Predicting Students Exam Performance using Wearable Sensors and Stress Level

THESIS

Submitted in University of Portsmouth

the Requirements for

the Degree of

MASTER OF SCIENCE (Data Analytics)

at the

University of Portsmouth
Faculty of Science

by

Yousif Aldossary

July 2024

Predicting Students Exam Performance using Wearable Sensors and Stress Level

THESIS

Submitted in University of Portsmouth

the Requirements for

the Degree of

MASTER OF SCIENCE (Data Analytics)

at the

University of Portsmouth
Faculty of Science

by

Yousif Aldossary

July 2024

Approved:

____

Department Chair Signature

____

Date

University ID: Up2145253

Approved by the Guidance Committee:

Major: Data Analytics

Supervisor
Mohamed Bader-El-Den
Doctor
in Data Science and Machine Learning

Declaration by student

I, Yousif Aldossary, hereby declare that the work presented herein is original work done by me and has not been published or submitted elsewhere for the requirement of a degree programme. Any literature date or work done by other and cited within this thesis has given due acknowledgement and listed in the reference section.

Acknowledgements

I express my sincere gratitude to Dr.Mohamed Bader-El-Den, the project supervisor, and member of the Data Science and Machine Learning team.

I am grateful to my supervisor for his continuous guidance and invertible suggestions throughout the research, for providing me with logistic support, and for his valuable suggestion to carry out my research successfully.

My utmost gratitude to Dr Hamidreza Khaleghzadeh, Senior Lecturer in data analytics, for providing me access to Jupyter Notebook through the university. For fast running RAM speed to work on large datasets.

Lastly, I would like to express my sincere appreciation to my parents for encouraging and supporting me throughout the study.

Yousif Aldossary

July 2024

I would like to dedicate this study to students who suffer from anxiety and stress. This study is to help other fellow students perform better not only in their studies but also in their future careers.

ABSTRACT

Predicting Students Exam Performance using Wearable Sensors and Stress Level

by

Yousif Aldossary

Advisor: Dr. Mohamed Bader-El-Den

Submitted in University of Portsmouth, the requirements for the degree of Master of Science in Data Analytics

This study examines how stress affects academic performance. Stress indicators such as heart rate, electrodermal activity and skin temperature using wearable devices. These metrics enabled a machine learning model to predict exam outcomes with 91.89% recall and accuracy, and 92.52% precision. Our findings reveal no direct correlation between test Stress and exam performance, supporting the idea that pre-existing knowledge takes precedence over stress levels in determining academic success. The study also highlighted the importance of fine-tuning methods, such as gradually defrosting, to improve model performance with accuracy, recall, and precision rates that exceed 90%. The study suggested that enhancing study strategies could be more beneficial than simply reducing stress. Future research should consider comprehensive support systems that can improve both learning and coping mechanisms.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

1.1.1 Stress in Education System

The anxiety and tension related to taking exams is a natural response to anticipatory events such as tests, exams, papers or presentations. A certain amount of stress can actually motivate and drive you to work harder. However, it becomes problematic when it interferes with your ability to achieve your academic and learning goals [3]. There is ample research indicating that exams generate stress for students, leading to increased worry, anxiety, depression, sleep loss, forgetfulness, irritability, feelings of being overwhelmed, tiredness, and a sense of loss of control [4]. In contrast, this is the concept of academic performance, which is essentially how well students perform in their subjects. Teachers use metrics such as classroom performance, graduation rates, and standardized test results to gauge student achievement. Academic success is the key to effective social development in young people, and those who perform well at school are more likely to adapt to adulthood and achieve professional and financial success [5]. Physiologically, stress can trigger specific mental and bodily responses. Although mild stress can improve cognitive tasks and performance, and especially high chronic stress can lead to neuropsychiatric disorders such as anxiety and depression. Hence, it is imperative that stressed students have access to effective relaxation programs and counseling services to better manage examination stress [6]. In the context of education, academic performance is a central aspect. In essence, it is the pivot around which the entire educational system revolves. The success or failure of any academic institution hinges on the academic performance of its students and is foundational to knowledge acquisition and skill development. Academic performance is evaluated based on grade assignments assigned by teachers and the achievement of educational goals set by students and teachers within a given time frame [7].

1.1.2 Wearable technology

1.1.2.1 Smart Watches Market Growth

Smartwatches are wearable gadgets equipped with a screen and various sensors, such as accelerometers, IR sensors, heart rate monitors, temperature sensors, and electrodermal activity EDA. They can connect to the Internet independently or via a smartphone, allowing them to run proprietary apps and third-party apps [8]. As they become more stylish, smartwatches are becoming both a tech device and a fashion accessory. It is clear why smartwatches, accounting for 42 percent of end-user spending on wearable devices in 2019 [10], are leading the pack. The forecasted 109.2 million global unit shipments of smartwatches in 2023 reflect ongoing demand and highlight the economic significance of these devices. In its current form, Apple, followed by Samsung and Garmin, dominates the smartwatch market [9].

1.1.2.2 How Wearable Sensors Measure Stress

Wearable AI is a cutting-edge technology that uses AI techniques to analyze large amounts of data, such as heart rate, heart rate variation, electrodermal activity (EDA) and skin temperature, collected by sensors in wearable devices. This allows for personalized feedback. Various types of wearable devices, such as on-body, near-body, in-body devices, and electronic textiles, can collect these biomarkers [11]. Wearable sensors can detect physical signs of stress, providing insight into an individual's emotional state. When a person feels threatened, their body reacts in various ways, including sweating, increased heart rate, rapid breathing, tense muscles, high blood pressure, and cooler skin on the hands and feet. Key methods for measuring stress involve observing changes in heart rate and skin electrical activity [13]. There are various methods to measure stress levels using wearable sensors. Some devices utilize heart rate, temperature, accelerometers, electrocardiogram (ECG), and electrodermal activity (EDA), while others may focus on body humidity or solely on variations in heart rate [12]. Through my research, I have simplified the most common factors in measuring stress: heart rate, EDA, and body temperature [1, 12, 13, 14, 15]. The operation of wearable sensors is illustrated in figures 1 and 2 below.

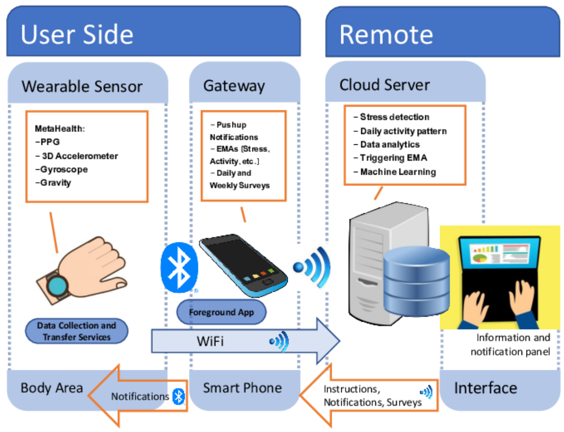


Figure 1.1: Overview of the system architecture.

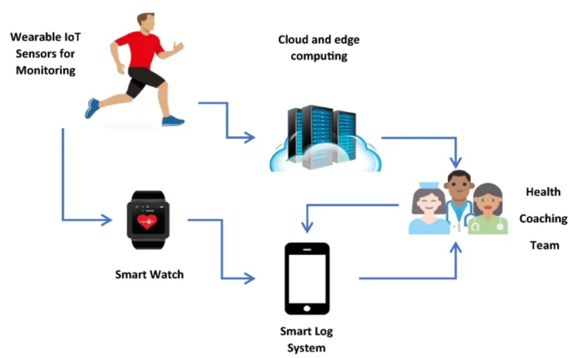


Figure 1.2: Using IoT architecture in health monitoring systems

In this study, we will concentrate on a specific wearable device, the Empatica E4. We chose this device because data supplied by Hosseini et al. and Amin et al. as cited in [17, 18] used E4 in their studies, which will be discussed in more detail in the section Materials and Methodology.

1.1.2.3 Why Empatica E4?

Empatica E4 allows for the acquisition and quantification of peripheral signals in a variety of contexts, from personal use to scientific research. These wearable devices (WDs) are both cheaper and more portable than medical-grade devices. it also provides Skin Temperature (ST) measurements that most WDs do not provide. However, they may produce lower quality data, which can be affected by body movements and data losses [19, 20]. Figure 1.3 shows a system comparison between Samsung Gear S family devices and Empatica E4 devices in more details [21].

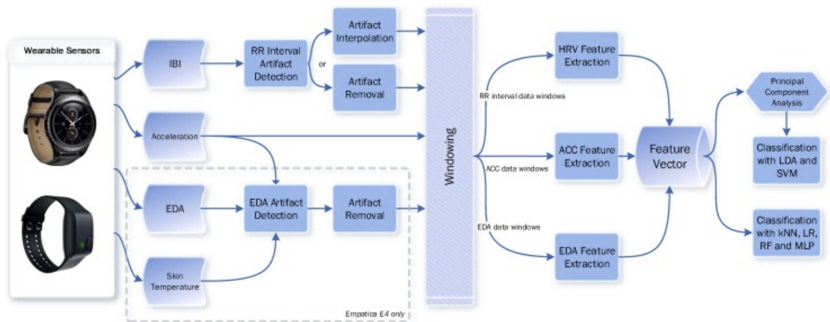


Figure 1.3: Stress level detection system for Samsung Gear S and S2 and Empatica E4.

1.2 Objectives

Research in epidemiology confirms that exposure to numerous stressful events and long-term high perceived stress can negatively impact both mental and physical health, and can even lead to premature death [23]. This link between exposure to stress and an increased risk of diseases has been observed in various types of stressors and a variety of aging-related health outcomes (e.g., cardiovascular disease, metabolic syndrome, mortality) [25, 26, 27].

Stress measurement predominantly falls into two categories: high-tech brain stimulus measurements and surveys [21]. However, with recent technological advancements, we can now use smartwatches for this purpose. Features such as mean amplitude, standard deviation, minimum and maximum values, RMS, delay between stimulus and response, number of peaks, peak height, rising time, recovery time, and the position of maximum and minimum are used to gauge user stress levels [22]. However, these measurements may not always be precise. For example, in a study involving nurses experiencing high stress levels during COVID-19 by Hosseini et al, wearable sensors sometimes recorded false data. An example of this is a high or moderate stress signal being shown when the nurse is simply taking a lunch break. Therefore, to ensure data accuracy, the data acquisition process included a questionnaire for nurses [16]. Hence, we can leverage the power of transfer learning using the nurses dataset mentioned in the example above to analyze and measure the stress levels of students during their examination periods. In doing so, we can gain a deeper understanding of how these stress levels can impact their overall performance. Ultimately, this could provide valuable insights into the correlation between stress and academic performance, enabling us to develop strategies to help students manage their stress more effectively and potentially enhance their academic achievement.

1.3 Limitation of the study

1. The accelerometer data is significantly larger than the other attributes. Also it is not a leading indicator of stress level, leading to its exclusion

2. While no specific times were given, all exams were stated to have started at 9 am. The mid-term lasted 1.5 hours and the final exam lasted 3 hours. Therefore, time was estimated, and we believe there is a possibility that some students may have left the exam early.

3. The dataset size is very small, containing only 10 participants. To improve the precision of our model for predicting exam performance, we require a

substantially larger dataset.

4. Additionally, this dataset does not provide demographic information about the participants. Such information could help determine the influence of factors like gender and environment on managing and maintaining stress levels.

5. The stress measurements were conducted on one subject by Amin et al. Although the exam difficulty level varies between subjects. Conducting stress measurements across multiple subjects' exams can significantly support the analysis of student performance and clarify our understanding of it. Also, consider students who major in other fields like education, science, or commerce.

6. The data was shared through PhysioNet [17]. Therefore, this study did not have control over data collections.

1.4 Hypothesis

The study aims to use **transfer learning to predict student exam performance based on their stress levels, using a dataset and three criteria provided by Hosseini et al. in their study on nurse stress.**

Accoridng LeBlanc, RV that elevated stress levels can adversely affect performance on tasks requiring divided attention, working memory, information retrieval, and decision making. These effects are influenced by an individual's assessment of the situation, the link between the stressor and the task, and factors such as coping styles, control locus, and social support [28]. Another study by Elias and her colleagues conducted in Malaysia opposing our theory. Their results revealed that undergraduate students generally experienced moderate levels of stress. Among all students, medical students had the highest levels of stress. Furthermore, the study found that first-year students had lower stress levels. Most of the sources of stress came from academic concerns. The study also discovered a significant, yet weak, negative correlation between undergraduate students' stress levels and their academic performance [29].

		Academic Achievement
Stress	Pearson correlation	-0.195
	Sig. (2-tailed)	0.000
	n	376

Table 1.1: Correlation between Stress and Academic Achievement

Degree Programme/Faculty	n	Mean	Standard Deviation
Medicine and Health Sciences	26	1051.80	342.29
Engineering	25	1025.00	306.17
Veterinary Medicine	26	1011.30	232.86
Economics and Management	25	980.56	373.95
Design and Architecture	25	954.96	285.08
Biotechnology and Biomolecular Sciences	24	948.83	241.11
Computer Science and Information Technology	25	928.80	338.62
Science	24	908.42	334.44
Human Ecology	25	893.20	289.84
Modern Language and Communication	25	889.28	239.64
Agriculture	26	881.92	187.23
Environmental Studies	24	863.04	339.45
Educational Studies	26	857.62	195.58
Forestry	24	706.29	221.30
Total	376	926.39	288.39

Table 1.2: Respondents' Mean Stress Score by Degree Programme

Sources of Stress	Percentage (%)
Starting of a new semester	100
Registration	100
Making new friends and get along with friends	83.2
Lack of sleep	75
Having two exams in a day	71.8
Concerns about own appearance	69.7
Undergoing the final week of a semester	68.1
Talking in front of a class	62.5
Having the sense of overload in school or work	58.2
Falling asleep in class	56.9
Changing in housing situation	56.1
Attending a class that one hates	54.8

Table 1.3: Respondents' High-ranked Sources of Stress

Elias and her team suggested that medical students typically experience more stress than the average individual. However, the impact of this stress on academic performance remains unclear. Thus, two other studies were conducted to gain further insight. One study by Singh, R. et al found that when stress levels changed, so did anxiety and saliva cortisol levels. However, these changes did not really affect the performance of medical students on the exam. Both males and females had similar changes in mood. This study suggests that because exams are stressful, those involved in medical training need to understand the negative effects of stress [6]. Another study by Kudachi, P. et al discovered that students with high anxiety had much higher cortisol levels (144.10±72.90 ng/ml, p¡0.05) during exams. Female students had even higher cortisol levels (203.10±81.20 ng/ml, p¡0.05) than male students (116.75±33.10 ng/ml), suggesting that females may experience more stress. Their study showed that elevated stress levels do not have a negative affect on exam performance [30]. Our study aims to predict the impact of stress on exam performance and estimate the anticipated range of exam results, expressed as percentage (%) using the New Zealand education system.

	New Zealand		America	
A+	9.0	90% - 100%	4.0	97% - 100%
A	8.0	85% - 89%	4.0	93% - 96%
A-	7.0	80% - 84%	3.7	90% - 92%
B+	6.0	75% - 79%	3.3	87% - 89%
B	5.0	70% - 74%	3.0	83% - 86%
B-	4.0	65% - 69%	2.7	80% - 82%
C+	3.0	60% - 64%	2.3	77% - 79%
C	2.0	55% - 59%	2.0	73% - 76%
C-	1.0	50% - 54%	1.7	70% - 72%
D+	0.0	45% - 49%	1.3	67% - 69%
D	0.0	40% - 44%	1.0	65% - 66%
D- / F	0.0	0% - 39%	0.0	Below 65%

Figure 1.4: The Differences Between the New Zealand and American University Grading Systems

Chapter 2

Literature Review

2.1 Review of Physiological and Biochemical Indicator of Stress

Stress is a known factor in several severe medical conditions and a trigger for acute cardiovascular events. It is also a root cause of several social challenges. As the global stress burden increases, so does interest in effective stress-monitoring solutions for preventive health, especially using wearable sensing technologies. Recent advancements in miniaturized and flexible biosensors have facilitated the development of connected wearable solutions to monitor stress. These can intervene in a timely manner to prevent the escalation of stress-induced medical conditions. This section reviews various physiological and chemical stress indicators commonly used for quantitative stress assessment and the associated sensing technologies.

2.1.1 Introduction

In recent years, there's been a significant increase in anxiety, depression, pathological stress, and other stress-related diseases. Stress generally harms both the physical and mental health of individuals [37, 38, 39]. Chronic stress, in particular, increases the risk of cardiovascular diseases [41], diabetes, stroke, and obesity [40, 44]. World Health Organization statistics show that stress is linked to various medical and social problems affecting the health and well-being of adults, children, and young people [45]. This has led to a growing interest in investigating the underlying mechanisms of stress and monitoring the various biophysiological and biochemical responses of the body to stress [42]. A reliable stress biomarker could enable accurate stress monitoring and potentially prevent pathological conditions in their early stages. During the past two decades, significant progress has been made in physiological and biochemical sensing technologies. These sensors offer an excellent platform for connected health solutions and preventive care for various stress-related conditions [46, 47].

