.

The output of the PPG sensor is digital and has a resolution of 0.9nW/Digit. Exposure to green light contains information about heartbeats, whereas exposure to red light assists in reducing noise or motion artefacts by dynamical compensation performed by the built-in firmware. The accuracy of Empatica E4 heart rate readings is highly comparable with the standard ambulatory monitory system. A detailed comparison of Empatica E4 readings with ambulatory monitoring systems is provided in [406].

The participants were asked to wear the watch on the non-dominant wrist. The watch can be operated in memory mode. In this mode, the data is stored on the built-in memory of the watch and once the session is completed, the data is uploaded to the 'Connect' cloud via personal computer or laptop using Empatica E4 manager software. Data can be

visualized in the cloud for visual analysis and could be downloaded from the cloud in .csv format. The data of each sensor, as well as estimated heart rate and inter-beat intervals, are downloaded as separate files. In our case, the start and end of each task were labelled by clicking the button on the watch, thus along with the physiological data, corresponding tags were also generated.

To induce stress, the participant performed the Stroop Colour-Word task, Trier Social Stress Test, and Hyperventilation Provocation Test tasks. From the start of the experiment to the end of the recording, each section was labelled using a built-in function (pressing the button on the watch once). Labelling helped us identify whether any stressor had a prolonged reaction even after the stimulus.

6.3.6 Data Analysis Matrices

Two statistical analyses are used to determine the utility of the stress-predict dataset:

6.3.6.1 Linear Mixed Model analysis

A linear mixed model was implemented for population-based analysis to determine the effect of stressors on HR and RR while accounting for the correlation of these variables within each person over time. A separate model was run for RR and HR, with random intercepts and slopes included for each participant to allow for within- and between-subject variability. The binary group variable is included as a fixed effect, and the coefficient of this variable in the model describes the average difference in the result between stress and normal situations. An interaction term between time and group (stress/normal) is also included, with the coefficient of this providing an estimate for the difference in change in outcome over time between stress and normal situations. Results are reported as coefficients, with 95% confidence intervals and *p*-values from linear mixed models.

6.3.6.2 Adaptive reference range analysis

For the development of an extensive understanding of changes in participants' responses over the study time, an individual-level statistical analysis was performed through the development of personalised adaptive referencing ranges, proposed by Davood *et al.* [394]. In this method, to see if there are any meaningful changes in a particular participant's response over time, individualised reference ranges are developed which adapt successively

whenever a new measurement is recorded for the individual. In the context of this work, adaptive reference ranges are generated sequentially according to normal physiological rates at each resting phase and then the following stress data are included for comparisons. Any value outside the developed adaptive reference range can be considered as an 'alert' that requires further consideration. The adaptive referencing range method works using a Bayesian framework and time-efficient approximate Expectation-Maximisation (EM).

6.4 Data Features included in Stress-Predict Dataset

The created dataset consists of physiological data collected from 35 students and employees of the University of Galway, Ireland, and the University Hospital Galway, Ireland. The participant performed three stress-inducing tasks along with 4 rest periods. All the readings have been tagged as the duration of the stress-inducing task and baseline/rest period. The time of each tag is available in the tag.csv file in the dataset. In all other files, the first row shows the timestamps in Unix timestamp UTC format, while the second column shows the sampling rate in Hz. The collected data starts from the third row and continues till the end. The data in the given Stress-Predict dataset is expressed as:

6.4.1 Blood Volume Pulse

This file has data collected by a PPG sensor. Typically, a BVP signal is obtained by passing the PPG through a high-pass filter. The cut-off frequency of this filter can be arbitrary but typically set between 0.05 and 0.5 Hz. The data in the file represent the BVP value calculated by the built-in algorithm in units of nanowatt units (nWatt). Figure 6.3 shows the typical PPG signal and its significant points.

In the PPG signal:

- The diastolic point is the local minima point, used to calculate the inter-beat-interval.
- The systolic point is a local maxima point, used to calculate the vasoconstriction of the participant.
- The presence of a dicrotic notch is observed in the study of different types of cardiac diseases.
- The dicrotic wave is the effect of the dicrotic notch and is referred to as the second wave.

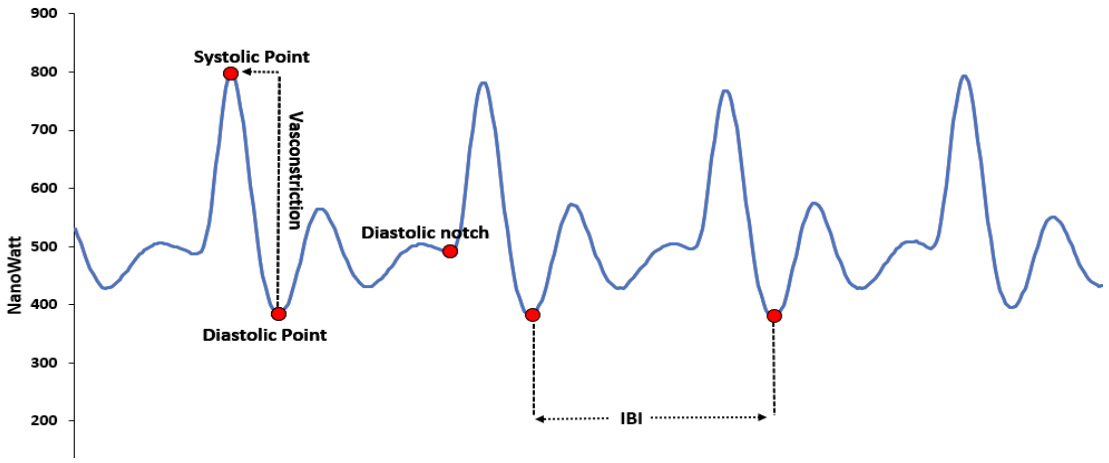


Figure 6.3 PPG signal obtained in typical condition, from the green and red light.

6.4.2 Inter-Beat-Intervals

The file contains time intervals between two consecutive heartbeats. The IBI values in the dataset are obtained by processing the BVP signal, with an algorithm that already eliminates the incorrect peaks in the signal generated due to noise. Figure 4 shows the PPG/BVP signal with some motion artefacts. The green dots show correct heartbeats while the red dots show incorrect heartbeats, corresponding to the time of movement. The timing of incorrect beats is not included in the IBI file, as demonstrated in Figure 6.4. The first row shows the timestamps in UNIX format, while the first column (excluding the first row) illustrates the time of detected inter-beat-intervals in seconds (s). The second column shows the distance in seconds (s) from the previous beat (the detected IBI); see Table 6.2.

Table 6.2 Inter-beat-intervals in IBI.csv file (unit = μS).

UNIX Start t	IBI
t0	d0
t1	d1
t2	d2

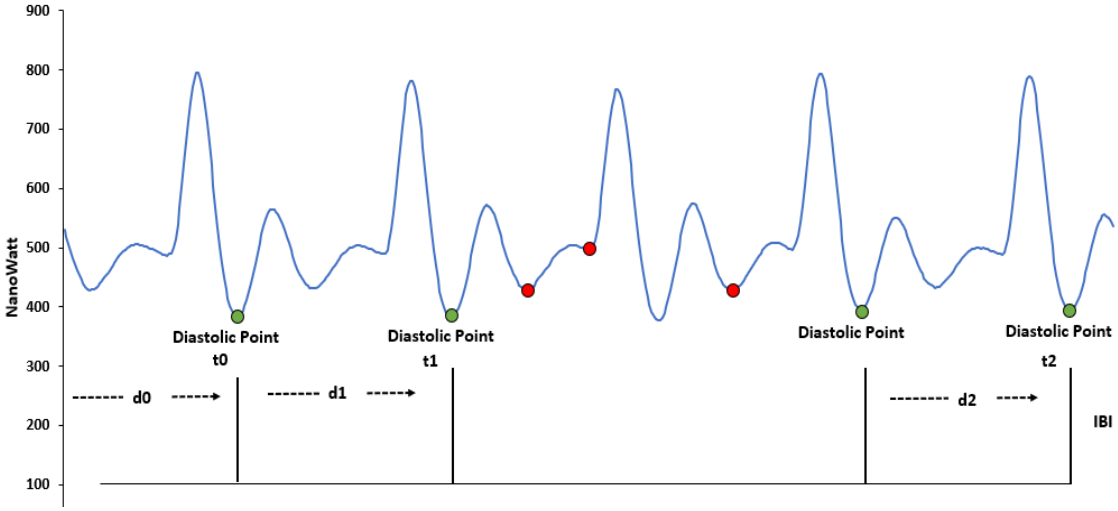


Figure 6.4 Inter-beat-intervals calculation. The green dots show valid peaks while red dots show the discarded peaks.

6.4.3 Heart Rate

The file has average heart rate values calculated by the watch from the raw BVP signal. The heart rate in this file is calculated with a 10-second sliding window. It is created only when the session is completed and uploaded to the Empatica *'connect'* cloud. The unit of each heart rate is beats per minute (BPM). Instantaneous heart rate can only be viewed during stream mode or online (in view session).

6.4.4 Labels

The file contains time marks when an event is marked. Each row corresponds to the physical button pressed on the watch. The time is presented in form of Unix timestamps in UTC and is synchronized with the initial time of the session indicated in the related data file.

6.4.5 Estimation of Respiratory Rate data

The created dataset is different from all existing public datasets, as it included information on the estimated RR for each participant. As explained in [393], the study proposed a novel RR estimating algorithm that worked on raw PPG signals. As BVP is a filtered form of raw PPG signal, the developed algorithm was able to estimate the RR of the participants during stress as well as rest/baseline time. The algorithm was implemented in three-fold steps (pre-processing, signal analysis, and post-processing).

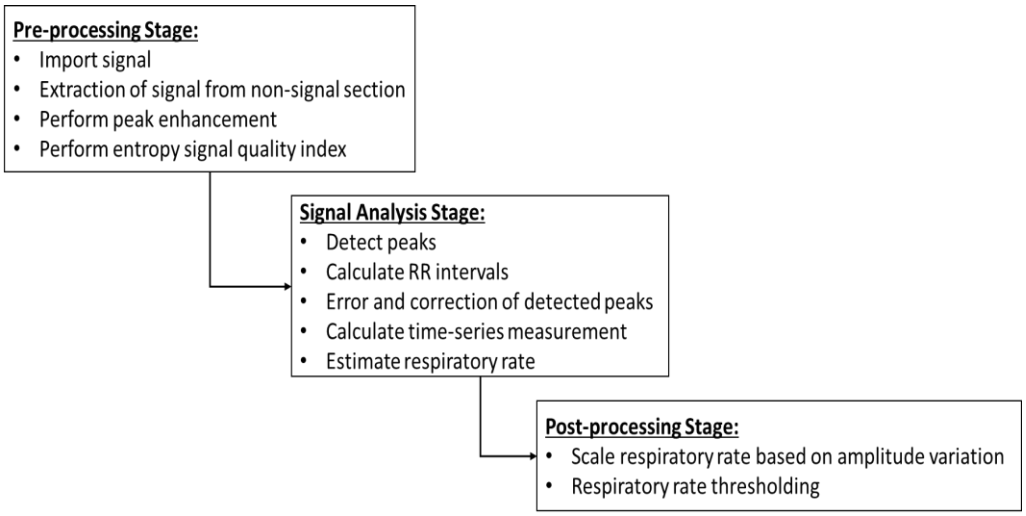


Figure 6.5 Pre-processing, signal analysis and post-processing steps of the RR estimation algorithm.

Figure 6.5 explains the steps of the proposed RR estimation algorithm. In the pre-processing stage, peak enhancement was performed to increase the signal-to-noise ratio and better signal information extraction. Peak detection, peak-to-peak interval, error and correction in peak detection, calculation of time-series measurement and estimation of RR were done during the signal analysis stage. Usually, the BVP waveform is synchronised with the respiratory cycle [48], thus there is an amplitude variation induced in the raw signal. In the post-processing stage, the estimated RR is scaled based on the range (maximum-minimum value) and defined window size of the signal. The data of the estimated respiratory rate is included in the dataset as a separate file.

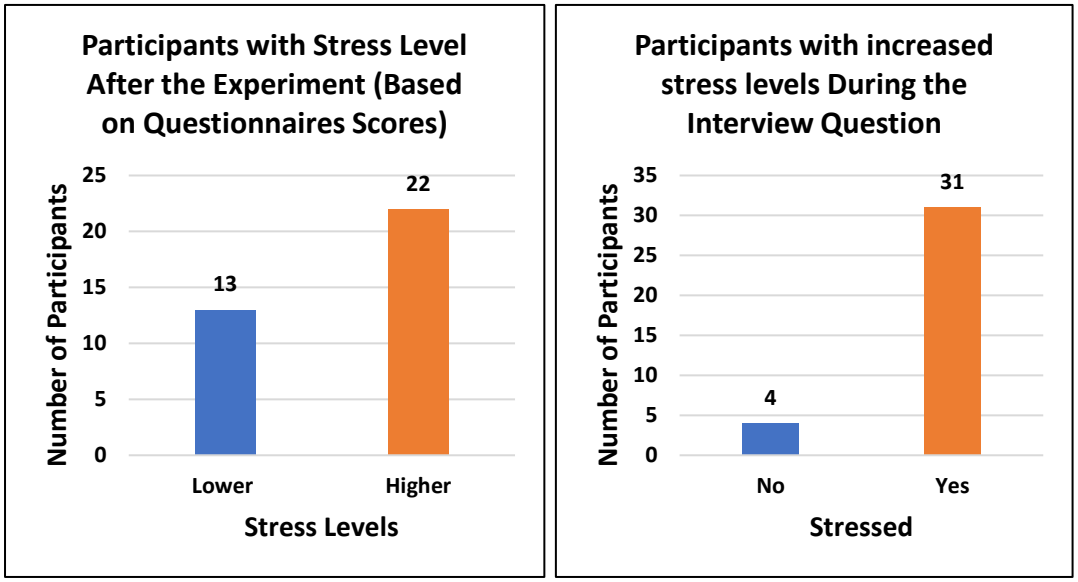
6.5 Analysis and Results

There were 25 women and 10 men participants (mean age = 32 ± 8.2 years) in the created dataset. The data collection protocol was not followed properly for one (1) participant and their data were removed from further analysis. The average number of entries per participant is presented in Table 6.3.

Table 6.3 The average number of entries (per participant).

Features	Samples
Blood Volume Pulse (BVP)	212234
Heart Rate (beats per min)	3308
Respiratory Rate (breaths per minute) [Calculated using sliding window of 10 sec]	3308

During the study, the participants were asked to fill out STAI and PSS questionnaires to rate their stress levels and were also asked during the Trier Social Stress Test that was they felt stressed out at any point in the study. Figure 6.6 shows the number of participants those had increased stress levels after the study (a) based on questionnaire scores and (b) based on the verbal query.



(a) (b)

Figure 6.6 Participants with increased stress levels (a) Based on Questionnaire score (b) Asked during Interview.

6.5.1 Population-based analysis using Linear Mixer Model

According to the results in Table 6.4, during the stress state, the HR was 1.40 beats per minute higher on average compared to normal state HR (95% CI 1.10, 1.71; $p < 0.001$). Participants also experienced a 5.05 bpm higher change in HR per hour compared to during a normal state (95% CI 4.36, 5.74 bpm/hour; $p < 0.001$).

Table 6.4 Linear Mixer Model Results for Heart Rate Parameter.

Predictors	Estimates	Confidence Intervals		p-value
		Lower	Higher	
(Intercept)	80.36	76.85	83.88	<0.001
Time	-2.65	-5.85	0.55	0.105
Group [Stress]	**1.40**	**1.10**	**1.71**	**<0.001**
Time * Group [Stress] (beats per hour)	**5.05**	**4.36**	**5.74**	**<0.001**
Observations	112472			

When exposed to stress, participants' average RR increased by 0.20 breaths per minute compared to their normal state (95% CI 0.16, 0.24 breaths per minute; $p < 0.001$), see Table 6.5. Participants also experienced -1.11 breaths per minute lower change in respiratory rate per hour compared to during a normal state. The drop in the RR can be related to sighs (deep breaths when under stress).

Table 6.5 Linear Mixer Model Results for Respiratory Rate Parameter.

Predictors	Estimates	Confidence Intervals		p-value
		Lower	Higher	
(Intercept)	12.68	11.98	13.39	<0.001
Time	-0.30	-1.02	0.41	0.408
Group [Stress]	**0.20**	**0.16**	**0.24**	**<0.001**
Time * Group [Stress] (breaths per hour)	**-1.11**	**-1.19**	**-1.03**	**<0.001**
Observations	112472			

Although the change in average HR and RR during the stress period is statistically significant (***$p<0.001$) when compared to the average value of the normal/baseline state, the difference may not be large enough for clinical decision-making.

6.5.2 Individual Participant's analysis using Adaptive Reference Range

In this method, each stress period is compared with the last reference range generated from the previous normal period to allow early detection of abrupt physiological changes.

Figure 6.7 (a) and (b) illustrate the developed reference ranges for HR and RR of participant 23, respectively. As can be seen from Figure 6.7(a), there are a number of atypical HR measurements for participant 23 in all three phases of this study. This is particularly true for the two Trier Social Stress Test and Hyperventilation Provocation Test sessions. This is where none of the respiratory rates goes outside the developed reference range during the Stroop colour test session, Figure 6.7(b).

It should be noted that the developed adaptive reference ranges are not a classification algorithm but are capable of triggering 'alerts' and should be used as an early warning system that warrants further attention and review.

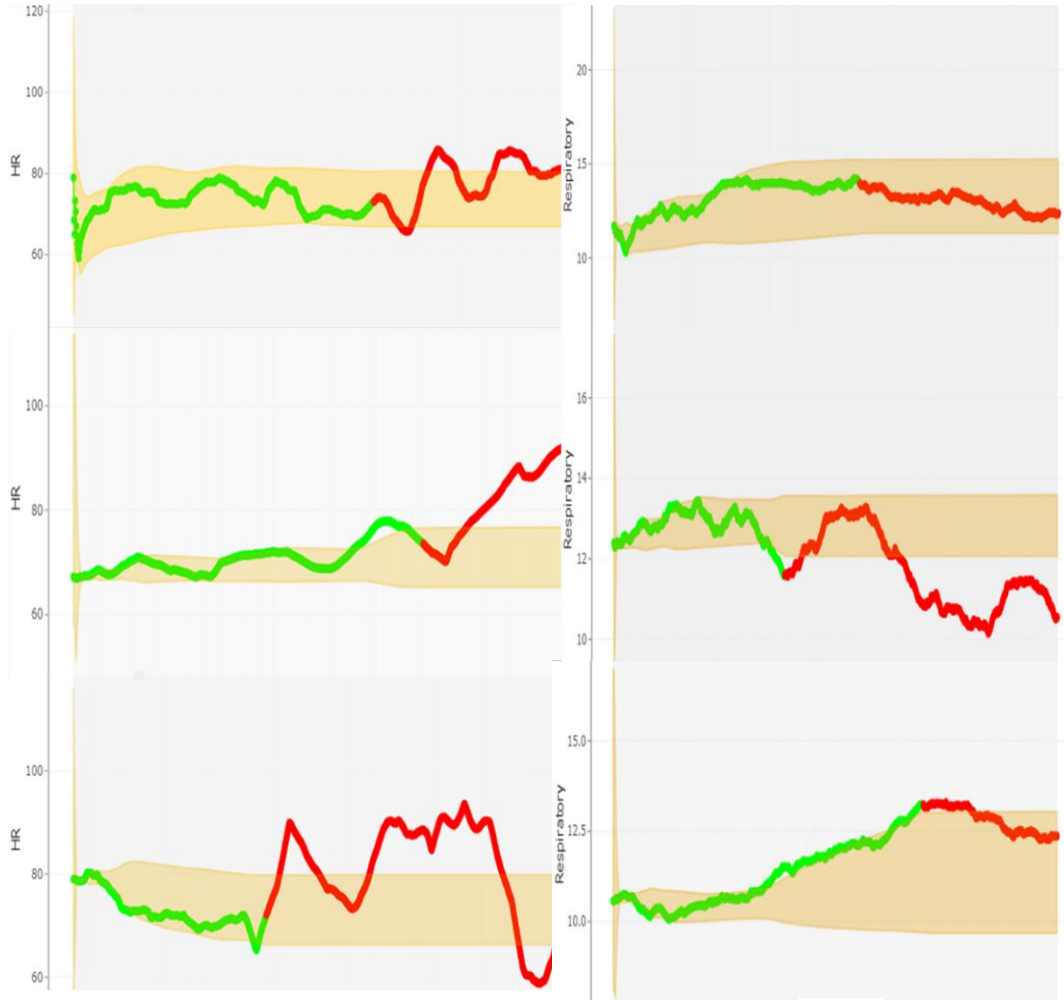


Figure 6.7 Statistical Analysis of Participant 23: Adaptive referencing range (shaded region) calculated by using approximate EM. (a) Heart rate reading: baseline (green) vs. stress (red) task (b) Respiratory rate: During each baseline (green) vs. stress (red) task.

Summative results of individual participants' analysis on HR and RR parameters are presented in Table 6.6.

Table 6.6 Summary of Statistical Analysis (adaptive reference range).

Heart Rate	**Respiratory Rate**	**Heart Rate**
Test	**Stress (Outside Baseline Values)**	**Stress (Outside Baseline Values)**
Stroop Test	24/34	19/34
Trier Social Stress Test	28/34	27/34

Hyperventilation Provocation Test	18/34	16/34

6.6 Discussion and Conclusion

The Stress-Predict dataset was developed using a commercially available wearable E4 watch by Empatica [407]. The purpose of developing this dataset was to analyse and identify different patterns of stress.

Most stress detection and monitoring studies report only the classification results and lack a statistical analysis of the extracted features. Moreover, the studies that report statistical analysis results perform a group analysis (considering all participants as a single group). The limitation of the group analysis is that, within a group (each participant), the variability of stress-related parameters is quite high. For example, the normal heart-rate values of one participant could overlap with the stressed heart-rate value of another participant and vice versa. The group analysis exploits this variability and thus results in biased outcomes. In this study, an individual analysis was also performed, along with a group analysis (linear mixed model), to get better insightful information. To validate that this difference in the readings is significant, the linear mixed model was implemented for group analysis while the development of a personalized adaptive reference range allowed for individual-level monitoring of heart rates and respiratory rates. Both models validated the hypothesis that the physiological data collected during stress and non-stress/baseline task are statistically differentiable. Table 6.7 provides a comparison of the proposed dataset with the state-of-the-art publicly available dataset.

Table 6.7 Summary: Comparison of Proposed Dataset with Existing State-Of-The-Art Datasets.

Study	Devices Used	No. of Subjects	Methods	Features	Limitations	Pros
[19]	RespiBAN and Empatica E4	15	BVP, EDA, EMG and Temperature sensors	Heart rate, skin conductance, respiratory rate, muscle activation and skin temperature	Uses a chest band. Subjects must remain immobile. Not a translational (practical) model	Data gather through chest bands is highly accurate. Respiratory rate data obtained by chest band

					(use of chest band) No justification for the selected sample size	
[331]	Video camera, computer logging and Kinect device	25	Task load, mental effort, emotion, and perceived stress questionnaires	Facial expression, computer logging and 3D body posture monitoring	Need control environment. Not a translational (practical) model (3D Kinect). No justification for the selected sample size	Provided subjective (personalized) results (stress versus work conditions)
[390]	Zephyr bio harness and Empatica E4	10	EDA, temperature, BVP, camera	Skin conductance and temperature, heart rate, respiratory rate, and hand movements	Not a translational (practical) model (use of chest band) No justification for the selected sample size	Data gathered through chest bands are highly accurate. Respiratory rate data obtained by chest band
[294]	Smart-phone and Wahoo chest belt	35	Number of calls, sleep length, distance, audio length, heart rate variation	Inter-beat-inter/heart rate	Not a translational (practical) model (use of chest belt) No justification for the selected sample size	Big data (4 months)
This Work	Empatica E4	35	BVP (wristband)	Heart rate and respiratory rate	Limited (approx. 60 min) data Small subject age window	A translational (practical) model Justification of selected sample size Only Empatica E4 dataset with respiratory rate data Provides subjective outcomes

The dataset is an open-access dataset named Stress-Predict dataset. The inclusion of an additional feature, i.e., respiratory rate data, along with stress and baseline labels within the dataset, makes the dataset more desirable and unique from all the other publicly available Empatica E4-based datasets. Additionally, the developed dataset will also help to evaluate proposed PPG-based feature extraction algorithms. The current dataset will certainly attract the attention and interest of researchers in the field of psychological, clinical, and biomedical research, as well as prevention, medicine, and connected health systems.

There were also some limitations to the study. First, at the start and end of the stress task, the HR and RR gradually changed, but there is no accurate way to determine this gradual

change. Thus, labelling was performed without considering these delayed changes. Secondly, sometimes there was more than one participant in the room where the study was conducted. The crosstalk, and especially the questions asked during the Trier Social Stress Test to induce stress, might be learned by the other participant during the resting/baseline period. Therefore, the effectiveness of the stress-inducing interview questions could have been decreased. Third, the interviewees were friendly and kind to the participant. They kept the overall interview environment friendly rather than mimicking a strict interview session, which might have resulted in less induced stress. Thus, there was less variation in the readings of stress versus non-stress parameters. To translate the proposed model to an ambulatory environment, the inclusion of activity data is also essential. In future studies, all these shortcomings should be considered to obtain an improved stress-monitoring model. Moreover, to obtain an accurate real-time stress monitoring system, accelerometer data might play an essential role in excluding the period of physical exercise, which causes changes in HR and RR. In future, the developed dataset might help in exploring, optimizing, and developing supervised and unsupervised machine-learning classifiers for the detection of physiological signal-based stress monitoring.

Supplementary Tables:

Descriptive Analysis: Variance in Heart Rate Readings During Each Task

Table S6.1. Stress vs non-stress based on a variance in the heart rate signal during each task – **Descriptive analysis**			
Participant	Stress or not (based on **heart rate** variance) Y for Yes; N of No		
	Stroop Colour Word	Interview	Hyperventilation
2	N	Y	N
3	N	N	N
4	Y	Y	Y (delayed)
5	Y	Y	N
6	N	N	Y (delayed)
7	Y	Y	Y (delayed)
8	N	Y	Y
9	Y	Y	Y (delayed)
10	Y	Y	Y (delayed)
11	Y	Y	Y
12	N	Y	Y (delayed)
13	N	N	Y
14	Y	Y	Y
15	Too Much Variation		
16	Too Much Variation		
17	Y	Y	Too much variation
18	Y	Y	Y
19	N	Y	N
20	N	Y	Y
21	Y	Y	Y (delayed)
22	Y	Y	Y (delayed)
23	Y	Y	Y (delayed)
24	Y	Y	too much variation
25	Y	Y	Y (delayed)
26	N	N	N
27	Y	Y	Y
28	Too Much Variation		
29	Too Much Variation		
30	Too Much Variation		
31	N	Y	Y (delayed)
32	Y	Y	Y (delayed)
33	N	Y	Y
34	Y	Y	Y (delayed)
35	Y	Y	N
Total (Y)	**18/34**	**25/34**	**8/34 (21/34 with delay)**

Statistical Analysis: Adaptive Reference Ranging (RR) Heart Rate Readings During Each Task

Adaptive Reference Range is generated sequentially according to normal heart rate and then the stress data are included for comparisons. In particular, each stress period is compared with the last RR generated from the previous normal period.