Stress is a disturbance in an individual's homeostatic balance, which triggers a coping mechanism known as the stress response [43]. Stress can be acute, an immediate reaction to a stressor, or chronic, a condition caused by continuous stress stimulus [48]. Chronic stress can reach a point where the body fails to maintain homeostatic balance, and the individual can no longer manage stressors. The activation of the stress response prompts various bodily changes due to the stimulation of the sympathetic nervous system and the inhibition of the parasympathetic system. Generally, the stress response involves the release of stress hormones that increase the level of alertness of the body. Consequently, there is an increase in heart rate, blood supply to the muscles, respiratory rate, skin temperature (due to heightened blood circulation), cognitive activity, and other responses. Hormonal responses specific to stress and other biomarkers impacted by the stress response are often used to quantitatively assess or monitor stress [39, 42, 43].

Most studies on stress monitoring reported in the literature typically follow a similar experimental approach. Sensors collect biophysiological data in both stress and non-stress states. Stress is initially induced in a controlled environment, such as a laboratory, or in real life, using methods such as mental arithmetic, the TSST or the Stroop test [49, 50]. Features are extracted from sensor data, and machine learning (ML) or pattern recognition is utilized to differentiate between stress and non-stress states (or baseline). Machine learning algorithms fall into two main categories. The first is supervised learning, where the model is provided with both input and classification labels for prediction and classification. The second is unsupervised learning, where no labels are given at the input stage, and the model is designed to group the input data according to inherent patterns or similarities. Typically, the sensor data is recorded on the device and then transferred to a computer or the cloud for processing and analysis. In certain scenarios, such as simulated driving, participants' wearable sensors are directly connected to a computer for real-time analysis during the experiment. Various machine learning techniques, including support vector machine (SVM), Bayesian networks (BN), artificial neural network (ANN), fuzzy logic, decision tree (DT), and other computer-aided diagnostics (CAD) tools, have been used for classification [51, 52, 53, 54]. A detailed review of these commonly used machine learning algorithms is provided in the following section.

A plethora of literature demonstrates the link between a higher heart rate and stress. This change in the heart rate alters the body's blood flow. An electrocardiogram (ECG) can monitor heart rate and its variability, while the blood volume pulse (BVP), derived from a photoplethysmography (PPG) signal, can measure the change in blood flow [55, 56]. Some studies discuss the sweat produced during stress, which alters the skin conductance measured by the electrodermal activity (EDA) device [53]. Muscle tension, another stress-related factor, can be monitored using electromyography (EMG) [57, 58]. Chronic stress can occasionally cause mild fever (between 37.2 and 37.8 °C), anxiety, and restlessness. Therefore, skin temperature (ST) and accelerometer (ACC) sensors can also assist in stress detection [59, 60, 61].

During periods of stress, the body prepares itself for a 'fight or flight' response, releasing catecholamines to manage the stress. Thus, evaluating plasma catecholamines can help measure stress [62]. Arginine vasopressin's (AVP) role in the acute stress response has been extensively discussed in literature. Copeptin, a stable biomarker of AVP release, increases significantly along with cortisol, prolactin, and adrenocorticotropic hormones, which all directly relate to the human body's stress response. Therefore, monitoring copeptin and prolactin hormone levels in the blood can help detect stress [63]. Alpha-amylase, one of the major salivary enzymes, is secreted in saliva in response to psychological stressors [64]. Cortisol, a primary stress hormone, is released into the bloodstream during stress and causes an increase in glucose levels [65]. Therefore, monitoring cortisol levels also aids in stress monitoring. All of these hormones can be measured using various available enzyme-linked immunosorbent assay (ELISA) kits.

There is a significant amount of literature on stress monitoring through physiological or biochemical responses in the human body. However, there is no clear agreement on the sensitivity and specificity of these biophysiological and biochemical responses for identifying stress. Factors such as stress response sensitivity, sensor sensitivity, type of stimulators, sample size, experimental design, and other variables may influence this [66]. Nonetheless, the sensitivity and specificity of measurable stress responses are vital for long-term stress monitoring in preventive and personalized care.

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 2.1 presents a list of biophysiological and biochemical indicators of stress.

S.No.	**Bio-signals**	**Units***
1	Skin conductance (also known as electrodermal Activity, EDA)	µS
2	Electrocardiography (ECG)	mV
3	Electroencephalograph (EEG)	µV
4	Respiration rate (Resp), blood pressure (BP) and blood volume pulse (BVP) using photoplethysmography (PPG)	Breaths/min, mmHg and mV
5	Skin temperature (ST)	°C
6	Electromyography (EMG)	µV
7	Plasma catecholamines, copeptin and prolactin, steroids samples	mcg/24-h, ng/mol, ng
9	Cortisol samples	nmol/L

Table 2.1: Most used bio-signals for stress monitoring.

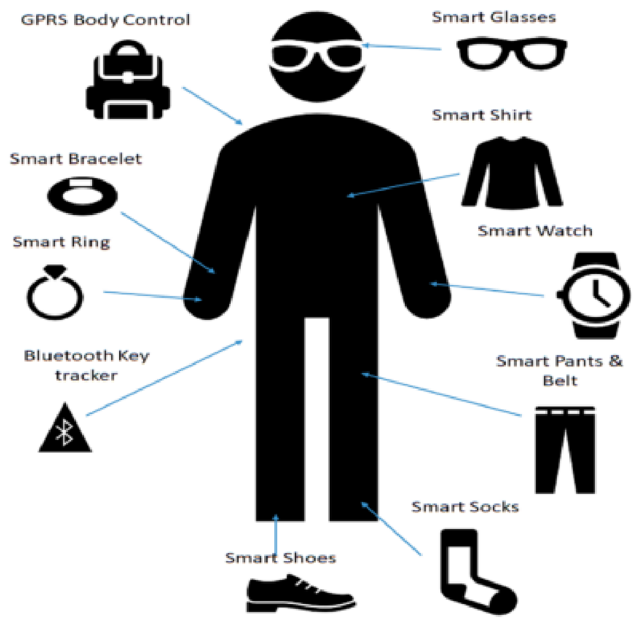


Figure 2.1: Possible sites for placement of smart sensors.

2.1.1.1 Stress Assessment Tests

Stress can be evaluated subjectively through structured questionnaires and self-reporting forms, a standard clinical practice. Alternatively, stress can be objectively assessed by measuring the body's various responses to it [66]. Common tools in clinical stress assessment are self-reported questionnaires, such as Cohen's Perceived Stress Scale (PSS), or self-reported visual scales such as the Visual Analogue Scale for Stress (VASS). Biomedical researchers often prefer biochemical markers like cortisol and amylase to detect stress. They typically induce a stress state in test subjects using the Trier Social Stress Test (TSST) [67, 68]. Meanwhile, numerous studies measure physiological signals in response to stress to assess it [69]. Table 2.2 presents details of commonly used stress assessment tests and questionnaires.

Test Name	**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).

Table 2.2: Stress assessment test and brief details.

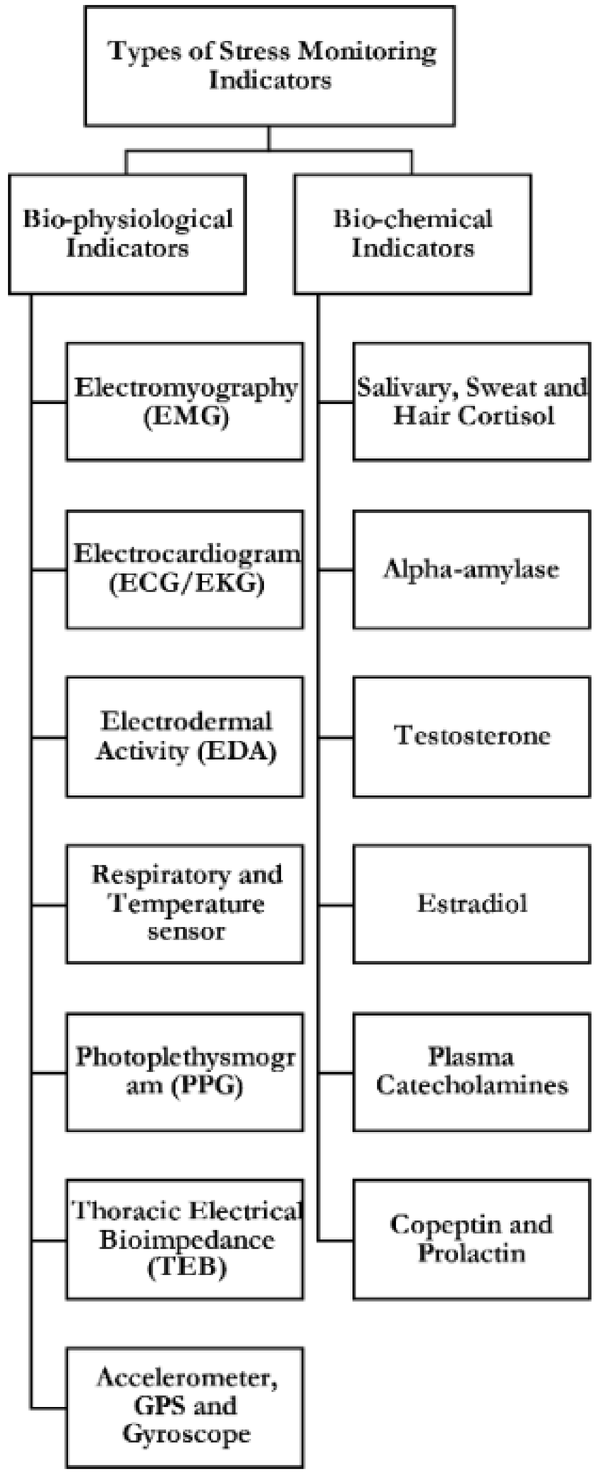


Figure 2.2: Types of stress indicators: physiological (left column) and biochemical (right column).

2.1.2 Biophysiological Indicators

Most literature studies associate an increase in heart rate, abrupt changes in skin conductance, respiratory rate, and blood pressure with stress, using these as ground truth. However, even when some studies use the same signals, stressors, and ground truth methods, the results reported often show significant differences. Figure 2.3 depicts the reported prediction accuracies versus the prediction algorithm, using different stress indicators/markers. Most authors used SVM (with various kernels) for prediction, but the highest prediction accuracy was achieved using an artificial neural network predictor [51, 70]. Table 2.3 summarizes the conclusions of different biophysiological indicators-based studies.

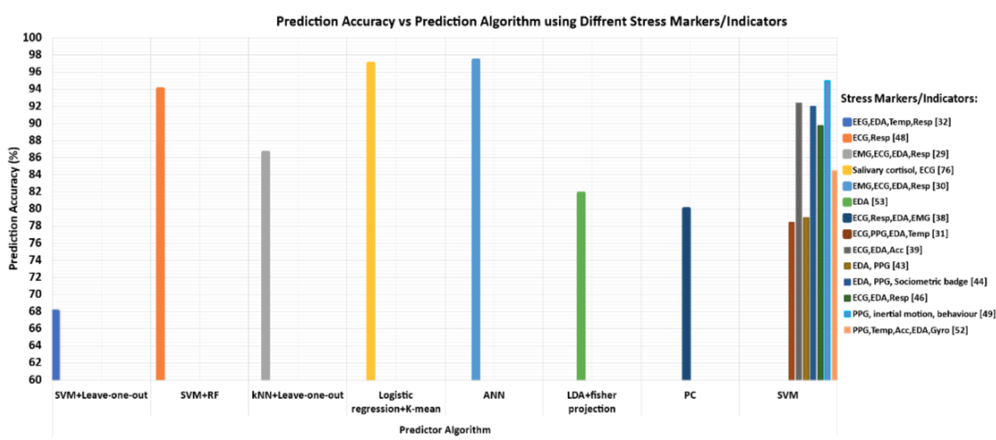


Figure 2.3: Reported prediction accuracies of various prediction algorithms using different stress indicators/markers. In the figure, SVM is a support vector machine, RF is random forest, kNN is a k-nearest neighbour, ANN is an artificial neural network, LDA is a linear discriminator analysis, and PC is a principal component analysis, [1].

Year	**Signals**	**No. of subjects**	**Stressors**	**Finings**	**Features extracting Methodology/ Classifier**
2020	Galvanic skin response (GSR)	10	Predefined PYSIONET dataset and driving on the highway, in the city	The authors developed a model that achieved 85.3% accuracy on PYSIONET data and 83.2% on a test set using cross-validation. They proposed its use in wearable GSR sensors to monitor driving stress in real-time.	GSR from Empatica E4 watch and binary logistic regression classifier with cross-validation.
2019	PPG, EDA, GSR, and ACC	21	Summer camp (training, the contest, and free day)	If sufficient individual data is available, design a person-based model. If not, cluster individuals by stress behavior to develop a model and improve overall classification accuracy.	Clustering models such as kNN
2019	EDA	1	Driving on the highway, in the city	The experiment concluded that road type and traffic conditions greatly impact driving stress. Using logistic regression as a classifier, the authors achieved 80.3% accuracy, 85% sensitivity, 78% specificity, and 70% positive predictivity.	EDA from Empatica E4 watch and logistic regression-based classifier
2018	ECG	1	Daily life stress	The change in the stress index matches up well with the person's work schedule. So, it can give a good way to compare stress levels between different people.	The stress index change aligns with the person's work routine. This can help compare stress levels among different individuals.

2018	PPG, Temp, ACC, and EDA	28	City Car Driving simulator	The accuracy was 68.31% when classifying into four states: normal, stressed, drowsy, and fatigued. When classifying into three states: normal, stressed, and either drowsy or fatigued, the accuracy increased to 84.46%.	Checked the gaps between pulses and compared them to normal pulse gaps. Used two methods, "Winner-takes-all" and "Max-wins voting", along with a tool called an "SVM classifier" to sort the data.
2018	EDA only (ECG, EMG, and respiratory)	11	Driving on the highway, in the city	After applying Fisher projection and linear discriminant analysis (LDA) to the dataset, the authors reported a classification accuracy of 81.82%.	Fisher's method and Simple Discriminant Analysis (LDA)
2017	ECG, EDA, and respiratory	14	Real driving environment	The SVM with a linear kernel, using all features, provided the best precision in classifying between drives. For classifying across different drives, the SVM with an RBF kernel achieved a precision score of 89.7%.	SVM with linear kernel, SVM with RBF kernel
2012	ECG, EDA, and ACC	20	Stroop color test and mental arithmetic problems based on Montreal Imaging Stress Task (MIST)	Using accelerometer data made the process of detecting stress in a mobile environment better (92.4%).	Decision tree classifier, 10-fold validation, and least complex classifier

2004	ECG, EMG, EDA, and respiratory	24	Some audio, visual, and cognitive stimuli	Got a 97.4% accuracy rate using 5-minute data intervals. Found the strongest connection between heart rate and skin sweat measurements.	Artificial neural networks (ANN)
Ref.	[21, 71, 72, 73, 74, 75, 76, 77, 87, 89]				

Table 2.3: 10 Biophysiological parameters by year.

2.1.3 Biochemical Indicators

Biochemical stress indicators offer enhanced detection and tracking, yet most require invasive measurements. In situations where non-invasive measurement is possible, extraction of hormones from the sample is time consuming and is not suitable for real-time stress monitoring devices. However, there's potential to combine certain biochemical indicators, like cortisol levels, with physiological signs such as heart rate, respiratory rate, and activity monitoring, to create a more reliable and precise stress monitoring device. Table 2.5 offers a summary of studies based on biochemical indicators and their findings.

Year	Signals	No. of subjects	Stressors	Finings
2019	ST, HR, pulse wave (EDA, ECG, PPG), copeptin, prolactin (blood), cortisol, and alpha-amylase (saliva)	40	Trier Social Stress Test (TSST)	There wasn't a strong link between physical signs and how stressed people felt. In addition, the highest level of alpha-amylase, a stress indicator, appears 10 to 15 minutes after a stressful event. So, it should be measured then. We measured alpha-amylase and another stress indicator, cortisol, in the morning when these levels can vary a lot in each person.
2019	PPG and endocrine (salivary) cortisol	32	Childhood Trauma Questionnaire (CTQ)	Early life stress didn't really change heart rate (an indicator related to our nervous system) and salivary cortisol (a hormone that indicates stress). But, the authors proposed that heart rate is a more accurate stress indicator than salivary cortisol. This is because heart rate is more responsive to how each person reacts to stress than salivary cortisol.
2018	Biochemical (salivary cortisol) and physiological (HRV measures) domains	30	Academic final examination, Psychological Stress Response Inventory	Salivary cortisol levels were negatively linked to the HRV parasympathetic indicator and positively to the sympathetic one. The mental stress index (MSI) was found to be highly sensitive to acute stress, predicting it with 97% accuracy.
2013	Sweat and saliva	17	Intense exercise	Intense exercise could increase the concentration of cortisol in hair, which was not decreased by hair washing.

Year	Signals	No. of subjects	Stressors	Finings
2012	Hair cortisol	-	Daily life stress (3	Found some areas the current literature doesn't cover: firstly, we need to understand how cortisol gets into hair, and secondly, we need to figure out what causes changes in hair cortisol, like the impact of washing hair.
2010	Salivary alphaamylase, plasma	33	College academic	The level of salivary alpha-amylase changed noticeably. We found a partial link between salivary alpha-amylase and factors like blood pressure, heart rate, and plasma catecholamines.
2004	Salivary cortisol (in the blood)	12	The psychosocial stress test, Trier social test (TSST), and reading test	Women who smoke have lower levels of a certain enzyme in their saliva than women who don't smoke. Men who smoke, on the other hand, have higher levels of this enzyme than men who don't smoke. The amount of a stress hormone in the saliva influences the relationship between a type of adrenaline and these enzymes. When a certain part of the nervous system is active, it reduces the total production and volume of saliva. Therefore, the amount of saliva and enzyme levels should be measured in relation to the amount of saliva produced.
Ref.	[78, 79, 80, 81, 82, 83, 84].			

Table 2.4: 7 Biochemical parameters sorted by year.

2.1.4 Conclusion

The two tables above summarize the types of stressors used in each reviewed study, along with the bio-signals collected to measure and monitor stress. It is notable that some studies used similar signals and stressors, yet reported significantly different results. The classification accuracy varied among studies, ranging from 80% to a high of 97.4%. Additionally, the studies by Kim, KH and colleagues, as well as Shi, Yuan and others, all used the same signal for stress detection. However, they reported classification accuracies of 78.4% and 68% respectively [85, 86]. Gjoreski 2017 says that just using body signals is not enough to accurately detect stress. You also need to know the situation the person is in [53]. Sun's team in 2010 found that knowing the situation improved accuracy to 92.4%. However, Li and Du managed to get 81.82% accuracy using only skin conductance, a type of body signal. Similarly, Amin's team in 2022 got 80% accuracy this way [15, 76]. Healey and Picard achieved a 97.4% accuracy using four different types of body signals and no context information [89].

Despite the high classification accuracies using biophysiological parameters, some studies, such as A. Arza et al and L. Bönke et al.,, have reported no clear correlation between perceived stress and these parameters [78, 79]. These studies suggest considering biochemical markers of stress when designing a stress monitoring system. Interestingly, L. Bönke et al. proposed that biophysiological markers of stress, like heart rate, could be a better indicator than biochemical markers like salivary cortisol, which is commonly used. However, in contrast to this, M. Uesato et al. and N. Ahmed et al. proposed a positive or partial correlation between salivary cortisol and physiological stress indicators, such as heart rate, heart rate variability, and respiratory rate [83, 84].

Figure 2.3 provides a visual of the reported accuracies. It is key to consider that there are various techniques for extracting features from a raw signal, as well as different methods for calculating accuracies. Tools such as the confusion matrix, specificity, sensitivity, recall, f-score, the area under the curve, positive predictive value, negative predictive value, and likelihood ratio (both positive and negative), are often used to assess the performance of different indicators and classifiers, as outlined in [90, 91, 92]. Given that authors may use different matrices for the extraction and classification of stress-relevant features in the reviewed literature, the reported accuracies may not be directly comparable.

The literature on physiological and biochemical stress markers provides variable and contradictory evidence. This leads to the conclusion that neither biomarker on its own can effectively monitor stress. Therefore, a more reliable approach is to combine physiological and chemical stress markers with contextual information to monitor stress. A multisensory platform that provides data-driven personal insights can help track and intervene in stress cases among the high-risk population. There is still a need for a new, more sensitive, and more specific stress monitoring system that can be easily implemented and adopted by medical professionals and home consumers.

2.2 Review of different machine learning algorithm used for stress classification

Cardiovascular activities are directly linked to the body's response under stress. Stress can be divided into two types based on intensity as well as the duration of symptoms: Acute stress (short-term) and Chronic stress (long-term). Both repeated acute stress and continuous chronic stress can contribute to inflammation in the circulatory system, potentially leading to heart attacks or strokes. In this study, we review some of the commonly used machine learning classification techniques applied to various stress-indicating parameters used in stress monitoring devices. These parameters include Photoplethysmography (PPG), Electrodermal Activity (EDA), Electrocardiograph (ECG), Electromyograph (EMG), Galvanic Skin Response (GSR), Heart Rate Variation (HRV), skin temperature, respiratory rate, Electroencephalograph (EEG), and salivary cortisol. The study also discusses the selection of a classifier, which depends on factors beyond just accuracy. These factors include the number of subjects involved in an experiment, the type of signal processing, and computational limitations.

2.2.1 Introduction

Stress can be categorized into two types: Physical Stress and Mental Stress.

- Physical stress often results from factors such as poor diet, sleep deprivation, overwork, or illness.
- Mental stress may be triggered by concerns about a loved one's illness, the death of close relations, retirement, or financial issues, such as losing a job. Generally, a lot of our stress comes from daily responsibilities.

The pressure and obligations of work, both mental and physical, might not always be apparent to us. In response to daily life stressors, our body automatically adjusts our blood pressure, respiration, heart rate, blood flow to muscles, and metabolism. This response aims to help our body react quickly and efficiently in high-pressure situations [37, 38].

The American Psychological Association (APA) links stress to six leading causes of death, including heart disease, depression, anxiety disorder, and diabetes. The Centre for Disease Control and Prevention reports that stress directly results in 110 million deaths every year - about 7 people every 2 seconds [134]. It is important to understand that without proper monitoring and management, stress becomes increasingly difficult to control. Many people are becoming depressed, easily angered, and are withdrawing from their social circles. However, these situations can be avoided if people understood how to cope with and overcome the effects of stress. Recognizing a high-stress state is difficult. One way to do this is by monitoring physiological indicators like increased heart rate, blood pressure, respiratory rate, sweaty hands, and a fast pulse. Unfortunately, many people struggle to recognize or absent-minded to these physical symptoms. For such individuals, stress monitoring devices can be useful. These devices alert users to increased stress levels in real time, allowing them to take countermeasures such as meditation or exercise. The following parameters, either individually or in combination, are considered for stress monitoring and will be discussed further:

- **EDA**: Electrodermal activity (EDA) measures changes in skin conductance caused by sweat gland activity during stress. When someone is stressed, the sympathetic nervous system produces sweat glands, resulting in increased skin conductance [93]. EDA sensors, usually placed on the fingers or palms, detect these changes. They track two primary components:

  - **Tonic EDA:** The baseline level of skin conductance is a slow-changing measure that indicates overall stress levels.

  - **Phasic EDA:** Rapid and transient changes in response to specific stimuli, reflecting immediate stress responses. [94]

  EDA, which provides information about the autonomic nervous system, is used in psychophysiological research to study emotional responses and physiological arousal to different stimuli.

  - Clinical settings for diagnosing stress-related disorders and managing stress in biofeedback therapy.
  - Human-Computer Interaction to Enhance User Experience by adapting to the user's emotional state.
  - Lie detection as an indicator of stress in polygraph tests. Consumer Research to measure emotional reactions to marketing stimuli.

  EDA limitation include non-specificity, individual differences, and influence of environmental factors like temperature [95]. The future of EDA includes wireless devices for continuous real-world monitoring and improved data interpretation when combined with other measures.

- **PPG:** Photoplethysmography (PPG) is a light-based plethysmogram used to detect changes in blood volume in the body's microvascular system. This is typically done using a pulse oximeter, which illuminates the skin with a light-emitting diode (LED) and calculates changes in light absorption through a photodiode. Since blood flow to the skin is modulated by various physiological systems, PPG can be used to monitor conditions such as hypovolemia, breathing, and other circulatory system conditions. In stress monitoring, a PPG signal from a pulse oximeter is used to calculate the body's oxygen level, also known as oxygen saturation. For a healthy individual, this oxygen saturation level ranges from 95 to 100%. A saturation level below 90% is considered abnormal and may require immediate medical attention [96]. It's worth noting that the PPG signal waveform can vary from patient to patient, and can also be impacted by the location and manner in which the pulse oximeter is attached.

- **ECG:** Electrocardiography (ECG, also known as EKG) is a method that records the heart's electrical activities over time. It uses electrodes placed on the skin, which detect small electrical changes caused by the depolarization and re-polarization of heart muscles with each heartbeat. ECG is often used to identify cardiac issues. For stress monitoring applications, the ECG signal is used to calculate Heart Rate Variation (HRV). When under stress, the heart rate can vary significantly, making HRV a useful monitoring tool [58].

- **EMG:** Electromyography (EMG) is a method used to record and assess the electrical activities produced by skeletal muscles. This process uses an instrument known as an electromyograph to produce a record called an electromyogram. EMG records the electrical activity of muscle cells when they are activated either electrically or neurologically. It allows for the analysis of medical abnormalities such as Polymyositis and Muscular Dystrophy, the order of muscle recruitment, and activation levels. In addition, EMG can also be used to study the biomechanics of animal or human movement [97].

- **GSR:** The Galvanic Skin Response (GSR) measures the skin's electrical conductance. It works by monitoring sweat glands, which release more sweat when the sympathetic nervous system is triggered by strong emotions. The signal is captured using two electrodes attached to two fingers on the same hand. GSR is commonly used to monitor sleep quality [99].

- **EEG:** Electroencephalography (EEG) is an electrophysiological method used to monitor and record the brain's electrical activities. More often than not, it is a non-invasive method that utilizes electrodes placed on the scalp. However, in certain cases, invasive electrodes may be used. EEG measures voltage fluctuations resulting from ionic current within the brain's neurons. These readings are analyzed over a specific period to make diagnoses based on spectral content or event-related potentials [98].

- **Respiratory Rate:** The respiration rate is the number of breaths a person takes in 60 seconds, usually determined by observing the rise and fall of the chest. Measuring this rate during sleep is challenging and cannot be achieved with standard lab devices. However, the respiration rate can be measured from blood volume pulses in two ways:

  1. by calculating the time difference between two successive heartbeats.
  2. by assessing the change in blood volume amplitude.

  The infrared sensor of the ring samples at 250 Hz from the finger's arteries and capillaries, avoiding disruption of the sleeping subject. The collected samples provide data on the inter-beat-interval, which yields information about heart rate variability, heart rate, and respiratory rate [100, 101].

- **Salivary Cortisol:** Cortisol is a steroid hormone produced in the zona fasciculata of the adrenal gland in response to stress. For an average person, the concentration of salivary cortisol ranges from 10.2 to 27.3 with an allowance of +/- 0.8 nmol/L in the morning. At night, it ranges from 2.2 to 4.1 with an allowance of +/- 0.2 nmol/L. High levels of cortisol in the body indicate prolonged stress, which can lead to an allostatic load. This can trigger various physical changes. Fluctuations in cortisol levels can result in mood disorders, anxiety disorders, illness, stomach pain, fear, and other physiological and psychological disorders. Collection protocols and approved methods for salivary cortisol are defined by Salimetrics USA [102, 103, 104].

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 the next chapter, 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

The research community frequently employs a 'Stroop test' to assess mental health. The Stroop test is a computer-based color-naming activity designed like a game. Subjects are asked to name the color, regardless of the written word. For example, if the word BLUE is written in PURPLE color, the subject should say PURPLE. Nagananda et al. [105] designed a Stroop test using five colors: BLUE, YELLOW, GREEN, RED, and PURPLE. They classified stress into low, medium, and high-stress levels using a simple neural network. Aside from the Stroop test, many researchers have designed their own test questionnaires. For example, Kallus et al. [106] designed a RESTQ that measures the frequency of stress and stress-related activities. They created five different versions of RESTQ, each tailored to different subjects. Each version had its own time frame and the output was indicated on a scale of 0 to 6, from never to always. Boynton et al. [107] presented a study on the selection, design, and development of a self-defined questionnaire. They argued that while anyone can create a list of questions, designing an effective and generalized questionnaire requires creativity and careful planning. They also discussed different aspects to consider when designing a questionnaire. They suggested using previously validated questionnaires whenever possible, rephrasing them for the targeted audience. They concluded that a well-explained and carefully designed questionnaire leads to better response rates.

2.2.3 Review of Machine Learning Algorithms for Stress Classification

The advancement in machine learning algorithms has significantly contributed to the development of tools that aid doctors in patient care and the prediction of mental disorders. Machine learning techniques are extensively used in complex health data analysis to establish decision boundaries. Supervised machine learning algorithms generate general hypotheses from externally supplied labeled features, which are then used to predict new incoming features [108]. The literature included in this review is evaluated on the following criteria:

- Frequently used classification methods
- Achieved good classification accuracy (>50%)
- Classification of physical and mental stress states

The selection of a learning algorithm is a crucial step. An algorithm is usually evaluated based on its prediction accuracy—the number of correctly predicted outcomes over the total number of prediction attempts. There are three ways to test a classifier. Firstly, by splitting the training set into training and evaluation sets with at least a 70% and 30% split. Secondly, through cross-validation where the training set is split into equally-sized, mutually exclusive subsets. For each subset, a classifier is trained on the union of all subsets. The average error rate, calculated from the error rate of each subset, determines the classifier's performance. A special type of cross-validation, Leave-one-out, is computationally expensive but is used for greater accuracy in the classifier's error rate. The error rate depends on factors such as the size of the training set, the dimension of the problem, the tuning of hyper-parameters, and the use of relative features of the problem. Lastly, classifier performance can be measured through a statistical comparison of accuracies when trained on specific datasets [109]. Machine learning methods can be categorized into five major types: Neural Newtork, Logic-based Algorithms, Perceptron-based Methods, Statistical Learning Techniques, and Support Vector Machines (SVM). Each type of algorithm has different sub-learning algorithms. Here is an example of a Decision Tree algorithm:

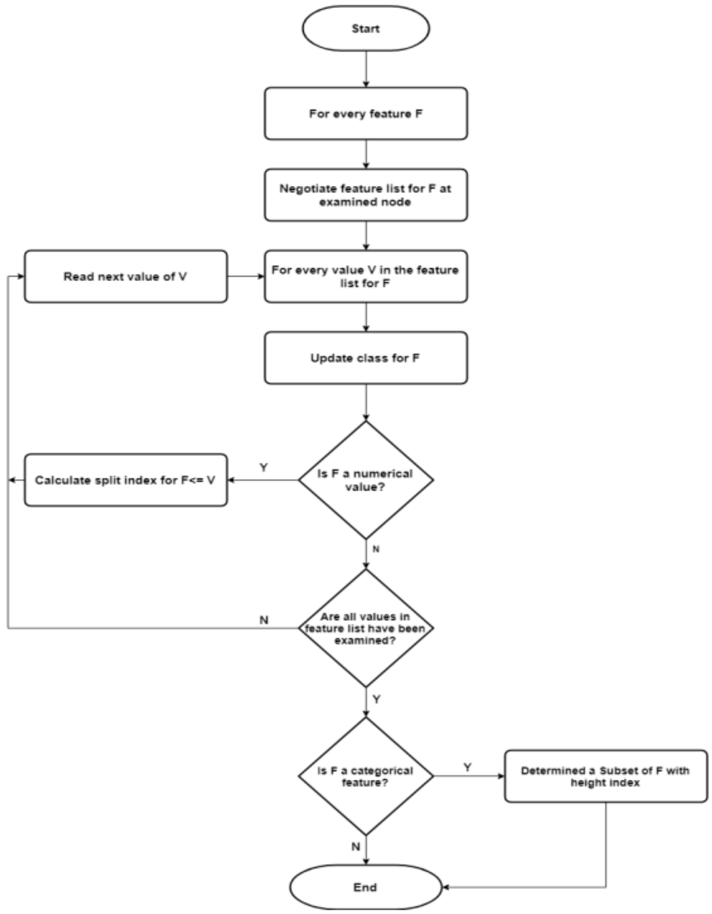


Figure 2.4: Flowchart of C.45 decision tree algorithm.

2.2.3.1 Decision Tree Classifier

A decision tree classifies input instances based on their feature values. Each node of the tree represents a classified feature from an input instance, and each branch represents an assumed nodal value. Classification begins at the root and sorts instances based on these feature values. The most efficient divider of the input training data becomes the root node of the tree. Figure 2.4 demonstrates the steps involved in classification using a decision tree. This divide-and-conquer method is both efficient and quick, making it a valuable classifier for large data sets with hundreds of thousands of input instances. A pseudo-code for designing a decision tree is provided in [110, 111].

2.2.3.2 Artificial Neural Network Classifier

An Artificial Neural Network (ANN) is used for classification when instances in the training dataset can not be linearly separated, as referenced in [121], [122]. [123] provides a comprehensive overview of ANN. The network is created by connecting numerous neurons (units), as depicted in Figure. 2.5.