Table S6.2. Stress vs no-stress based on heart rate values outside the adaptive reference range during each task **– Statistical analysis**

Participant	Stress or not (based on **heart rate** outside normal/baseline value) Y for Yes; N of No		
	Stroop Colour Word	Interview	Hyperventilation
2	N	Y	N
3	Y	Y	Y
4	Y	Y	N
5	Y	Y	N
6	Y	Y	N
7	Y	Y	N
8	N	Y	Y
9	Y	Y	Y
10	N	Y	Y
11	N	N	Y
12	N	N	N
13	N	N	N
14	Y	Y	N
15	N	N	N
16	Y	N	N
17	N	Y	N
18	Y	Y	N
19	N	Y	N
20	N	Y	N
21	Y	N	N
22	Y	Y	N
23	Y	Y	Y
24	N	Y	Y
25	Y	Y	Y
26	N	N	N
27	Y	Y	N
28	N	N	N
29	Y	Y	N
30	Y	Y	N
31	Y	Y	N
32	Y	Y	N
33	N	Y	Y
34	Y	N	N
35	Y	Y	N
Total (Y)	**20/34**	**25/34**	**9/34**

Descriptive Analysis: Variance in Respiratory Rate Readings During Each Task

Table S6.3. Stress vs no-stress based on the variance in the respiratory rate signal during each task – **Descriptive analysis**			
Participant	Stress or not (based on **respiratory rate** variance) Y for Yes; N of No		
	Word Stroop Test	Interview	Hyperventilation
2	Y	Y	Y
3	N	Y	Y
4	Y	Y	N
5	Y	Y	N
6	Y	Y	N
7	Y	Y	Y
8	Too Much Variation	Y	
9	Y	Y	Y (delayed)
10	Y	Y	Y (delayed)
11	Too Much Variation		
12	N	N	Too Much Variation
13	Y	N	Y
14	Y	N	Y
15	Y	Y	N
16	Y + Delay	N	N
17	Too Much Variation		
18	Too Much Variation		
19	N	Y	Y (delayed)
20	N	Y	N
21	Y	Y	N
22	Y	Too Much Variation	Too Much Variation
23	Y	Y	Y
24	Y	N	Y
25	Y	Y	Y
26	Y (delayed)	Y (delayed)	Y (delayed)
27	Y (delayed)	Y	Y (delayed)
28	Y	Y	Too Much Variation
29	Too Much Variation		
30	Y	Y	Y
31	N	Y	Y (delayed)
32	Y	Y	Y (delayed)
33	Y	Y	Y
34	Y	Y	Y (delayed)
35	Too Much Variation	Y	Too Much Variation
Total (Y)	**20/34 (23/34 with delay)**	**23/34 (24/34 with delay)**	**10/34 (18/34 with delay)**

Statistical Analysis: Adaptive Reference Ranging Respiratory Rate Readings During Each Task

In this report, adaptive reference ranges (RR) are generated for each phase of the study. By phase, it means each round of the Normal –> Stress period. In particular reference ranges were calculated for each Normal phase and then Stress data were compared with the last reference range calculated from Normal data.

Table S6.4. Stress vs no-stress based on respiratory rate values outside the adaptive reference range during each task **– Statistical analysis**

Participant	Stress or not (based on **respiratory rate** outside normal/baseline value) Y for Yes; N of No		
	Word Stroop Test	Interview	Hyperventilation
2	Y	Y	Y
3	Y	Y	N
4	Y	N	N
5	N	Y	N
6	N	Y	Y
7	Y	Y	Y
8	N	Y	Y
9	Y	Y	N
10	Y	Y	Y
11	Y	Y	N
12	Y	Y	Y
13	Y	Y	Y
14	N	N	Y
15	Y	Y	Y
16	Y	Y	N
17	Y	Y	N
18	Y	Y	N
19	Y	Y	N
20	N	N	Y
21	Y	Y	N
22	N	N	N
23	N	Y	Y
24	Y	N	N
25	Y	Y	N
26	Y	Y	N
27	N	N	Y
28	Y	Y	N
29	Y	Y	N
30	Y	Y	Y
31	Y	Y	Y
32	Y	Y	Y
33	N	Y	Y
34	Y	Y	Y
35	N	Y	Y
Total (Y)	**24/34**	**28/34**	**18/34**

Chapter 7

Improved Stress Classification using Automatic Feature Selection from Heart Rate and Respiratory Rate Time Signals: Application to Stress-predict Dataset

The conventional statistical features (such as mean, variance, standard deviation, mean absolute deviation) of respiratory and heart rate do help in distinguishing between stress versus baseline readings but the reported accuracy (in literature and determined by experimentation in chapter 3 and Chapter 4) is low i.e., around 75%. There could be two major reasons behind the lower classification performance of the classifiers:

1. Lack of features that could translate/show the well-distinguishable patterns between stress and baseline readings.
2. The presence of highly correlated features within the dataset results in compromised generalizability and overfitting.

Thus, this chapter[7] proposed an optimized feature engineering algorithm to determine well-distinguishable features with high classification accuracy, using the Stress-Predict dataset (explained in Chapter 6). The algorithm calculates a list of 1578 features of heart rate and respiratory rate (combined) using the python library (tsfresh). These features are then shortlisted to the more specific time-series features using Principal Component Analysis (PCA) and correlation ranking techniques. A comparative study of conventional statistical features versus correlation-based selected features was performed using linear (logistic

[7] *The following body of the chapter is exact copy of the preprint paper which is published in Applied Sciences (2022). I am the first lead author in the paper, which is co-authored with my supervisors. I implement all the machine learning algorithms for comparison and led all parts of the work with the support of my supervisors.*

regression), ensemble (random forest), and clustering (k-nearest neighbours) predictive models.

Overall, the classification accuracy increased drastically (over 95%) as compared to the conventional statistical feature's 67.4%., when correlation-based time-series features are used. The respiratory signal pattern, based on the number of peaks that is greater than its 3 neighbours, was determined to be the best distinguishable time-series feature. This finding is perfectly correlated with the previous studies (presented in Chapter 3, Chapter 4, and Chapter 5) and is true, as the breathing pattern is supposed to vary significantly during stress conditions when compared to baseline/normal conditions. This work completes the final objective of the thesis i.e., providing a solution that calculates the well-distinguishable time-series features that could provide high accuracy while classifying stress conditions from no-stress/baseline condition and can be characterized as a stress response from any physiological stimuli.

7.1 Improved Stress Classification using Automatic Feature Selection from Heart Rate and Respiratory Rate Time Signals

Abstract: For accurate stress monitoring, it is essential that stress-related features are not only informative but also well-distinguishable and interpretable by the classification models. Time-series features are the characteristics of data collected periodically over time. The calculation of time-series features helps in understanding the underlying patterns and structure of the data as well as in visualising the data. The manual calculation and selection of time-series features, from a large temporal dataset, is a time-consuming process. It requires researchers to consider several signal-processing algorithms and time-series analysis methods to identify and extract meaningful features from the given time-series data. These features are the core of a machine learning-based predictive model and are designed to describe the informative characteristics of the time-series signal. Recently, a lot of work has been done on automating the extraction and selection of times-series features. In this chapter, a correlation-based time-series feature selection algorithm is proposed and evaluated on the stress-predict dataset. The algorithm calculates a list of 1578 features of heart rate and respiratory rate signals (combined) using the *tsfresh* library. These features are then shortlisted to the more specific time-series features using Principal Component Analysis (PCA) and Pearson, Kendall and Spearman correlation ranking techniques. A comparative study of conventional statistical features (such as, mean, standard deviation, median, and mean absolute deviation) versus correlation-based selected features is performed using linear (logistic regression), ensemble (random forest), and clustering (k-nearest neighbours) predictive models. The correlation-based selected features achieved higher classification performance with an accuracy of 98.6% as compared to the conventional statistical feature's 67.4%. The outcome of the proposed study suggests that it is vital to have better analytical features rather than conventional statistical features for accurate stress classification.

Keywords: Time-series, distinctive features, respiratory rate, heart rate, feature engineering, stress, classification.

7.2 Introduction

In recent years, sensor technologies have been developed significantly to help generate large data at relatively low-cost [3]. A time-series data is a sequence of measurements/observations taken sequentially in time [408]. In the context of stress, time-series data are referred to as a collection of data points over time that measure the psychological or physiological responses of an individual to any applied stressor. These data points are collected at regular intervals (per second, per minute or per hour) and include measurement of respiratory rate, heart rate, cortisol levels, blood pressure or a self-reporting stress level. The time-series data are analysed to understand the stress patterns over time, which provides better insights into how an individual's stress levels change in response to different stressors and interventions. The learnings can be then used to develop algorithms or classification models to predict individuals' stress conditions based on their response to stress.

The field of the Internet of Things (IoT) [409], precision medicine [410], and industry 4.0 [411] produce advanced large temporally annotated data. The analysis of this large temporal data, as it is, is a dilemma for researchers and data scientists. Thus, encourages the reduction of large time-series data into smaller series, capturing ample characteristics of primary data, for enhanced analysis. The resulting time series is the basis of machine learning applications, such as analysis of heartbeat [412] and respiratory rate [393], identification of high-risk patients who are at increased risk of infection [413], optimization of a production line [414], incident detection over cloud [415]. The reduction of large data to feature-based representation is crucial, as the implementation of the machine learning algorithm is straightforward, but the selection of well-discriminating features is a challenging task [4].

Considering T as a set of time-series data, where $T = \{F_i\}_{i=1}^{n}$. To use T set as an input to supervised or unsupervised classification algorithms, each time-series F_i must be mapped to a well-defined feature vector with X dimensions ($\overrightarrow{F_i} = (F_{i,1}, F_{i,2}, \dots, F_{i,X})$) [416]. The most efficient and effective way of feature extraction is to characterize the time-series data into a distribution of data, stationarity, correlation properties, entropy, and non-linear time-series analysis [417]. The extraction of only significant features is vital for both regression

and classification tasks, as the irrelevant features will weaken the algorithm's ability to generalize beyond the training set and cause overfitting [418].

7.2.1 Related Work

Many existing stress classification studies use trivial feature extraction and selection methods. These studies use either raw data (data collected through the sensors) [419], [420] or common features of the collected data [421]–[423] such as rate of change, mean, standard deviation, variance, mean absolute deviation, and skewness. This set of features does not describe the characteristics of the dataset fully and is also unable to be generalised to another time-series dataset. This is because the underlying patterns and characteristics of a particular dataset are specific to itself and cannot be applied to another dataset. Furthermore, the above-mentioned features are also affected by the conditions under which data are collected and pre-processing steps performed on them.

Christ *et al.* [416] proposed *tsfresh*, a machine-learning time-series feature extraction library that has been used in several studies. The library uses the method called AutoTS (Automatic Time-Series Feature Extraction) and is based on some pre-defined feature estimation algorithms. It estimates the trends, seasonality, periodicity, and volatility of the time-series data and applies the feature selection method to select the most relevant features for further modelling or analysis. Several studies use the *tsfresh* library for feature engineering. Ouyang *et al.* [424] used the *tsfresh* feature extraction library to detect anomalous power consumption by users. They extracted 794 features that were used as input to the supervised binary classification (to detect abnormalities). However, the authors did not perform any feature selection, which makes their approach computationally expensive and not feasible to be implemented in real-time. Zhang *et al.* [425] proposed unsupervised anomaly detection using DBSCAN and feature engineering. The authors used the *tsfresh* library in their feature engineering process. They performed features selected based on Maximum-Relevance Minimum-redundancy and variance technique along with Maximal Information Coefficient. However, the proposed approach works best with historical data (as the calculation of relevance, redundancy and information coefficient is performed on complete data) and cannot be applied to real-world or streaming data. In the field of healthcare, Liu *et al.* [426] recommended a solution for the classification of flawed sensors using *tsfresh* feature extraction and selection. The algorithm automatically calculated and selected

features using univariate hypothesis tests with a controlled false discovery rate [427]. The selected features were fed to Long-Short-Term Memory (LSTM) model for classification. However, the extracted features were still over hundreds of features.

Thus, further research is required to obtain a well-generalizable feature engineering algorithm that can calculate and provide well-distinguishable features from large time-series data for an accurate and efficient classification (monitoring) system.

7.2.1.1 Motivation and contribution

For extremely large data, the current automated feature estimation algorithms are not able to capture sufficient valuable information about the feature dynamics [428]. The research aims of this study are to implement and explore the efficacy of heart rate and respiratory rate signal-based (time-series) features extraction algorithm for accurate stress classification, using a stress-predict dataset [429]. The study also determines the best (well-distinguishable) time-series feature from respiratory and heart rate signals for accurate stress monitoring. The algorithm calculates several time-series features using the *tsfresh* library and then performs feature reduction using principal component analysis (PCA) and correlation co-efficient analysis (Pearson, Spearman, and Kendall) to shortlist the most discriminative features.

For the validation of the proposed method, a combination of different extracted features is fed into supervised linear, ensemble and clustering classifiers. The proposed method of time-series features estimation/extraction and feature selection, due to its fast computation and selection of well-distinguishable stress features, can potentially be deployed on photoplethysmography (PPG) sensor-based watches and can detect the anomalies (stress) in real-time.

The rest of the chapter is organised as Section 7.3 discusses stress-predict dataset and methods implemented for time-series features extraction and selection; Section 7.4 reports the detailed analysis and results of supervised machine learning classification and provides discussion around the results; Section 7.5 concludes the chapter and provides future direction towards the development of the reliable stress-monitoring device.

7.3 Material and Methods

7.3.1 Stress-Predict dataset

The stress-predict dataset consists of BVP signal, Inter-beat-intervals, heart rate, respiratory rate, and accelerometer data collected from 35 healthy volunteers who performed three stress-inducing tasks (i.e., Stroop colour word test, interview session, and hyperventilation period) with baseline/normal period. Empatica E4 was used to collect all the information while the overall study lasted for 60 minutes per participant [429]. A brief introduction of the dataset is presented as follows:

The study was designed to be a cross-sectional study that collected data on exposure and outcome in a short time window, in a controlled laboratory setting. The study aimed to understand the behaviour, attitudes, and prevalence to estimate health needs.

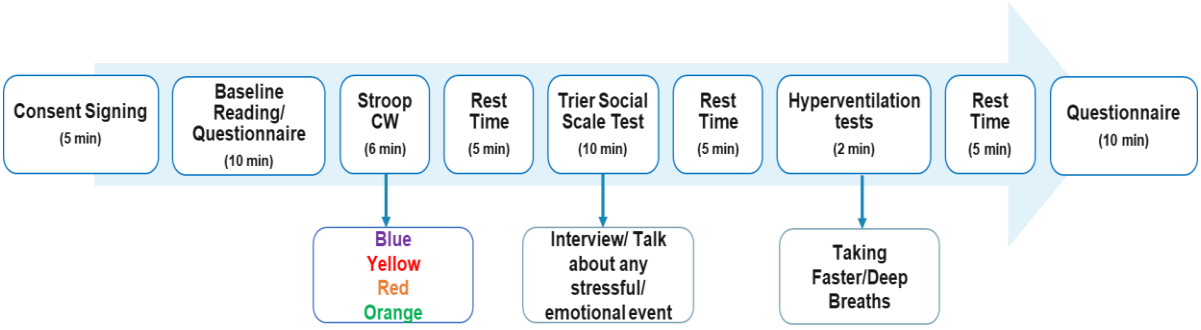


Figure 7.1 Study Protocol of the stress monitoring study including 3 stress-inducing tasks/sessions, 2 self-reporting questionnaires sessions and in-between rest sessions.

7.3.1.1 Study methodology and protocol

The study took 60 minutes per participant and was completely non-invasive. The protocol followed is illustrated in Figure 7.1. The protocol and clinical study were approved by Clinical Research Ethics Committee, Merlin Park Hospital, Galway, Ireland as*: "Stress levels monitoring using sensor-derived signals from non-invasive wearable device and dataset development (Ref: C.A. 2731)"*. For sample size calculation, authors followed sensitivity analysis outcomes reported in [283].

7.3.1.2 Data acquisition

The study used an Empatica E4 wrist-worn watch to measure the physiological changes based on the PPG signal. Labelling of the data was done using tags generated by pressing the button on the watch at the start and end of each task.

Table 7.2 summarizes the number of data entries per participant for each signal recorded. The total recording time reported is approximately around 50 minutes. The heart rate and respiratory rate signals are generated using 10 seconds windows (averaged over 10 seconds) with a window step (slide) of 1 second. Figure 7.2 shows the number of participants who reported higher stress levels during the study based on respiratory and heart rate signal gradients.

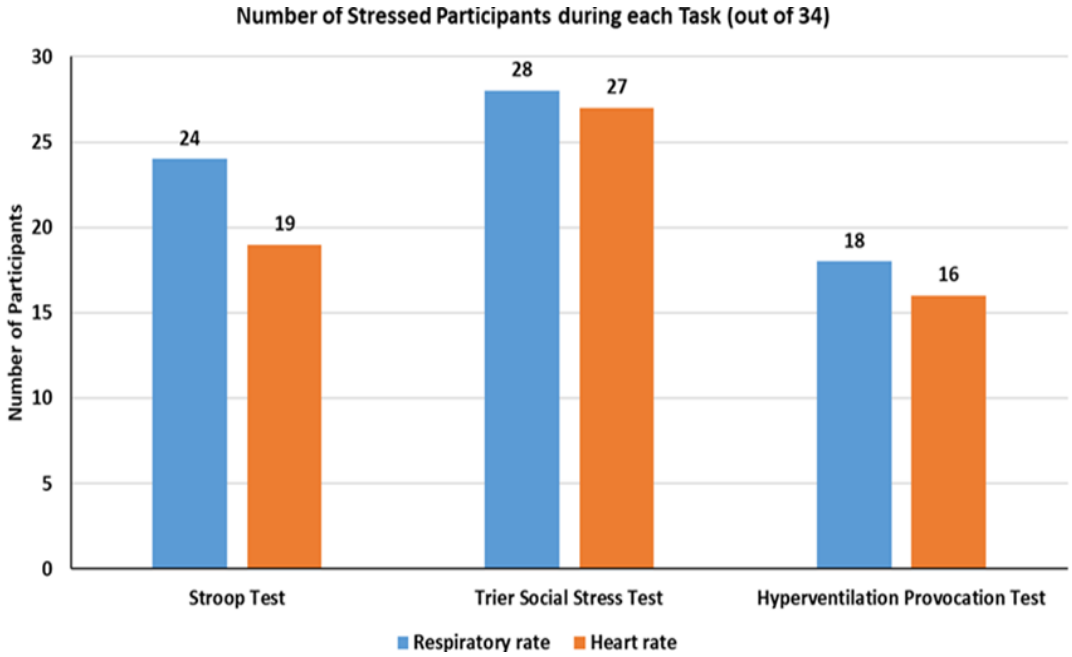


Figure 7.2 Participants with increased stress levels (a) Asked during Interview (b) Based on Questionnaire score.

Table 7.1 The Average Number of Entries (Per Participant)

Time Series Signals	Data Points	Recording Time
Blood Volume Pulse (BVP)	212234	~ 50 minutes
Heart Rate (beats per min)	3308	
Respiratory Rate (breaths per minute)	3308	

7.3.2 Feature extraction and selection

Recently, efforts are made to automate the time-series feature extraction methods and calculate hundreds of different features [418], [430]. However, these high dimensional features lead to challenges when calculating, predicting, storing, and even understanding the correlation of data with the target/outcome [431]. A common technique used for time-series feature extraction is windowing (data is divided into smaller windows and features

are extracted for each window) [10]. Feature such as mean can reduce the signal noise but averages the overall signal. On the contrary, some features such as maximum can be unduly affected by noise [430]. As automated feature extraction has limitations and motivates the need for a systematic process of feature selection along with feature extraction [416]; thus, this study implements a three-stage feature extraction and selection algorithm. Each stage is explained in the following subsection, illustrated in Figure 7.3.

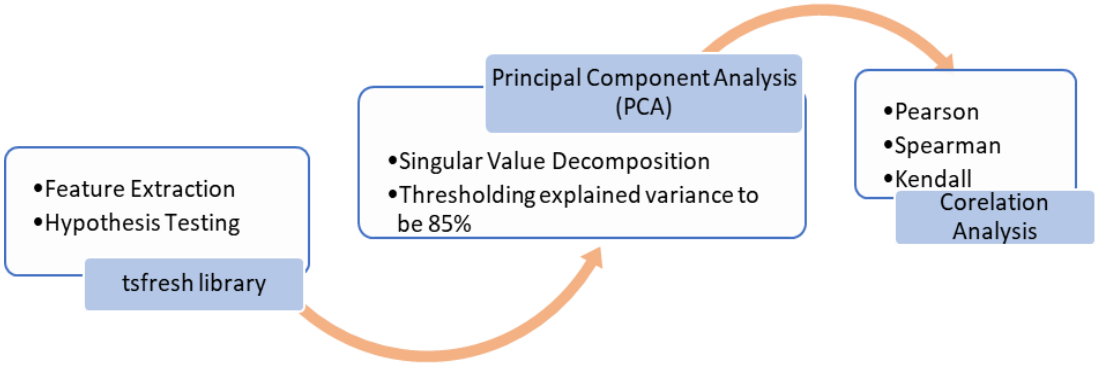


Figure 7.3 Stages of feature extraction and feature selection. tsfresh library calculates and shortlist the hundreds of time-series features, PCA is applied to reduce the feature dimension, to select well-distinguishable features correlation coefficients are calculated using the three methods.

7.3.2.1 tsfresh library

Features were extracted using a Python library *tsfresh* [432]. The library is composed of a combination of 63 time-series characterization techniques. The library calculates 794 features (based on estimating trends, seasonality and periodicity of the data) by default and shortlists the features based on automatically configured hypothesis tests [416]. The library uses standard APIs of time-series (pandas) and machine learning (scikit-learn) packages and provides exploratory analyses. A list of the calculated features and their respective runtime is documented in [433].

7.3.2.2 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a statistical method that is generally used to reduce the dimensionality of high-dimensional data [115]. PCA projects the multidimensional data into a new reduced linear coordinate system using Singular Value Decomposition (SVD). The coordinates of this system represent the largest aggregate of variance within a time-

series dataset and are useful for better visualization and interpretation of complex multivariable data. The dimensionality reduction so helps to observe trends, clusters, and outliers within data [434]. In the proposed study, principal components were selected based on the explained variance of the features. All features that had explained the proportion of variance exceeding 85% were selected for further analysis.

7.3.2.3 Correlation analysis

The *tsfresh* and PCA eliminate calculated time-series features based on the hypothesis testing (feature vs target significance) and explain the variance of the features. For a classification problem, it is vital to remove the highly correlated features as they can introduce bias in the training of the model, makes the model computationally expensive (as the model learns the same information after skimming several different correlated features), reduces the precision of coefficient estimation and effects the interpretability [435], [436]. Thus, this study makes use of correlation analysis to ensure the selection of weakly correlated, well-distinguishable features for an accurate, precise, and easily interpretable classification model. The three commonly used methods for correlation analysis are described below:

7.3.2.3.1 Pearson correlation coefficient

The Pearson correlation coefficient [437], [438] is a statistical test that measures the ratio between the covariance of two features and their standard deviations. The coefficients show the magnitude of correlation/association and the direction (positive or negative) of the relationship. The value of the Pearson coefficient varies between -1 and +1 and is calculated as:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 (y_i - \bar{y})^2}} \qquad (7.1)$$

In eq. 7.1, r is the Pearson correlation coefficient, x_i and y_i is the i^{th} value of two features while $\bar{x}$ and $\bar{y}$ is the mean value of the x and y features.

7.3.2.3.2 Spearman ranking correlation coefficient

The Spearman ranking correlation [439], [440] is a statistical test that measures how closely two features fluctuate. The features are ranked based on their similarity and dissimilarity and then correlation/association between the features is calculated using the following equation (eq. 7.2):

$$\rho = 1 - \frac{6 \sum {d_i}^2}{n(n^2-1)} \qquad (7.2)$$

Where ρ is the Spearman coefficient, d_i is the difference between each feature rank while n is the total number of observations. The value of the Spearman coefficient also varies between -1 to +1.

7.3.2.3.3 Kendall ranking correlation coefficient

The Kendall ranking correlation [441], [442] analysis is a statistical method of measuring the rank association of the two measured features. The correlation is determined based on the concordance and discordance values and is determined as:

$$\tau = \frac{2\,(n_c - n_d)}{n(n-1)} \qquad (7.3)$$

In eq. (7.3), τ is the Kendall coefficient, n_c is the number of concordant values (i.e., $x_2 - x_1$ and $y_2 - y_1$ has the same sign), n_d is the number of discordant values (i.e., $x_2 - x_1$ and $y_2 - y_1$ has an opposite sign) while n is the total number of observations and x,y are the two features. The correlation coefficient calculated using the Kendall method could vary between -1 and +1.

Table 7.3 provides a guideline [443] on the interpretation of the correlation coefficient and the association of features with each other. In this study, all the features with a correlation coefficient between -0.4 to 0.4 were chosen for further analysis.

Table 7.2 Correlation Coefficients and Their Interpretation

Correlation Coefficient Value	Association
+1.0	Perfect positive
+0.8 to +1.0	Very strong positive
+0.6 to +0.8	Strong positive
+0.4 to +0.6	Moderate positive
+0.2 to +0.4	Weak positive
0.0 to +0.2	Very weak positive
0.0 to -0.2	Very weak negative
-0.2 to -0.4	Weak negative
-0.4 to -0.6	Moderate negative
-0.6 to -0.8	Strong negative
-0.8 to -1.0	Very strong negative
-1.0	Perfect negative

7.3.2.4 Machine learning classification

Machine learning classifiers are computer-based models that can learn and adapt without any explicit instructions using statistics and algorithms rules. The machine learning classifier must be generalizable (measurement of the trained classifier's ability to classify the unseen data accurately). A generalized model presents the best trade-off between bias and variance and provides the best prediction performance [444]. In this study, three commonly used supervised machine learning classifiers i.e., logistic regression classifier, random forest classifier and k-nearest neighbour classifier are implemented [445], [446]. Each of these classifiers is representative of their classification categories (linear, ensemble, and clustering). The selection of these classifiers is based on their simplicity, efficiency, interpretability, robustness, and regularization when used for binary classification of categorical-natured data, in this case, stress versus non-stress conditions.

7.3.2.5 Data Split for Training and Testing

For the classification analysis, the dataset needs to be divided into test and train sets. Splitting the dataset helps to evaluate the performance of the model on unseen data. The training set will allow the model to fit and adjust the weights of the model while the test set evaluates the model performance on the new dataset, prevents overfitting, and ensures the model's generalizability. The stress-predict dataset is an imbalanced dataset, with more baseline readings than stress readings. As a result, conventional classification methods focus on minimizing the error rate and exhibit a bias towards the minority class. Also, the random split of the imbalanced data might have negligible or no data from the minority class, thus resulting in biased classification results. The solution to the problem is the use of a stratified k-fold classification split. Stratified sampling ensures that splitting is performed randomly and that the same imbalance class distribution is maintained for each subset (fold). Thus, to get an unbiased model performance, stratified 10-fold cross-validation was implemented.

7.3.2.6 Performance validation methods

The performance of the classifier is validated based on accuracy, standard deviation, precision, recall, f1-score, sensitivity, and specificity. These matrices are described in [283], [388]. *Accuracy* is defined as the ability of the classifier to correctly predict the label of the data point within the test dataset. *Precision* is the classifier's ability to predict a data point belonging to a certain class, while *Recall* is the classifier's ability to identify all the data points

within a certain class. The *F1-score* is a combination of precision and recall using a harmonic mean. The *Sensitivity* of a classifier is metric that shows its ability to predict true positives within each class, and *Specificity* is the evaluation metric that measures the ability to predict true negatives with each class. Table 7.4 shows the confusion matrix used to determine true positive and true negative readings.

Table 7.3 Confusion Matrix

		Actual Labels	
		Positive	**Negative**
Predicted Labels	**Positive**	True Positive	False Positive
	Negative	False Negative	True Negative

7.4 Results and discussions

Figure 7.3 demonstrates the steps of feature extraction and selection. For the stress-predict dataset, the *tsfresh* library calculates 1578 trends, seasonality, periodicity, and volatility-based features for heart rate (789) and respiratory rate (789) signals, combined. The hypothesis test (*p*-value) is performed within the library to check the independence between each feature and label (target variable) and selects 314 features out of 1578 features. For further dimensionality reduction, PCA using singular value decomposition (SVD) was performed. For comparison, PCA resulted in 37 features when implemented on a full feature set (1578), while selecting only 19 features with the feature set obtained after Kruskal-Wallis's hypothesis test (314). As the selected features might still have correlated features, a correlation analysis was performed using Pearson, Kendall, and Spearman methods to determine the most specific features of heart rate and respiratory rate signals to accurately distinguish stress conditions.

7.4.1.1 Correlation analysis

Table 7.5 summarises the number of calculated and selected features at each stage.

Table 7.4 Calculated Features and Correlation Analysis Results

		Correlation Analysis		
Features	**Total**	**Pearson**	**Kendall**	**Spearman**
Full	1578	148	450	201
Filtered (p-test)	314	3	2	1
PCA on Full	37	3	2	1
PCA on filtered	19	3	2	1

Correlation analysis shortlisted similar features even though they were provided with a different number of features (Filtered, PCA on full and PCA on filtered features). The selected features are tabulated and described in table 7.6, detailed in [447].

Table 7.5 Calculated Features and Description

Selected features	Description	Correlation method
Number of peaks in baseline versus the number of peaks in stress periods (In respiratory signal)	This feature calculates the number of peaks that is greater than its *n* neighbours (left and right)	Pearson, Kendall, Spearman
Changes in the variance with higher and lower quantile ranges (In respiratory signal)	This feature fixes a corridor given by lower and higher quantiles and then calculates the variance of the absolute change of the time series inside that corridor.	Pearson
Lag in partial correlation (In Heart rate signal)	This feature calculates the value of partial autocorrelation at the given lag	Pearson
Coefficient of the imaginary part after Fast Fourier Transform (FFT) (In respiratory signal)	This feature calculates the value of the imaginary part of the Fourier coefficient.	Kendall

It can be noted that all three (Pearson, Kendall, and Spearman) correlation analysis methods resulted in selecting the number of peaks within the time series specifically of respiratory rate as the most well-distinguishable feature for accurate stress monitoring. This finding is perfectly correlated with our previously published literature [283], [292], [393], [429] and is true, as the breathing pattern is supposed to vary significantly during stress conditions when compared to baseline/normal conditions.