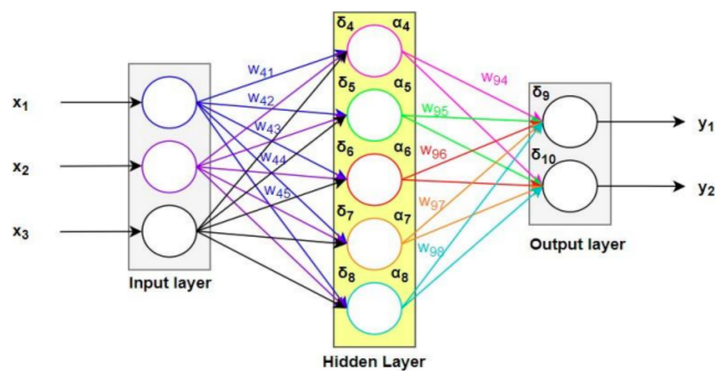


Figure 2.5: Artificial Neural Network

The neurons in the network are divided into three layers: an input layer that receives information from the training dataset; an output layer that delivers processed results (usually probabilities); and a hidden layer located between the input and output layers. If the communication between network neurons is one-way—from input to output—it is known as a feed-forward network. The outcome of an ANN relies on three factors: the network architecture, the weights associated with each neuron, and the input paired with activation functions used for each neuron.

All weights are updated to bring the outcome, which is the result of the classifier, closer to the desired output. The most frequently used weight-update algorithm is the backpropagation (BP) algorithm.

2.2.3.3 Transfer Learning

Transfer learning is a research-proven machine learning method that optimizes the knowledge gained from one task to improve performance on a different but related task. Instead of constructing a model from the ground up, a pre-trained model initially created for a specific task is repurposed as a foundation for another task. This technique is especially helpful when the second task has scarce data, which would make it difficult to train a model from scratch [112].

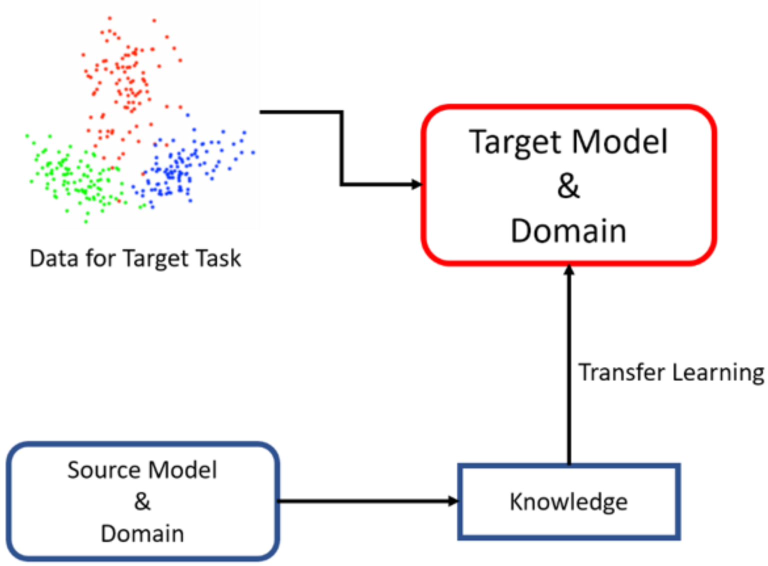


Figure 2.6: Knowledge from an existing task acts as an additional input when learning a target task

Pre-trained models play a vital role in transfer learning. These models have been trained on vast datasets, often for tasks similar to the current one. For instance, models trained on ImageNet for image classification or on large text corpora for natural language processing (NLP) tasks are frequently repurposed. Feature extraction is a key strategy in transfer learning. The pre-trained model serves as a fixed feature extractor, using its initial layers to obtain meaningful features from new data. These features are then input into a new model specifically trained for the new task. This method considerably trims down training time, as the foundational features have already been learned by the pre-trained model. Fine-tuning enhances feature extraction. It involves unfreezing some of the layers of the pre-trained model and training both the pre-trained and new layers on the new task. This allows the model to adapt more closely to the specific details of the new task while still benefiting from the generalized knowledge encoded in the pre-trained model [113].

Transfer learning has broad applications across various fields. In computer vision, pre-trained models like VGG, ResNet, and Inception, initially trained on ImageNet, are frequently used for tasks such as object detection, image segmentation, and other image classification tasks. In natural language processing, models such as BERT, GPT, and ELMo, pre-trained on extensive text datasets, are often fine-tuned for specific tasks like text classification, translation, and sentiment analysis [114]. Transfer learning offers numerous benefits. It significantly cuts down training time, as the initial training phase on the pre-trained model is already finished. Models using pre-trained networks often achieve improved performance, particularly when the new task has limited data. Moreover, transfer learning allows for the creation of high-performance models with relatively small amounts of training data compared to training a model from scratch. However, transfer learning does have its challenges. Domain mismatch can happen if the original task and the new task are too dissimilar, thus limiting the benefits of transfer learning. Fine-tuning complexity is another challenge as deciding which layers to freeze and which to fine-tune can be complicated and may need extensive experimentation [115].

Transfer learning is a significant breakthrough in the field of machine learning, enabling robust model development even with limited data. By leveraging knowledge from pre-trained models, it reduces training time, enhances performance, and expands the range of problems machine learning can address.

2.2.3.4 Bayesian Network Classifier

A Bayesian network is used to depict the probability relationships of input instances or features in graph form. The structure of this network is a directed acyclic graph (DAG), with a one-to-one correspondence between its nodes. The arcs in the DAG illustrate the interplay between different features. If there's no arc showing causal influences between features or no descendant nodes from a particular node (feature), we can identify conditional independence. Usually, the process of learning a Bayesian network involves two main steps: firstly, understanding the Directed Acyclic Graph (DAG) structure and how it's used to create the Bayesian Network (BN) structure, followed by determining the BN parameters. You can find a pseudo-code for training the BN in reference [111]. The structure of a general Bayesian network is illustrated in Figure 2.7.

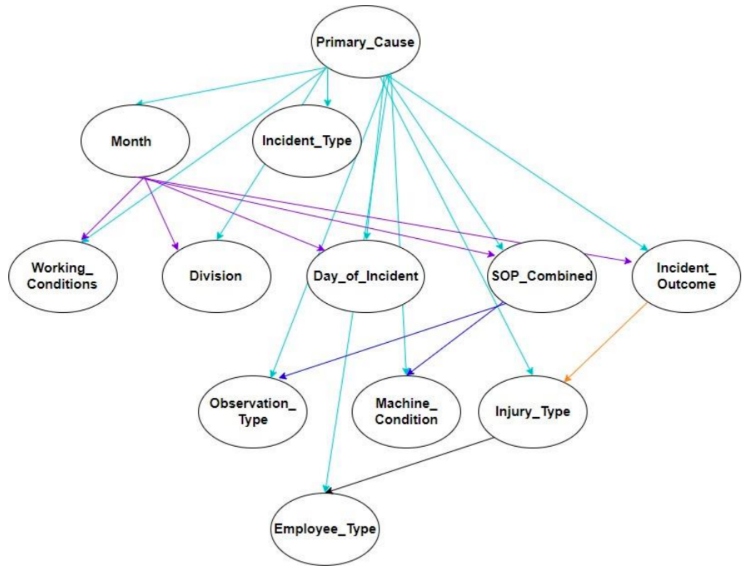


Figure 2.7: Structure of Bayesian Network.

2.2.3.5 k-Nearest Neighbor Classifier

The k-Nearest Neighbor (kNN) is a fundamental instance-based learning algorithm. Its operation involves classifying all instances within a database into one group. When a new instance or feature is introduced, the classifier examines its properties and assigns it to the most similar group, or the 'nearest neighbour.' Figure 2.8 provides a flowchart illustrating the operation of the kNN classifier. To achieve precise classification, initializing a value to k is an essential step in the kNN classifier [116].

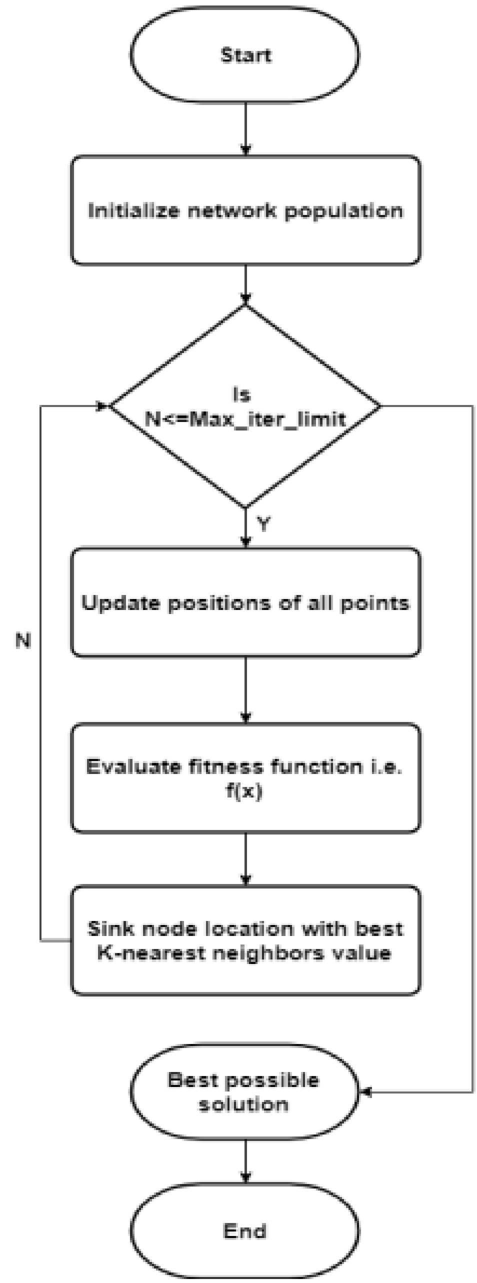


Figure 2.8: Flowchart of kNN classifier

2.2.3.6 Principal Component Analysis

Principal Component Analysis, or PCA, is a common technique used to simplify complex datasets before applying any learning or classification algorithms. Essentially, PCA converts data into a new, lower-dimensional coordinate system. In this new system, the first axis corresponds to the first principal component, which accounts for the largest amount of variance in the dataset. As Figure 2.9 shows, calculating just two principal components can capture the variance of the entire dataset, with each component being independent of the others [117, 118].

2.2.3.7 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a popular choice for dimensionality reduction. It is typically used as a pre-processing step in pattern classification and various machine learning applications. The LDA process, as described in [118], involves several steps. First, calculate the d-dimensional mean vectors for each class in your dataset. Then, compute the in-between-class scatter matrix and the within-class scatter matrix. Next, compute the eigenvectors and their corresponding eigenvalues for both scatter matrices. Select your linear discriminants for the new feature subspace by sorting the eigenvectors in descending order using the eigenvalues. Finally, transform the samples onto the new subspace by executing a matrix multiplication. In essence, LDA calculates an additional axis that maximizes the separation between two classes, as shown in Figure 2.9. For reference, see Figure 2.10 for an illustration of two principal components of a dataset with variables X1 and X2.

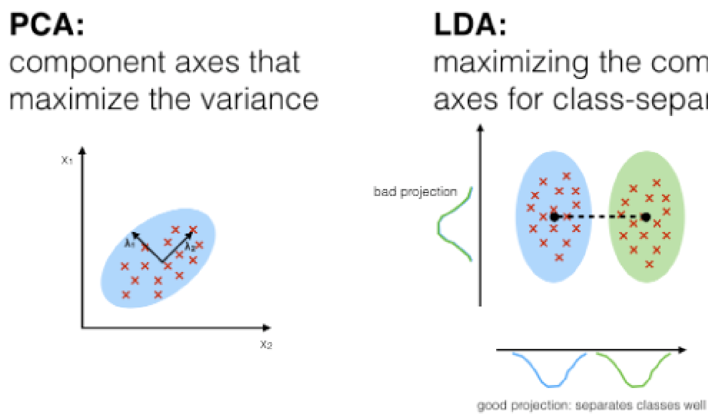


Figure 2.9: PCA for dimensionality reduction followed by an LDA

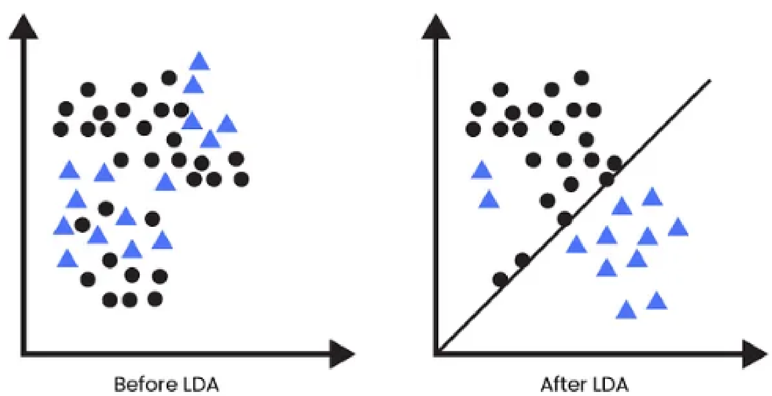


Figure 2.10: A maximum separation between the two classes for LDA

2.2.3.8 Logistic Regression

Logistic Regression is a straightforward machine learning algorithm, typically employed for binary classification problems. It is easy to implement, and often serves as a starting point for two-class classifiers. Logistic Regression comes in three variants:

- Binary Logistic Regression: Used when the target feature has two possible outcomes. Examples include spam or not spam emails, or diabetic or not diabetic.

- Multinomial Logistic Regression: Used when the target feature has three or more nominal categories. For example, predicting the type of clothing.
- Ordinal Logistic Regression: Used when the target feature has three or more ordinal categories, such as rating a product between 1 to 5.

Logistic Regression estimates and outlines the relationship between independent and dependent binary features within a dataset. Figure 2.11 illustrates a classification boundary calculated by a logistic regression algorithm [119].

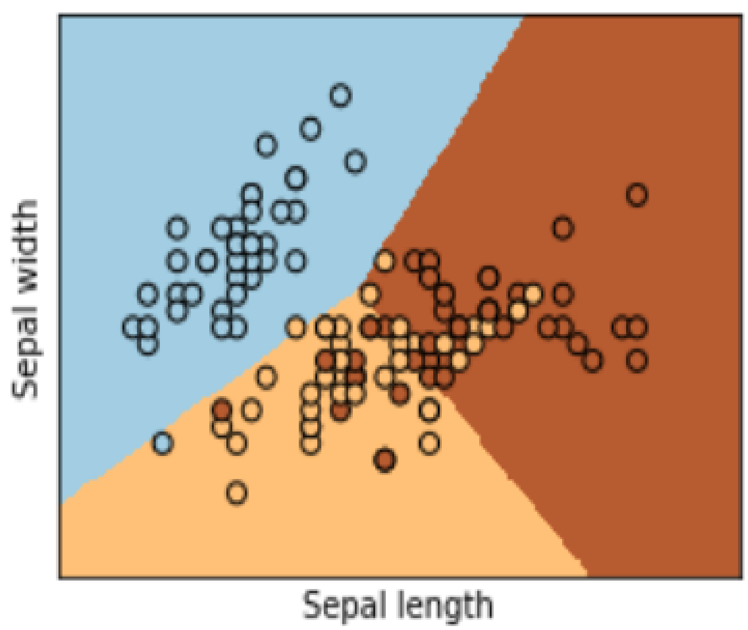


Figure 2.11: Classification using Logistic Regression on Iris dataset.

2.2.3.9 Random Forest

Random Forest is a type of supervised machine learning algorithm. It creates a collection of decision trees, also known as a forest, where each tree is slightly different from the others. When training, it uses a method called bagging, which combines several models to improve the overall result. Adding a new tree to the random forest introduces more randomness. Instead of searching for the most important feature for node splitting, it selects the best feature from a random subset. This increases diversity and typically results in a better model. The unique aspect of Random Forest is that it only considers a random subset of features for splitting a node. This makes the model's trees more random by using random thresholds for each feature, rather than the optimal threshold value. Figure 2.12 illustrates the classification of an input instance (feature signal) using a Random Forest classifier [120].

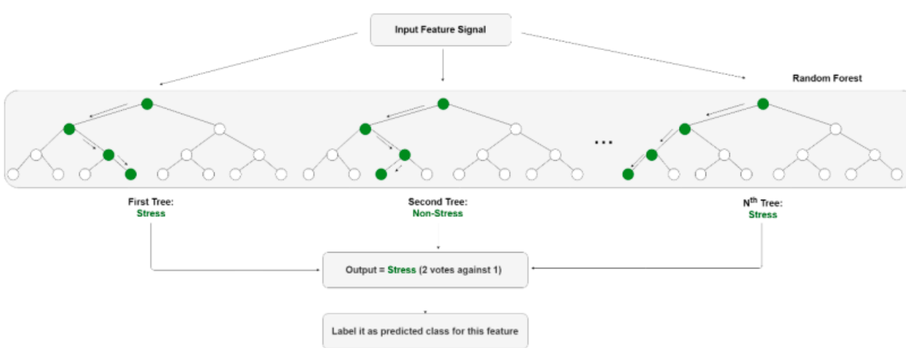


Figure 2.12: Simplified Random Forest Classification of stress and non- stress.