As most of the shortlisted features belong to the respiratory rate signal, this study also performed a univariable time-series correlation analysis on the heart rate signal (feeding only the heart rate signal along with labels to the algorithm) to determine the most specific heart rate-related features. The Pearson, Kendall, and Spearman correlation analysis method determined that the '*number_cwt_peaks_n_5'* feature (number of peaks that are at enough width (time) scale (here 5) and have high signal-to-noise ratio) is the most specific feature of heart rate to distinguish stress from the baseline readings.

7.4.1.2 Machine learning classifications

For classification analysis, the commonly used statistical features (mean, standard deviation, median, median absolute deviation) and the selected features after correlation analysis were used to train supervised machine learning classifiers. For supervised learning, logistic

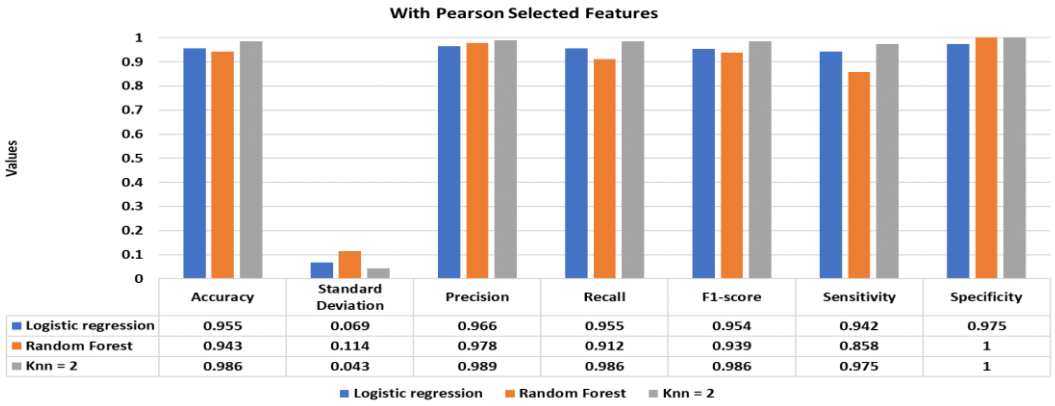
regression, random forest, and K-nearest neighbours (KNN) were selected from linear, ensemble and clustering models, respectively. The results of each classification analysis are reported as follows:

7.4.1.3 Standard Statistical Features

Using standard statistical features, the highest classification performance was achieved using the logistic regression model with an accuracy of 67.4%. Figure 7.4 illustrates the accuracy, standard deviation, precision, recall, f1-score, specificity, and sensitivity of the reported classifiers.

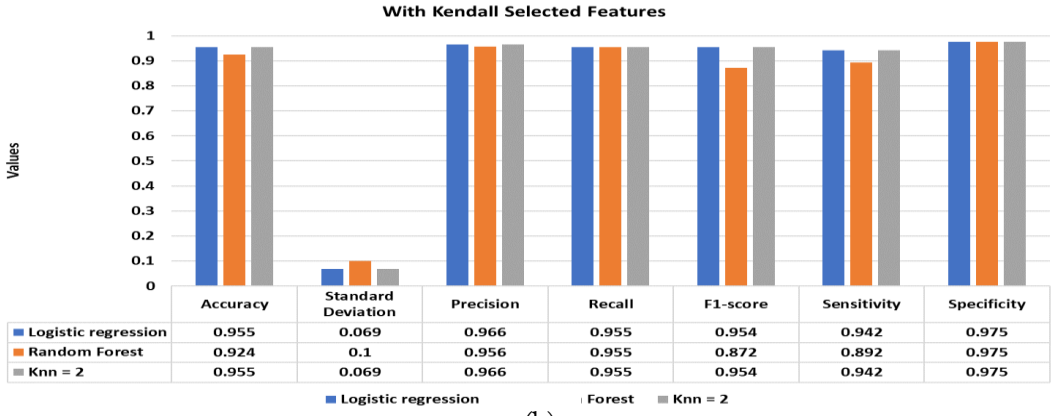
	Accuracy	Standard Deviation	Precision	Recall	F1-score	Sensitivity	Specificity
Logistic regression	0.674	0.15	0.697	0.674	0.661	0.658	0.683
Random Forest	0.664	0.14	0.524	0.619	0.659	0.558	0.717
Knn = 2	0.574	0.17	0.582	0.574	0.526	0.308	0.817

Figure 7.4 Standard statistical features-based stress versus baseline classification using logistic regression, random forest, and K-nearest neighbours classifiers.



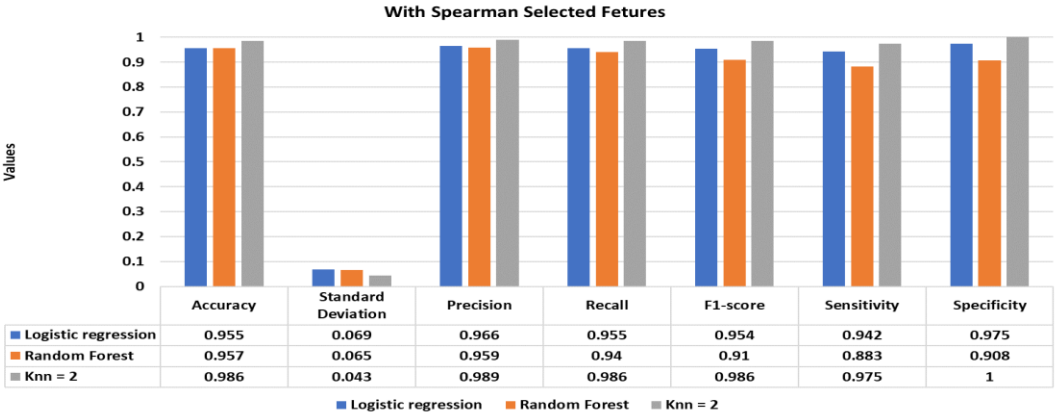
	Accuracy	Standard Deviation	Precision	Recall	F1-score	Sensitivity	Specificity
Logistic regression	0.955	0.069	0.966	0.955	0.954	0.942	0.975
Random Forest	0.943	0.114	0.978	0.912	0.939	0.858	1
Knn = 2	0.986	0.043	0.989	0.986	0.986	0.975	1

(a)



	Accuracy	Standard Deviation	Precision	Recall	F1-score	Sensitivity	Specificity
Logistic regression	0.955	0.069	0.966	0.955	0.954	0.942	0.975
Random Forest	0.924	0.1	0.956	0.955	0.872	0.892	0.975
Knn = 2	0.955	0.069	0.966	0.955	0.954	0.942	0.975

(b)



	Accuracy	Standard Deviation	Precision	Recall	F1-score	Sensitivity	Specificity
Logistic regression	0.955	0.069	0.966	0.955	0.954	0.942	0.975
Random Forest	0.957	0.065	0.959	0.94	0.91	0.883	0.908
Knn = 2	0.986	0.043	0.989	0.986	0.986	0.975	1

(c)

Figure 7.5 Shortlisted features-based stress versus baseline classification using logistic regression, random forest, and K-nearest neighbours classifiers. (a) Using Pearson shortlisted features (b) using Kendall shortlisted features and (c) using Spearman shortlisted features.

7.4.1.4 Selected features after correlation analysis

Figure 7.5 illustrates the classification performance of the supervised classifiers using Pearson (figure 7.5(a)), Kendall (figure 7.5(b)) and Spearman (figure 7.5(c)) selected features. The inclusion of the selected features with the standard statistical features improves the classification performance significantly. The best classification performance is achieved using Pearson and Spearman-based features with a classification accuracy of 98.6% using the KNN classifier. Moreover, the other performance matrices such as standard deviation, precision, recall, f1-score, sensitivity, and specificity of the models have also improved drastically achieving values well above 95%.

7.4.1.5 Summary

Automated feature extraction and selection help in the development of a highly accurate classification model that could be generalizable to new, unseen time-series data. Time-series feature engineering is a substantial component of machine learning classification analytics. The irrelevant features within the training dataset overfit the model to a specific dataset and are not well generalizable. Thus, systematic time-series feature engineering allows automation of the overall classification process and easy-off the difficulties faced during manual feature estimation and selection. In the context of stress classification, feature engineering plays a vital role in improving classification performance. The careful selection and estimation of the time-series features do help in achieving higher classification accuracy with better interpretability of the classifier's decision and achieved results. The dimensionality reduction also helps the predictive model to be computationally efficient especially if required to run on resource-constraint devices.

7.5 Conclusion

In this study, three-fold feature extraction and selection steps are proposed. In the first step, the *tsfresh* library is used to calculate 1578 time-series features of heart rate and respiratory rate (789 features each) signals, which are then shortlisted to 314 features after the hypothesis test. In the second stage, PCA is applied to further reduce the feature dimensions from 314 to only 19 feature components. To detect and eliminate the most correlated features with the estimated feature list, a correlation analysis (with a threshold coefficient value of ± 0.4) is performed using three different methods. The Pearson, Kendall and Spearman correlation analysis determined the count of peaks within the respiratory rate

reading to be the best and well-distinguishable feature among all other heart rate and respiratory rate-related features. For the univariate (heart rate signal) analysis, the number of CWT peaks was the most specific feature to distinguish the stress state from the baseline state.

Furthermore, this study also trained and validated different supervised machine-learning classification models using the K-fold cross-validation technique. The performance of the classification models has been measured in terms of classification accuracy, standard deviation (of the model's accuracy), precision, recall, f1-score, sensitivity, and specificity. The general statistical features (mean, standard deviation, median, mean absolute deviation) that are frequently used in the literature give only an accuracy of 67.4%. The proposed correlation-based time-series feature selection algorithm has resulted in more accurate classification performance compared to conventional statistical features. The time-series correlation analysed feature set when used in conjunction with the statistical features improves the performance of the classifiers significantly and resulted in high-stress classification accuracies, the highest being 98.6% using the KNN classifier.

Future work includes the translation of the proposed algorithm as an online feature learning system for real-time scenarios. The objective will be to update the selected features based on the updated data received. This would eventually lead to a more robust and accurate stress detection system. Additionally, there is a need for dynamic thresholding for PCA as different time-series features (such as heart rate, respiratory rate, skin conductance, muscle activation, and skin temperature) have different PCA subspaces. Thus, requires the estimation of best-suited thresholding levels when applying PCA. Furthermore, a comparison of other supervised and unsupervised machine learning classification models is also the prospect of future work.

Chapter 8

Conclusion and Future Work

This chapter summarises the conclusions and findings of the thesis. Section 8.1 summarizes the motivation and findings while the future work to extend and improve the findings of this thesis is discussed in Section 8.2.

8.1 Conclusion

Stress arises from events that threaten the homeostatic stability of a person [13]. According to the World Health Organisation, stress is directly related to several social and mental problems and these problems seriously affect the health of not only adults but also youngsters [42]. Stress can be acute (state immediate response to the stressor) or chronic (caused by a constant stimulus) [44]. The reasons behind the extensive stress are complex personal, social, and diversity in expressing stress, ecological environment, as well as multiple transactions of humans due to his/her environments. Even though stress is a routine trait of life, nowadays, if it becomes continuous and increasing, an individual might show problematic symptoms which threaten their health as well as people in their surroundings.

Prolonged exposure to stress triggers the adrenocorticotrophic hormonal (ACTH) system and causes the release of cortisol hormones from the adrenal cortex. The human biological system is very complex, and stress evokes different physiological and cognitive reactions in the human body [448]. Therefore, stress markers established until now do not provide any reliable assessment of the quantitative stress response. To the best of our knowledge, no easily applicable and repeatable method can compare the stress response levels of one person in different situations. Moreover, the stress response of two different persons is also distinct.

In chapter 2 (section 2.1), a literature review was performed to find approximated quantitative measures of a person's homeostatic imbalance. This review aimed to provide an overview of the most promising techniques, currently used for stress detection and the challenges associated with them. The review also provides a feasibility report about

measuring different physiological and biochemical levels, a correlation of perceived stress with different indicators, and methods/devices used in the laboratory as well as in the ambulatory environment for stress detection. This work met the first objective of the thesis i.e., review and determine the biophysiological and biochemical indicators of stress.

Chapter 2 (section 2.2) provides a literature review of different machine learning algorithms and their reported classification accuracies. It can also be observed that there is a significant difference in the prediction accuracies in stress classification using a similar set of parameters/indicators along with the same models. The possible reasons for this variation in prediction accuracies can be due to the variations in the experimental setup (real and controlled environment), use of different features extracted from the raw data (time- and frequency-domain features), different lengths of data recordings, different placement of sensors (chest worn, wrist-worn, and foot-worn), the different number of subjects recruited for the experiment, and different stress-perceived questionnaire for labelling.

Theses psychometric questionnaires used for labelling the stress period/instances are not designed to be used in conventional applications. Moreover, these questionnaires are subject to individual variability and depend upon the person's perception of their condition. Among the research community in the medical field, there is no agreement on the reference criteria for monitoring stress levels and measurement techniques. This lack of standard criteria for stress evaluation is due to the variability in stimuli of stress, to which each human reacts differently. Additionally, the available literature aimed to address one or few stress responses in an individual study rather than systematically describing the physiological stress response. The literature review completed the second objective of the thesis i.e., Review the different machine learning algorithms used as predictive models for stress detection and identify the shortcomings resulting in different predictive accuracies.

Chapter 2 (section 2.3) suggested that the use of biochemical indicators may result in better and more promising results in the detection of stress, but one of the biggest shortcomings of biochemical markers of stress levels is their correlation with the intensity of perceived stress. The reason is that this relationship between biochemical hormones and stress is both complex and understudied. The detection of cortisol level was shortlisted as a single biochemical indicator that can provide better detection of stress among all other biochemical indicators. This work met the third objective of the thesis i.e., Investigating the

different detection methods of shortlisting the biochemical indicators to get more sensitive and specific indicators of stress and are less affected by other factors.

In summary, the literature review helped in enlisting different physiological and biochemical indicators of stress along with their sensory technologies. The review also enlisted different machine learning algorithms used as predictive models for stress monitoring. From the literature surveys, it can be concluded that a system with a combination of physiological and biochemical stress biomarkers detection can be a more reliable solution for stress monitoring.

There are several different physiological and biochemical indicators of stress, such as heart rate, respiratory rate, skin temperature, skin conductance, muscle activation, cortisol levels and many others. The development of a system that could detect all these indicators is unrealistic. Thus, a comparative analysis was performed in chapter 3 to determine the relative sensitivity and specificity of these stress-related biophysiological and biochemical indicators of stress in the literature.

The analysis has been performed by applying a t-test and deviance analysis to validate the hypothesis that the physiological data for each variable for the stress and non-stress (baseline) states is statistically differentiable, and logistic regression was applied to identify the strongest predictor of stress. The results of two types of statistical analysis and classification model suggest that respiratory rate is the strongest (stand-alone) predictor of stress compared to other commonly used physiological variables while Heart rate (RRI) emerged as the second-best predictor of stress. The prediction model, consisting of the combination of respiratory rate, heart rate and heart rate variation, derived from a single sensor (PPG), gives accurate classification results as a combination of heart rate, respiratory rate, RR interval, heart rate variability, skin conductance and muscle activation. The latter is a more complex sensory system, prone to motion artefacts. This work met the fourth objective of investigating and shortlisting the biophysiological indicators as well as predictive algorithms to get more sensitive and specific indicators of stress (that are less affected by other factors).

While the work of the previous chapter identified the most effective physiological signals (HR, PPG), there is still a challenge of the accurate and robust measurement devices to

record these signals and in particular labelling the recorded data. Commonly, different types of self-reporting questionnaires are used to label the perceived stress instances. These questionnaires only capture stress levels at a specific point in time. Moreover, self-reporting is subjective and prone to inaccuracies. Also, the existing literature lacks the evidence to suggest whether unsupervised learning classification methods are either feasible or not for stress monitoring devices, specifically regarding stress management in an ambulatory environment. A study considering the above-mentioned gap was designed (presented in chapter 4) in such a way that it not only explores unsupervised machine learning classification methods for stress detection but also provides evidence that with some improvement the unsupervised learning methods have great potential to replace the supervised learning classification methods for the development of non-invasive, continuous, and robust detection and monitoring of physiological or pathological stress.

This study provides a fair comparison of unsupervised and supervised learning algorithms to bring the attention/focus of the researcher community towards unsupervised learning methods for stress monitoring. This study explores the potential feasibility of unsupervised learning clustering classifiers such as (1) Affinity Propagation, (2) Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), (3) K-mean, (4) Mini-Batch K-mean, (5) Mean Shift, (6) Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and (7) Ordering Points to Identify the Clustering Structure (OPTICS) for implementation in stress monitoring wearable devices. The classification results of unsupervised machine learning classifiers are found comparable to supervised machine learning classifiers on two publicly available datasets, i.e., the SWELL-KW dataset and Stress recognition in the automobile driver's dataset. The findings enhance our understanding of the feasibility of unsupervised learning classifiers in wearable devices. Furthermore, these findings also confirmed that respiratory rate and heart rate are the potential indicators for accurate stress detection and back the results found in chapter 3. This work was also performed under the fourth objective of the thesis i.e., Investigate and shortlist the biophysiological indicators as well as predictive algorithms to get more sensitive and specific indicators of stress (that are less affected by other factors).

The respiratory rate and heart rate were found to be the most effective stress-related signals in Chapter 3. Both respiratory rate and heart rate can easily be extracted using a single

wearable sensor i.e., a PPG sensor. The easiest way of recording a PPG signal is through watches. Watches are easy to wear, at feasible locations (at the wrist), have longer battery life and recorded data can easily be accessed for analysis. Most smartwatches use a photoplethysmography signal to extract only the heart rate, as it is the count of the number of peaks in the signal (per minute) [352], [353]. While few algorithms have been proposed in the literature to extract respiratory rates from PPG signals, each algorithm has certain limitations. Such as:

- Estimation of respiratory rate from a PPG signal achieved using digital signal processing (DSP) techniques is highly dependent on the cut-off frequency of the filter.
- The analytical methods are very sensitive to noise and result in very poor respiratory rate detection in presence of motion artefacts.
- Time-frequency analysis-based methods such as Wavelet transform addresses most of the common problems of filtering and analytical methods. It is less sensitive to noise and motion artefacts but requires the selection of more than one parameter, such as the mother wavelet function and the total number of decomposition levels, which in practice are unknown [354], [355].

Other key challenges in respiratory rate estimation are coping with respiratory-induced amplitude variability caused due changes in intra-thoracic pressure during inhale and exhale cycle, and estimation of optimal window size for segmentation of the signal as a shorter window size provides better real-time performance and high resolution while larger window size provides better accuracy [357]. Thus, in Chapter 5, a novel PPG-based respiratory rate estimation algorithm is developed considering all the major limitations including noise, motion artefacts, poor signal quality, the effect of window size, and the cut-off frequencies of the filters. The algorithm is based on three steps:

- The pre-processing steps perform signal extraction, filtration, and peak enhancement for increasing the signal-to-noise ratio.
- The signal analysis stage, peak detection, peak-to-peak interval, error in peak detection and correction, calculation of different time-series measurements and estimation of respiratory rate are performed.

- As the amplitude of the PPG signal is affected by respiration, thus in the final stage, scaling is performed based on the induced amplitude variation.

The estimated respiratory rate is compared with the reference respiration data provided in the publicly available dataset called BIDMC dataset. The results achieved by the proposed algorithm endorse the integration of the proposed algorithm into a commercially available pulse oximetry device would expand its functionality from the measurement of oxygen saturation level (SpO2) and heart rate to the continuous measurement of the respiratory rate with great efficiency in the clinical setting as well as in the ambulatory home-based environment. This work met the fifth objective of the thesis which was to develop and evaluate an algorithm that can extract the stress-specific information from a physiological signal, should be able to deal with low-quality signal and deals with other co-founding factors to detect and classify stress accurately.

To make the developed respiratory rate estimation algorithm more generalizable, it needed to be tested on another stress-related dataset. The availability of open-source public datasets in the field of biomedical research is always a challenge. This might be because of data privacy and effort put into the manual annotation of the signal by experts which incur high costs and time consumption. Thus, in Chapter 6, a clinical study was performed to develop a stress-predict dataset for the identification and analysis of different patterns of stress-induced while performing specific tasks. Along with dataset development, this study primarily focused on further validation of the developed algorithm (Chapter 5) and a comparative analysis of the features collected in terms of prediction specificity and sensitivity for stress monitoring. The results of statistical analysis (linear mixer model and adaptive reference range) showed that, in the Stress-Predict dataset, the respiratory rate is the best predictor of the stress state.

The stress-predict dataset is an open-access dataset and is made publicly available. The inclusion of an additional feature, i.e., respiratory rate data along with stress and baseline labels within the dataset, makes this dataset more desirable and unique from all the other publicly available Empatica E4-based datasets and will certainly attract the attention and interest of researchers in the all the field of psychological, clinical, and biomedical research, as well as prevention, medicine, and connected health systems.

Furthermore, the developed respiratory rate estimation algorithm was evaluated on the dataset and was successful in estimating accurate respiratory rate (when compared to breathing rate during the hyperventilation period). Moreover, the stress versus baseline state classification was also performed and the highest accuracy of 67.4% was achieved.

While the conventional statistical features (such as mean, variance, standard deviation, and mean absolute deviation) of respiratory and heart rate do help in distinguishing between stress versus baseline readings but the reported accuracy (determined by experimentation in Chapter 3 and Chapter 4) is low i.e., around 75%. The stress classification is highly dependent on the classification model and features fed to the model. For accurate stress monitoring, it is essential that these features are not only informative but also well-distinguishable and interpretable by the classification models. Chapter 7 implemented a time-series feature engineering algorithm that extracts well-distinguishable time-series features from a physiological signal and helped in achieving high-stress detection performance. The algorithm extracts 1578 features using the *tsfresh* library on the stress-predict dataset. These features are then reduced to more specific time-series features using principal component analysis (PCA) along with Pearson, Kendall and Spearman correlation ranking techniques. The comparative analysis of conventional features versus correlation-based features is performed using logistic regression, random forest, and k-nearest neighbour models. The outcome of the proposed study concludes that it is vital to have better analytical features rather than conventional statistical features for accurate stress classification.

The work presented in chapters 6 and 7 was performed to achieve the sixth and final objective of the thesis, which was to investigate and provide a solution of identification and calculation of discriminative features of the stressed condition and should easily be characterized as a stress response from all other physiological stimuli.

This thesis provides feasibility studies about different biophysiological and biochemical indicators of stress and determines the most sensitive and specific indicators of stress that are less likely to be affected by another stimulus. Moreover, a novel PPG-based respiratory rate estimation algorithm was developed to accurately estimate breathing rate even from low-quality, distorted PPG signals. Additionally, a clinical study was also conducted to propose a new dataset, named Stress-Predict Dataset, and further evaluate the developed

estimation algorithm. Finally, different time-series feature engineering parameters were investigated to find the best parameter which improves the overall performance of stress detection. The work presented in this thesis is a stepping-stone toward the clinical evaluation of stress monitoring technology.

8.2 Future Work

The work presented in this thesis can be further extended to the potential of designing a non-invasive, continuous, and robust wearable real-time stress monitoring device for the detection and prediction of biophysiological and psychological stress based on photoplethysmography (PPG) signals. The suggestions for future work in the field of stress monitoring are presented in this section.

The presented work can be extended in several ways to further accelerate the translation of stress-monitoring prototypes from the research to patients. The recommendations related to future work are listed below:

1. The field of detecting biochemical indicators for stress monitoring is relatively novel and immature. Many studies reported a good sensitivity and specificity of cortisol sampling and showed comparable accuracies in cortisol detection when compared to gold-standard or conventional methods. Thus, in future, a wearable biochemical sensor could be developed but there is a need of determining the true relationship between stress and cortisol, through accurate, reliable, and valid analysis. These analyses should be based on robust methodologies and should be well-documented, transparent, and easily understandable by the intended audience. Additionally, the data used for the investigation should be of high quality, which includes proper collection, cleaning, and pre-processing in such a way that their results provide valuable insights about the research question. Furthermore, the effect of sex, ethnicity, and treatments on the association between stress and cortisol levels must also need to be investigated.
2. Currently, in the respiratory rate estimation algorithm, the scaling value to cope with the amplitude variation of the PPG signal is the only hyperparameter that might need to be determined empirically. The default method of scaling does work for most of the PPG data but may require improvement in some cases. In future,

the scaling technique can be improved which will eventually improve the estimation accuracy furthermore. To determine a more generalizable scaling value automatically for accurate estimation of the respiratory rate, the algorithm needs to be evaluated on adequate datasets having the following characteristics:

a. Large and diverse dataset: the dataset should have large enough numbers of data points to represent the sampled population and should be diverse to test the algorithm's performance on different types of inputs.
b. Balanced: the dataset should have balanced distributions to avoid bias in the algorithm's performance.
c. Clean and pre-processed: the dataset should be clean of any errors and should have suitable data format for the algorithm testing.
d. Annotated: The dataset should be well annotated with labels and other relevant information.
e. Representative and publicly available: The dataset should be publicly available so that other researchers could evaluate their algorithms and compare them with existing ones.

The dataset with such characteristics gives more confidence in the results and allows the algorithm to be tested in different scenarios.

3. The classification results indicate that unsupervised machine learning classifiers can achieve comparable performance in terms of classification accuracy, precision, recall and F1-score, without any training phase which is usually time-consuming and inaccurate. The findings enhance our understanding of the feasibility of unsupervised learning classifiers in wearable devices. These unsupervised machine learning algorithms require further investigation and modification to discover the hidden patterns and structures within the data which will help these algorithms to surpass the accuracies provided by supervised learning algorithms. The modification includes the use of clustering, dimensionality reduction, generative models, anomaly detection, self-organizing maps, and/or deep learning.
4. The inclusion of an additional feature, i.e., respiratory rate data along with stress and baseline labels within the dataset, makes the dataset unique from all the other publicly available Empatica E4-based datasets. There were also some limitations to the study.

a. First, at the start and end of the stress task, the heart rate and respiratory rate change gradually, but there is no accurate way to determine this gradual change. Thus, labelling is performed without considering these (actual) delayed changes.
b. Secondly, sometimes there is more than one participant in the room where the study was conducted. The crosstalk and, especially, the questions asked during the interview to induce stress were learned by the other participant while the resting/baseline period. Therefore, the effectiveness of the stress-inducing interview questions was decreased.
c. Third, the interviewees were friendly and kind to the participant. They kept the overall interview environment friendly rather than mimicking a strict interview session which might have resulted in less induced stress. Thus, there was less variation in the readings of stress versus non-stress parameters.

All these factors must have been related to the (relatively low) classification accuracies achieved. In future studies, all these shortcomings should be considered to obtain an improved stress monitoring model. In future, a standard respiratory and heart rate measuring chest-worn band could be used to accurately detect the gradual changes in these parameters at the start and end of each stress-inducing task. Furthermore, data recording could be done in two separate rooms if more than one participant is interviewed. Efforts should also be made to strict the overall environment to get the participants pressured and stressed by either asking rapid questions or by asking them to perform a difficult task such as talking in a non-native language or giving them complex mathematics questions. All these steps will benefit in collecting high-quality stress-induced data which will eventually help in the development of an accurate stress monitoring device.

8.2.1 Towards future stress monitoring clinical device

Future stress-monitoring clinical devices will have the following features:

1. Non-invasive and continuous: will be able to monitor biophysiological and biochemical signals (such as respiratory rate, heart rate and cortisol levels) continuously and without any surgery, as illustrated in this thesis.
2. Multimodal sensing: will be able to utilize optical, electrical, and mechanical sensors to measure the stress-specific indicators (determine in this thesis) and provide comprehensive and accurate data.
3. Real-time analysis: will analyse the measured signal (as proposed in the thesis) and provide instant interventions to users on the go, allowing them to react immediately and reduce their stress either by meditation or talking to their loved ones.
4. Personalised: will have built-in unsupervised (investigated in this thesis) or reinforcement learning algorithm in the device, enabling it to learn from the user's unique stress response and make a recommendation when stress is detected.
5. Wearable and portable: this will be a small (wrist-worn) device that can easily be worn for a long time, will have long battery life and is easy to carry in daily routine.
6. Last but not the least, Affordable and accessible: will be affordable and accessible to a large range of users and could be easily used by clinicians as well as non-clinicians.