2.2.3.10 Ada-Boost

Boosting refers to a group of techniques that construct a robust classifier from a set of weaker classifiers. To identify a weak classifier, a different machine learning algorithm with a varied distribution is employed. This results in a new weak classification rule at every iteration. Ultimately, a boosting algorithm is created by combining all the newly generated weak classification rules into a strong predictive rule. The selection of the appropriate distribution involves a few steps:

- **Step 1:** Provide all distributions to the base learner and give equal weights to every observation.
- **Step 2:** If the first base learner yields a prediction error, focus more on the observations leading to this error. Then, introduce a new base learner.
- **Step 3:** Continue repeating Step 2 until the base learning limit is reached or the desired level of accuracy is obtained.

Adaboost is often paired with decision trees. The Adaboost model is progressively developed, with the weight of each training instance (or feature) updated to influence the learning carried out by the subsequent tree. After the first tree is generated, the tree's performance is weighted for each training instance, determining the consideration the next instance will receive from the upcoming tree. After all trees are created, a prediction is made for the new data [121]. Figure 2.13 illustrates classification via the Ada-boost technique .

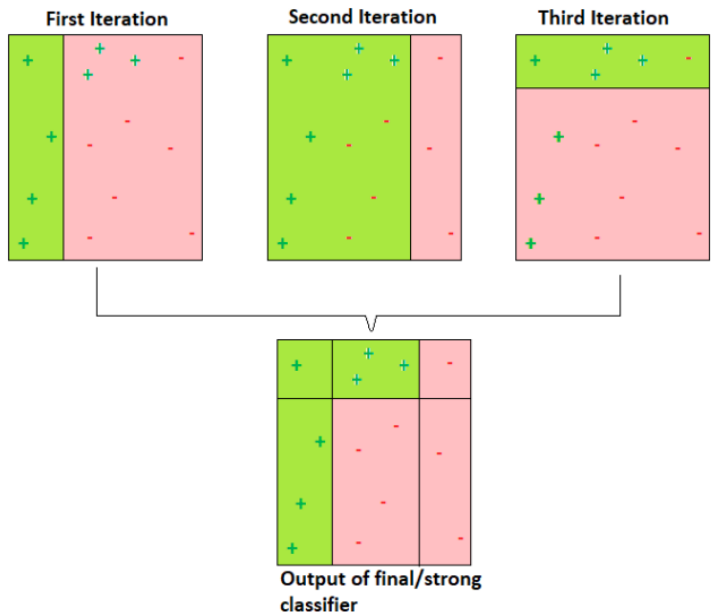


Figure 2.13: AdaBoost Classifier

2.2.3.11 Support Vector Machine Classifier

The Support Vector Machine (SVM) technique is a precise strategy for tackling classification issues. It is built around the concept of margins, dividing data into two classes on either side of a hyperplane. The SVM classifier is binary, so for multi-class classification, we train multiple machines. The goal of SVM is to maximize the margin between instances of two classes and minimize the generalization error which is a common issue in other classifiers. Figure 2.14 provides a visual of two distinct feature sets classified using SVM. Data points that are on the margin of an optimized hyperplane are termed as support vector points. The linear combinations of these points solve the classification problem, while all other points are disregarded. In mathematical terms, SVM uses the Quadratic Problem (QP) with N dimensions, where N represents training samples [121].

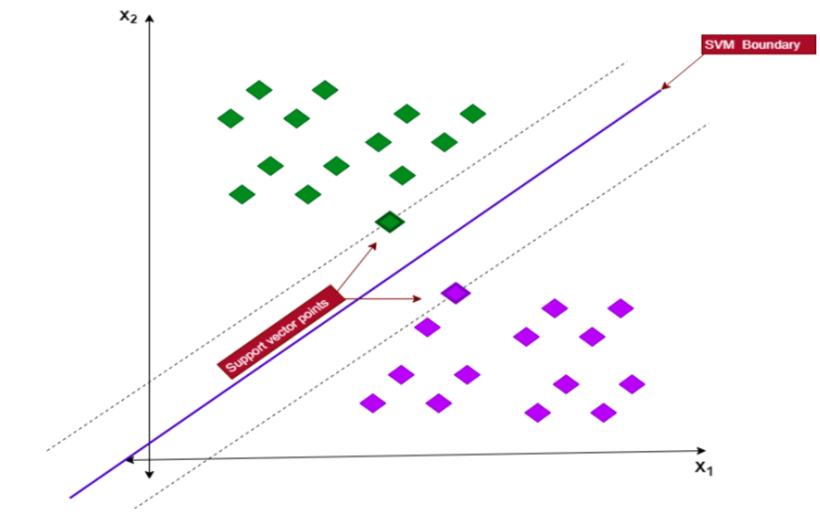


Figure 2.14: Maximum margin hyperplane for SVM trained model.

2.2.3.12 Algorithms reviewed for the task of stress level monitoring.

The table below provides information on 10 different machine learning algorithms. These algorithms, discussed in this review study, use various parameters to train and predict stress levels. All considered parameters are correlated with stress.

Method	Ref	No. of subjects	Signals/Devices Used	Accuracy (%)
AdaBoost	[122]	15	Motion as well as physiological data using sensors	77.2
AdaBoost	[123]	18	Physiological and social responses	94.3
ANN	[124]	20	ECG and PPG	98.8
ANN	[47]	55	Leaf node evidence, context, workload, and student health trait.	85.3
KNN	[123]	20	ECG and PPG	97.5
KNN	[125]	18	Physiological and social responses	87.3
Transfer Learning	[126]	20	physiological sensors (HR, HRV EDA)	93.9
Transfer Learning	[127]	296	EDA, PPG and ST signals	86.76
Decision Tree	[128]	34	Blood Volume Pulse, GSR, Skin Temperature and Pupil Diameter	88.02
Decision Tree	[129]	42	Heart Rate Variability (HRV)	79*
LDA	[130]	21	Heart Rate (HR) and Pulse Transit Time (PTTsensors	60.8
Logistic Regression	[131]	42	Electroencephalogram (EEG) signals	89*
Naïve Bayes	[128]	33	Blood Volume Pulse, GSR, Skin Temperature and Pupil Diameter	78.65
Random Forest	[122]	15	Motion as well as physiological data using sensors.	77.21
Random Forest	[132]	20	Heart Rate Variation (HRV) and simple heart rate signals.	97.2
SVM	[128]	32	Blood Volume Pulse, GSR, Skin Temperature and Pupil Diameter	90.1
SVM	[133]	15	ECG, EEG and EDA signals + Salivary	86

Table 2.5: Comparison Table of Machine Learning Approach for Stress Monitoring and Identification.

2.3 Chapter Conclusion

This study delves into the categorization of stress states and their impact on students' exam performance. We have explored 10 distinct machine learning algorithms that utilize various parameters to train and predict stress, all of which are outlined in Table 2.5. All the parameters taken into account have a correlation with stress. Some are unique parameters like GSR and HR, and some work hand in hand with other parameters to monitor and recognize stress, like Skin temperature (ST) and EMG. However, stress cannot be fully defined by the parameters in Table 2.3 alone. We also need contextual information to interpret the sensor data and understand the circumstances at the time of data collection. This context can be gathered using mobile phones (IoT-based) or computers, which devices that are frequently used in our daily lives. Machine learning algorithms such as Random Forest, kNN, Ada-Boost, and Support Vector Machine were successful in recognizing and classifying a subject's stress state, with an accuracy exceeding 90%. The SVM classifier consistently demonstrated a higher classification rate and can be considered a top choice for stress monitoring procedures. However, it is important to consider the number of subjects and type of sensors required for training and testing the classifier. The question of which classification algorithm is the best fit for stress monitoring devices remains, as perceptions about the use of machine learning algorithms vary from person to person. It is important to set a balance between computation time, accuracy, and device cost when selecting an algorithm.

Chapter 3

Material and Methodology

3.1 Reviewing Related Studies

Here are detailed summaries of two significant studies that were instrumental to this research. These studies not only contributed to the foundation of our knowledge but also provided the critical data used in our analysis and findings.

3.1.1 Review of wearable exam stress dataset for predicting cognitive performance in real-world settings

The study by Amin al. 2022 uses data from wearable devices to measure stress levels among students during exams, with the aim of predicting their grades based on certain features. It involved 11 students (nine boys, two girls). Before participating, each student read and signed a consent form. They collected data from three tests: two midterms and a final exam. The midterms lasted 1.5 hours each, while the final exam took 3 hours. Five minutes before each test, students attached a wearable device known as "Empatica E4" to their hand to record body signals. One student required additional time and completed the test in a separate room. They excluded this student's data as it differed from the others. During the first midterm, they primarily collected data on body signals and movements. Additional data was collected during subsequent tests. This study primarily focuses on EDA and secondarily on ACC signals. Amin and his team utilized the Phasic and Tonic method, which is used to measure stress levels by only EDA and time. Stress features of a basic trend line were analyzed using **KNN**, achieving **70-80% accuracy** in predicting high vs. low grades. This suggests that wearable devices could be used to understand stress. The study acknowledges limitations such as a small sample size and manual filter adjustment. The researchers believe future research should consider optimal filtering based on motion level and explore more physiological features with AI techniques to overcome these limitations [15].

3.1.2 Review of stress detection of nurses in a hospital

Hosseini et al 2022 detailed a study that developed a multimodal sensor dataset aimed at detecting stress in nurses in a hospital setting. The researchers used a machine learning model known as Random Forest to identify and categorize stress levels. Three types of signals were utilized in the process: Electrodermal Activity (EDA), Skin temperature, and Heart Rate. The stress detection process involved three steps. Initially, the team took sets of data, or "windows", from these signals every 10 seconds, overlapping by 5 seconds. Various measurements from these windows were then used to create a profile for each, including elements like the average, minimum, and maximum of each signal. They also considered the average of the last 10 windows for context, giving them 48 pieces of information for each window. Next, they used **the Random Forest model**, which they had trained and optimized using the Affective Road dataset, to categorize the stress level for each window into three categories: **no stress (0), low stress (1), or high stress (2)**. Lastly, they implemented a sliding window-based system from the Ruptures package to break down the continuous stress signals into discrete sessions, each having a start time, end time, and stress label. To validate the stress detection algorithm, the nurses were asked to fill out a survey after stressful events. This was designed not to add to their stress, so it was given at the end of their shift. The questionnaire was chosen based on previous research about stress nurses experience and from direct conversations with them. The full dataset, which includes physiological signals from Empatica E4 wristbands worn by individual participants and stress survey responses, is accessible on Dryad. The document also provides a detailed breakdown of the dataset and the information contained in each file. The study found that signals from EDA, particularly the average, lowest, and highest within a certain time period, were better at telling if a person was stressed. In contrast, Heart Rate and changes in Heart Rate were not as effective at detecting stress. The dataset collected in this study could be beneficial for various research areas such as signal processing and machine learning, to create better stress detection models. It could also help understand links between movement patterns and stress levels. Moreover, those in human resources, human factors, or organizational psychology might find the survey data and biometric signals useful. Despite the valuable insights, the study also acknowledges several limitations, many of which were due to the unique circumstances of the COVID-19 pandemic. Specifically, this research can help determine whether exam stress negatively affects students' performance or has no effect at all [16].

3.2 Wearable Technology Description

Both studies used the same device "Empatica wristband" type E4 to measure stress level. An E4 wristband device (Empatica Inc., Milano, Italy) that collects physiological data such as EDA, Heart Rate, skin temperature, and accelerometer data from the right wrist of the subject. EDA is measured via E4's silver (Ag) electrode (valid range [0.01–100]), while Heart Rate is measured via E4's PPG sensor. The E4 wristband is powered by a rechargeable lithium battery and transmits data to the subject's smartphone, using Bluetooth, in near-real-time. All the data collected from the E4 wristband and the sampling frequencies are presented in Table 3.1.

Signal	Abbreviation	Frequency
electrodermal activity	EDA	4.0 Hz
Heart Rate	HR	1.0 Hz
skin temperature	ST	1.0 Hz
accelerometer	ACC	32 Hz
inter-beat interval	IBI	64 Hz
blood volume pulse	BVP	64 Hz

Table 3.1: Empatica E4 signal output

In Hosseini et al. 2022, physiological data is then transmitted to the data collection app on a nurse's phone in near real-time [2]. the nurses can also tag the data using the tag button on the E4 device to indicate an undetected stress experience, and this is also transmitted to the data collection app . Nurses accessed the survey instrument through their mobile phone to validate the stress response. The survey instrument has the list of events detected (by the model) along with the time these events occurred. The survey instrument allows them to provide context for the event. Because some events may not have been detected, nurses were asked to report any stress events undetected by the model through the survey. Nurses also had an additional mechanism to report the events from the wristwatch by pushing the E4 button to tag stress events. However, no one used the watch to report additional stress events [16].

3.3 Data Collections

During stress, the heart rate typically increases, leading to increased blood flow within the body. This change can be measured through blood volume pulse (BVP). Stress also triggers sweat release, altering skin conductance properties. In some individuals, chronic stress results in a low-grade fever (between 37.2 to 37.7 Celsius) and can cause anxiety and restlessness. Therefore, temperature (TEMP) sensors and accelerometer (ACC) readings can also be utilized to monitor stress. Below is a list of attributes collected from the related studies:

- **Temperature:** Temperature data in Celsius (°C).
- **Electrodermal activities:** Data from the skin activity sensor in microsiemens.
- **Blood volume pulse:** Data from the heart rate sensor.
- **(Accelerometer data:** Data from the 3-axis movement sensor. x, y, and z axis.
- **inter beat intervals data:** Time between heart beats from the BVP signal.
- **Heart rate:** Average heart rate from the BVP signal.

This study will also utilize the same attributes from Hosseini et al., 2022. It will use Electrodermal Activity (EDA) signals, skin temperature, and heart rate to determine students' stress levels during exams.

3.4 Machine Learning

This study will utilize a transfer learning algorithm to ascertain student stress levels. Subsequently, we will correlate these stress levels with student scores. Amin et al. 2022 provided student results for the midterms and final exam, with midterm marks of 100 and the final exam of 200 [15]. To make the scores comparable, we will categorize them into three categories according to the New Zealand educational system (Figure 1.4): "Perfect" (75 and above), "Pass" (50-74), and "Fail" (below 50). Note that the American system considers scores below 65% as a fail. Here is our process for using the nurses' data to train the students' dataset:

- **Pre-processing:**
  - Remove Nulls or Duplicates if they constitute less than 20% of the data, otherwise the average would suffice.
  - Remove duplicates as this is a very large dataset, and we should focus on unique values.
  - Match the attributes by removing unrelated attribute.
  - Identify the class imbalance.
  - Address class imbalance using Sklearn SMOTETomek.
- **Modeling:**

  - Train the model.
  - Save the model.
  - Evaluate accuracy, precision, and recall.

- **Predict Students' Stress Levels:**
  - Correlate stress levels with exam scores.

- **Train the Students' Data to Determine Exam Performance:**
  - Use Fine-Tuning methods:
    * Freezing Last Layer.
    * Change Learning Rate.
    * Gradually Freezing.
  - Compare Fine-Tuning Methods using accuracy, precision, and recall.

- **Reverse the method to train the nurses data to determine exam performance.**

A transfer learning technique will correlate students' stress levels, indicated by heart rate, skin conductance, temperature, to identify exam performance. After preprocessing the data and addressing class imbalance, the model will be trained and evaluated. Then, exam performance model will be predicted using fine-tuning methods.

Chapter 4

Results

4.1 Pre-processing Data

The data has been filtered for feature extraction and application of Machine Learning algorithms. The nurses' data is organized, and stress levels are categorized into three groups. Duplicate entries do not significantly affect the results. Conversely, the students' data must be merged and organized for all participants. This is necessary because the accelerometer sensor data does not match the remaining attributes. But since we are not using accelerometer data from both sources, we can safely discard them. Next, we examined the class imbalance, as depicted in Figure 4.1. The figure shows a significant difference between high-stress inputs and the other two attributes. This imbalance is expected, as nurses are more likely to experience high stress, especially during the COVID-19 pandemic. Also, It is more likely to experience no stress levels than low stress levels. Additionally, we will train the database using the label attributes, as displayed in the graph.

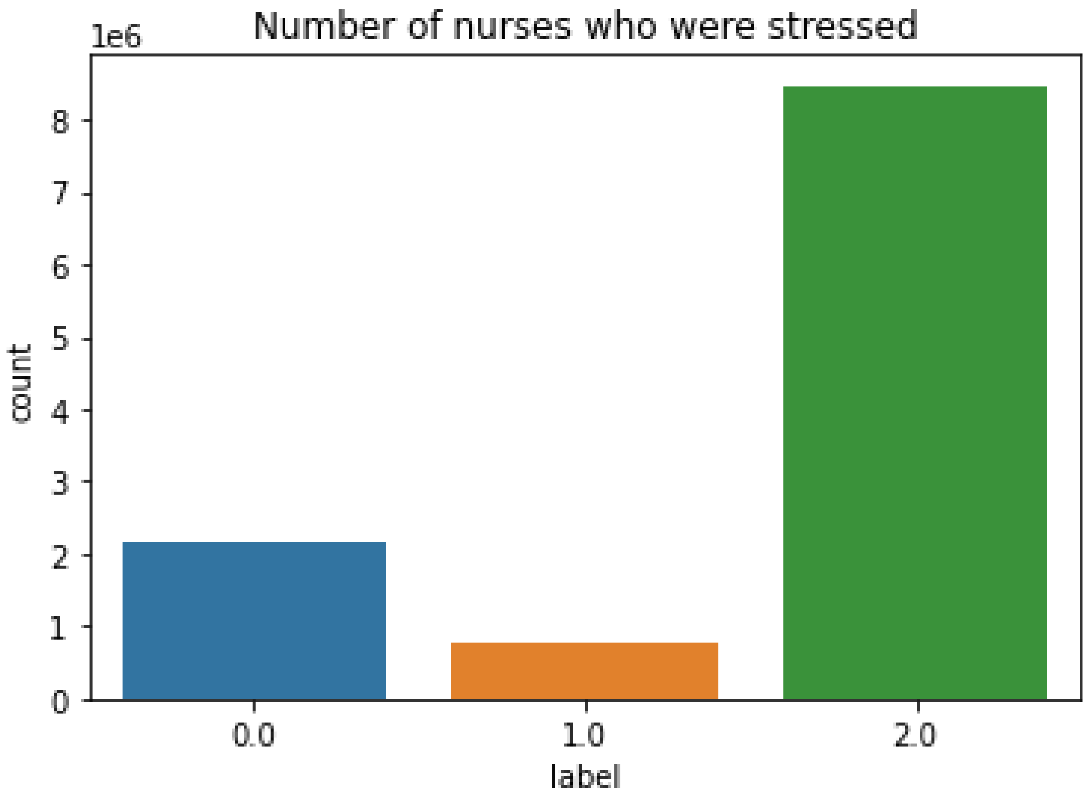


Figure 4.1: Imbalance in the Nurses' Data Label Class

For the student data, we removed the 'id' and 'type of exam' attributes, as they are not necessary. Additionally, the scores are based on the exam type: mid-term scores are out of 100, and final scores are out of 200. We categorized the scores into performance levels: anything 75 or above is considered "2" indicating a perfect score, scores between 50-74 are considered "1" indicating a pass or good score, and anything below 50 is considered "0" indicating a fail. Then, we can discard the "score" attribute as it is no longer necessary. For both datasets, we have set the time in seconds, minutes, and hours. Now that we have identified and pre-processed the issues in both datasets, we can apply machine learning.

4.2 Transfer Learning

4.2.1 Training the nurses dataset using ANN

First, we train the model using standard ANN methods, including rule and softmax. Next, we apply RobustScaler, which uses the median and the interquartile range (IQR) for scaling. We use RobustScaler to transform the dataset to ensure that features with outliers do not dominate the learning process. Figure 4.3 illustrates the accuracy of our data, using a confusion matrix to understand how data are labeled. The confusion matrix, presented in Figure 4.2, includes True positive (TP), False negative (FN), False positive (FP), and True negative (TN).

The results display a prediction **accuracy of 99.40%** for stress levels. The **recall**, which is the ratio of correctly predicted positive observations to all observations in the actual class,

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)}$$

and **precision**, which is the ratio of correctly predicted positive observations to the total predicted positive observations,

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)}$$

both resulted in 99.40%.

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

Figure 4.2: Confusion Matrix

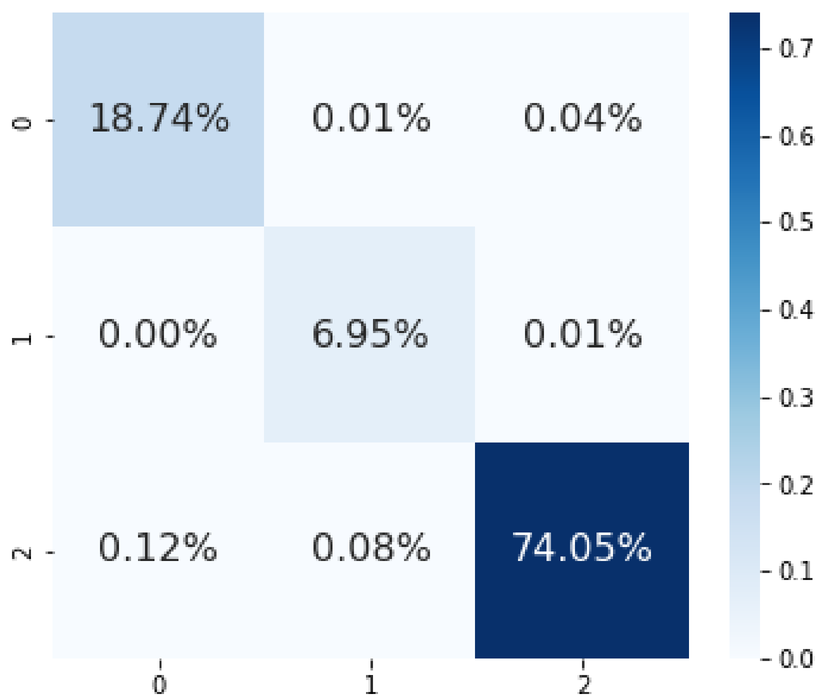


Figure 4.3: The model results in a confusion matrix.

4.2.2 Features model impact in the nurses data

To better understand feature importance and model decision-making, we used SHAP values, as shown in Figure 4.4. These plots revealed that time(h) was important in predicting stress. The other significant attributes were TEMP and EDA, aligning with the results of our feature selection methods.

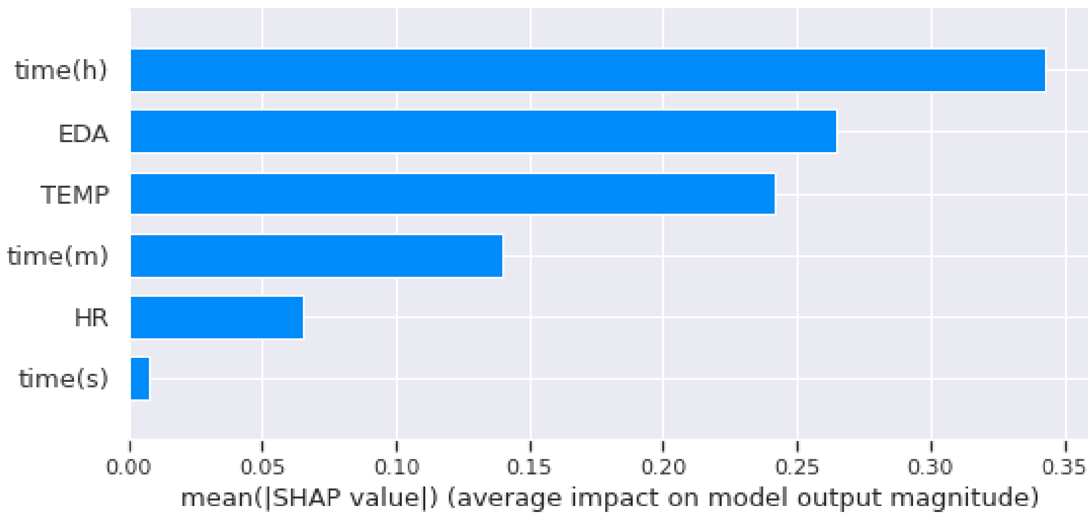


Figure 4.4: Feature influences with SHAP on both classes, with ANN model

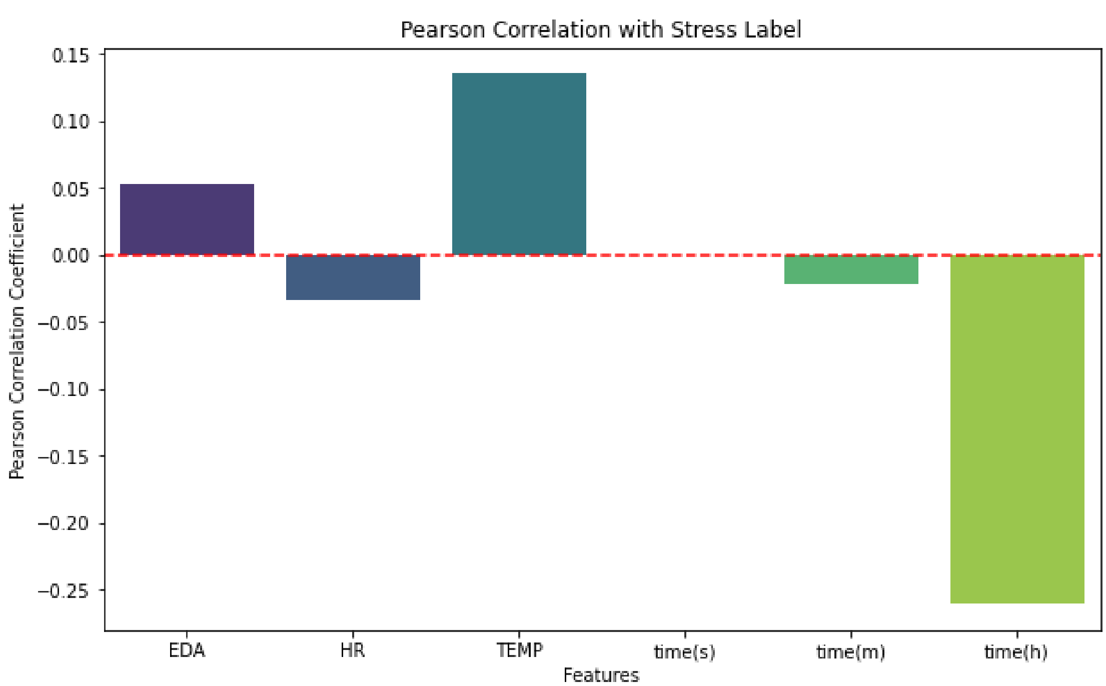


Figure 4.5: Person's Correlation for stress levels.

4.2.3 Comparing Transfer Learning with other models

We utilized the sklearn library for generating models such as Decision Tree and SVM. Additionally, we used statsmodels.api to obtain metrics such as accuracy, recall, and the F-test. As shown in Figure 4.6, the Random Forest model achieved an accuracy of 100%, and the KNN model scored 99.98%. This indicates that our model ranks third in comparison to the current models. However, all of these models, including ours, exhibit a high accuracy rate of approximately 99%. Thus, we can confidently anticipate that our model will produce promising predictions.

	Accuracy_train	Accuracy_test	ROC_AUC_test	F1_test	MCC_test
random_forest	1.000000	0.999996	1.000000	0.999996	0.999989
knn	0.999850	0.999138	0.999998	0.999139	0.997892
decision_tree	0.884960	0.834372	0.965925	0.844467	0.674328
mlp	0.793534	0.722155	0.893049	0.744087	0.512418
logistic	0.563793	0.456754	0.713478	0.514044	0.228119

Figure 4.6: Comparing other models against tranfer learning model

4.2.4 Applying Transfer Learning to students data

After training our model, we utilize the transfer learning model to enhance students' data for predicting exam performance. This requires Fine-Tuning techniques. We have selected three types for this task as detailed in Table 4.1: updating the last layer, employing different learning rates, and gradually unfreezing. The most effective was gradual unfreezing, yielding a score of 99.71% in recall, precision, and accuracy. A different learning rate resulted in approximately 86% accuracy and recall, but achieved 88.86% in precision. Updating the last layer had a poor performance, with only 55% accuracy and recall, and 66.85% precision.

	Updating last Layer	Different Learning Rates	Gradually Unfreezing
Accuracy	55.4%	86.05%	99.71%
Precision	66.85%	88.86%	99.71%
Recall	55.04%	86.51%	99.71%
F1-score	57.48%	87.16%	99.71%

Table 4.1: The results of each Fine-Tuning method

4.3 Exam performance under stress

4.3.1 Predicting students' stress level

The data presented in Table 4.2 highlight a significant observation that all students, regardless of their academic performance, experienced a comparable level of stress. This striking similarity in stress levels among students suggests that there may not be a substantial correlation between the level of stress experienced and the results of exams. However, while stress may not have an immediate and noticeable impact on academic outcomes, it is important to emphasize that repeated exposure to high levels of stress can lead to serious long-term health consequences, such as acute cardiovascular events.

Performance	Score	Prediction of Stress level
0	36.21	1.748
1	63.64	1.715
2	85.52	1.764

Table 4.2: Average score and stress level within each performance category

4.3.2 Predicting exam performance using the nurses' dataset

Table 4.3 presents the average stress levels nurses experienced during the experiments and their expected exam performance. The prediction indicates that even nurses under high stress can perform well, as evidenced by subject 83 and BG. However, other high-stress subjects, like CE, may not perform as well. On the other hand, Table 4.4 focuses on the relationship between nurses' stress levels and their expected performance. The data does not show any significant difference between no stress and high stress levels.

id	label	Predicted$_{P}erformance$
15	1.897453	1.203411
83	1.750970	1.659708
94	0.502339	1.099018
5C	0.944572	1.116654
6B	1.737068	1.017596
6D	0.298707	1.139123
7A	1.666811	1.233481
7E	1.075752	0.966068
8B	1.347546	1.025967
BG	1.930595	1.699058
CE	1.976414	0.928320
DF	1.856234	1.112073
E4	1.611198	1.294014
EG	2.000000	1.045276
F5	1.869921	1.230876

Table 4.3: The average expected score and stress level for each nursing subject.

Stress Level	Expected exam performance
0	1.175448
1	1.102262
2	1.249154

Table 4.4: Average stress and average expected exam scores

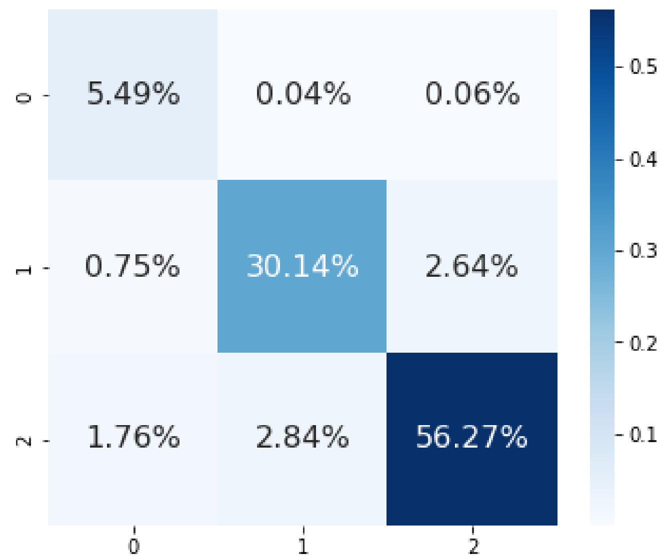


Figure 4.7: Confusion matrix for predicting exam performance using the nurses dataset.

4.3.3 Training the nurses dataset.

As indicated in Table 4.5, the model accurately predicted the exam performance using the nurses' dataset, with an accuracy and recall of 91.8% and a precision of 92.52%. This suggests that our model can be applied to other datasets with a prediction accuracy of almost 91%.

	Gradually Unfreezing
Accuracy	91.89%
Precision	92.52%
Recall	91.89%
F1-score	92.05%

Table 4.5: The results of each Fine-Tuning method

Chapter 5

Discussion

Test stress may not significantly impact exam results, possibly due to a student's pre-existing knowledge. We found no direct correlation between stress and exam performance. Training in study strategies can enhance achievement, while interventions targeting stress and anxiety might not address knowledge gaps. Several other studies suggest a negative correlation, with less anxious individuals often performing better. However, less prepared students may be more anxious, which could skew the correlations. It is possible that only acute stressors impact performance, and coping styles, skills, and social support might mitigate such stress effects. We addressed class imbalance in our study through data manipulation and feature selection. Our model achieved recall and precision scores of 91.89% and 92.52%, respectively, indicating the effectiveness of the model. For optimal use of transfer learning to train other datasets, it is highly recommended to use fine-tuning methods, particularly gradual unfreezing, to achieve the best results.

5.1 Final determination of the impact of anxiety and stress on exam results

It was initially hypthosized that highly anxious students would perform worse in exams. However, our study's findings suggest that test anxiety does not significantly impact exam results, possibly due to the influence of pre-existing knowledge. Researchers Theobald, Maria, Jasmin Breitwieser, and Garvin Brod from Germany proposed that well-managed knowledge could negate the severe impacts of anxiety or stress on exam results [32]. Their research involved 309 medical students and used mock and actual exams to test if students with high test anxiety and high working memory capacity performed better due to less interference. They did not find a direct correlation between test anxiety and lower final exam scores. Instead, their findings suggest that students with test anxiety might struggle more with encoding, storing, and organizing information rather than simply recalling it. These students reported weak study skills and the study found they learned less during exam preparation, potentially leading to increased anxiety and lower scores. However, they concluded that test anxiety might not affect exam performance as significantly as previously assumed. Further research with diverse samples is still necessary to fully understand these relationships. Their study suggests to support students with high test anxiety, it might be more beneficial to provide strategies for more effective learning, rather than merely reducing anxiety during exams. Training on study strategies could enhance academic achievement, while interventions targeting test anxiety might not compensate for knowledge gaps. Future research should focus on interventions that improve learning during exam preparation or even earlier.