Some of the features discussed above have already been developed and are available but still further technological advancement and research is needed to have a multimodal sensory wearable device. At present, there is no single best analytical method for the development of ambulatory stress monitoring technology, and thus, this field appears to be relatively immature and engrossing. In the future years, with the help of the proposed algorithms and stress-predict dataset presented in this thesis, I am hopeful to overcome the remaining procedural and technological hurdles and make a true ambulatory bio-physiological and biochemical diagnostic wearable stress monitoring device for a home as well as a clinic-based healthcare system.

References

[1] F. M. Reis, L. M. Coutinho, S. Vannuccini, S. Luisi, and F. Petraglia, "Is stress a cause or a consequence of endometriosis?," *Reprod. Sci.*, vol. 27, pp. 39–45, 2020.

[2] U. Pluntke, S. Gerke, A. Sridhar, J. Weiss, and B. Michel, "Evaluation and classification of physical and psychological stress in firefighters using heart rate variability," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 2207–2212.

[3] L. Richard, T. Hurst, and J. Lee, "Lifetime exposure to abuse, current stressors, and health in federally qualified health center patients," *J. Hum. Behav. Soc. Environ.*, vol. 29, no. 5, pp. 593–607, 2019.

[4] G. S. Everly and J. M. Lating, "The anatomy and physiology of the human stress response," in *A clinical guide to the treatment of the human stress response*, Springer, 2019, pp. 19–56.

[5] E. Hemmingsson, "Early childhood obesity risk factors: socioeconomic adversity, family dysfunction, offspring distress, and junk food self-medication," *Curr. Obes. Rep.*, vol. 7, no. 2, pp. 204–209, 2018.

[6] D. Carneiro, P. Novais, J. C. Augusto, and N. Payne, "New Methods for Stress Assessment and Monitoring at the Workplace," *IEEE Trans. Affect. Comput.*, vol. 14, no. 8, pp. 1–1, 2017, doi: 10.1109/taffc.2017.2699633.

[7] H. S. Executive, "Work-related ill health and occupational disease in Great Britain." [Online]. Available: https://www.hse.gov.uk/statistics/causdis/.

[8] P. Verma and S. K. Sood, "A comprehensive framework for student stress monitoring in fog-cloud IoT environment: m-health perspective," *Med. Biol. Eng. Comput.*, vol. 57, no. 1, pp. 231–244, 2019, doi: 10.1007/s11517-018-1877-1.

[9] K. L. Tamashiro, R. R. Sakai, C. A. Shively, I. N. Karatsoreos, and L. P. Reagan, "Chronic stress, metabolism, and metabolic syndrome," *Stress*, vol. 14, no. 5, pp. 468–474, 2011.

[10] A. Golgouneh and B. Tarvirdizadeh, "Fabrication of a portable device for stress monitoring using wearable sensors and soft computing algorithms," *Neural Comput. Appl.*, pp. 1–23, 2019.

[11] P. Suresh, A. Matthews, and I. Coyne, "Stress and stressors in the clinical environment: a

comparative study of fourth-year student nurses and newly qualified general nurses in Ireland," *J. Clin. Nurs.*, vol. 22, no. 5–6, pp. 770–779, 2013.

[12] J. Aguiló *et al.*, "Project ES3: attempting to quantify and measure the level of stress," *Rev Neurol*, vol. 61, no. 9, pp. 405–415, 2015.

[13] B. S. McEwen, "Stressed or stressed out: what is the difference?," *J. Psychiatry Neurosci.*, vol. 30, no. 5, p. 315, 2005.

[14] E. G. Brown, A.-M. Creaven, and S. Gallagher, "Loneliness and cardiovascular reactivity to acute stress in younger adults," *Int. J. Psychophysiol.*, vol. 135, pp. 121–125, 2019.

[15] M. al'Absi, D. Hatsukami, G. L. Davis, and L. E. Wittmers, "Prospective examination of effects of smoking abstinence on cortisol and withdrawal symptoms as predictors of early smoking relapse," *Drug Alcohol Depend.*, vol. 73, no. 3, pp. 267–278, 2004.

[16] S. Ranabir and K. Reetu, "Stress and hormones," *Indian J. Endocrinol. Metab.*, vol. 15, no. 1, p. 18, 2011.

[17] L. Petrakova, K. Boy, L. Mittmann, L. Möller, H. Engler, and M. Schedlowski, "Salivary alpha-amylase and noradrenaline responses to corticotropin-releasing hormone administration in humans," *Biol. Psychol.*, vol. 127, pp. 34–39, 2017.

[18] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *J. Biomed. Inform.*, vol. 59, pp. 49–75, 2016.

[19] P. Schmidt, R. Duerichen, K. Van Laerhoven, C. Marberger, and A. Reiss, "Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection," pp. 400–408, 2018, doi: 10.1145/3242969.3242985.

[20] M. J. H. Begemann, E. J. R. Florisse, R. Van Lutterveld, M. Kooyman, and I. E. Sommer, "Efficacy of EEG neurofeedback in psychiatry: A comprehensive overview and meta-analysis," *Transl. Brain Rhythm.*, vol. 1, no. 1, pp. 19–29, 2016.

[21] J. Sánchez-Molina, J. J. Robles-Pérez, and V. J. Clemente-Suárez, "Assessment of psychophysiological response and specific fine motor skills in combat units," *J. Med. Syst.*, vol. 42, no. 4, pp. 1–7, 2018.

[22] R. Delgado-Moreno, J. J. Robles-Pérez, and V. J. Clemente-Suárez, "Combat stress decreases memory of warfighters in action," *J. Med. Syst.*, vol. 41, no. 8, pp. 1–7, 2017.

[23] V. J. Clemente-Suárez, J. J. Robles-Pérez, and J. Fernández-Lucas, "Psychophysiological response in parachute jumps, the effect of experience and type of jump," *Physiol. \& Behav.*, vol. 179, pp. 178–183, 2017.

[24] K. Hovsepian, M. Al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "cStress: towards a gold standard for continuous stress assessment in the mobile environment," in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, 2015, pp. 493–504.

[25] J. Rodr\'\iguez-Arce, L. Lara-Flores, O. Portillo-Rodr\'\iguez, and R. Mart\'\inez-Méndez, "Towards an anxiety and stress recognition system for academic environments based on physiological features," *Comput. Methods Programs Biomed.*, vol. 190, p. 105408, 2020.

[26] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Continuous stress detection using a wrist device: in laboratory and real life," in *proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct*, 2016, pp. 1185–1193.

[27] L. Han, Q. Zhang, X. Chen, Q. Zhan, T. Yang, and Z. Zhao, "Detecting work-related stress with a wearable device," *Comput. Ind.*, vol. 90, pp. 42–49, 2017, doi: 10.1016/j.compind.2017.05.004.

[28] M. Chesnut *et al.*, "Stress markers for mental states and biotypes of depression and anxiety: A scoping review and preliminary illustrative analysis," *Chronic Stress*, vol. 5, p. 24705470211000336, 2021.

[29] J. M. Peake, G. Kerr, and J. P. Sullivan, "A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations," *Front. Physiol.*, vol. 9, no. JUN, pp. 1–19, 2018, doi: 10.3389/fphys.2018.00743.

[30] T. Hao, K. N. Walter, M. J. Ball, H.-Y. Chang, S. Sun, and X. Zhu, "StressHacker: towards practical stress monitoring in the wild with Smartwatches," in *AMIA Annual Symposium Proceedings*, 2017, vol. 2017, p. 830.

[31] S. Lee *et al.*, "Mental stress assessment using ultra short term HRV analysis based on non-linear method," *Biosensors*, vol. 12, no. 7, p. 465, 2022.

[32] D. American Psychiatric Association, A. P. Association, and others, *Diagnostic and statistical manual of mental disorders: DSM-5*, vol. 5, no. 5. American psychiatric association Washington, DC, 2013.

[33] K. M. Parker and S. A. Smith, "Aquatic-aerobic exercise as a means of stress reduction during pregnancy," *J. Perinat. Educ.*, vol. 12, no. 1, p. 6, 2003.

[34] M. E. Colbaugh, *Whirl-Winded: Stressed and Coping*. Indiana University of Pennsylvania, 2014.

[35] B. Østerås, H. Sigmundsson, and M. Haga, "Psychometric properties of the perceived stress questionnaire (PSQ) in 15--16 years old Norwegian adolescents," *Front. Psychol.*, vol. 9, p. 1850, 2018.

[36] C. Emmanouil, F. Bacopoulou, D. Vlachakis, G. P. Chrousos, and C. Darviri, "Validation of the Stress in Children (SiC) Questionnaire in a Sample of Greek Pupils," *J. Mol. Biochem.*, vol. 9, no. 1, p. 74, 2020.

[37] K. Plarre *et al.*, "Continuous inference of psychological stress from sensory measurements collected in the natural environment," in *Proceedings of the 10th ACM/IEEE international conference on information processing in sensor networks*, 2011, pp. 97–108.

[38] Y. S. Can, B. Arnrich, and C. Ersoy, "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey," *J. Biomed. Inform.*, vol. 92, p. 103139, 2019.

[39] S. Gradl, M. Wirth, R. Richer, N. Rohleder, and B. M. Eskofier, "An overview of the feasibility of permanent, real-time, unobtrusive stress measurement with current wearables," in *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 2019, pp. 360–365.

[40] M. Kivimäki and A. Steptoe, "Effects of stress on the development and progression of cardiovascular disease," *Nat. Rev. Cardiol.*, vol. 15, no. 4, p. 215, 2018.

[41] A. Mariotti, "The effects of chronic stress on health: new insights into the molecular mechanisms of brain--body communication," *Futur. Sci. OA*, vol. 1, no. 3, 2015.

[42] W. H. Organization and others, "World Health Statistics data visualizations dashboard," *Obtenido World Heal. Stat. data Vis. dashboard http//apps. who. int/gho/data/node. sdg*, pp. 3–4.

[43] A. Golgouneh and B. Tarvirdizadeh, "Fabrication of a portable device for stress monitoring using wearable sensors and soft computing algorithms," *Neural Comput. Appl.*, vol. 32, no. 11, pp. 7515–7537, 2020.

[44] D. H. Hellhammer, A. A. Stone, J. Hellhammer, and J. Broderick, "Measuring stress," *Encycl. Behav. Neurosci.*, vol. 2, pp. 186–191, 2010.

[45] B. Cvetković *et al.*, "Real-time physical activity and mental stress management with a

wristband and a smartphone," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 2017, pp. 225–228.

[46] F. Seoane *et al.*, "Wearable biomedical measurement systems for assessment of mental stress of combatants in real time," *Sensors (Switzerland)*, vol. 14, no. 4, pp. 7120–7141, 2014, doi: 10.3390/s140407120.

[47] C. D. Katsis, N. S. Katertsidis, and D. I. Fotiadis, "An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders," *Biomed. Signal Process. Control*, vol. 6, no. 3, pp. 261–268, 2011.

[48] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *J. Biomed. Inform.*, vol. 73, pp. 159–170, 2017, doi: 10.1016/j.jbi.2017.08.006.

[49] E. L. van den Broek, F. van der Sluis, and T. Dijkstra, "Cross-validation of bimodal health-related stress assessment," *Pers. Ubiquitous Comput.*, vol. 17, no. 2, pp. 215–227, 2013.

[50] S.-H. Seo, J.-T. Lee, and M. Crisan, "Stress and EEG," *Converg. hybrid Inf. Technol.*, vol. 27, 2010.

[51] P. Zontone, A. Affanni, R. Bernardini, A. Piras, and R. Rinaldo, "Stress detection through electrodermal activity (EDA) and electrocardiogram (ECG) analysis in car drivers," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.

[52] A. Affanni, "Wireless sensors system for stress detection by means of ECG and EDA acquisition," *Sensors*, vol. 20, no. 7, p. 2026, 2020.

[53] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Trans. Affect. Comput.*, 2019.

[54] S. Pourmohammadi and A. Maleki, "Stress detection using ECG and EMG signals: A comprehensive study," *Comput. Methods Programs Biomed.*, vol. 193, p. 105482, 2020.

[55] T. Oka, "Psychogenic fever: how psychological stress affects body temperature in the clinical population," *Temperature*, vol. 2, no. 3, pp. 368–378, 2015.

[56] K. A. Herborn *et al.*, "Skin temperature reveals the intensity of acute stress," *Physiol. Behav.*, vol. 152, pp. 225–230, 2015.

[57] F. de Arriba-Pérez, J. M. Santos-Gago, M. Caeiro-Rodr\'\iguez, and M. Ramos-Merino,

"Study of stress detection and proposal of stress-related features using commercial-off-the-shelf wrist wearables," *J. Ambient Intell. Humaniz. Comput.*, vol. 10, no. 12, pp. 4925–4945, 2019.

[58] H. Reims, K. Sevre, E. Fossum, A. Høieggen, I. Eide, and S. Kjeldsen, "Plasma catecholamines, blood pressure responses and perceived stress during mental arithmetic stress in young men," *Blood Press.*, vol. 13, no. 5, pp. 287–294, 2004.

[59] J. B. Drummond *et al.*, "Copeptin response to hypoglycemic stress is linked to prolactin activation in children," *Pituitary*, vol. 23, no. 6, pp. 681–690, 2020.

[60] B. Valentin *et al.*, "Cortisol and alpha-amylase as stress response indicators during pre-hospital emergency medicine training with repetitive high-fidelity simulation and scenarios with standardized patients," *Scand. J. Trauma. Resusc. Emerg. Med.*, vol. 23, no. 1, pp. 1–8, 2015.

[61] L. Thau and S. Sharma, "Physiology, cortisol," *StatPearls [Internet]*, 2020.

[62] N. Hjortskov, A. H. Garde, P. Ørbæk, and Å. M. Hansen, "Evaluation of salivary cortisol as a biomarker of self-reported mental stress in field studies," *Stress Heal. J. Int. Soc. Investig. Stress*, vol. 20, no. 2, pp. 91–98, 2004.

[63] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Med. Biol. Eng. Comput.*, vol. 42, no. 3, pp. 419–427, 2004, doi: 10.1007/BF02344719.

[64] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss, "Activity-aware mental stress detection using physiological sensors," in *International conference on Mobile computing, applications, and services*, 2010, pp. 282–301.

[65] L. lan Chen, Y. Zhao, P. fei Ye, J. Zhang, and J. zhong Zou, "Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers," *Expert Syst. Appl.*, vol. 85, pp. 279–291, 2017, doi: 10.1016/j.eswa.2017.01.040.

[66] J. Kim, J. Park, and J. Park, "Development of a statistical model to classify driving stress levels using galvanic skin responses," *Hum. Factors Ergon. Manuf. Serv. Ind.*, vol. 30, no. 5, pp. 321–328, 2020.

[67] J. Healey and R. Picard, "SmartCar: detecting driver stress," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, 2000, vol. 4, pp. 218–221.

[68] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, 2005.

[69] R. Mahmoud, T. Shanableh, I. P. Bodala, N. V Thakor, and H. Al-Nashash, "Novel classification system for classifying cognitive workload levels under vague visual stimulation," *IEEE Sens. J.*, vol. 17, no. 21, pp. 7019–7028, 2017.

[70] Y. Shi *et al.*, "Personalized stress detection from physiological measurements," *Int. Symp. Qual. Life Technol.*, pp. 28–29, 2010, doi: 10.1.1.207.8062.

[71] J. Wijsman, B. Grundlehner, H. Liu, H. Hermens, and J. Penders, "Towards mental stress detection using wearable physiological sensors," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 1798–1801, 2011, doi: 10.1109/IEMBS.2011.6090512.

[72] V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, and O. M. Mozos, "Stress detection using wearable physiological sensors," in *International work-conference on the interplay between natural and artificial computation*, 2015, pp. 526–532.

[73] M. Choi, G. Koo, M. Seo, and S. W. Kim, "Wearable Device-Based System to Monitor a Driver's Stress, Fatigue, and Drowsiness," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 3, pp. 634–645, 2018, doi: 10.1109/TIM.2017.2779329.

[74] C. H. Vinkers *et al.*, "The effect of stress on core and peripheral body temperature in humans," *Stress*, vol. 16, no. 5, pp. 520–530, 2013.

[75] M. Uesato *et al.*, "Salivary amylase activity is useful for assessing perioperative stress in response to pain in patients undergoing endoscopic submucosal dissection of gastric tumors under deep sedation," *Gastric Cancer*, vol. 13, no. 2, pp. 84–89, 2010.

[76] E. Ullmann, A. Barthel, K. Petrowski, T. Stalder, C. Kirschbaum, and S. R. Bornstein, "Pilot study of adrenal steroid hormones in hair as an indicator of chronic mental and physical stress," *Sci. Rep.*, vol. 6, no. October 2015, pp. 1–7, 2016, doi: 10.1038/srep25842.

[77] A. Arza *et al.*, "Measuring acute stress response through physiological signals: towards a quantitative assessment of stress," *Med. Biol. Eng. Comput.*, vol. 57, no. 1, pp. 271–287, 2019, doi: 10.1007/s11517-018-1879-z.

[78] E. Cashdan, "Hormones and Competitive Aggression in Women," *Aggress. Behav.*, vol. 29, no. 2, pp. 107–115, 2003, doi: 10.1002/ab.10041.

[79] N. Rohleder, U. M. Nater, J. M. Wolf, U. Ehlert, and C. Kirschbaum, "Psychosocial stress-

induced activation of salivary alpha-amylase: An indicator of sympathetic activity?," *Ann. N. Y. Acad. Sci.*, vol. 1032, pp. 258–263, 2004, doi: 10.1196/annals.1314.033.

[80] A. Tasaka, Y. Tahara, T. Sugiyama, and K. Sakurai, "Influence of chewing rate on salivary stress hormone levels.," *Nihon Hotetsu Shika Gakkai Zasshi*, vol. 52, no. 4, pp. 482–7, 2008, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19037143.

[81] A. Goyal, S. Singh, D. Vir, and D. Pershad, "Automation of stress recognition using subjective or objective measures," *Psychol. Stud. (Mysore).*, vol. 61, no. 4, pp. 348–364, 2016.

[82] A. P. Allen, P. J. Kennedy, J. F. Cryan, T. G. Dinan, and G. Clarke, "Biological and psychological markers of stress in humans: Focus on the Trier Social Stress Test," *Neurosci. Biobehav. Rev.*, vol. 38, pp. 94–124, 2014.

[83] Y. S. Can, N. Chalabianloo, D. Ekiz, and C. Ersoy, "Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study," *Sensors*, vol. 19, no. 8, p. 1849, 2019, doi: 10.3390/s19081849.

[84] O. V. Bitkina, J. Kim, J. Park, J. Park, and H. K. Kim, "Identifying traffic context using driving stress: A longitudinal preliminary case study," *Sensors (Switzerland)*, vol. 19, no. 9, pp. 1–16, 2019, doi: 10.3390/s19092152.

[85] T. Li, Y. Chen, and W. Chen, "Daily Stress Monitoring Using Heart Rate Variability of Bathtub ECG Signals," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2018-July, pp. 2699–2702, 2018, doi: 10.1109/EMBC.2018.8512767.

[86] Y. Liu and S. Du, "Psychological stress level detection based on electrodermal activity," *Behav. Brain Res.*, vol. 341, no. November 2017, pp. 50–53, 2018, doi: 10.1016/j.bbr.2017.12.021.

[87] D. S. Lee, T. W. Chong, and B. G. Lee, "Stress Events Detection of Driver by Wearable Glove System," *IEEE Sens. J.*, vol. 17, no. 1, pp. 194–204, 2017, doi: 10.1109/JSEN.2016.2625323.

[88] O. M. Mozos *et al.*, "Stress Detection Using Wearable Physiological and Sociometric Sensors," *Int. J. Neural Syst.*, vol. 27, no. 02, p. 1650041, 2017, doi: 10.1142/s0129065716500416.

[89] I. Mohino-Herranz, R. Gil-Pita, J. Ferreira, M. Rosa-Zurera, and F. Seoane, "Assessment of mental, emotional and physical stress through analysis of physiological signals using

smartphones," *Sensors (Switzerland)*, vol. 15, no. 10, pp. 25607–25627, 2015, doi: 10.3390/s151025607.

[90] A. Muaremi, A. Bexheti, F. Gravenhorst, B. Arnrich, and G. Troster, "Monitoring the impact of stress on the sleep patterns of pilgrims using wearable sensors," *2014 IEEE-EMBS Int. Conf. Biomed. Heal. Informatics, BHI 2014*, pp. 185–188, 2014, doi: 10.1109/BHI.2014.6864335.

[91] P. KARTHIKEYAN, M. MURUGAPPAN, and S. YAACOB, "Detection of Human Stress Using Short-Term Ecg and Hrv Signals," *J. Mech. Med. Biol.*, vol. 13, no. 02, p. 1350038, 2013, doi: 10.1142/s0219519413500383.

[92] S. A. Hosseini and M. A. Khalilzadeh, "Emotional stress recognition system using EEG and psychophysiological signals: Using new labelling process of EEG signals in emotional stress state," *2010 Int. Conf. Biomed. Eng. Comput. Sci. ICBECS 2010*, pp. 1–6, 2010, doi: 10.1109/ICBECS.2010.5462520.

[93] J. Choi and R. Gutierrez-Osuna, "Using heart rate monitors to detect mental stress," in *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*, 2009, pp. 219–223.

[94] L. Bönke *et al.*, "Examining the effect of Early Life Stress on autonomic and endocrine indicators of individual stress reactivity," *Neurobiol. Stress*, vol. 10, no. November 2018, p. 100142, 2019, doi: 10.1016/j.ynstr.2018.100142.

[95] K. Sunthad, Y. Niitsu, M. Inoue, and T. Yokemura, "Brain's Stress Observation System Using 2-Channels NIRS Based on Classroom Activity," in *2019 IEEE International Conference on Consumer Electronics (ICCE)*, 2019, pp. 1–4.

[96] W. Wu, S. Pirbhulal, H. Zhang, and S. C. Mukhopadhyay, "Quantitative Assessment for Self-Tracking of Acute Stress based on Triangulation Principle in a Wearable Sensor System," *IEEE J. Biomed. Heal. Informatics*, vol. PP, no. c, p. 1, 2018, doi: 10.1109/JBHI.2018.2832069.

[97] E. Russell, G. Koren, M. Rieder, and S. H. M. Van Uum, "The detection of cortisol in human sweat: implications for measurement of cortisol in hair," *Ther. Drug Monit.*, vol. 36, no. 1, pp. 30–34, 2014.

[98] E. Russell, G. Koren, M. Rieder, and S. Van Uum, "Hair cortisol as a biological marker of chronic stress: current status, future directions and unanswered questions.,"

Psychoneuroendocrinology, vol. 37, no. 5, pp. 589–601, 2012, doi: 10.1016/j.psyneuen.2011.09.009.

[99] N. Ahmed, B. De La Torre, and N. G. Wahlgren, "Salivary cortisol, a biological marker of stress, is positively associated with 24-hour systolic blood pressure in patients with acute ischaemic stroke," *Cerebrovasc. Dis.*, vol. 18, no. 3, pp. 206–213, 2004, doi: 10.1159/000079943.

[100] P. Redon, A. Shahzad, T. Iqbal, and W. Wijns, "Development of a New Detection Algorithm to Identify Acute Coronary Syndrome Using Electrochemical Biosensors for Real-World Long-Term Monitoring," *Bioengineering*, vol. 8, no. 2, p. 28, 2021.

[101] S. Minaee, "popular machine learning metrics. part 1: Classification & regression evaluation metrics,"," *Mediu. url https//towardsdatascience. com/20-popular-machine-learning-metrics-part-1-classification-regressionevaluation-metrics-1ca3e282a2ce*, 2019.

[102] J. Jordan, "Evaluating a machine learning model."," *DATA Sci. url https//www.jeremyjordan.me/evaluating-a-machine-learning-model/*, 2017.

[103] M. C. Staff, "Stress symptoms: Effects on your body and behavior." 2016.

[104] A. Reisner, P. A. Shaltis, D. McCombie, H. H. Asada, D. S. Warner, and M. A. Warner, "Utility of the photoplethysmogram in circulatory monitoring," *J. Am. Soc. Anesthesiol.*, vol. 108, no. 5, pp. 950–958, 2008.

[105] Z. Villines and S. Flack, "Why do we use pulse oximetry?" 2017.

[106] O. Crew, "How And Why To Track Respiratory Rate Trends With the Oura Wellness Ring." 2017.

[107] M. H. Laudat, S. Cerdas, C. Fournier, D. Guiban, B. Guilhaume, and J. P. Luton, "Salivary cortisol measurement: a practical approach to assess pituitary-adrenal function," *J. Clin. Endocrinol. \& Metab.*, vol. 66, no. 2, pp. 343–348, 1988.

[108] P. Braveman and L. Gottlieb, "The social determinants of health: it's time to consider the causes of the causes," *Public Health Rep.*, vol. 129, no. 1_suppl2, pp. 19–31, 2014.

[109] S. Salimetrics USA, "SALIVARY CORTISOL." 2019.

[110] F. N. U. Apoorvagiri and M. S. Nagananda, "Quantization of mental stress using various physiological markers," 2015.

[111] K. W. Kallus and M. Kellmann, *The recovery-stress questionnaires: user manual*. Pearson London, UK:, 2016.

[112] P. M. Boynton and T. Greenhalgh, "Selecting, designing, and developing your questionnaire," *Bmj*, vol. 328, no. 7451, pp. 1312–1315, 2004.

[113] E. G. Pintelas, T. Kotsilieris, I. E. Livieris, and P. Pintelas, "A review of machine learning prediction methods for anxiety disorders," in *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, 2018, pp. 8–15.

[114] R. R. Bouckaert, "Choosing between two learning algorithms based on calibrated tests," in *ICML*, 2003, vol. 3, pp. 51–58.

[115] T. Iqbal, A. Elahi, A. Shahzad, and W. Wijns, "Review on Classification Techniques used in Biophysiological Stress Monitoring," *arXiv Prepr. arXiv2210.16040*, 2022.

[116] Y. Ye, T. Hu, A. Nassehi, S. Ji, and H. Ni, "Context-aware manufacturing system design using machine learning," *J. Manuf. Syst.*, vol. 65, pp. 59–69, 2022.

[117] G. R. S. Reddy, B. Srinivasulu, M. Roshini, and V. R. Lakshmi, "The Study of Supervised Classification Techniques in Machine Learning using Keras," *i-Manager's J. Futur. Eng. Technol.*, vol. 15, no. 3, p. 20, 2020.

[118] C. Dadhirao and R. Sangam, "Localization techniques using machine learning algorithms," *Archit. Wirel. Networks Solut. Secur. Issues*, pp. 175–193, 2021.

[119] S. K. Murthy, "Automatic construction of decision trees from data: A multi-disciplinary survey," *Data Min. Knowl. Discov.*, vol. 2, no. 4, pp. 345–389, 1998.

[120] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artif. Intell. Rev.*, vol. 26, no. 3, pp. 159–190, 2006.

[121] A. Guarino and G. Spagnuolo, "Automatic features extraction of faults in PEM fuel cells by a siamese artificial neural network," *Int. J. Hydrogen Energy*, vol. 46, no. 70, pp. 34854–34866, 2021.

[122] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, "Introduction to machine learning, neural networks, and deep learning," *Transl. Vis. Sci. \& Technol.*, vol. 9, no. 2, p. 14, 2020.

[123] G. P. Zhang, "Neural networks for classification: a survey," *IEEE Trans. Syst. Man, Cybern.*

Part C (Applications Rev., vol. 30, no. 4, pp. 451–462, 2000.

[124] H. Parvin, M. Mohamadi, S. Parvin, Z. Rezaei, and B. Minaei, "Nearest cluster classifier," in *International Conference on Hybrid Artificial Intelligence Systems*, 2012, pp. 267–275.

[125] J. Brownlee, "Learning Vector Quantization for Machine Learning," *Online) http//machinelearningmastery. com/learning-vector-quantization-for-machine-learning/(12 April 2017)*, 2016.

[126] A. Ahmad and R. Yusof, "A modified kohonen self-organizing map (KSOM) clustering for four categorical data," *J. Teknol.*, vol. 78, no. 6–13, 2016.

[127] S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, "An overview of principal component analysis," *J. Signal Inf. Process.*, vol. 4, no. 3B, p. 173, 2013.

[128] D. Rana, S. P. Jena, and S. K. Pradhan, "Performance Comparison of PCA and LDA with Linear Regression and Random Forest for IRIS Flower Classification," *PalArch's J. Archaeol. Egypt/Egyptology*, vol. 17, no. 9, pp. 2353–2360, 2020.