Furthermore, a 2023 study by John Jerrim also found a negative correlation, with less anxious individuals often performing better. However, he also suggested that less prepared students tend to be more anxious, which could potentially skew this relationship [33]. Shah et al. (2010) have also reported a negative correlation between perceived stress and academic performance, which was not significant [34]. The authors suggested that only acute stressors may be responsible for affecting academic performance. In addition, students who are striving hard to perform well in the exam may be stressed. Their Individual coping styles and skills along with their access to different forms of social support may play a role in negating the effect of stress on academic performance. Sansgiry et al. (2006) also found an insignificant negative correlation between academic performance and test anxiety [35]. The authors suggested an efficient counseling service in their institution as one of the causes of not having a significant association between performance and test anxiety. According to them, such programs are efficient in improving academic performance, which has also been supported by others [36]. Therefore, it is best to develop interventions that facilitate effective knowledge acquisition during exam preparation, or even earlier. This can improve the academic prospects of students who experience high test anxiety.

5.2 Model explainability

5.2.1 Feature Selections

Data manipulation and feature selection can effectively address unbalanced datasets. An accelerometer tracks and understands motion during stressful events. Hosseini et al. (2022) noted an instance where the model indicated high stress, but the subject was merely having a lunch break. This observation led to the creation of a survey to better determine stress levels. However, in our case, we do not require an accelerometer. Additionally, inter-beat interval data and Blood Volume Pulse were used to calculate average heart rate. Thus, we can focus primarily on heart rate, disregarding the other two features. In Figure 4.4, time plays a significant role in our model. If removed, the accuracy, recall, and precision would all drop to roughly 82% — similar to the results by Amin et al. (2022) — when determining students' stress levels. The same applies to predicting performance, with an accuracy of up to 80%. Time is important to establish a baseline and allow the algorithm to detect significant elevations in EDA, HR, and temperature. However, it is important to note that heart rate and time in minutes and seconds did not significantly contribute to stress identification in our model.

5.2.2 Class Imbalance

Class imbalance is a frequent issue with large datasets. For example, Figure 4.1 in Hosseini et al. (2022) highlights an elevated stress level amongst most nurses due to the COVID-19 pandemic. This was mainly due to the stress of adhering to COVID-19 precautions and restrictions while caring for patients. To address class imbalance, we utilized SMOTETomek, which merges over- and under-sampling techniques. Because of class imbalance, the model's accuracy isn't the only indicator of its quality. Therefore, the use of recall and precision becomes crucial. In terms of predicting exam performance, our model achieved recall and precision scores of 91.89% and 92.52%, respectively. This suggests that our model is effective for learning from other data sets to predict exam results based on the selected characteristics EDA, Temp, HR, and Time. On the other hand, our training for the stress model resulted in approximately 99.40% for both precision and recall. Therefore, we can also train other datasets to identify stress levels.

When examining the student data, it is important to take into account the sample size. Currently, we have data from only 10 participants, stemming from 3 exams of undisclosed subjects. For a more comprehensive understanding, it would be beneficial to have data from a larger participant pool across various subjects, such as medicine, math, or physics. Thus, we should aim to obtain a larger sample size and data from multiple subjects to more accurately assess exam performance.

5.2.3 Transfer Learning Model

Our model demonstrated strong performance in terms of accuracy, recall, and precision. However, to apply this model learning to new datasets, appropriate fine-tuning methods are essential. We chose three fine-tuning methods for our experiment, as outlined in Table 4.1 for predicting stress levels. The most commonly used method is updating the last layer, but it under-performed in comparison to the other two - Different Learning Rates and Gradually Unfreezing. Different Learning Rates achieved a score of about 86%, while Gradually Unfreezing scored 99.71% across all three criteria. Therefore, to achieve the best results when training new datasets, it is recommended to employ fine-tuning, particularly the Gradually Unfreezing method. With Gradually Unfreezing identified as the most effective fine-tuning method, we can apply this approach to predict exam performance, where it achieved a score of roughly 92% as shown in Table 4.5. Consequently, we can leverage our transfer model to train for exam performance or stress level measurements, achieving accuracy, recall, and precision rates that exceed 90%.

Chapter 6

Conclusions

This study investigated the impact of stress on academic performance, focusing on the relationship between physiological stress markers and exam results. The primary aim was to understand whether test anxiety significantly affects academic outcomes and to evaluate the efficacy of machine learning models in predicting exam performance under stress. A central component of the research was the development and application of a machine learning model, specifically designed to analyze data from wearable devices that monitor physiological indicators of stress. The model incorporated transfer learning techniques to enhance its predictive accuracy. Transfer learning allowed the model to leverage knowledge gained from one dataset and apply it to another, improving the robustness and adaptability of the predictions. Key findings indicate that there is no direct correlation between test anxiety and exam performance. This aligns with the hypothesis that pre-existing knowledge plays a more significant role in determining exam outcomes than anxiety levels. Studies cited support these results, suggesting that students with high working memory capacity and test anxiety can manage to perform well by effectively controlling knowledge interference. The machine learning model demonstrated high effectiveness, with recall and precision scores of 91.89% and 92.52%, respectively. This indicates the model's strong capability to accurately predict exam performance based on stress levels measured by wearable devices. The sample size from Amin et all [17] is currently limited, comprising just 10 participants and 3 exams from unidentified subjects. To massively improve the robustness of our analysis, we should aim to increase this sample size and include participants from a wider range of subject areas. Fine-tuning methods, especially gradual unfreezing, were pivotal in achieving these high performance metrics, highlighting the importance of tailored model training strategies for specific datasets. The results suggest that interventions aimed at improving study strategies could be more effective than those solely focusing on reducing test anxiety. By enhancing learning techniques, students may better prepare for exams, thereby indirectly reducing anxiety and improving performance. Future research should explore diverse samples and interventions that target effective learning methods, providing a more comprehensive understanding of the relationship between stress, anxiety, and academic performance. Our study emphasized the complexity of the relationship between stress, anxiety, and academic performance. While acute stressors might impact performance, the overall influence of anxiety on exam results is mitigated by factors such as pre-existing knowledge and study habits. The research suggests that less anxious individuals tend to perform better, yet this could be influenced by their preparedness rather than anxiety alone. Therefore, future studies should focus on developing comprehensive support systems that enhance students' learning processes and coping mechanisms, providing a holistic approach to academic success.

Bibliography

[1] T. Iqbal, A. J. Simpkin, D. Roshan, N. Glynn, J. Killilea, J. Walsh, G. Molloy, S. Ganly, H. Ryman, E. Coen *et al.*, "Stress monitoring using wearable sensors: a pilot study and stress-predict dataset," *Sensors*, vol. 22, no. 21, p. 8135, 2022.

[2] S. Campanella, A. Altaleb, A. Belli, P. Pierleoni, and L. Palma, "A method for stress detection using empatica e4 bracelet and machine-learning techniques," *Sensors*, vol. 23, no. 7, p. 3565, 2023.

[3] S. Ali, "Exam stress and its impact on academic performance," 2022. [Online]. Available: https://www.fccollege.edu.pk/wp-content/uploads/Final-S.-Basit-Ali.pdf

[4] N. Šimić and I. Manenica, "Exam experience and some reactions to exam stress," *Human physiology*, vol. 38, pp. 67–72, 2012.

[5] K. R. Wentzel and A. Wigfield, "Academic and social motivational influences on students' academic performance," *Educational psychology review*, vol. 10, pp. 155–175, 1998.

[6] R. Singh, M. Goyal, S. Tiwari, A. Ghildiyal, S. M. Nattu, and S. Das, "Effect

of examination stress on mood, performance and cortisol levels in medical students," *Indian J Physiol Pharmacol*, vol. 56, no. 1, pp. 48–55, 2012.

[7] A. Abaidoo, *Factors contributing to academic performance of students in a Junior High School*. Grin Verlag, 2018.

[8] B. Nascimento, T. Oliveira, and C. Tam, "Wearable technology: What explains continuance intention in smartwatches?" *Journal of Retailing and Consumer Services*, vol. 43, pp. 157–169, 2018.

[9] R. Llamas, J. Ubrani, and M. Shirer, "Idc reports strong growth in the worldwide wearables market, led by holiday shipments of smartwatches, wrist bands, and ear-worn devices," *Repéré à https://www. idc. com/getdoc. Jsp*, 2019.

[10] L. Goasduff, "Gartner says global end-user spending on wearable devices to total $52 billion in 2020. gartner,''$ 2019.

[11] S. Elzeiny and M. Qaraqe, "Machine learning approaches to automatic stress detection: A review," in *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, 2018, pp. 1–6.

[12] A. Tazarv, S. Labbaf, S. M. Reich, N. Dutt, A. M. Rahmani, and M. Levorato, "Personalized stress monitoring using wearable sensors in everyday settings," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 7332–7335.

[13] A. Hasanbasic, M. Spahic, D. Bosnjic, V. Mesic, O. Jahic *et al.*, "Recognition of stress levels among students with wearable sensors," in *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*. IEEE, 2019, pp. 1–4.

[14] A. A. Al-Atawi, S. Alyahyan, M. N. Alatawi, T. Sadad, T. Manzoor, M. Farooq-i Azam, and Z. H. Khan, "Stress monitoring using machine learning, iot and wearable sensors," *Sensors*, vol. 23, no. 21, p. 8875, 2023.

[15] M. R. Amin, D. S. Wickramasuriya, and R. T. Faghih, "A wearable exam stress dataset for predicting grades using physiological signals," in *2022 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*. IEEE, 2022, pp. 30–36.

[16] S. Hosseini, R. Gottumukkala, S. Katragadda, R. T. Bhupatiraju, Z. Ashkar, C. W. Borst, and K. Cochran, "A multimodal sensor dataset for continuous stress detection of nurses in a hospital," *Scientific Data*, vol. 9, no. 1, p. 255, 2022.

[17] M. R. Amin, D. S. Wickramasuriya, and R. T. Faghih, "A wearable exam stress dataset for predicting cognitive performance in real-world settings," 2022. [Online]. Available: https://physionet.org/content/wearable-exam-stress/1.0.0/

[18] S. Hosseini and R. Gottumukkala, "Nurse stress prediction wearable sensors," 2023. [Online]. Available: https://www.kaggle.com/dsv/7125235

[19] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti, "Empatica e3—a wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition," in *2014 4th international conference on wireless mobile communication and healthcare-transforming healthcare through innovations in mobile and wireless technologies (MOBIHEALTH)*. IEEE, 2014, pp. 39–42.

[20] A. Bizzego, G. Gabrieli, C. Furlanello, and G. Esposito, "Comparison of wearable and clinical devices for acquisition of peripheral nervous system signals," *Sensors*, vol. 20, no. 23, p. 6778, 2020.

[21] 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.

[22] Y. S. Can, B. Arnrich, and C. Ersoy, “Stress detection in daily life scenarios using smart phones and wearable sensors: A survey,” *Journal of biomedical informatics*, vol. 92, p. 103139, 2019.

[23] E. S. Epel, A. D. Crosswell, S. E. Mayer, A. A. Prather, G. M. Slavich, E. Puterman, and W. B. Mendes, “More than a feeling: A unified view of stress measurement for population science,” *Frontiers in neuroendocrinology*, vol. 49, pp. 146–169, 2018.

[24] R. M. Kaplan, M. L. Spittel, and D. H. David, *Population health: behavioral and social science insights*. Government Printing Office, 2015.

[25] G. Miller, E. Chen, and S. W. Cole, “Health psychology: Developing biologically plausible models linking the social world and physical health,” *Annual review of psychology*, vol. 60, pp. 501–524, 2009.

[26] W. T. Boyce, “Epigenomics and the unheralded convergence of the biological and social sciences,” *Population Health: Behavioral and Social Science Insights. Rockville, MD: Agency for Healthcare Research and Quality and Office of Behavioral and Social Sciences Research, National Institutes of Health*, pp. 219–232, 2015.

[27] B. S. McEwen, “The brain on stress: How behavior and the social environment “get under the skin.”,” *Population Health: Behavioral and Social Science Insights*, vol. 233, 2015.

[28] V. R. LeBlanc, “The effects of acute stress on performance: implications for health professions education,” *Academic Medicine*, vol. 84, no. 10, pp. S25–S33, 2009.

[29] H. Elias, W. S. Ping, and M. C. Abdullah, "Stress and academic achievement among undergraduate students in universiti putra malaysia," *Procedia-Social and Behavioral Sciences*, vol. 29, pp. 646–655, 2011.

[30] P. Kudachi, R. Latti, and S. S. Goudar, "Effect of examination stress on the academic performance of first year medical students," *Biomedicine*, vol. 28, no. 2, pp. 142–144, 2008.

[31] S. Karingula. (2017) Nz uni vs usa college ?! [Online]. Available: https://www.theinsideword.ac.nz/2017/08/nz-uni-vs-usa-college/

[32] M. Theobald, J. Breitwieser, and G. Brod, "Test anxiety does not predict exam performance when knowledge is controlled for: Strong evidence against the interference hypothesis of test anxiety," *Psychological Science*, vol. 33, no. 12, pp. 2073–2083, 2022.

[33] J. Jerrim, "Test anxiety: Is it associated with performance in high-stakes examinations?" *Oxford Review of Education*, vol. 49, no. 3, pp. 321–341, 2023.

[34] M. Shah, S. Hasan, S. Malik, and C. T. Sreeramareddy, "Perceived stress, sources and severity of stress among medical undergraduates in a pakistani medical school," *BMC medical education*, vol. 10, pp. 1–8, 2010.

[35] S. S. Sansgiry, M. Bhosle, and K. Sail, "Factors that affect academic performance among pharmacy students," *American journal of pharmaceutical education*, vol. 70, no. 5, 2006.

[36] K. Henning, S. Ey, and D. Shaw, "Perfectionism, the impostor phenomenon and psychological adjustment in medical, dental, nursing and pharmacy students," *Medical education*, vol. 32, no. 5, pp. 456–464, 1998.

[37] L. Richard, T. Hurst, and J. Lee, "Lifetime exposure to abuse, current stressors, and health in federally qualified health center patients," *Journal of Human Behavior in the Social Environment*, vol. 29, no. 5, pp. 593–607, 2019.

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

[39] E. Hemmingsson, "Early childhood obesity risk factors: socioeconomic adversity, family dysfunction, offspring distress, and junk food self-medication," *Current obesity reports*, vol. 7, pp. 204–209, 2018.

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

[41] M. Kivimäki and A. Steptoe, "Effects of stress on the development and progression of cardiovascular disease," *Nature Reviews Cardiology*, vol. 15, no. 4, pp. 215–229, 2018.

[42] J. Aguiló, P. Ferrer-Salvans, A. García-Rozo, A. Armario, Á. Corbi, F. J. Cambra, R. Bailón, A. González-Marcos, G. Caja, S. Aguiló *et al.*, "Project es3: attempting to quantify and measure the level of stress," *Rev Neurol*, vol. 61, no. 9, pp. 405–415, 2015.

[43] B. S. McEwen, "Stressed or stressed out: what is the difference?" *Journal of Psychiatry and Neuroscience*, vol. 30, no. 5, pp. 315–318, 2005.

[44] A. Mariotti, "The effects of chronic stress on health: new insights into the molecular mechanisms of brain–body communication," *Future science OA*, vol. 1, no. 3, 2015.

[45] W. H. Organization *et al.*, "World health statistics data visualizations dashboard," *Tobacco Control*, 2018.

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

[47] P. Verma and S. K. Sood, "A comprehensive framework for student stress monitoring in fog-cloud iot environment: m-health perspective," *Medical & biological engineering & computing*, vol. 57, pp. 231–244, 2019.

[48] D. Hellhammer, A. Stone, J. Hellhammer, and J. Broderick, "Measuring stress," *Encyclopedia of behavioral neuroscience*, vol. 2, pp. 186–191, 2010.

[49] B. Cvetković, M. Gjoreski, J. Šorn, P. Maslov, M. Kosiedowski, M. Bogdański, A. Stroiński, and M. Luštrek, "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.

[50] F. Seoane, I. Mohino-Herranz, J. Ferreira, L. Alvarez, R. Buendia, D. Ayllón, C. Llerena, and R. Gil-Pita, "Wearable biomedical measurement systems for assessment of mental stress of combatants in real time," *Sensors*, vol. 14, no. 4, pp. 7120–7141, 2014.