[129] A. Navlani, "Understanding logistic regression in python," *Link https//www. datacamp. com/community/tutorials/understanding-logistic-regressionpython\# comments. Udgivet*, vol. 7, 2018.

[130] L. Kotthoff, C. Thornton, H. H. Hoos, F. Hutter, and K. Leyton-Brown, "Auto-WEKA: Automatic model selection and hyperparameter optimization in WEKA," in *Automated Machine Learning*, Springer, Cham, 2019, pp. 81–95.

[131] S. Sayad, "No Title," *Classification*. [Online]. Available: http://www.saedsayad.com/classification.htm.

[132] A. A. Johari, M. H. Abd Wahab, and A. Mustapha, "Two-Class Classification: Comparative Experiments for Chronic Kidney Disease," in *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, 2019, pp. 789–792.

[133] Xiaoharper, "ML Studio (Classic): Two-class decision jungle - azure," *ML Studio (classic): Two-Class Decision Jungle - Azure | Microsoft Docs*. [Online]. Available: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-decision-jungle.

[134] W. Lin, Z. Wu, L. Lin, A. Wen, and J. Li, "An ensemble random forest algorithm for insurance big data analysis," *Ieee access*, vol. 5, pp. 16568–16575, 2017.

[135] J. Yan, Z. Zhang, K. Lin, F. Yang, and X. Luo, "A hybrid scheme-based one-vs-all decision

trees for multi-class classification tasks," *Knowledge-Based Syst.*, vol. 198, p. 105922, 2020.

[136] Y. Freund, R. Schapire, and N. Abe, "A short introduction to boosting," *Journal-Japanese Soc. Artif. Intell.*, vol. 14, no. 771–780, p. 1612, 1999.

[137] S. R. Gomes *et al.*, "A comparative approach to email classification using Naive Bayes classifier and hidden Markov model," in *2017 4th International Conference on Advances in Electrical Engineering (ICAEE)*, 2017, pp. 482–487.

[138] A. Bilski, "A review of artificial intelligence algorithms in document classification," *Int. J. Electron. Telecommun.*, vol. 57, pp. 263–270, 2011.

[139] M. Singh and A. Bin Queyam, "A Novel Method of Stress Detection using Physiological Measurements of Automobile Drivers," *Int. J. Electron. Eng.*, no. 2, pp. 13–20, 2013.

[140] D. Cogan, M. B. Pouyan, M. Nourani, and J. Harvey, "A wrist-worn biosensor system for assessment of neurological status," *2014 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC 2014*, pp. 5748–5751, 2014, doi: 10.1109/EMBC.2014.6944933.

[141] F. P. Akbulut and A. Akan, "A smart wearable system for short-term cardiovascular risk assessment with emotional dynamics," *Measurement*, vol. 128, pp. 237–246, 2018.

[142] A. Barreto, J. Zhai, and M. Adjouadi, "Non-intrusive Physiological Monitoring for Automated Stress Detection in Human-Computer Interaction," *Human–Computer Interact.*, pp. 29–38, 2007, doi: 10.1007/978-3-540-75773-3_4.

[143] R. Castaldo, W. Xu, P. Melillo, L. Pecchia, L. Santamaria, and C. James, "Detection of mental stress due to oral academic examination via ultra-short-term HRV analysis," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 3805–3808.

[144] F. Mokhayeri, M. R. Akbarzadeh-T, and S. Toosizadeh, "Mental stress detection using physiological signals based on soft computing techniques," *2011 18th Iran. Conf. Biomed. Eng. ICBME 2011*, no. December, pp. 232–237, 2011, doi: 10.1109/ICBME.2011.6168563.

[145] D. F. Dinges *et al.*, "Optical computer recognition of facial expressions associated with stress induced by performance demands," *Aviat. Space. Environ. Med.*, vol. 76, no. 6, pp. B172--B182, 2005.

[146] H. Jebelli, S. Hwang, and S. H. Lee, "EEG-based workers' stress recognition at construction sites," *Autom. Constr.*, vol. 93, no. April, pp. 315–324, 2018, doi:

10.1016/j.autcon.2018.05.027.

[147] R. R. Singh, S. Conjeti, and R. Banerjee, "Biosignal based on-road stress monitoring for automotive drivers," *2012 Natl. Conf. Commun. NCC 2012*, pp. 1–5, 2012, doi: 10.1109/NCC.2012.6176845.

[148] S. C. Pauws, M. Biehl, and others, "Insightful stress detection from physiology modalities using learning vector quantization," *Neurocomputing*, vol. 151, pp. 873–882, 2015.

[149] C. Dobbins and S. Fairclough, "Detecting Negative Emotions during Real-Life Driving via Dynamically Labelled Physiological Data," *2018 IEEE Int. Conf. Pervasive Comput. Commun. Work. PerCom Work. 2018*, pp. 830–835, 2018, doi: 10.1109/PERCOMW.2018.8480369.

[150] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 410–417, 2009.

[151] A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel, and A. S. Malik, "Machine learning framework for the detection of mental stress at multiple levels," *IEEE Access*, vol. 5, pp. 13545–13556, 2017.

[152] R. A. Calvo, I. Brown, and S. Scheding, "Effect of experimental factors on the recognition of affective mental states through physiological measures," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5866 LNAI, pp. 62–70, 2009, doi: 10.1007/978-3-642-10439-8_7.

[153] H. M. Khan, B. Ahmed, J. Choi, and R. Gutierrez-Osuna, "Using an ambulatory stress monitoring device to identify relaxation due to untrained deep breathing," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 1744–1747, 2013, doi: 10.1109/EMBC.2013.6609857.

[154] O. Oti, I. Azimi, and A. Anzanpour, "IoT-based Healthcare System for Real-time Maternal Stress Monitoring," no. September, 2018.

[155] T. Salafi and J. C. Y. Kah, "Design of unobtrusive wearable mental stress monitoring device using physiological sensor," *IFMBE Proc.*, vol. 52, pp. 11–14, 2015, doi: 10.1007/978-3-319-19452-3_4.

[156] S. Betti *et al.*, "Evaluation of an integrated system of wearable physiological sensors for stress monitoring in working environments by using biological markers," *IEEE Trans. Biomed.*

Eng., vol. 65, no. 8, pp. 1748–1758, 2017.

[157] S. H. Sunwoo *et al.*, "Chronic and acute stress monitoring by electrophysiological signals from adrenal gland," *Proc. Natl. Acad. Sci.*, vol. 116, no. 4, pp. 1146–1151, 2019.

[158] I. Berger, M. Werdermann, S. R. Bornstein, and C. Steenblock, "The adrenal gland in stress--Adaptation on a cellular level," *J. Steroid Biochem. Mol. Biol.*, vol. 190, pp. 198–206, 2019.

[159] J.-H. Lee and H.-I. Jung, "Biochip technology for monitoring posttraumatic stress disorder (PTSD)," *BioChip J.*, vol. 7, no. 3, pp. 195–200, 2013.

[160] Z. Djuric *et al.*, "Biomarkers of psychological stress in health disparities research," *Open Biomark. J.*, vol. 1, p. 7, 2008.

[161] A. Kaushik, A. Vasudev, S. K. Arya, S. K. Pasha, and S. Bhansali, "Recent advances in cortisol sensing technologies for point-of-care application," *Biosens. Bioelectron.*, vol. 53, pp. 499–512, 2014.

[162] M. Holleman, S. A. Vreeburg, J. J. M. Dekker, and B. W. J. H. Penninx, "The relationships of working conditions, recent stressors and childhood trauma with salivary cortisol levels," *Psychoneuroendocrinology*, vol. 37, no. 6, pp. 801–809, 2012.

[163] A. S. Zainol Abidin *et al.*, "Current and potential developments of cortisol aptasensing towards point-of-care diagnostics (POTC)," *Sensors*, vol. 17, no. 5, p. 1180, 2017.

[164] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'Trier Social Stress Test'--a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1–2, pp. 76–81, 1993.

[165] B. M. Kudielka, A. Buske-Kirschbaum, D. H. Hellhammer, and C. Kirschbaum, "HPA axis responses to laboratory psychosocial stress in healthy elderly adults, younger adults, and children: impact of age and gender," *Psychoneuroendocrinology*, vol. 29, no. 1, pp. 83–98, 2004.

[166] R. Gatti, G. Antonelli, M. Prearo, P. Spinella, E. Cappellin, and F. Elio, "Cortisol assays and diagnostic laboratory procedures in human biological fluids," *Clin. Biochem.*, vol. 42, no. 12, pp. 1205–1217, 2009.

[167] A. Levine, O. Zagoory-Sharon, R. Feldman, J. G. Lewis, and A. Weller, "Measuring cortisol in human psychobiological studies," *Physiol. \& Behav.*, vol. 90, no. 1, pp. 43–53, 2007.

[168] M. Akinola and W. B. Mendes, "Stress-induced cortisol facilitates threat-related decision making among police officers.," *Behav. Neurosci.*, vol. 126, no. 1, p. 167, 2012.

[169] A. D. Clements, "Salivary cortisol measurement in developmental research: where do we go from here?," *Dev. Psychobiol.*, vol. 55, no. 3, pp. 205–220, 2013.

[170] V. L. Kallen, J. H. Stubbe, H. J. Zwolle, and P. Valk, "Capturing effort and recovery: reactive and recuperative cortisol responses to competition in well-trained rowers," *BMJ open Sport \& Exerc. Med.*, vol. 3, no. 1, p. e000235, 2017.

[171] M. V. Rosati *et al.*, "Plasma cortisol concentrations and lifestyle in a population of outdoor workers," *Int. J. Environ. Health Res.*, vol. 21, no. 1, pp. 62–71, 2011.

[172] M. R. Sladek, L. D. Doane, L. J. Luecken, and N. Eisenberg, "Perceived stress, coping, and cortisol reactivity in daily life: A study of adolescents during the first year of college," *Biol. Psychol.*, vol. 117, pp. 8–15, 2016.

[173] K. Wingenfeld, M. Schulz, A. Damkroeger, C. Philippsen, M. Rose, and M. Driessen, "The diurnal course of salivary alpha-amylase in nurses: An investigation of potential confounders and associations with stress," *Biol. Psychol.*, vol. 85, no. 1, pp. 179–181, 2010.

[174] J.-H. Lee, Y. Hwang, K.-A. Cheon, and H.-I. Jung, "Emotion-on-a-chip (EOC): Evolution of biochip technology to measure human emotion using body fluids," *Med. Hypotheses*, vol. 79, no. 6, pp. 827–832, 2012.

[175] M. Debono *et al.*, "Modified-release hydrocortisone to provide circadian cortisol profiles," *J. Clin. Endocrinol. \& Metab.*, vol. 94, no. 5, pp. 1548–1554, 2009.

[176] E. R. De Kloet, M. Joëls, and F. Holsboer, "Stress and the brain: from adaptation to disease," *Nat. Rev. Neurosci.*, vol. 6, no. 6, pp. 463–475, 2005.

[177] B. S. McEwen, "Cortisol, Cushing's syndrome, and a shrinking brain—new evidence for reversibility," *J. Clin. Endocrinol. \& Metab.*, vol. 87, no. 5, pp. 1947–1948, 2002.

[178] O. Edwards, J. M. Galley, R. J. Courtenay-Evans, J. Hunter, and A. Tait, "Changes in cortisol metabolism following rifampicin therapy," *Lancet*, vol. 304, no. 7880, pp. 549–551, 1974.

[179] F. Holsboer and M. Ising, "Stress hormone regulation: biological role and translation into therapy," *Annu. Rev. Psychol.*, vol. 61, pp. 81–109, 2010.

[180] C. W. Le Roux, G. A. Chapman, W. M. Kong, W. S. Dhillo, J. Jones, and J. Alaghband-Zadeh, "Free cortisol index is better than serum total cortisol in determining hypothalamic-pituitary-adrenal status in patients undergoing surgery," *J. Clin. Endocrinol. \& Metab.*, vol. 88, no. 5, pp. 2045–2048, 2003.

[181] K. Hogenelst, M. Soeter, and V. Kallen, "Ambulatory measurement of cortisol: Where do we stand, and which way to follow?," *Sens. Bio-Sensing Res.*, vol. 22, p. 100249, 2019.

[182] M. Frasconi, M. Mazzarino, F. Botrè, and F. Mazzei, "Surface plasmon resonance immunosensor for cortisol and cortisone determination," *Anal. Bioanal. Chem.*, vol. 394, no. 8, pp. 2151–2159, 2009.

[183] G. A. Akceoglu, Y. Saylan, and F. Inci, "A Snapshot of Microfluidics in Point-of-Care Diagnostics: Multifaceted Integrity with Materials and Sensors," *Adv. Mater. Technol.*, p. 2100049, 2021.

[184] M. Yaneva, G. Kirilov, and S. Zacharieva, "Midnight salivary cortisol, measured by highly sensitive electrochemiluminescence immunoassay, for the diagnosis of Cushing's syndrome," *Open Med.*, vol. 4, no. 1, pp. 59–64, 2009.

[185] A. J. Steckl and P. Ray, "Stress Biomarkers in Biological Fluids and Their Point-of-Use Detection," *ACS Sensors*, vol. 3, no. 10, pp. 2025–2044, 2018, doi: 10.1021/acssensors.8b00726.

[186] O. Parlak, "Portable and wearable real-time stress monitoring: A critical review," *Sensors and Actuators Reports*, vol. 3, p. 100036, 2021.

[187] J. Zhang *et al.*, "Wearable biosensors for human fatigue diagnosis: A review," *Bioeng. \& Transl. Med.*, p. e10318, 2022.

[188] M. Zea *et al.*, "Electrochemical sensors for cortisol detections: Almost there," *TrAC Trends Anal. Chem.*, vol. 132, p. 116058, 2020.

[189] A. Kaushik, A. Vasudev, S. K. Arya, S. K. Pasha, and S. Bhansali, "Recent advances in cortisol sensing technologies for point-of-care application," *Biosens. Bioelectron.*, vol. 53, pp. 499–512, 2014, doi: 10.1016/j.bios.2013.09.060.

[190] M. D. VanBruggen, A. C. Hackney, R. G. McMurray, and K. S. Ondrak, "The relationship between serum and salivary cortisol levels in response to different intensities of exercise," *Int. J. Sports Physiol. Perform.*, vol. 6, no. 3, pp. 396–407, 2011.

[191] J. Bakusic, S. De Nys, M. Creta, L. Godderis, and R. C. Duca, "Study of temporal variability of salivary cortisol and cortisone by LC-MS/MS using a new atmospheric pressure ionization source," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019.

[192] N. El-Farhan, D. A. Rees, and C. Evans, "Measuring cortisol in serum, urine and saliva--are

our assays good enough?," *Ann. Clin. Biochem.*, vol. 54, no. 3, pp. 308–322, 2017.

[193] A. Hodes *et al.*, "Hair cortisol in the evaluation of Cushing syndrome," *Endocrine*, vol. 56, no. 1, pp. 164–174, 2017.

[194] B. Sauvé, G. Koren, G. Walsh, S. Tokmakejian, and S. H. M. Van Uum, "Measurement of cortisol in human hair as a biomarker of systemic exposure," *Clin. Investig. Med.*, pp. E183--E191, 2007.

[195] A. Ghemigian *et al.*, "Cushing's disease--Same condition, different scenarios," *Arch. Balk. Med. Union*, vol. 53, no. 1, pp. 135–139, 2018.

[196] G. Cizza and K. I. Rother, "Cortisol binding globulin: more than just a carrier?," *J. Clin. Endocrinol. \& Metab.*, vol. 97, no. 1, pp. 77–80, 2012.

[197] M. Venugopal, S. K. Arya, G. Chornokur, and S. Bhansali, "A realtime and continuous assessment of cortisol in ISF using electrochemical impedance spectroscopy," *Sensors Actuators A Phys.*, vol. 172, no. 1, pp. 154–160, 2011.

[198] A. El-Laboudi, N. S. Oliver, A. Cass, and D. Johnston, "Use of microneedle array devices for continuous glucose monitoring: a review," *Diabetes Technol. \& Ther.*, vol. 15, no. 1, pp. 101–115, 2013.

[199] P. M. Wang, M. Cornwell, and M. R. Prausnitz, "Minimally invasive extraction of dermal interstitial fluid for glucose monitoring using microneedles," *Diabetes Technol. \& Ther.*, vol. 7, no. 1, pp. 131–141, 2005.

[200] P. Khanna, J. A. Strom, J. I. Malone, and S. Bhansali, "Microneedle-based automated therapy for diabetes mellitus," *J. Diabetes Sci. Technol.*, vol. 2, no. 6, pp. 1122–1129, 2008.

[201] I. Schmalbach *et al.*, "Cortisol reactivity in patients with anorexia nervosa after stress induction," *Transl. Psychiatry*, vol. 10, no. 1, pp. 1–15, 2020.

[202] A. Siddiqui, N. G. Desai, S. B. Sharma, M. Aslam, U. K. Sinha, and S. V Madhu, "Association of oxidative stress and inflammatory markers with chronic stress in patients with newly diagnosed type 2 diabetes," *Diabetes. Metab. Res. Rev.*, vol. 35, no. 5, p. e3147, 2019.

[203] G. Grossi, A. Perski, M. Ekstedt, T. Johansson, M. Lindström, and K. Holm, "The morning salivary cortisol response in burnout," *J. Psychosom. Res.*, vol. 59, no. 2, pp. 103–111, 2005.

[204] L. H. Powell *et al.*, "Physiologic markers of chronic stress in premenopausal, middle-aged

women," *Psychosom. Med.*, vol. 64, no. 3, pp. 502–509, 2002.

[205] L. C. Carlesso, J. A. Sturgeon, and A. J. Zautra, "Exploring the relationship between disease-related pain and cortisol levels in women with osteoarthritis," *Osteoarthr. Cartil.*, vol. 24, no. 12, pp. 2048–2054, 2016.

[206] S. Balters, J. W. Geeseman, A.-K. Tveten, H. P. Hildre, W. Ju, and M. Steinert, "Mayday, Mayday, Mayday: Using salivary cortisol to detect distress (and eustress!) in critical incident training," *Int. J. Ind. Ergon.*, vol. 78, p. 102975, 2020.

[207] P. Ethridge, N. Ali, S. E. Racine, J. C. Pruessner, and A. Weinberg, "Risk and resilience in an acute stress paradigm: Evidence from salivary cortisol and time-frequency analysis of the reward positivity," *Clin. Psychol. Sci.*, vol. 8, no. 5, pp. 872–889, 2020.

[208] K. Mizuhata, H. Taniguchi, M. Shimada, N. Hikita, and S. Morokuma, "Effects of Breastfeeding on Stress Measured by Saliva Cortisol Level and Perceived Stress," *Asian/Pacific Isl. Nurs. J.*, vol. 5, no. 3, p. 128, 2020.

[209] K. Jansakova, K. Kyselicova, H. Celusakova, G. Repiska, and D. Ostatnikova, "The effect of saliva stimulation on the secretion of cortisol during stress and physiological conditions," *Clin. STUDY*, vol. 428, p. 430, 2020.

[210] S. Metz *et al.*, "Blunted salivary cortisol response to psychosocial stress in women with posttraumatic stress disorder," *J. Psychiatr. Res.*, vol. 130, pp. 112–119, 2020.

[211] R. H. Pompon, A. N. Smith, C. Baylor, and D. Kendall, "Exploring associations between a biological marker of chronic stress and reported depression and anxiety in people with aphasia," *J. Speech, Lang. Hear. Res.*, vol. 62, no. 11, pp. 4119–4130, 2019.

[212] J. Vliegenthart, G. Noppe, E. F. C. Van Rossum, J. W. Koper, H. Raat, and E. L. T. den Akker, "Socioeconomic status in children is associated with hair cortisol levels as a biological measure of chronic stress," *Psychoneuroendocrinology*, vol. 65, pp. 9–14, 2016.

[213] X. Chen *et al.*, "Caregivers' hair cortisol: a possible biomarker of chronic stress is associated with obesity measures among children with disabilities," *BMC Pediatr.*, vol. 15, no. 1, pp. 1–13, 2015.

[214] P. Henley *et al.*, "Hair cortisol as a biomarker of stress among a first nation in Canada," *Ther. Drug Monit.*, vol. 35, no. 5, pp. 595–599, 2013.

[215] L. E. Bautista, P. K. Bajwa, M. M. Shafer, K. M. C. Malecki, C. A. McWilliams, and A.

Palloni, "The relationship between chronic stress, hair cortisol and hypertension," *Int. J. Cardiol. Hypertens.*, vol. 2, p. 100012, 2019.

[216] S. L. Moch, V. R. Panz, B. I. Joffe, I. Havlik, and J. D. Moch, "Longitudinal changes in pituitary-adrenal hormones in South African women with burnout," *Endocrine*, vol. 21, no. 3, pp. 267–272, 2003.

[217] J. K. Cremeans-Smith, K. Greene, and D. L. Delahanty, "Physiological indices of stress prior to and following total knee arthroplasty predict the occurrence of severe post-operative pain," *Pain Med.*, vol. 17, no. 5, pp. 970–979, 2016.

[218] S. Khoromi *et al.*, "Effects of chronic osteoarthritis pain on neuroendocrine function in men," *J. Clin. Endocrinol. \& Metab.*, vol. 91, no. 11, pp. 4313–4318, 2006.

[219] C. D. Butts *et al.*, "Urine cortisol concentration as a biomarker of stress is unrelated to IVF outcomes in women and men," *J. Assist. Reprod. Genet.*, vol. 31, no. 12, pp. 1647–1653, 2014.

[220] C.-L. Du, M. C. Lin, L. Lu, and J. J. Tai, "Correlation of occupational stress index with 24-hour urine cortisol and serum DHEA sulfate among city bus drivers: a cross-sectional study," *Saf. Health Work*, vol. 2, no. 2, pp. 169–175, 2011.

[221] C. Shimanoe *et al.*, "Perceived stress, depressive symptoms, and cortisol-to-cortisone ratio in spot urine in 6878 older adults," *Psychoneuroendocrinology*, vol. 125, p. 105125, 2021.

[222] G. M. Hall, D. Peerbhoy, A. Shenkin, C. J. R. Parker, and P. Salmon, "Relationship of the functional recovery after hip arthroplasty to the neuroendocrine and inflammatory responses," *Br. J. Anaesth.*, vol. 87, no. 4, pp. 537–542, 2001.

[223] L. M. Oswald, P. Zandi, G. Nestadt, J. B. Potash, A. E. Kalaydjian, and G. S. Wand, "Relationship between cortisol responses to stress and personality," *Neuropsychopharmacology*, vol. 31, no. 7, pp. 1583–1591, 2006.

[224] M. S. Herbert *et al.*, "Ethnicity, cortisol, and experimental pain responses among persons with symptomatic knee osteoarthritis," *Clin. J. Pain*, vol. 33, no. 9, p. 820, 2017.

[225] R. J. McQuaid, O. A. McInnis, A. Paric, F. Al-Yawer, K. Matheson, and H. Anisman, "Relations between plasma oxytocin and cortisol: the stress buffering role of social support," *Neurobiol. Stress*, vol. 3, pp. 52–60, 2016.

[226] E. Ortega, I. Gálvez, M. D. Hinchado, J. Guerrero, L. Mart\'\in-Cordero, and S. Torres-Piles, "Anti-inflammatory effect as a mechanism of effectiveness underlying the clinical

benefits of pelotherapy in osteoarthritis patients: regulation of the altered inflammatory and stress feedback response," *Int. J. Biometeorol.*, vol. 61, no. 10, pp. 1777–1785, 2017.

[227] \cSükrü Burak Tönük, E. Serin, F. F. Ayhan, and Z. R. Yorgancioglu, "The effects of physical therapeutic agents on serum levels of stress hormones in patients with osteoarthritis," *Medicine (Baltimore).*, vol. 95, no. 35, 2016.

[228] A. G. Bertollo *et al.*, "Stress and serum cortisol levels in major depressive disorder: a cross-sectional study," *AIMS Neurosci.*, vol. 7, no. 4, p. 459, 2020.

[229] N. Schaffter *et al.*, "Serum cortisol as a predictor for posttraumatic stress disorder symptoms in post-myocardial infarction patients," *J. Affect. Disord.*, 2021.

[230] P. Pearlmutter *et al.*, "Sweat and saliva cortisol response to stress and nutrition factors," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, 2020.

[231] S. N. Doan, G. DeYoung, T. E. Fuller-Rowell, C. Liu, and J. Meyer, "Investigating relations among stress, sleep and nail cortisol and DHEA," *Stress*, vol. 21, no. 2, pp. 188–193, 2018.

[232] H. Wu, K. Zhou, P. Xu, J. Xue, X. Xu, and L. Liu, "Associations of perceived stress with the present and subsequent cortisol levels in fingernails among medical students: A prospective pilot study," *Psychol. Res. Behav. Manag.*, vol. 11, pp. 439–445, 2018, doi: 10.2147/PRBM.S181541.

[233] V. L. Wester and E. F. C. van Rossum, "Clinical applications of cortisol measurements in hair," *Eur. J. Endocrinol.*, vol. 173, no. 4, pp. M1--M10, 2015.

[234] B. R. Walker, "Glucocorticoids and cardiovascular disease," *Eur. J. Endocrinol.*, vol. 157, no. 5, pp. 545–559, 2007.

[235] L. Manenschijn *et al.*, "High long-term cortisol levels, measured in scalp hair, are associated with a history of cardiovascular disease," *J. Clin. Endocrinol. \& Metab.*, vol. 98, no. 5, pp. 2078–2083, 2013.

[236] S. Feller, M. Vigl, M. M. Bergmann, H. Boeing, C. Kirschbaum, and T. Stalder, "Predictors of hair cortisol concentrations in older adults," *Psychoneuroendocrinology*, vol. 39, pp. 132–140, 2014.

[237] D. Pereg *et al.*, "Hair cortisol and the risk for acute myocardial infarction in adult men," *Stress*, vol. 14, no. 1, pp. 73–81, 2011.

[238] T. Stalder *et al.*, "Cortisol in hair and the metabolic syndrome," *J. Clin. Endocrinol. \& Metab.*,

vol. 98, no. 6, pp. 2573–2580, 2013.

[239] G. Noppe, E. F. C. Van Rossum, J. Vliegenthart, J. W. Koper, and E. L. T. Van Den Akker, "Elevated hair cortisol concentrations in children with adrenal insufficiency on hydrocortisone replacement therapy," *Clin. Endocrinol. (Oxf).*, vol. 81, no. 6, pp. 820–825, 2014.

[240] V. L. Wester *et al.*, "Long-term cortisol levels measured in scalp hair of obese patients," *Obesity*, vol. 22, no. 9, pp. 1956–1958, 2014.

[241] J. O. Younge *et al.*, "Cortisol levels in scalp hair of patients with structural heart disease," *Int. J. Cardiol.*, vol. 184, pp. 71–78, 2015.

[242] S. Braig *et al.*, "Determinants of maternal hair cortisol concentrations at delivery reflecting the last trimester of pregnancy," *Psychoneuroendocrinology*, vol. 52, pp. 289–296, 2015.

[243] S. Kalra, A. Einarson, T. Karaskov, S. Van Uum, and G. Koren, "The relationship between stress and hair cortisol in healthy pregnant women," *Clin. Investig. Med.*, pp. E103--E107, 2007.

[244] J. Karlén, J. Ludvigsson, A. Frostell, E. Theodorsson, and T. Faresjö, "Cortisol in hair measured in young adults-a biomarker of major life stressors?," *BMC Clin. Pathol.*, vol. 11, no. 1, pp. 1–6, 2011.

[245] T. Stalder *et al.*, "Cortisol in hair, body mass index and stress-related measures," *Biol. Psychol.*, vol. 90, no. 3, pp. 218–223, 2012.

[246] Y. Dowlati *et al.*, "Relationship between hair cortisol concentrations and depressive symptoms in patients with coronary artery disease," *Neuropsychiatr. Dis. Treat.*, vol. 6, p. 393, 2010.

[247] H. M. Burke, M. C. Davis, C. Otte, and D. C. Mohr, "Depression and cortisol responses to psychological stress: a meta-analysis," *Psychoneuroendocrinology*, vol. 30, no. 9, pp. 846–856, 2005.

[248] M. A. B. Veldhorst *et al.*, "Increased scalp hair cortisol concentrations in obese children," *J. Clin. Endocrinol. \& Metab.*, vol. 99, no. 1, pp. 285–290, 2014.

[249] S. Steudte *et al.*, "Hair cortisol as a biomarker of traumatization in healthy individuals and posttraumatic stress disorder patients," *Biol. Psychiatry*, vol. 74, no. 9, pp. 639–646, 2013.

[250] L. Wang *et al.*, "Linking hair cortisol levels to phenotypic heterogeneity of posttraumatic

stress symptoms in highly traumatized chinese women," *Biol. Psychiatry*, vol. 77, no. 4, pp. e21--e22, 2015.

[251] L. Manenschijn *et al.*, "Long-term cortisol in bipolar disorder: associations with age of onset and psychiatric co-morbidity," *Psychoneuroendocrinology*, vol. 37, no. 12, pp. 1960–1968, 2012.

[252] S. Steudte, I.-T. Kolassa, T. Stalder, A. Pfeiffer, C. Kirschbaum, and T. Elbert, "Increased cortisol concentrations in hair of severely traumatized Ugandan individuals with PTSD," *Psychoneuroendocrinology*, vol. 36, no. 8, pp. 1193–1200, 2011.

[253] X. Weng, Z. Fu, C. Zhang, W. Jiang, and H. Jiang, "A Portable 3D Microfluidic Origami Biosensor for Cortisol Detection in Human Sweat," *Anal. Chem.*, vol. 94, no. 8, pp. 3526–3534, 2022.

[254] S. Zhang, A. Garcia-D'Angeli, J. P. Brennan, and Q. Huo, "Predicting detection limits of enzyme-linked immunosorbent assay (ELISA) and bioanalytical techniques in general," *Analyst*, vol. 139, no. 2, pp. 439–445, 2014.

[255] R. Minic and I. Zivkovic, "Optimization, Validation and Standardization of ELISA," in *Norovirus*, IntechOpen, 2020.

[256] S. K. Arya, A. Dey, and S. Bhansali, "Polyaniline protected gold nanoparticles based mediator and label free electrochemical cortisol biosensor," *Biosens. Bioelectron.*, vol. 28, no. 1, pp. 166–173, 2011.

[257] T. Vural, Y. T. Yaman, S. Ozturk, S. Abaci, and E. B. Denkbas, "Electrochemical immunoassay for detection of prostate specific antigen based on peptide nanotube-gold nanoparticle-polyaniline immobilized pencil graphite electrode," *J. Colloid Interface Sci.*, vol. 510, pp. 318–326, 2018.

[258] G. A. Posthuma-Trumpie, J. Korf, and A. van Amerongen, "Lateral flow (immuno) assay: its strengths, weaknesses, opportunities and threats. A literature survey," *Anal. Bioanal. Chem.*, vol. 393, no. 2, pp. 569–582, 2009.

[259] B. G. Andryukov, I. N. Lyapun, E. V Matosova, and L. M. Somova, "Biosensor technologies in medicine: from detection of biochemical markers to research into molecular targets," *Современные технологии в медицине*, vol. 12, no. 6 (eng), 2020.

[260] H.-K. Oh, K. Kim, J. Park, H. Jang, and M.-G. Kim, "Advanced trap lateral flow immunoassay sensor for the detection of cortisol in human bodily fluids," *Sci. Rep.*, vol. 11,

no. 1, pp. 1–12, 2021.

[261] P. Hampitak *et al.*, "A point-of-care immunosensor based on a quartz crystal microbalance with graphene biointerface for antibody assay," *ACS sensors*, vol. 5, no. 11, pp. 3520–3532, 2020.

[262] T. Ito, N. Aoki, S. Kaneko, and K. Suzuki, "Highly sensitive and rapid sequential cortisol detection using twin sensor QCM," *Anal. Methods*, vol. 6, no. 18, pp. 7469–7474, 2014.

[263] M. Falk, C. Psotta, S. Cirovic, and S. Shleev, "Non-Invasive Electrochemical Biosensors Operating in Human Physiological Fluids," *Sensors*, vol. 20, no. 21, p. 6352, 2020.

[264] Y.-H. Kim *et al.*, "Direct immune-detection of cortisol by chemiresistor graphene oxide sensor," *Biosens. Bioelectron.*, vol. 98, pp. 473–477, 2017.

[265] C. Tlili, N. V Myung, V. Shetty, and A. Mulchandani, "Label-free, chemiresistor immunosensor for stress biomarker cortisol in saliva," *Biosens. Bioelectron.*, vol. 26, no. 11, pp. 4382–4386, 2011.

[266] S. Jo *et al.*, "Localized surface plasmon resonance aptasensor for the highly sensitive direct detection of cortisol in human saliva," *Sensors Actuators B Chem.*, vol. 304, p. 127424, 2020.

[267] M. Yamaguchi *et al.*, "Immunosensor with fluid control mechanism for salivary cortisol analysis," *Biosens. Bioelectron.*, vol. 41, pp. 186–191, 2013.

[268] A. F. D. Cruz, N. Norena, A. Kaushik, and S. Bhansali, "A low-cost miniaturized potentiostat for point-of-care diagnosis," *Biosens. Bioelectron.*, vol. 62, pp. 249–254, 2014.

[269] A. Vasudev, A. Kaushik, Y. Tomizawa, N. Norena, and S. Bhansali, "An LTCC-based microfluidic system for label-free, electrochemical detection of cortisol," *Sensors Actuators B Chem.*, vol. 182, pp. 139–146, 2013.

[270] A. Kaushik *et al.*, "Electrochemical sensing method for point-of-care cortisol detection in human immunodeficiency virus-infected patients," *Int. J. Nanomedicine*, vol. 10, p. 677, 2015.

[271] W. Leung *et al.*, "One-step quantitative cortisol dipstick with proportional reading," *J. Immunol. Methods*, vol. 281, no. 1–2, pp. 109–118, 2003.

[272] E. A. Shirtcliff, R. L. Buck, M. J. Laughlin, T. Hart, C. R. Cole, and P. D. Slowey, "Salivary cortisol results obtainable within minutes of sample collection correspond with traditional immunoassays," *Clin. Ther.*, vol. 37, no. 3, pp. 505–514, 2015.

[273] S. Choi, S. Kim, J.-S. Yang, J.-H. Lee, C. Joo, and H.-I. Jung, "Real-time measurement of human salivary cortisol for the assessment of psychological stress using a smartphone," *Sens. Bio-Sensing Res.*, vol. 2, pp. 8–11, 2014.

[274] M. Zangheri *et al.*, "A simple and compact smartphone accessory for quantitative chemiluminescence-based lateral flow immunoassay for salivary cortisol detection," *Biosens. Bioelectron.*, vol. 64, pp. 63–68, 2015.

[275] A. Singh, A. Kaushik, R. Kumar, M. Nair, and S. Bhansali, "Electrochemical sensing of cortisol: a recent update," *Appl. Biochem. Biotechnol.*, vol. 174, no. 3, pp. 1115–1126, 2014.

[276] C. J. Cook, "Rapid noninvasive measurement of hormones in transdermal exudate and saliva," *Physiol. \& Behav.*, vol. 75, no. 1–2, pp. 169–181, 2002.

[277] M. Yamaguchi, S. Yoshikawa, Y. Tahara, D. Niwa, Y. Imai, and V. Shetty, "Point-of-use measurement of salivary cortisol levels," in *SENSORS, 2009 IEEE*, 2009, pp. 343–346.

[278] E. Panfilova, "Development of a Prototype Lateral Flow Immunoassay of Cortisol in Saliva for Daily Monitoring of Stress," *Biosensors*, vol. 11, no. 5, p. 146, 2021.

[279] K. Kosicka, A. Siemi\katkowska, A. Szpera-Goździewicz, M. Krzyścin, G. Br\keborowicz, and F. Główka, "High-performance liquid chromatography methods for the analysis of endogenous cortisol and cortisone in human urine: Comparison of mass spectrometry and fluorescence detection," *Ann. Clin. Biochem.*, vol. 56, no. 1, pp. 82–89, 2019.

[280] M. Yamaguchi, H. Katagata, Y. Tezuka, D. Niwa, and V. Shetty, "Automated-immunosensor with centrifugal fluid valves for salivary cortisol measurement," *Sens. bio-sensing Res.*, vol. 1, pp. 15–20, 2014.

[281] B. J. Sanghavi *et al.*, "Aptamer-functionalized nanoparticles for surface immobilization-free electrochemical detection of cortisol in a microfluidic device," *Biosens. Bioelectron.*, vol. 78, pp. 244–252, 2016.

[282] O. Parlak, S. T. Keene, A. Marais, V. F. Curto, and A. Salleo, "Molecularly selective nanoporous membrane-based wearable organic electrochemical device for noninvasive cortisol sensing," *Sci. Adv.*, vol. 4, no. 7, p. eaar2904, 2018.

[283] T. Iqbal *et al.*, "A Sensitivity Analysis of Biophysiological Responses of Stress for Wearable Sensors in Connected Health," *IEEE Access*, vol. 9, pp. 93567–93579, 2021.

[284] E. S. Epel *et al.*, "More than a feeling: A unified view of stress measurement for population

science," *Front. Neuroendocrinol.*, vol. 49, pp. 146–169, 2018.

[285] A. Tawakol *et al.*, "Relation between resting amygdalar activity and cardiovascular events: a longitudinal and cohort study," *Lancet*, vol. 389, no. 10071, pp. 834–845, 2017.

[286] M. Kusserow, O. Amft, and G. Tröster, "Monitoring stress arousal in the wild," *IEEE Pervasive Comput.*, vol. 12, no. 2, pp. 28–37, 2012.

[287] J. S. Sandhu, M. Paul, and H. Agnihotri, "Biofeedback approach in the treatment of generalized anxiety disorder," *Iran. J. Psychiatry*, vol. 2, no. 3, pp. 90–95, 2007.

[288] S. Klangphukhiew, R. Srichana, and R. Patramanon, "Cortisol stress biosensor based on molecular imprinted polymer," in *Multidisciplinary Digital Publishing Institute Proceedings*, 2017, vol. 1, no. 4, p. 538.

[289] J. Botelho *et al.*, "Stress, salivary cortisol and periodontitis: A systematic review and meta-analysis of observational studies," *Arch. Oral Biol.*, vol. 96, pp. 58–65, 2018.

[290] A. Zamkah, T. Hui, S. Andrews, N. Dey, F. Shi, and R. S. Sherratt, "Identification of suitable biomarkers for stress and emotion detection for future personal affective wearable sensors," *Biosensors*, vol. 10, no. 4, p. 40, 2020.

[291] E. Iob and A. Steptoe, "Cardiovascular disease and hair cortisol: a novel biomarker of chronic stress," *Curr. Cardiol. Rep.*, vol. 21, no. 10, pp. 1–11, 2019.

[292] T. Iqbal, A. Elahi, P. Redon, P. Vazquez, W. Wijns, and A. Shahzad, "A Review of Biophysiological and Biochemical Indicators of Stress for Connected and Preventive Healthcare," *Diagnostics*, vol. 11, no. 3, 2021, doi: 10.3390/diagnostics11030556.

[293] F. Sun, C. Kuo, H. Cheng, and S. Buthpitiya, "Activity-Aware Mental Stress Detection," pp. 282–301, 2012.

[294] A. Muaremi, B. Arnrich, and G. Tröster, "Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep," *Bionanoscience*, vol. 3, no. 2, pp. 172–183, 2013, doi: 10.1007/s12668-013-0089-2.

[295] K. Lai, S. N. Yanushkevich, and V. P. Shmerko, "Intelligent Stress Monitoring Assistant for First Responders," *IEEE Access*, vol. 9, pp. 25314–25329, 2021.

[296] E. Smets *et al.*, "Large-scale wearable data reveal digital phenotypes for daily-life stress detection," *NPJ Digit. Med.*, vol. 1, no. 1, pp. 1–10, 2018.

[297] Z. Wang and S. Fu, "An analysis of pilot's physiological reactions in different flight phases," in *International Conference on Engineering Psychology and Cognitive Ergonomics*, 2014, pp. 94–103.

[298] A. Reiss, I. Indlekofer, P. Schmidt, and K. Van Laerhoven, "Deep PPG: large-scale heart rate estimation with convolutional neural networks," *Sensors*, vol. 19, no. 14, p. 3079, 2019.

[299] Y. Jiang, W. Li, M. S. Hossain, M. Chen, A. Alelaiwi, and M. Al-Hammadi, "A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition," *Inf. Fusion*, vol. 53, pp. 209–221, 2020.

[300] C. K. Aridas, S. Karlos, V. G. Kanas, N. Fazakis, and S. B. Kotsiantis, "Uncertainty based under-sampling for learning naive Bayes classifiers under imbalanced data sets," *IEEE Access*, vol. 8, pp. 2122–2133, 2019.

[301] M. T. Uddin and S. Canavan, "Synthesizing physiological and motion data for stress and meditation detection," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 2019, pp. 244–247.

[302] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, 2010.

[303] P. Hamilton, "Open source ECG analysis," in *Computers in cardiology*, 2002, pp. 101–104.

[304] X. Tang, Q. Hu, and W. Tang, "A real-time QRS detection system with PR/RT interval and ST segment measurements for wearable ECG sensors using parallel delta modulators," *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 4, pp. 751–761, 2018.

[305] C. A. Ledezma and M. Altuve, "Optimal data fusion for the improvement of QRS complex detection in multi-channel ECG recordings," *Med. \& Biol. Eng. \& Comput.*, vol. 57, no. 8, pp. 1673–1681, 2019.

[306] S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A survey on ECG analysis," *Biomed. Signal Process. Control*, vol. 43, pp. 216–235, 2018.

[307] G. Pope *et al.*, "An ultra-low resource wearable EDA sensor using wavelet compression," in *2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2018, pp. 193–196.

[308] B. Lamichhane, U. Großekathöfer, G. Schiavone, and P. Casale, "Towards stress detection in real-life scenarios using wearable sensors: normalization factor to reduce variability in stress physiology," in *eHealth 360°*, Springer, 2017, pp. 259–270.

[309] J. Choi, B. Ahmed, and R. Gutierrez-Osuna, "Development and evaluation of an ambulatory stress monitor based on wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 2, pp. 279–286, 2011.

[310] S. Kim, W. Rhee, D. Choi, Y. J. Jang, and Y. Yoon, "Characterizing driver stress using physiological and operational data from real-world electric vehicle driving experiment," *Int. J. Automot. Technol.*, vol. 19, no. 5, pp. 895–906, 2018.

[311] J. Wijsman, B. Grundlehner, J. Penders, and H. Hermens, "Trapezius muscle EMG as predictor of mental stress," *ACM Trans. Embed. Comput. Syst.*, vol. 12, no. 4, pp. 1–20, 2013.

[312] D. Huysmans *et al.*, "Unsupervised learning for mental stress detection-exploration of self-organizing maps," *Proc. Biosignals 2018*, vol. 4, pp. 26–35, 2018.

[313] R. Li and Z. Liu, "Stress detection using deep neural networks," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 11, pp. 1–10, 2020.

[314] A. D. McDonald, F. Sasangohar, A. Jatav, and A. H. Rao, "Continuous monitoring and detection of post-traumatic stress disorder (PTSD) triggers among veterans: a supervised machine learning approach," *IISE Trans. Healthc. Syst. Eng.*, vol. 9, no. 3, pp. 201–211, 2019.

[315] D. Leightley, V. Williamson, J. Darby, and N. T. Fear, "Identifying probable post-traumatic stress disorder: applying supervised machine learning to data from a UK military cohort," *J. Ment. Heal.*, vol. 28, no. 1, pp. 34–41, 2019.

[316] K. M. Dalmeida and G. L. Masala, "Hrv features as viable physiological markers for stress detection using wearable devices," *Sensors*, vol. 21, no. 8, p. 2873, 2021.

[317] K. Wang and P. Guo, "An ensemble classification model with unsupervised representation learning for driving stress recognition using physiological signals," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3303–3315, 2020.

[318] E. Vildjiounaite, J. Kallio, J. Mäntyjärvi, V. Kyllönen, M. Lindholm, and G. Gimel'farb, "Unsupervised stress detection algorithm and experiments with real life data," in *EPIA Conference on Artificial Intelligence*, 2017, pp. 95–107.

[319] F. Larradet, R. Niewiadomski, G. Barresi, D. G. Caldwell, and L. S. Mattos, "Toward emotion recognition from physiological signals in the wild: approaching the methodological issues in real-life data collection," *Front. Psychol.*, vol. 11, p. 1111, 2020.

[320] P. Adams *et al.*, "Towards personal stress informatics: comparing minimally invasive

techniques for measuring daily stress in the wild," in *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, 2014, pp. 72–79.

[321] C. Maaoui and A. Pruski, "Unsupervised stress detection from remote physiological signal," in *2018 IEEE International Conference on Industrial Technology (ICIT)*, 2018, pp. 1538–1543.

[322] G. Rescioa, A. Leonea, and P. Sicilianoa, "Unsupervised-based framework for aged worker's stress detection," *Work. Artif. Intell. an Ageing Soc.*, 2020.

[323] J. Ramos, J.-H. Hong, and A. K. Dey, "Stress Recognition-A Step Outside the Lab.," in *PhyCS*, 2014, pp. 107–118.

[324] L. Fiorini, G. Mancioppi, F. Semeraro, H. Fujita, and F. Cavallo, "Unsupervised emotional state classification through physiological parameters for social robotics applications," *Knowledge-Based Syst.*, vol. 190, p. 105217, 2020.

[325] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science (80-.).*, vol. 315, no. 5814, pp. 972–976, 2007.

[326] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: an efficient data clustering method for very large databases," *ACM sigmod Rec.*, vol. 25, no. 2, pp. 103–114, 1996.

[327] D. Sculley, "Web-scale k-means clustering," in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 1177–1178.

[328] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, and others, "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *kdd*, 1996, vol. 96, no. 34, pp. 226–231.

[329] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: Ordering points to identify the clustering structure," *ACM Sigmod Rec.*, vol. 28, no. 2, pp. 49–60, 1999.

[330] J. A. Healey, "Wearable and automotive systems for affect recognition from physiology," Massachusetts Institute of Technology, 2000.

[331] S. Koldijk, M. Sappelli, S. Verberne, M. A. Neerincx, and W. Kraaij, "The swell knowledge work dataset for stress and user modeling research," in *Proceedings of the 16th international conference on multimodal interaction*, 2014, pp. 291–298.

[332] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215--e220, 2000.

[333] A. Widyanti, A. Johnson, and D. de Waard, "Adaptation of the rating scale mental effort (RSME) for use in Indonesia," *Int. J. Ind. Ergon.*, vol. 43, no. 1, pp. 70–76, 2013.

[334] T.-M. Bynion and M. T. Feldner, "Self-assessment manikin," *Encycl. Personal. Individ. Differ.*, pp. 4654–4656, 2020.

[335] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002.

[336] P. C. Chaitra and R. S. Kumar, "A review of multi-class classification algorithms," *Int. J. Pure Appl. Math*, vol. 118, no. 14, pp. 17–26, 2018.

[337] S. Koldijk, M. A. Neerincx, and W. Kraaij, "Detecting work stress in offices by combining unobtrusive sensors," *IEEE Trans. Affect. Comput.*, vol. 9, no. 2, pp. 227–239, 2016.

[338] M. A. Russo, D. M. Santarelli, and D. O'Rourke, "The physiological effects of slow breathing in the healthy human," *Breathe*, vol. 13, no. 4, pp. 298–309, 2017.

[339] R. Yousefi and M. Nourani, "Separating arterial and venous-related components of photoplethysmographic signals for accurate extraction of oxygen saturation and respiratory rate," *IEEE J. Biomed. Heal. informatics*, vol. 19, no. 3, pp. 848–857, 2014.

[340] L. Al-Ghussain, S. El Bouri, H. Liu, D. Zheng, and others, "Clinical evaluation of stretchable and wearable inkjet-printed strain gauge sensor for respiratory rate monitoring at different measurements locations," *J. Clin. Monit. Comput.*, pp. 1–10, 2020.

[341] N. A. Nayan, R. Jaafar, and N. S. Risman, "Development of respiratory rate estimation technique using electrocardiogram and photoplethysmogram for continuous health monitoring," *Bull. Electr. Eng. Informatics*, vol. 7, no. 3, pp. 487–494, 2018.

[342] R. B. Prasetiyo, K.-S. Choi, and G.-H. Yang, "Design and implementation of respiration rate measurement system using an information filter on an embedded device," *Sensors*, vol. 18, no. 12, p. 4208, 2018.

[343] C. X. Pan, B. C. Palathra, and W. F. Leo-To, "Management of respiratory symptoms in those with serious illness," *Med. Clin.*, vol. 104, no. 3, pp. 455–470, 2020.

[344] J. C. Collins and R. J. Moles, "Management of respiratory disorders and the pharmacist's role: Cough, colds, and sore throats and allergies (including eyes)," *Encycl. Pharm. Pract. Clin. Pharm.*, p. 282, 2019.

[345] N. J. Brendish *et al.*, "Routine molecular point-of-care testing for respiratory viruses in adults

presenting to hospital with acute respiratory illness (ResPOC): a pragmatic, open-label, randomised controlled trial," *Lancet Respir. Med.*, vol. 5, no. 5, pp. 401–411, 2017.

[346] A. D. Bedoya, M. E. Clement, M. Phelan, R. C. Steorts, C. O'Brien, and B. A. Goldstein, "Minimal impact of implemented early warning score and best practice alert for patient deterioration," *Crit. Care Med.*, vol. 47, no. 1, p. 49, 2019.

[347] K. V. Madhav, M. R. Ram, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "Robust extraction of respiratory activity from PPG signals using modified MSPCA," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 5, pp. 1094–1106, 2013.

[348] E. R. S. Sheffield, "The Global Impact of Respiratory Disease," *Forum Int. Respir. Soc.*, vol. Second Edi, 2017.

[349] WHO, "Technical Seminar- Actue Respiratory Infections," 2010.

[350] K. H. Chon, S. Dash, and K. Ju, "Estimation of respiratory rate from photoplethysmogram data using time--frequency spectral estimation," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 2054–2063, 2009.

[351] H. Kim, J.-Y. Kim, and C.-H. Im, "Fast and robust real-time estimation of respiratory rate from photoplethysmography," *Sensors*, vol. 16, no. 9, p. 1494, 2016.

[352] G. Liu, D. Wu, Z. Mei, Q. Zhu, and L. Wang, "Automatic detection of respiratory rate from electrocardiogram, respiration induced plethysmography and 3D acceleration signals," *J. Cent. South Univ.*, vol. 20, no. 9, pp. 2423–2431, 2013.

[353] Q. Qananwah, A. Dagamseh, H. Alquran, K. S. Ibrahim, M. Alodat, and O. Hayden, "A comparative study of photoplethysmogram and piezoelectric plethysmogram signals," *Phys. Eng. Sci. Med.*, vol. 43, no. 4, pp. 1207–1217, 2020.

[354] W. K. Ngui, M. S. Leong, L. M. Hee, and A. M. Abdelrhman, "Wavelet analysis: mother wavelet selection methods," in *Applied mechanics and materials*, 2013, vol. 393, pp. 953–958.

[355] K. T. Sweeney *et al.*, "Identification of sleep apnea events using discrete wavelet transform of respiration, ecg and accelerometer signals," in *2013 IEEE International Conference on Body Sensor Networks*, 2013, pp. 1–6.

[356] P. Dehkordi, A. Garde, B. Molavi, J. M. Ansermino, and G. A. Dumont, "Extracting instantaneous respiratory rate from multiple photoplethysmogram respiratory-induced variations," *Front. Physiol.*, vol. 9, p. 948, 2018.

[357] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multiparameter respiratory rate estimation from the photoplethysmogram," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 1946–1953, 2013.

[358] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "HeartPy: A novel heart rate algorithm for the analysis of noisy signals," *Transp. Res. part F traffic Psychol. Behav.*, vol. 66, pp. 368–378, 2019.

[359] C. Park, H. Shin, and B. Lee, "Blockwise PPG enhancement based on time-variant zero-phase harmonic notch filtering," *Sensors*, vol. 17, no. 4, p. 860, 2017.

[360] D. Luguern *et al.*, "Wavelet Variance Maximization: A contactless respiration rate estimation method based on remote photoplethysmography," *Biomed. Signal Process. Control*, vol. 63, p. 102263, 2021.

[361] R. Katiyar, V. Gupta, and R. B. Pachori, "FBSE-EWT-based approach for the determination of respiratory rate from PPG signals," *IEEE Sensors Lett.*, vol. 3, no. 7, pp. 1–4, 2019.

[362] Q. Zhang, Q. Xie, M. Wang, and G. Wang, "Motion artifact removal for PPG signals based on accurate fundamental frequency estimation and notch filtering," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 2965–2968.

[363] M. A. Motin, C. K. Karmakar, and M. Palaniswami, "Selection of empirical mode decomposition techniques for extracting breathing rate from PPG," *IEEE Signal Process. Lett.*, vol. 26, no. 4, pp. 592–596, 2019.

[364] R. Lei, B. W.-K. Ling, P. Feng, and J. Chen, "Estimation of Heart Rate and Respiratory Rate from PPG Signal Using Complementary Ensemble Empirical Mode Decomposition with both Independent Component Analysis and Non-Negative Matrix Factorization," *Sensors*, vol. 20, no. 11, p. 3238, 2020.

[365] S. Xiao, P. Yang, L. Liu, Z. Zhang, and J. Wu, "Extraction of Respiratory Signals and Respiratory Rates from the Photoplethysmogram," in *EAI International Conference on Body Area Networks*, 2020, pp. 184–198.

[366] H. Yang, M. Li, D. He, X. Che, and X. Qin, "Respiratory Rate Estimation from the Photoplethysmogram Combining Multiple Respiratory-induced Variations Based on SQI," in *2019 IEEE International Conference on Mechatronics and Automation (ICMA)*, 2019, pp. 382–

387.

[367] B. Roy, A. Roy, J. K. Chandra, and R. Gupta, "i-PRExT: Photoplethysmography Derived Respiration Signal Extraction and Respiratory Rate Tracking Using Neural Networks," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2020.

[368] J. Deny, E. Muthukumaran, S. Ramkumar, and S. Kartheesawaran, "Extraction of respiratory signals and motion artifacts from ppg signal using modified multi scale principal component analysis," *Int. J. Pure Appl. Math.*, vol. 119, no. 12, pp. 13719–13727, 2018.

[369] A. Cicone and H.-T. Wu, "How nonlinear-type time-frequency analysis can help in sensing instantaneous heart rate and instantaneous respiratory rate from photoplethysmography in a reliable way," *Front. Physiol.*, vol. 8, p. 701, 2017.

[370] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, K. Sivani, and K. A. Reddy, "ICA-based improved DTCWT technique for MA reduction in PPG signals with restored respiratory information," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 10, pp. 2639–2651, 2013.

[371] H. Kang, "The prevention and handling of the missing data," *Korean J. Anesthesiol.*, vol. 64, no. 5, p. 402, 2013.

[372] M. Peeters, M. Zondervan-Zwijnenburg, G. Vink, and R. de Schoot, "How to handle missing data: A comparison of different approaches," *Eur. J. Dev. Psychol.*, vol. 12, no. 4, pp. 377–394, 2015.

[373] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "Analysing noisy driver physiology real-time using off-the-shelf sensors: Heart rate analysis software from the taking the fast lane project," *J. Open Res. Softw.*, vol. 7, no. 1, 2019.

[374] G. Slapničar, N. Mlakar, and M. Luštrek, "Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network," *Sensors*, vol. 19, no. 15, p. 3420, 2019.

[375] N. Selvaraj, Y. Mendelson, K. H. Shelley, D. G. Silverman, and K. H. Chon, "Statistical approach for the detection of motion/noise artifacts in Photoplethysmogram," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 4972–4975.

[376] M. Elgendi, "Optimal signal quality index for photoplethysmogram signals," *Bioengineering*, vol. 3, no. 4, p. 21, 2016.

[377] H.-T. Wu and E. Z. Soliman, "A new approach for analysis of heart rate variability and QT variability in long-term ECG recording," *Biomed. Eng. Online*, vol. 17, no. 1, pp. 1–14, 2018.

[378] E. J. Argüello-Prada, "The mountaineer's method for peak detection in photoplethysmographic signals," *Rev. Fac. Ing. Univ. Antioquia*, no. 90, pp. 42–50, 2019.

[379] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Trans. audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, 1967.

[380] P. K. Rahi and R. Mehra, "Analysis of power spectrum estimation using welch method for various window techniques," *Int. J. Emerg. Technol. Eng.*, vol. 2, no. 6, pp. 106–109, 2014.

[381] D. J. Meredith, D. Clifton, P. Charlton, J. Brooks, C. W. Pugh, and L. Tarassenko, "Photoplethysmographic derivation of respiratory rate: a review of relevant physiology," *J. Med. Eng. \& Technol.*, vol. 36, no. 1, pp. 1–7, 2012.

[382] M. A. F. Pimentel *et al.*, "Toward a robust estimation of respiratory rate from pulse oximeters," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1914–1923, 2016.

[383] L. Nilsson, A. Johansson, and S. Kalman, "Monitoring of respiratory rate in postoperative care using a new photoplethysmographic technique," *J. Clin. Monit. Comput.*, vol. 16, no. 4, pp. 309–315, 2000.

[384] S. G. Fleming and L. Tarassenko, "A comparison of signal processing techniques for the extraction of breathing rate from the photoplethysmogram," *Int. J. Biol. Med. Sci*, vol. 2, no. 4, pp. 232–236, 2007.

[385] "basic statistics and data presentation," *Basic statistics and data presentation*. The United States Food and Drug Administration is a federal agency of the Department of Health and Human Services, [Online]. Available: https://www.fda.gov/media/73535/download.

[386] R. S. Reis, A. A. Hino, and C. R. Añez, "Perceived stress scale," *J. Heal. Psychol*, vol. 15, pp. 107–114, 2010.

[387] H. \THórarinsdóttir *et al.*, "The validity of daily self-assessed perceived stress measured using smartphones in healthy individuals: cohort study," *JMIR mHealth uHealth*, vol. 7, no. 8, p. e13418, 2019.

[388] T. Iqbal, A. Elahi, W. Wijns, and A. Shahzad, "Exploring Unsupervised Machine Learning Classification Methods for Physiological Stress Detection," *Front. Med. Technol.*, vol. 4, 2022.

[389] F. Rahimi Sardo *et al.*, "Recent Progress of Triboelectric Nanogenerators for Biomedical Sensors: From Design to Application," *Biosensors*, vol. 12, no. 9, pp. 1–25, 2022, doi: 10.3390/bios12090697.

[390] N. El Haouij, J.-M. Poggi, S. Sevestre-Ghalila, R. Ghozi, and M. Ja\"\idane, "AffectiveROAD system and database to assess driver's attention," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, 2018, pp. 800–803.

[391] J. Healey and R. Picard, "SmartCar: detecting driver stress," pp. 218–221, 2002, doi: 10.1109/icpr.2000.902898.

[392] S. Hosseini *et al.*, "A multimodal sensor dataset for continuous stress detection of nurses in a hospital," *Sci. Data*, vol. 9, no. 1, pp. 1–13, 2022.

[393] T. Iqbal, A. Elahi, S. Ganly, W. Wijns, and A. Shahzad, "Photoplethysmography-Based Respiratory Rate Estimation Algorithm for Health Monitoring Applications," *J. Med. Biol. Eng.*, pp. 1–11, 2022.

[394] D. Roshan *et al.*, "A comparison of methods to generate adaptive reference ranges in longitudinal monitoring," *PLoS One*, vol. 16, no. 2, p. e0247338, 2021.

[395] P. S. O'Súilleabháin, B. M. Hughes, A. M. Oommen, L. Joshi, and S. Cunningham, "Vulnerability to stress: Personality facet of vulnerability is associated with cardiovascular adaptation to recurring stress," *Int. J. Psychophysiol.*, vol. 144, pp. 34–39, 2019.

[396] F. Scarpina and S. Tagini, "The stroop color and word test," *Front. Psychol.*, vol. 8, p. 557, 2017.

[397] E. C. Helminen, M. L. Morton, Q. Wang, and J. C. Felver, "Stress reactivity to the trier social stress test in traditional and virtual environments: a meta-analytic comparison," *Psychosom. Med.*, vol. 83, no. 3, pp. 200–211, 2021.

[398] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *J. Health Soc. Behav.*, pp. 385–396, 1983.

[399] E.-H. Lee, "Review of the psychometric evidence of the perceived stress scale," *Asian Nurs. Res. (Korean. Soc. Nurs. Sci).*, vol. 6, no. 4, pp. 121–127, 2012.

[400] C. D. Spielberger, R. Gorsuch, R. Lushene, P. Vagg, and G. Jacobs, "Manual for the Stait-Trait Anxiety Inventory Consulting Psychologists Press: Palo Alto," *CA, USA*, 1983.

[401] R. D. Riley *et al.*, "Calculating the sample size required for developing a clinical prediction

model," *Bmj*, vol. 368, 2020.

[402] T. Vallès-Català, A. Pedret, D. Ribes, D. Medina, and M. Traveria, "Effects of stress on performance during highly demanding tasks in student pilots," *Int. J. Aerosp. Psychol.*, vol. 31, no. 1, pp. 43–55, 2021.

[403] V. Chandra, A. Priyarup, and D. Sethia, "Comparative Study of Physiological Signals from Empatica E4 Wristband for Stress Classification," in *International Conference on Advances in Computing and Data Sciences*, 2021, pp. 218–229.

[404] M. Kim, J. Kim, K. Park, H. Kim, and D. Yoon, "Comparison of Wristband Type Devices to Measure Heart Rate Variability for Mental Stress Assessment," in *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, 2021, pp. 766–768.

[405] A. Giorgi *et al.*, "Wearable technologies for mental workload, stress, and emotional state assessment during working-like tasks: A comparison with laboratory technologies," *Sensors*, vol. 21, no. 7, p. 2332, 2021.

[406] A. A. T. Schuurmans *et al.*, "Validity of the Empatica E4 Wristband to Measure Heart Rate Variability (HRV) Parameters: a Comparison to Electrocardiography (ECG)," *J. Med. Syst.*, vol. 44, no. 11, 2020, doi: 10.1007/s10916-020-01648-w.

[407] "E4 wristband data." 2022, [Online]. Available: https://support.empatica.com/hc/en-us/sections/200582445-E4-wristband-data.

[408] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley \& Sons, 2015.

[409] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Futur. Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, 2013.

[410] F. S. Collins and H. Varmus, "A new initiative on precision medicine," *N. Engl. J. Med.*, vol. 372, no. 9, pp. 793–795, 2015.

[411] M. Hermann, T. Pentek, and B. Otto, "Design principles for industrie 4.0 scenarios," in *2016 49th Hawaii international conference on system sciences (HICSS)*, 2016, pp. 3928–3937.

[412] B. D. Fulcher, M. A. Little, and N. S. Jones, "Highly comparative time-series analysis: the empirical structure of time series and their methods," *J. R. Soc. Interface*, vol. 10, no. 83, p. 20130048, 2013.

[413] J. Wiens, E. Horvitz, and J. Guttag, "Patient risk stratification for hospital-associated c. diff as a time-series classification task," *Adv. Neural Inf. Process. Syst.*, vol. 25, 2012.

[414] M. Christ, F. Kienle, and A. W. Kempa-Liehr, "Time series analysis in industrial applications," 2016.

[415] M. M. Saad, T. Iqbal, H. Ali, M. F. Bulbul, S. Khan, and C. Tanougast, "Incident Detection over Unified Threat Management platform on a cloud network," in *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, 2019, vol. 2, pp. 592–596.

[416] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series feature extraction on basis of scalable hypothesis tests (tsfresh--a python package)," *Neurocomputing*, vol. 307, pp. 72–77, 2018.

[417] B. D. Fulcher, "Feature-based time-series analysis," in *Feature engineering for machine learning and data analytics*, CRC Press, 2018, pp. 87–116.

[418] M. Christ, A. W. Kempa-Liehr, and M. Feindt, "Distributed and parallel time series feature extraction for industrial big data applications," *arXiv Prepr. arXiv1610.07717*, 2016.

[419] M. A. Rassam, M. A. Maarof, and A. Zainal, "Adaptive and online data anomaly detection for wireless sensor systems," *Knowledge-Based Syst.*, vol. 60, pp. 44–57, 2014.

[420] A. Fawzy, H. M. O. Mokhtar, and O. Hegazy, "Outliers detection and classification in wireless sensor networks," *Egypt. Informatics J.*, vol. 14, no. 2, pp. 157–164, 2013.

[421] G. Jäger *et al.*, "Assessing neural networks for sensor fault detection," in *2014 IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA)*, 2014, pp. 70–75.

[422] G. R. Abuaitah and B. Wang, "Data-centric anomalies in sensor network deployments: Analysis and detection," in *2012 IEEE 9th International Conference on Mobile Ad-Hoc and Sensor Systems (MASS 2012)*, 2012, pp. 1–6.

[423] A. Rahman, D. V Smith, and G. Timms, "A novel machine learning approach toward quality assessment of sensor data," *IEEE Sens. J.*, vol. 14, no. 4, pp. 1035–1047, 2013.

[424] Z. Ouyang, X. Sun, and D. Yue, "Hierarchical time series feature extraction for power consumption anomaly detection," in *Advanced Computational Methods in Energy, Power, Electric Vehicles, and Their Integration*, Springer, 2017, pp. 267–275.

[425] W. Zhang, X. Dong, H. Li, J. Xu, and D. Wang, "Unsupervised detection of abnormal electricity consumption behavior based on feature engineering," *IEEE Access*, vol. 8, pp. 55483–55500, 2020.

[426] G. Liu, L. Li, L. Zhang, Q. Li, and S. S. Law, "Sensor faults classification for SHM systems using deep learning-based method with Tsfresh features," *Smart Mater. Struct.*, vol. 29, no. 7, p. 75005, 2020.

[427] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Stat.*, pp. 1165–1188, 2001.

[428] S. Simmons, L. Jarvis, D. Dempsey, and A. W. Kempa-Liehr, "Data Mining on Extremely Long Time-Series," in *2021 International Conference on Data Mining Workshops (ICDMW)*, 2021, pp. 1057–1066.

[429] T. Iqbal *et al.*, "Stress Monitoring Using Wearable Sensors: A Pilot Study and Stress-Predict Dataset," *Sensors*, vol. 22, no. 21, p. 8135, 2022, doi: 10.3390/s22218135.

[430] B. D. Fulcher and N. S. Jones, "Highly comparative feature-based time-series classification," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 12, pp. 3026–3037, 2014.

[431] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, no. Mar, pp. 1157–1182, 2003.

[432] "Release v0.11.0 · blue-yonder/tsfresh," *GitHub*. Accessed: Dec. 08, 2022. [Online]. Available: https://github.com/blue-yonder/tsfresh/releases/tag/v0.11.0.

[433] "Overview on extracted features - tsfresh," *GitHub*. Accessed: Dec. 08, 2022. [Online]. Available: http://tsfresh.readthedocs.io/en/latest/text/list_of_features.html.

[434] I. T. Jollife and J. Cadima, "Principal component analysis: A review and recent developments," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0202.

[435] M. Vettoretti and B. Di Camillo, "A variable ranking method for machine learning models with correlated features: In-silico validation and application for diabetes prediction," *Appl. Sci.*, vol. 11, no. 16, 2021, doi: 10.3390/app11167740.

[436] L. Toloşi and T. Lengauer, "Classification with correlated features: Unreliability of feature ranking and solutions," *Bioinformatics*, vol. 27, no. 14, pp. 1986–1994, 2011, doi: 10.1093/bioinformatics/btr300.

[437] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*, Springer, 2009, pp. 1–4.

[438] F. Z. Okwonu, B. L. Asaju, and F. I. Arunaye, "Breakdown analysis of pearson correlation coefficient and robust correlation methods," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 917, no. 1, p. 12065.

[439] M. Lobo and R. D. Guntur, "Spearman's rank correlation analysis on public perception toward health partnership projects between Indonesia and Australia in East Nusa Tenggara Province," in *Journal of Physics: Conference Series*, 2018, vol. 1116, no. 2, p. 22020.

[440] J. Hauke and T. Kossowski, "Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data," *Quaest. Geogr.*, vol. 30, no. 2, p. 87, 2011.

[441] K. H. Hamed, "The distribution of Kendall's tau for testing the significance of cross-correlation in persistent data," *Hydrol. Sci. J.*, vol. 56, no. 5, pp. 841–853, 2011.

[442] M.-T. Puth, M. Neuhäuser, and G. D. Ruxton, "Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits," *Anim. Behav.*, vol. 102, pp. 77–84, 2015.

[443] M. M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi Med. J.*, vol. 24, no. 3, pp. 69–71, 2012.

[444] G. Vos, K. Trinh, Z. Sarnyai, and M. R. Azghadi, "Machine Learning for Stress Monitoring from Wearable Devices: A Systematic Literature Review," *arXiv Prepr. arXiv2209.15137*, 2022.

[445] P. T. Noi and M. Kappas, "Comparison of random forest," *K-Nearest Neighbor Support Vector Mach. Classif. L. Cover Classif. Using Sentin. Imagery, MDPI*, 2018.

[446] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A comparative analysis of logistic regression, random forest and KNN models for the text classification," *Augment. Hum. Res.*, vol. 5, pp. 1–16, 2020.

[447] "Feature extraction package - tsfresh," *GitHub*. Accessed: Dec. 08, 2022. [Online]. Available: https://tsfresh.readthedocs.io/en/latest/api/tsfresh.feature_extraction.html.

[448] B. S. Oken, I. Chamine, and W. Wakeland, "A systems approach to stress, stressors and resilience in humans," *Behav. Brain Res.*, vol. 282, pp. 144–154, 2015.

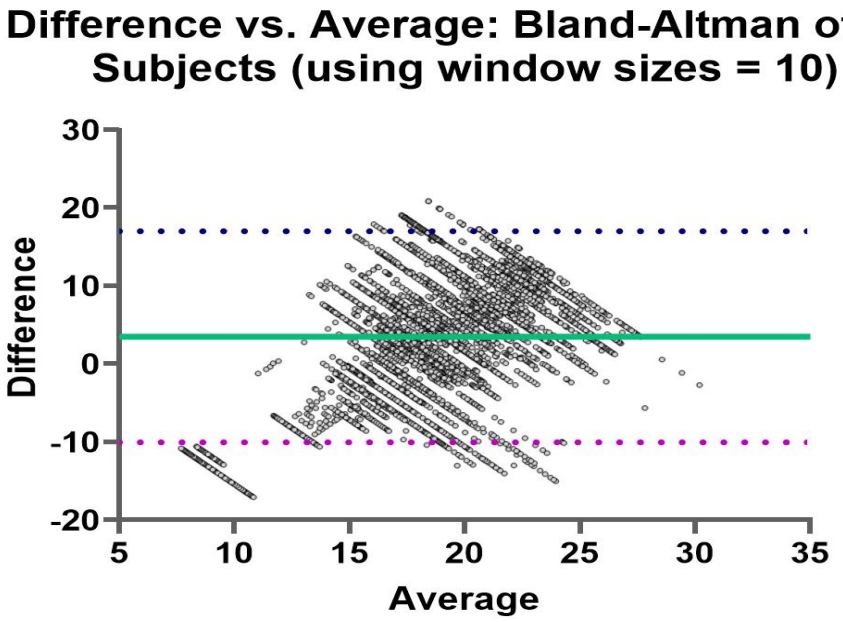
Appendix

10.1 Chapter 3: A Sensitivity Analysis of Biophysiological Responses of Stress for Wearable Sensors in Connected Health

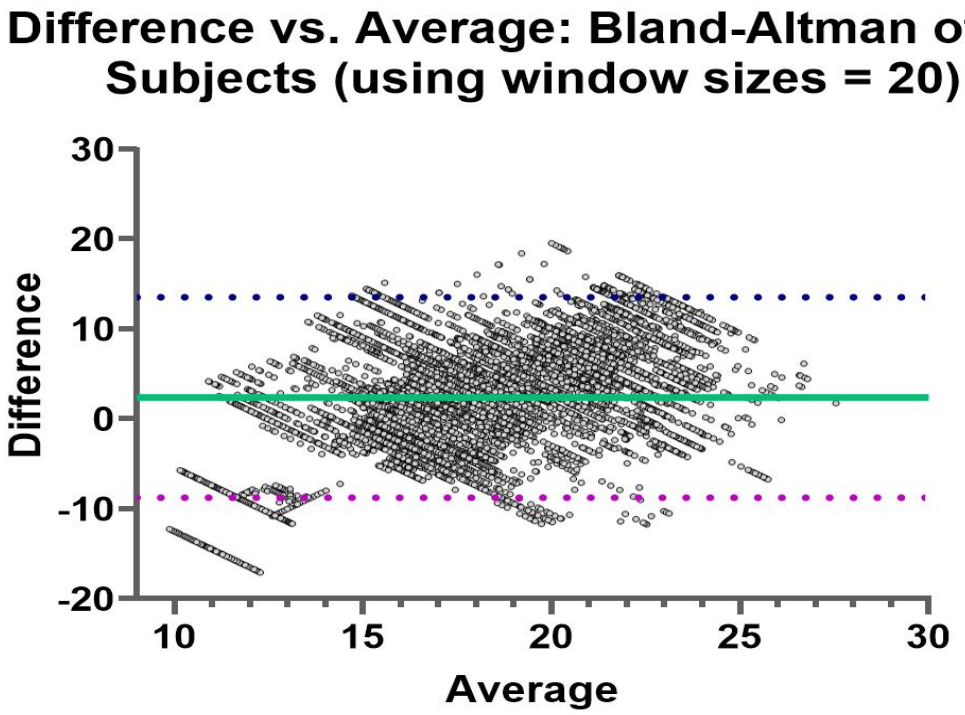
Table Chapter 3 S1. Logistic regression classifier results (70-30 Split)						
Features	**Test- Train Split**	**Accuracy (%)**	**Sensitivity**	**Specificity**	**LR+**	**LR-**
EMG	30-70 %	64.91	1.00	0.00	1.00	inf
EDA		63.82	0.93	0.09	1.03	0.73
RspR		77.21	0.87	0.59	2.11	0.22
RRI		76.40	0.91	0.49	1.78	0.18
HR		77.10	0.89	0.56	2.12	0.21
HRV		66.73	1.00	0.51	2.12	0.00
HR +RspR		83.82	0.90	0.72	3.25	0.14
HR+ RspR+ HRV		88.88	1.00	0.83	6.00	0.00
All combined		88.90	1.00	0.86	7.00	0.00

Table Chapter 3 S2. Logistic regression classifier results (4-fold Cross Validation)										
Features	**Test-Train Split**	**Accuracy (%)**	**Sensitivity**	**Specificity**	**95% Confidence Intervals of sensitivity and specificity**		**LR+**	**LR-**	**Variance**	**Standard deviance**
					Lower	Upper				
EMG	4-fold Cross Validation	64.96	1.00	0.00	[1.00, 0.00]	[1.00, 0.00]	1.0	Inf	0.00	0.00
EDA		59.23	0.87	0.88	[0.86, 0.08]	[0.86, 0.088]	0.95	1.55	0.01	0.10
RspR		77.00	0.87	0.59	[0.87, 0.59]	[0.87, 0.59]	2.11	0.22	0.00	0.01
RRI		71.37	0.85	0.47	[0.75, 0.49]	[0.75, 0.48]	1.59	0.33	0.01	0.10
HR		73.55	0.85	0.53	[0.85, 0.53]	[0.85, 0.53]	1.80	0.29	0.01	0.09
HR +RspR		79.89	0.86	0.69	[0.86, 0.69]	[0.86, 0.69]	2.79	0.21	0.00	0.06
HR+ RspR +HRV		82.14	0.79	0.86	[0.57, 0.67]	[1.00, 1.04]	5.51	0.25	0.03	0.18
All combined		82.14	0.79	0.86	[0.57, 0.67]	[1.00, 1.04]	5.51	0.25	0.03	0.18

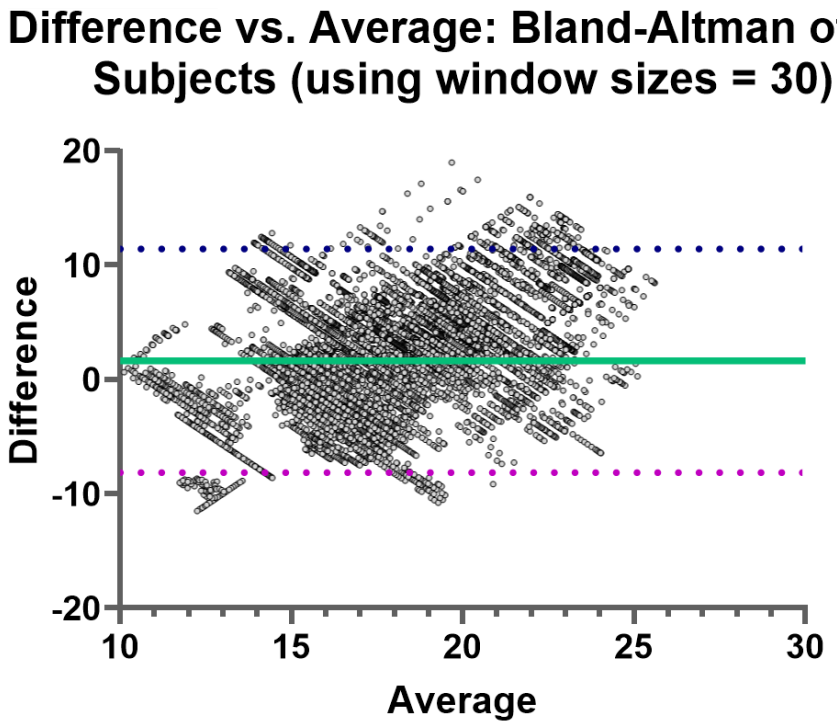
10.2 Chapter 5: Photoplethysmography-Based Respiratory Rate Estimation Algorithm for Health Monitoring Applications



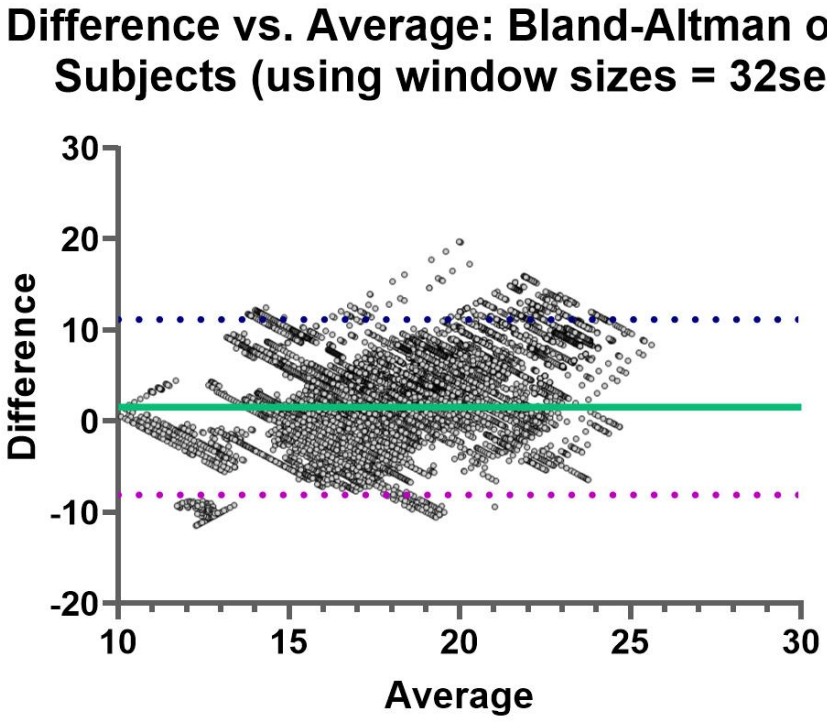
Chapter 5 S 1. *Bland-Altman Plot of respiratory rate estimation using 10 seconds window size*



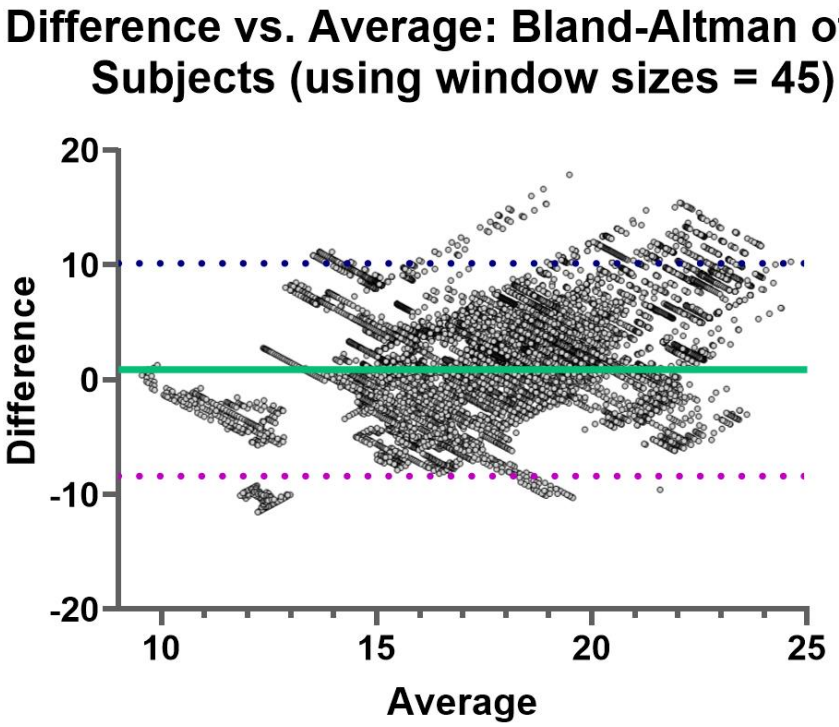
Chapter 5 S 2. *Bland-Altman Plot of respiratory rate estimation using 20 seconds window size*



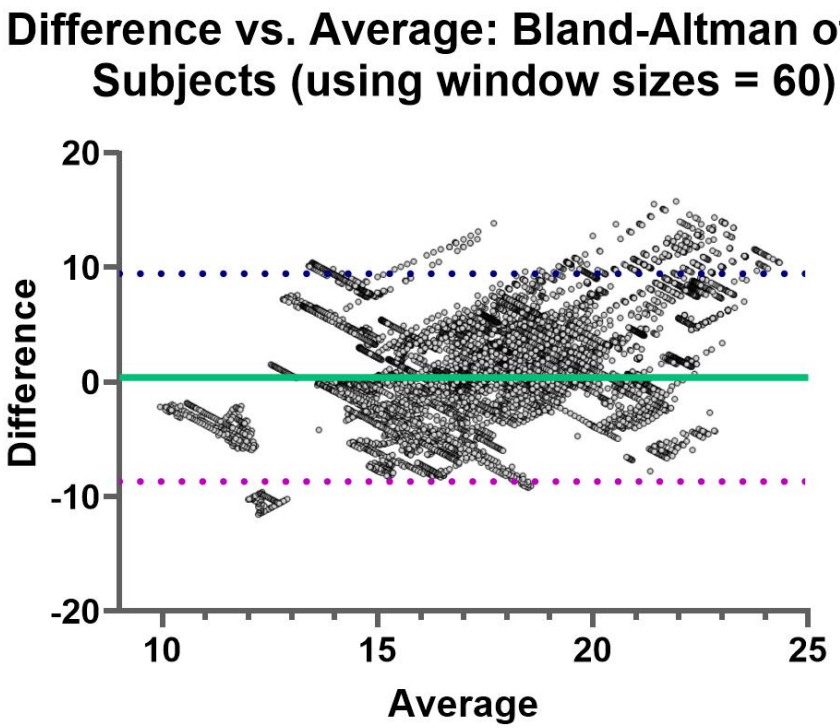
Chapter 5 S 3. *Bland-Altman Plot of respiratory rate estimation using 30 seconds window size*



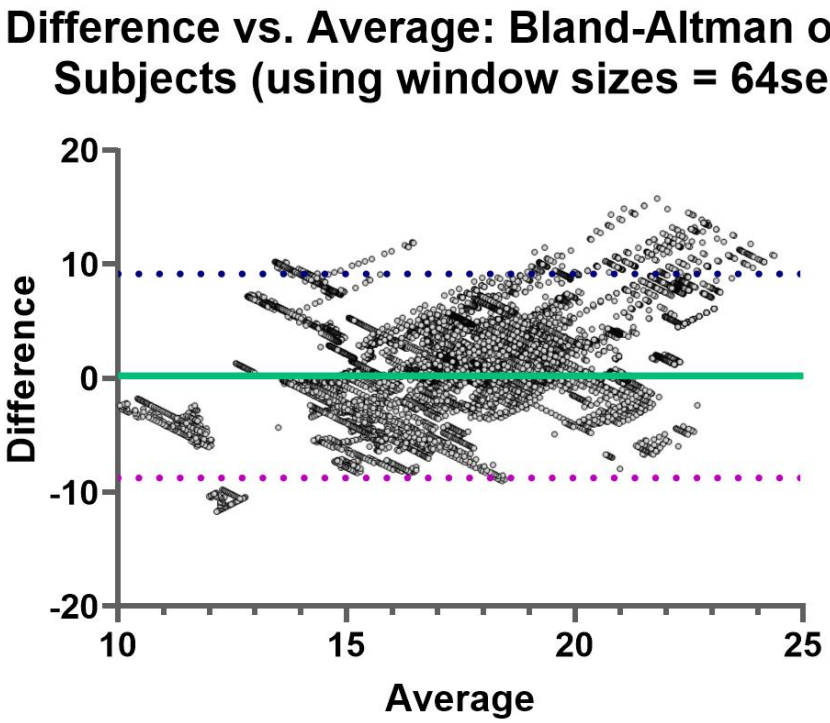
Chapter 5 S 4. *Bland-Altman Plot of respiratory rate estimation using 32 seconds window size*



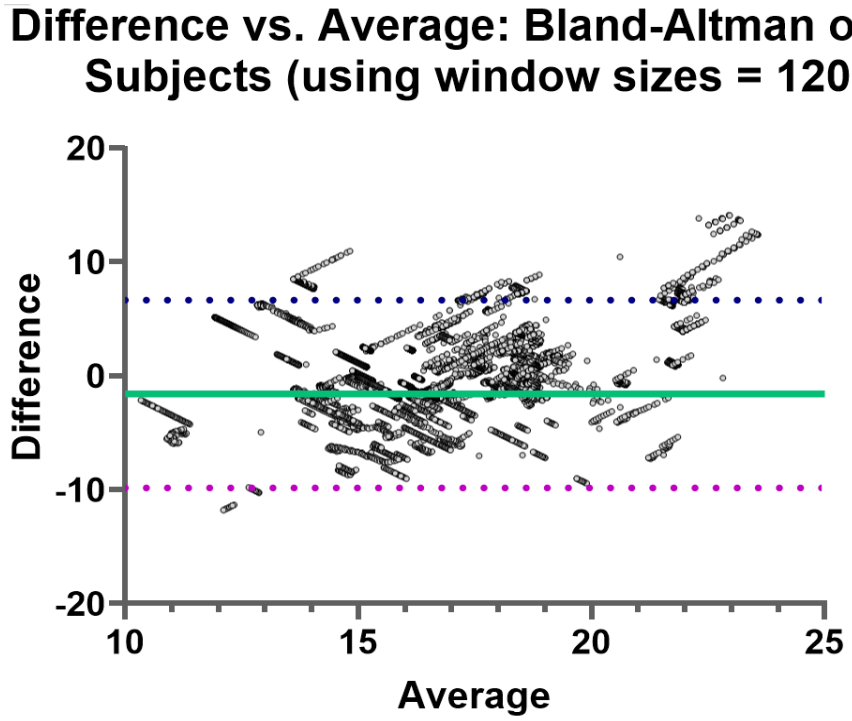
Chapter 5 S 5. *Bland-Altman Plot of respiratory rate estimation using 45 seconds window size*



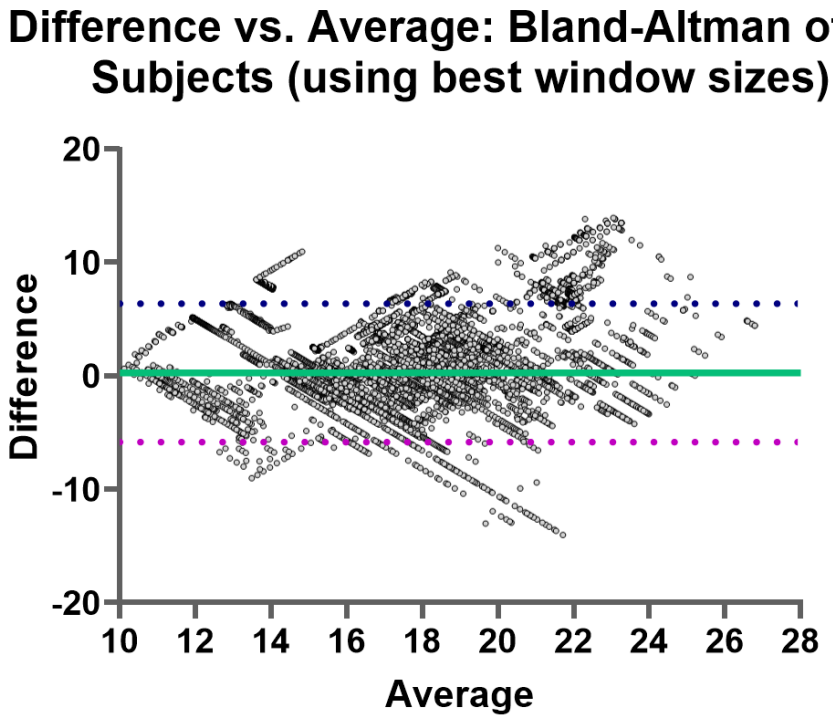
Chapter 5 S 6. *Bland-Altman Plot of respiratory rate estimation using 60 seconds window size*



Chapter 5 S 7. *Bland-Altman Plot of respiratory rate estimation using 64 seconds window size*



Chapter 5 S 8. *Bland-Altman Plot of respiratory rate estimation using 120 seconds window size*



Chapter 5 S 9. *Bland-Altman Plot of respiratory rate estimation using best window size for each subject*

Table Chapter 5 A1. Bland-Altman plot: Bias values along with upper and lower limits of agreement				
Window Size	**Bias Value**	**Standard Deviation of Bias**	**Limit of agreement**	
			Lower	**Upper**
10	3.49	6.90	-10.03	17.01
20	2.38	5.69	-8.77	13.52
30	1.62	4.99	-8.06	11.40
32	1.55	4.91	-8.07	11.17
45	0.87	4.72	-8.38	10.13
60	0.38	4.63	-8.68	9.45
64	0.21	4.56	-8.73	9.15
90	-0.45	4.47	-9.20	8.31
120	-1.60	4.20	-9.83	6.64
Best window sizes	0.25	3.11	-5.84	6.35
Note: The negative bias value indicates the average reference respiratory rate was higher than the average estimated respiratory rate.				

Table Chapter 5 S1. Error in Respiratory Rate Estimation using Different Window Sizes for each subject

S1	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.23	6.39	1.78	0.84	**0.58**	1.11	2.37	1.62	0.64
RMSE	7.31	6.62	2.24	1.14	**0.79**	1.68	3.08	2.15	0.93
S2	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**

MAE	7.95	3.30	1.76	0.88	0.50	0.68	1.34	1.56	**0.43**
RMSE	8.12	3.45	1.92	1.00	0.63	0.91	1.66	1.69	**0.57**
S3	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.41	2.16	0.61	1.15	1.73	2.64	3.46	**0.60**	1.90
RMSE	7.49	2.52	0.88	1.27	1.83	2.69	3.56	**0.79**	1.99
S4	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.11	5.28	4.14	3.07	2.45	**1.83**	2.48	3.85	2.25
RMSE	7.83	5.85	5.00	3.76	3.27	**2.61**	2.64	4.64	3.05
S5	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	12.14	8.26	7.12	6.39	5.44	4.42	**3.40**	7.00	5.19
RMSE	12.38	8.63	7.50	6.85	5.73	4.72	**4.00**	7.38	5.42
S6	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	3.94	**1.39**	3.06	4.20	4.78	5.46	6.04	3.29	4.91
RMSE	4.00	**1.46**	3.09	4.22	4.80	5.47	6.06	3.31	4.93
S7	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE		8.15	6.50	5.32	4.58	3.29	**3.06**	6.26	4.36
RMSE		8.17	6.53	5.37	4.72	3.83	**3.28**	6.29	4.54
S8	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE		4.66	3.20	2.12	1.54	1.59	2.56	2.96	**1.48**
RMSE		4.72	3.27	2.21	1.63	1.66	2.98	3.03	**1.54**
S9	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE		8.51	6.78	5.51	4.71	3.35	**3.11**	6.54	4.49
RMSE		8.54	6.82	5.57	4.85	3.95	**3.31**	6.58	4.66
S10	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.51	6.09	4.37	3.12	2.34	**1.61**	1.96	4.08	2.08
RMSE	8.56	6.43	4.67	3.33	2.53	**1.74**	2.01	4.33	2.27
S11	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.06	2.80	1.52	**1.00**	1.17	1.90	2.40	1.35	1.30
RMSE	7.41	3.08	2.00	**1.68**	1.86	2.45	2.83	1.87	1.97
S12	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**1.75**	3.54	4.89	5.96	6.46	7.18	7.65	5.11	6.56
RMSE	**2.31**	3.86	5.17	6.14	6.61	7.26	7.73	5.37	6.70
S14	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	5.40	4.58	3.24	2.29	**1.70**	2.24	3.12	3.07	1.73

RMSE	6.44	5.74	4.30	3.10	**2.31**	2.56	3.83	4.11	2.24
S15	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	5.40	1.70	1.17	1.36	1.54	1.68	1.83	**1.15**	1.56
RMSE	5.76	2.23	1.61	1.61	1.81	1.83	1.97	**1.53**	1.81
S16	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**1.14**	3.11	4.42	5.30	5.76	6.31	6.69	4.60	5.87
RMSE	**1.62**	3.21	4.45	5.32	5.77	6.32	6.71	4.62	5.88
S17	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	9.01	7.82	6.35	4.76	4.07	2.94	**1.88**	6.04	3.90
RMSE	9.14	8.17	6.65	4.95	4.26	3.23	**2.47**	6.31	4.10
S18	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**6.97**	9.39	10.19	10.72	10.91	10.98	10.79	10.29	10.93
RMSE	**7.04**	9.41	10.20	10.73	10.92	11.00	10.83	10.30	10.94
S19	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	6.96	2.15	**2.12**	3.00	3.78	4.35	4.43	2.29	3.87
RMSE	7.80	2.77	**2.62**	3.35	4.01	4.54	4.60	2.74	4.08
S20	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	11.39	7.46	5.78	4.60	3.89	2.98	**1.89**	5.55	3.74
RMSE	11.50	7.68	5.93	4.70	3.96	3.15	**2.39**	5.69	3.80
S21	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**2.18**	2.37	3.11	3.84	4.26	4.90	5.56	3.22	4.37
RMSE	**2.96**	2.75	3.49	4.10	4.46	5.05	5.68	3.59	4.56
S22	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**2.37**	4.17	5.81	6.90	7.48	8.04	8.32	6.03	7.59
RMSE	**3.05**	4.22	5.84	6.92	7.49	8.05	8.33	6.05	7.60
S23	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE		14.07	7.88	6.73	7.06	6.98	**4.24**	6.97	6.83
RMSE		14.11	8.34	7.84	8.02	7.36	**5.08**	8.08	7.73
S24	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	3.80	**1.14**	1.40	2.20	2.81	3.82	5.15	1.44	3.00
RMSE	4.33	**1.54**	1.67	2.34	2.91	3.95	5.44	1.68	3.09
S25	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	3.23	**0.85**	1.64	2.51	3.00	3.43	3.84	1.80	3.09

RMSE	3.85	**1.14**	1.88	2.65	3.09	3.51	3.93	2.01	3.18
S26	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	13.84	9.62	7.51	5.59	5.52	**4.31**	4.59	6.63	5.10
RMSE	13.99	9.88	7.98	7.26	7.10	**5.77**	5.59	7.66	6.75
S27	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	16.34	11.48	9.96	9.09	8.43	7.59	**6.58**	9.88	8.21
RMSE	16.41	11.56	10.05	9.22	8.55	7.78	**6.86**	10.00	8.33
S28	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	6.66	4.99	3.62	2.07	1.39	**1.25**	1.92	3.27	1.23
RMSE	6.89	5.57	4.12	2.34	1.60	**1.35**	2.10	3.76	1.46
S29	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	6.21	4.37	2.83	1.93	1.29	**1.12**	2.02	2.92	1.15
RMSE	6.98	4.90	3.28	2.48	1.74	**1.45**	2.22	3.53	1.61
S30	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.42	3.98	2.34	1.43	1.05	1.00	1.86	2.14	**0.95**
RMSE	7.76	4.31	2.72	1.71	1.35	1.37	2.53	2.52	**1.24**
S31	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	8.71	3.69	1.96	0.92	0.54	0.73	1.72	1.72	**0.51**
RMSE	8.81	3.86	2.17	1.11	0.67	0.96	2.04	1.93	**0.60**
S32	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	2.62	**2.53**	3.94	5.16	5.91	6.80	7.78	4.17	6.09
RMSE	2.89	**2.68**	4.10	5.23	5.96	6.87	7.92	4.32	6.13
S34	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	10.10	5.93	4.13	2.93	2.32	**1.46**	1.59	3.84	2.16
RMSE	10.39	6.17	4.39	3.18	2.56	**1.84**	1.68	4.07	2.40
S35	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	6.52	3.01	1.91	1.11	**0.96**	1.20	1.63	1.73	0.98
RMSE	6.90	3.66	2.83	1.75	**1.31**	1.39	1.98	2.64	1.27
S36	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.98	4.79	2.76	1.38	**0.70**	0.72	1.49	2.48	0.60
RMSE	8.09	4.89	2.88	1.56	**0.95**	0.89	1.87	2.61	0.83
S37	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.36	2.91	1.26	**0.72**	0.90	1.54	2.55	1.08	1.01

RMSE	7.52	3.32	1.66	**0.86**	1.10	1.94	3.01	1.46	1.23
S38	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.74	6.40	5.20	3.57	2.67	1.57	**1.06**	4.85	2.41
RMSE	9.00	6.99	6.03	4.16	3.16	1.84	**1.24**	5.70	2.83
S39	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	10.24	8.77	7.03	5.74	5.05	4.13	**2.94**	6.77	4.91
RMSE	11.12	9.10	7.36	5.93	5.16	4.33	**3.38**	7.07	5.04
S40	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	3.53	3.60	3.80	2.34	1.72	1.70	2.77	3.72	**1.58**
RMSE	4.46	4.83	5.11	3.05	2.30	2.06	2.97	4.97	**2.12**
S41	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	9.30	6.02	4.55	4.57	4.68	5.27	**4.30**	4.32	4.67
RMSE	9.88	6.48	5.24	5.42	5.62	6.10	**4.87**	5.02	5.64
S42	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	1.59	1.09	0.90	0.75	**0.73**	0.96	1.43	0.86	0.75
RMSE	1.81	1.22	1.01	0.84	**0.79**	1.00	1.58	0.94	0.80
S43	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.59	3.45	1.99	1.20	**1.02**	1.45	2.27	1.77	1.02
RMSE	7.92	4.03	2.46	1.44	**1.34**	1.88	2.65	2.21	1.38
S44	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**1.35**	1.47	1.64	1.74	1.68	1.83	1.49	1.60	1.64
RMSE	**2.05**	2.58	2.99	3.19	2.84	2.79	1.91	2.91	2.74
S45	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	6.15	8.79	9.07	8.33	8.31	**5.27**	5.78	8.49	7.23
RMSE	7.55	9.62	9.76	9.34	9.26	**7.75**	6.57	9.35	8.61
S46	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**3.99**	4.24	4.32	4.24	4.06	4.16	4.56	4.33	4.05
RMSE	**5.22**	5.29	5.27	5.05	4.63	4.49	4.69	5.26	4.57
S47	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**2.62**	3.16	3.40	3.57	3.84	4.37	5.30	3.46	3.94
RMSE	**2.94**	3.28	3.46	3.61	3.87	4.48	5.53	3.50	3.97
S48	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	2.69	2.66	**2.70**	2.71	2.80	3.15	3.72	2.71	2.88

RMSE	3.14	3.04	**3.03**	3.00	3.08	3.51	4.12	3.04	3.18
S49	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	8.33	6.80	4.98	3.57	2.81	**1.47**	1.66	4.73	2.51
RMSE	9.34	7.06	5.28	3.88	3.11	**1.87**	1.72	5.02	2.85
S50	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	12.08	10.33	8.67	8.11	7.50	6.49	**4.79**	8.98	7.34
RMSE	12.12	10.45	8.83	8.34	7.69	6.77	**5.38**	9.24	7.54
S51	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE		7.51	9.64	8.35	7.54	6.11	**3.88**	9.39	7.30
RMSE		8.20	9.67	8.40	7.67	6.59	**5.26**	9.42	7.47
S52	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**4.91**	5.16	5.27	5.37	5.41	5.52	5.72	5.30	5.43
RMSE	**4.96**	5.18	5.29	5.37	5.41	5.53	5.74	5.31	5.44
S53	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	2.88	2.10	1.77	1.83	1.96	2.02	2.48	**1.77**	2.03
RMSE	3.89	2.94	2.25	2.13	2.21	2.24	2.82	**2.19**	2.26

Table Chapter 5 S2. Error in Respiratory Rate Estimation using Different Window Sizes for each subject (removing points with SQI = nan)

S1	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.23	6.37	1.74	0.84	**0.61**	1.18	2.46	1.62	0.64
RMSE	7.31	6.60	2.22	1.13	**0.83**	1.77	3.16	2.15	0.93
S2	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.95	3.30	1.76	0.88	0.50	0.68	1.34	1.56	**0.43**
RMSE	8.12	3.45	1.92	1.00	0.63	0.91	1.66	1.69	**0.57**
S3	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.41	2.16	0.61	1.15	1.73	2.64	3.46	**0.60**	1.90
RMSE	7.49	2.52	0.88	1.27	1.83	2.69	3.56	**0.79**	1.99
S4	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.11	5.28	4.14	3.07	2.45	**1.83**	2.48	3.85	2.25
RMSE	7.83	5.85	5.00	3.76	3.27	**2.61**	2.64	4.64	3.05
S5	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	12.14	8.01	6.64	5.78	5.11	4.42	**3.40**	6.55	5.01
RMSE	12.38	8.26	6.83	6.04	5.31	4.72	**4.00**	6.76	5.21
S6	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	3.94	**1.39**	3.06	4.20	4.78	5.46	6.04	3.29	4.91
RMSE	4.00	**1.46**	3.09	4.22	4.80	5.47	6.06	3.31	4.93
S7	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE		8.15	6.50	5.32	4.58	3.29	**3.06**	6.26	4.36
RMSE		8.17	6.53	5.37	4.72	3.83	**3.28**	6.29	4.54
S8	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE		4.66	3.20	2.12	1.54	1.59	2.56	2.96	**1.48**
RMSE		4.72	3.27	2.21	1.63	1.66	2.98	3.03	**1.54**
S9	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE		8.51	6.78	5.51	4.71	3.35	**3.11**	6.54	4.49
RMSE		8.54	6.82	5.57	4.85	3.95	**3.31**	6.58	4.66
S10	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.88	5.92	4.15	3.04	2.41	1.81	**1.55**	3.95	2.27
RMSE	8.81	6.22	4.39	3.23	2.60	1.90	**1.56**	4.17	2.44
S11	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**

MAE	7.06	2.80	1.52	**1.00**	1.17	1.90	2.40	1.35	1.30
RMSE	7.41	3.08	2.00	**1.68**	1.86	2.45	2.83	1.87	1.97
S12	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**1.75**	3.54	4.89	5.96	6.46	7.18	7.65	5.11	6.56
RMSE	**2.31**	3.86	5.17	6.14	6.61	7.26	7.73	5.37	6.70
S14	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	5.58	**2.33**	2.54					2.83	
RMSE	6.42	**3.00**	3.30					3.56	
S15	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	5.44	1.70	1.15	1.33	1.52	1.71	1.91	**1.15**	1.56
RMSE	5.82	2.21	1.59	1.59	1.79	1.86	2.06	**1.53**	1.82
S16	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**1.14**	3.16	4.44	5.31	5.76	6.31	6.69	4.62	5.87
RMSE	**1.62**	3.25	4.47	5.33	5.77	6.32	6.71	4.64	5.88
S17	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	9.03	7.82	6.42	4.83	4.07	2.94	**1.88**	6.12	3.90
RMSE	9.17	8.18	6.73	5.01	4.26	3.23	**2.47**	6.39	4.10
S18	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**6.97**	9.39	10.19	10.72	10.91	10.98	10.79	10.29	10.93
RMSE	**7.04**	9.41	10.19	10.73	10.92	11.00	10.83	10.30	10.94
S19	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	6.78	**2.14**	2.17	3.02	3.74	4.27	4.28	2.29	3.87
RMSE	7.65	**2.73**	2.65	3.35	3.96	4.46	4.46	2.74	4.08
S20	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	11.46	7.34	5.68	4.54	3.89	2.98	**1.88**	5.45	3.74
RMSE	11.56	7.54	5.83	4.64	3.96	3.15	**2.39**	5.59	3.80
S21	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**2.02**	2.34	3.30	4.17	4.77	5.53	5.83	3.44	4.95
RMSE	**2.65**	2.74	3.63	4.40	4.91	5.60	5.92	3.75	5.06
S22	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**2.37**	4.17	5.81	6.90	7.48	8.04	8.32	6.03	7.59
RMSE	**3.05**	4.22	5.84	6.92	7.49	8.05	8.33	6.05	7.60
S23	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE		13.87	9.09	8.11	6.63	6.01	**3.99**	8.96	6.28

RMSE		13.92	9.43	8.57	7.41	6.25	**4.79**	9.54	7.05
S24	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	3.80	**1.14**	1.40	2.20	2.81	3.82	5.15	1.44	3.00
RMSE	4.33	**1.54**	1.67	2.34	2.91	3.95	5.44	1.68	3.09
S25	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	3.15	**0.85**	1.64	2.58	3.13	3.70	4.58	1.80	3.23
RMSE	3.74	**1.15**	1.89	2.72	3.22	3.76	4.60	2.01	3.32
S26	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	13.86	9.68	7.99	6.82	6.37	4.91	**2.86**	6.63	5.93
RMSE	14.01	9.93	8.43	8.06	7.56	5.85	**3.01**	7.66	7.17
S27	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	16.39	11.50	9.95	8.99	8.57	8.04	**7.60**	9.88	8.49
RMSE	16.45	11.57	10.01	9.06	8.63	8.10	**7.71**	10.00	8.55
S28	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	6.66	4.99	3.62	2.07	1.39	1.25	1.92	3.27	**1.23**
RMSE	6.89	5.57	4.12	2.34	1.60	1.35	2.10	3.76	**1.46**
S29	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	6.30	4.34	2.60	1.49	0.85	0.77	2.75	2.40	**0.68**
RMSE	7.03	4.76	2.92	1.76	0.99	1.06	2.80	2.68	**0.80**
S30	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.47	3.86	2.26	1.42	1.05	1.00	1.87	2.08	**0.95**
RMSE	7.81	4.18	2.66	1.71	1.35	1.37	2.53	2.47	**1.24**
S31	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	8.71	3.69	1.96	0.92	0.54	0.73	1.72	1.72	**0.51**
RMSE	8.81	3.86	2.17	1.11	0.67	0.96	2.04	1.93	**0.60**
S32	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	2.62	**2.53**	3.94	5.16	5.91	6.80	7.78	4.17	6.09
RMSE	2.89	**2.68**	4.10	5.23	5.96	6.87	7.92	4.32	6.13
S34	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	10.21	5.88	4.09	2.77	2.03	**1.07**	1.21	3.80	1.83
RMSE	10.50	6.09	4.34	3.02	2.25	**1.34**	1.23	4.01	2.03
S35	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	6.59	2.99	1.60	**0.79**	0.85	1.46	2.73	1.73	0.92

RMSE	6.96	3.50	1.98	**0.96**	1.01	1.67	2.74	2.64	1.08
S36	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.98	4.79	2.76	1.38	**0.70**	0.72	1.49	2.48	0.60
RMSE	8.09	4.89	2.88	1.56	**0.95**	0.89	1.87	2.61	0.83
S37	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.38	2.95	1.34	**0.73**	0.90	1.75	2.95	1.08	1.01
RMSE	7.53	3.33	1.72	**0.88**	1.13	2.21	3.36	1.46	1.27
S38	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.66	6.24	5.02	3.40	2.46	1.23	**1.15**	4.85	2.17
RMSE	8.91	6.83	5.91	4.06	3.01	1.43	**1.35**	5.70	2.63
S39	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	10.24	8.77	7.03	5.74	5.05	4.13	**2.94**	6.77	4.91
RMSE	11.12	9.10	7.36	5.93	5.16	4.33	**3.38**	7.07	5.04
S40	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	3.03	3.09	3.07	2.24	0.85	0.47	2.89	3.03	**0.65**
RMSE	3.73	3.69	3.49	2.51	0.88	0.47	2.90	3.41	**0.67**
S41	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	9.37	5.95	4.45	4.44	4.73	4.68	**3.17**	4.21	4.77
RMSE	9.94	6.42	5.19	5.34	5.69	5.32	**3.69**	4.96	5.76
S42	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	1.59	1.09	0.90	0.75	**0.73**	0.96	1.43	0.86	0.75
RMSE	1.81	1.22	1.01	0.84	**0.79**	1.00	1.58	0.94	0.80
S43	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	7.59	3.45	1.99	1.20	**1.02**	1.45	2.27	1.77	1.02
RMSE	7.92	4.03	2.46	1.44	**1.34**	1.88	2.65	2.21	1.38
S44	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	1.35	1.25	1.16	0.91	0.80	**0.67**	0.94	1.60	0.76
RMSE	1.99	2.02	1.83	1.35	1.18	**1.08**	1.07	2.91	1.14
S45	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	6.15	8.79	9.07	8.33	8.31	**5.27**	5.78	8.49	7.23
RMSE	7.55	9.62	9.76	9.34	9.26	**7.75**	6.57	9.35	8.61
S46	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**3.99**	4.24	4.32	4.24	4.06	4.16	4.56	4.33	4.05

RMSE	**5.22**	5.29	5.27	5.05	4.63	4.49	4.69	5.26	4.57
S47	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**2.62**	3.16	3.38	3.57	3.84	4.37	5.30	3.46	3.94
RMSE	**2.94**	3.27	3.43	3.61	3.87	4.48	5.53	3.50	3.97
S48	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	2.67	2.81	2.81	**2.79**	2.80	3.11	3.44	2.82	2.87
RMSE	3.12	3.14	3.15	**3.12**	3.14	3.52	3.91	3.17	3.22
S49	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	9.21	7.05	4.82	2.70	2.25	**1.63**	1.84	4.28	1.73
RMSE	10.11	7.53	5.37	2.88	3.41	**1.88**	1.85	4.51	1.94
S50	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	12.10	10.30	8.66	7.65	7.01	6.04	**4.79**	8.72	6.83
RMSE	12.14	10.40	8.81	7.79	7.14	6.32	**5.38**	8.92	6.97
S51	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE		7.51	9.64	8.35	7.54	6.11	**3.88**	9.39	7.30
RMSE		8.20	9.67	8.40	7.67	6.59	**5.26**	9.42	7.47
S52	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	**4.91**	5.16	5.28	5.37	5.41	5.52	5.73	5.31	5.43
RMSE	**4.96**	5.19	5.30	5.37	5.41	5.53	5.74	5.32	5.44
S53	**WIN_10**	**WIN_20**	**WIN_30**	**WIN_45**	**WIN_60**	**WIN_90**	**WIN_120**	**WIN_32**	**WIN_64**
MAE	2.88	2.10	1.77	1.83	1.96	2.02	2.48	**1.77**	2.03
RMSE	3.89	2.94	2.25	2.13	2.21	2.24	2.82	**2.19**	2.26

Ethics Approval Letter:

Saolta
Grúpa Ollscoile Cúram Sláinte
University Health Care Group

Ospidéal na h-Ollscoile, Gaillimh
University Hospital Galway
GALWAY UNIVERSITY HOSPITALS

Clinical Research Ethics Committee
Room 59
1st Floor
HR Building
Merlin Park Hospital
Galway.

20th January, 2022.

Mr. Talha Iqbal
Doctoral Researcher
2nd Floor
Lambe Institute of Translational Research
University College Hospital
Galway.

Ref: C.A. 2731 ***Stress Levels Monitoring Using Sensor-Derived Signals from Non-Invasive Wearable Device and Dataset Development***

Dear Mr. Iqbal,

The Clinical Research Ethics Committee approved and ratified the above submission at it's meeting on Wednesday 19th January, 2022.

The date of this letter is the date of authorization of the study.

Please keep a copy of this signed approval letter in your study master file for audit purposes. The study must be carried out in accordance with General Data Protection Regulation and Health Research Regulation 2018.

You should note that ethical approval will lapse if you do not adhere to the following conditions:

1. Submission of an Annual Progress Report/Annual Renewal Survey (due annually from the date of this approval letter). **We would encourage you to keep note of this date as the CREC will not issue a reminder.**

2. Report unexpected adverse events, serious adverse events or any event that may affect ethical acceptability of the study.

3. Submit any change to study documentation (minor or major) to CREC for review and approval. Amendments must be submitted on an amendment application form and revised study documents must clearly highlight the changes and contain a new version number and date. Amendments cannot be implemented without written approval from CREC.

4. Notify CREC of discontinuation of the study

5. Submit an End of Trial Declaration Form and Final Study Report/Study Synopsis when the study has been completed.

This application has been reviewed from an Ethical perspective. It remains the responsibility of the Principal Investigator to ensure that data processes are compliant with National Data Protection Regulations and local policies.

Yours sincerely,

B. Gerard Loftus FRCPI, MD
Emeritus Professor of Paediatrics, NUI,Galway
Chair, Galway Clinical Research Ethics Committee.