[51] L. Han, Q. Zhang, X. Chen, Q. Zhan, T. Yang, and Z. Zhao, "Detecting work-related stress with a wearable device," *Computers in Industry*, vol. 90, pp. 42–49, 2017.

[52] 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," *Biomedical Signal Processing and Control*, vol. 6, no. 3, pp. 261–268, 2011.

[53] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *Journal of biomedical informatics*, vol. 73, pp. 159–170, 2017.

[54] E. L. van den Broek, F. van der Sluis, and T. Dijkstra, "Cross-validation of bimodal health-related stress assessment," *Personal and Ubiquitous Computing*, vol. 17, pp. 215–227, 2013.

[55] 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)*. IEEE, 2019, pp. 1–5.

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

[57] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE transactions on affective computing*, vol. 13, no. 1, pp. 440–460, 2019.

[58] S. Pourmohammadi and A. Maleki, "Stress detection using ecg and emg signals: A comprehensive study," *Computer methods and programs in biomedicine*, vol. 193, p. 105482, 2020.

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

[60] K. A. Herborn, J. L. Graves, P. Jerem, N. P. Evans, R. Nager, D. J. McCafferty, and D. E. McKeegan, "Skin temperature reveals the intensity of acute stress," *Physiology & behavior*, vol. 152, pp. 225–230, 2015.

[61] F. de Arriba-Pérez, J. M. Santos-Gago, M. Caeiro-Rodríguez, and M. Ramos-Merino, "Study of stress detection and proposal of stress-related features using commercial-off-the-shelf wrist wearables," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, pp. 4925–4945, 2019.

[62] 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 pressure*, vol. 13, no. 5, pp. 287–294, 2004.

[63] J. B. Drummond, B. S. Soares, W. Pedrosa, E. L. Vieira, A. L. Teixeira, M. Christ-Crain, and A. Ribeiro-Oliveira, "Copeptin response to hypoglycemic stress is linked to prolactin activation in children," *Pituitary*, vol. 23, pp. 681–690, 2020.

[64] B. Valentin, O. Grottke, M. Skorning, S. Bergrath, H. Fischermann, D. Rörtgen, M.-T. Mennig, C. Fitzner, M. P. Müller, C. Kirschbaum *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," *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 23, pp. 1–8, 2015.

[65] L. Thau, J. Gandhi, and S. Sharma, "Physiology, cortisol," 2019.

[66] 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 and Health: Journal of the International Society for the Investigation of Stress*, vol. 20, no. 2, pp. 91–98, 2004.

[67] 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," *Neuroscience & biobehavioral reviews*, vol. 38, pp. 94–124, 2014.

[68] 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," *Biological Psychology*, vol. 127, pp. 34–39, 2017.

[69] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *Journal of biomedical informatics*, vol. 59, pp. 49–75, 2016.

[70] T. Iqbal, "Signal processing and machine learning algorithms for stress monitoring using wearable sensor technologies," 2023.

[71] J. Kim, J. Park, and J. Park, "Development of a statistical model to classify driving stress levels using galvanic skin responses," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 30, no. 5, pp. 321–328, 2020.

[72] 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*, vol. 19, no. 9, p. 2152, 2019.

[73] T. Li, Y. Chen, and W. Chen, "Daily stress monitoring using heart rate variability

of bathtub ecg signals," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 2699–2702.

[74] M. Choi, G. Koo, M. Seo, and S. W. Kim, "Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 3, pp. 634–645, 2017.

[75] D. S. Lee, T. W. Chong, and B. G. Lee, "Stress events detection of driver by wearable glove system," *IEEE Sensors Journal*, vol. 17, no. 1, pp. 194–204, 2016.

[76] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss, "Activity-aware mental stress detection using physiological sensors," in *Mobile Computing, Applications, and Services: Second International ICST Conference, MobiCASE 2010, Santa Clara, CA, USA, October 25-28, 2010, Revised Selected Papers 2*. Springer, 2012, pp. 282–301.

[77] J. Healey and R. Picard, "Smartcar: detecting driver stress," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 4. IEEE, 2000, pp. 218–221.

[78] A. Arza, J. M. Garzón-Rey, J. Lázaro, E. Gil, R. Lopez-Anton, C. de la Camara, P. Laguna, R. Bailon, and J. Aguiló, "Measuring acute stress response through physiological signals: towards a quantitative assessment of stress," *Medical & biological engineering & computing*, vol. 57, pp. 271–287, 2019.

[79] L. Bönke, S. Aust, Y. Fan, K. Wirth, E. Khawli, A. Stevense, A. Herrera, A. Loayza, M. Bajbouj, and S. Grimm, "Examining the effect of early life stress on autonomic and endocrine indicators of individual stress reactivity," *Neurobiology of stress*, vol. 10, p. 100142, 2019.

[80] 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 journal of biomedical and health informatics*, vol. 23, no. 2, pp. 703–713, 2018.

[81] E. Ullmann, A. Barthel, K. Petrowski, T. Stalder, C. Kirschbaum, and S. Bornstein, “Pilot study of adrenal steroid hormones in hair as an indicator of chronic mental and physical stress,” *Scientific reports*, vol. 6, no. 1, p. 25842, 2016.

[82] 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.

[83] M. Uesato, Y. Nabeya, T. Akai, M. Inoue, Y. Watanabe, H. Kawahira, T. Mamiya, Y. Ohta, R. Motojima, A. Kagaya *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, pp. 84–89, 2010.

[84] 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,” *Cerebrovascular diseases*, vol. 18, no. 3, pp. 206–213, 2004.

[85] Y. Shi, M. H. Nguyen, P. Blitz, B. French, S. Fisk, F. De la Torre, A. Smailagic, D. P. Siewiorek, M. Al’Absi, E. Ertin *et al.*, “Personalized stress detection from physiological measurements,” in *International symposium on quality of life technology*, vol. 1, 2010, pp. 28–29.

[86] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical and biological engineering and computing*, vol. 42, pp. 419–427, 2004.

[87] Y. Liu and S. Du, "Psychological stress level detection based on electrodermal activity," *Behavioural brain research*, vol. 341, pp. 50–53, 2018.

[88] L.-l. Chen, Y. Zhao, P.-f. Ye, J. Zhang, and J.-z. Zou, "Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers," *Expert Systems with Applications*, vol. 85, pp. 279–291, 2017.

[89] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.

[90] 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.

[91] S. Minaee, "popular machine learning metrics. part 1: Classification & regression evaluation metrics," in *Towards Data Science. URL https://towardsdatascience.com/20-popular-machine-learning-metrics-part-1-classification-regression-evaluation-metrics-1ca3 Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI*, vol. 19, 2019.

[92] J. Jordan, "Evaluating a machine learning model. 2017," *Reference Source*, 2022.

[93] W. Boucsein, *Electrodermal activity.* Springer Science & Business Media, 2012.

[94] M. E. Dawson, A. M. Schell, D. L. Filion *et al.*, "The electrodermal system," *Handbook of psychophysiology*, vol. 2, pp. 200–223, 2007.

[95] H. D. Critchley, "Electrodermal responses: what happens in the brain," *The Neuroscientist*, vol. 8, no. 2, pp. 132–142, 2002.

[96] Z. Villines and S. Flack, "Why do we use pulse oximetry," *Medical News Today*, 2017.

[97] M. Paoletti, A. Belli, L. Palma, M. Vallasciani, and P. Pierleoni, "A wireless body sensor network for clinical assessment of the flexion-relaxation phenomenon," *Electronics*, vol. 9, no. 6, p. 1044, 2020.

[98] D. Schomer, "Niedermeyer's electroencephalography; schomer, dl, lopes da silva, fh, eds," 2017.

[99] P. Venables and M. Christie, "Electrodermal activity (1980) techniques in psychophysiology," *Martion I., and*, pp. 3–67.

[100] G. B. Drummond, D. Fischer, and D. Arvind, "Current clinical methods of measurement of respiratory rate give imprecise values," *ERJ open research*, vol. 6, no. 3, 2020.

[101] T. Iqbal, A. Elahi, A. Shahzad, and W. Wijns, "Review on classification techniques used in biophysiological stress monitoring," *arXiv preprint arXiv:2210.16040*, 2022.

[102] M. Laudat, S. Cerdas, C. Fournier, D. Guiban, B. Guilhaume, and J. Luton, "Salivary cortisol measurement: a practical approach to assess pituitary-adrenal function," *The Journal of Clinical Endocrinology & Metabolism*, vol. 66, no. 2, pp. 343–348, 1988.

[103] P. Braveman and L. Gottlieb, "The social determinants of health: it's time to consider the causes of the causes," *Public health reports*, vol. 129, no. 1_suppl2, pp. 19–31, 2014.

[104] S. Salimetrics, ""salivary cortisol"," 2019.

[105] M. Nagananda *et al.*, "Quantization of mental stress using various physiological markers," *arXiv preprint arXiv:1504.03343*, 2015.

[106] K. W. Kallus and M. Kellmann, *The recovery-stress questionnaires: user manual.* Pearson London, UK:, 2016.

[107] P. M. Boynton and T. Greenhalgh, "Selecting, designing, and developing your questionnaire," *Bmj*, vol. 328, no. 7451, pp. 1312–1315, 2004.

[108] 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.

[109] R. R. Bouckaert, "Choosing between two learning algorithms based on calibrated tests," in *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, 2003, pp. 51–58.

[110] S. K. Murthy, "Automatic construction of decision trees from data: A multi-disciplinary survey," *Data mining and knowledge discovery*, vol. 2, pp. 345–389, 1998.

[111] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review

of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, pp. 159–190, 2006.

[112] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning in international conference on artificial neural networks," *Cham, Switzerland: Springer*, pp. 270–279, 2018.

[113] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," *arXiv preprint arXiv:1801.06146*, 2018.

[114] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[115] D. Sarkar, R. Bali, and T. Ghosh, *Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras*. Packt Publishing Ltd, 2018.

[116] H. Parvin, M. Mohamadi, S. Parvin, Z. Rezaei, and B. Minaei, "Nearest cluster classifier," in *Hybrid Artificial Intelligent Systems: 7th International Conference, HAIS 2012, Salamanca, Spain, March 28-30th, 2012. Proceedings, Part I 7*. Springer, 2012, pp. 267–275.

[117] S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, "An overview of principal component analysis," *Journal of Signal and Information Processing*, vol. 4, 2020.

[118] 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 Journal of Archaeology of Egypt/Egyptology*, vol. 17, no. 9, pp. 2353–2360, 2020.

[119] A. Navlani, "Understanding logistic regression in python," *Link https//www. datacamp. com/community/tutorials/understanding-logistic-regressionpython\# comments. Udgivet*, vol. 7, 2018.

[120] 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. 16 568–16 575, 2017.

[121] S. R. Gomes, S. G. Saroar, M. Mosfaiul, A. Telot, B. N. Khan, A. Chakrabarty, and M. Mostakim, "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)*. IEEE, 2017, pp. 482–487.

[122] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.

[123] O. M. Mozos, V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, R. Dobrescu, and J. M. Ferrandez, "Stress detection using wearable physiological and sociometric sensors," *International journal of neural systems*, vol. 27, no. 02, p. 1650041, 2017.

[124] M. J. Begemann, E. J. Florisse, R. van Lutterveld, M. Kooyman, and I. E. Sommer, "Efficacy of eeg neurofeedback in psychiatry: A comprehensive overview and meta-analysis," *Translational Brain Rhythmicity*, vol. 1, no. 1, pp. 19–29, 2016.

[125] D. Cogan, M. B. Pouyan, M. Nourani, and J. Harvey, "A wrist-worn biosensor system for assessment of neurological status," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 5748–5751.

[126] K. Woodward, E. Kanjo, D. J. Brown, and T. McGinnity, "On-device transfer learning for personalising psychological stress modelling using a convolutional neural network," *arXiv preprint arXiv:2004.01603*, 2020.

[127] J. Wu, Y. Zhang, and X. Zhao, "Stress detection using wearable devices based on transfer learning," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2021, pp. 3122–3128.

[128] A. Barreto, J. Zhai, and M. Adjouadi, "Non-intrusive physiological monitoring for automated stress detection in human-computer interaction," in *Human–Computer Interaction: IEEE International Workshop, HCI 2007 Rio de Janeiro, Brazil, October 20, 2007 Proceedings 4*. Springer, 2007, pp. 29–38.

[129] 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)*. IEEE, 2016, pp. 3805–3808.

[130] C. Dobbins and S. Fairclough, "Detecting negative emotions during real-life driving via dynamically labelled physiological data," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 2018, pp. 830–835.

[131] 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. 13 545–13 556, 2017.

[132] O. Oti, I. Azimi, A. Anzanpour, A. M. Rahmani, A. Axelin, and P. Liljeberg, "Iot-based healthcare system for real-time maternal stress monitoring," in *Proceedings of*

the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies, 2018, pp. 57–62.

[133] S. Betti, R. M. Lova, E. Rovini, G. Acerbi, L. Santarelli, M. Cabiati, S. Del Ry, and F. Cavallo, "Evaluation of an integrated system of wearable physiological sensors for stress monitoring in working environments by using biological markers," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 8, pp. 1748–1758, 2017.

[134] C. D. B. S. J. B. K. C. N. D. B. L. F. B. L. B. S. B. A. B. K. K. M. S. Norman B. A., Suzanne B J and A. V., "Stress in america: Our health at risk," *American Psychology Association*, 2011.

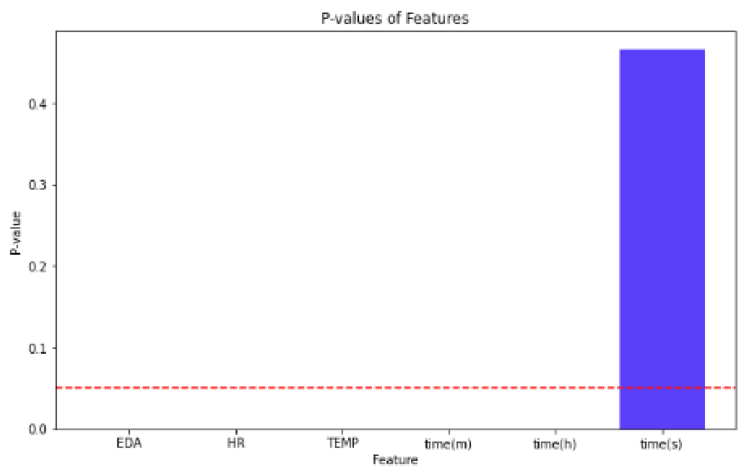
Appendix A

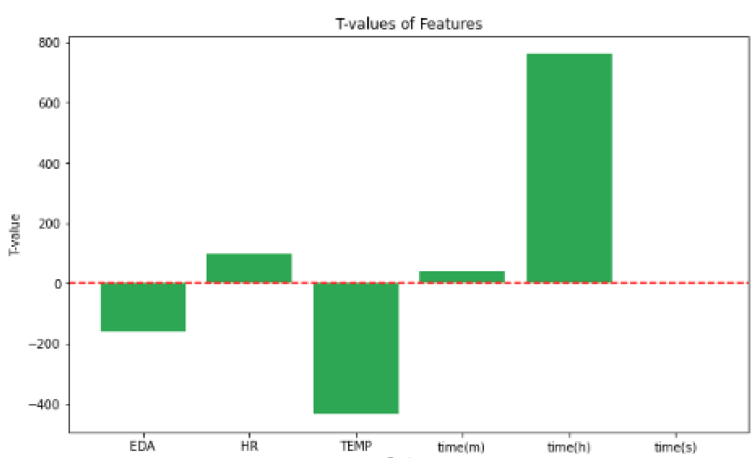
APPENDIX A: Modeling

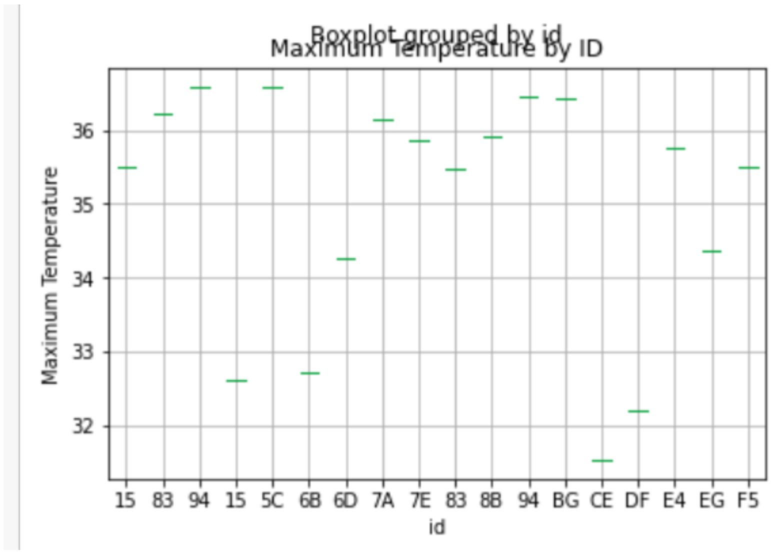
:

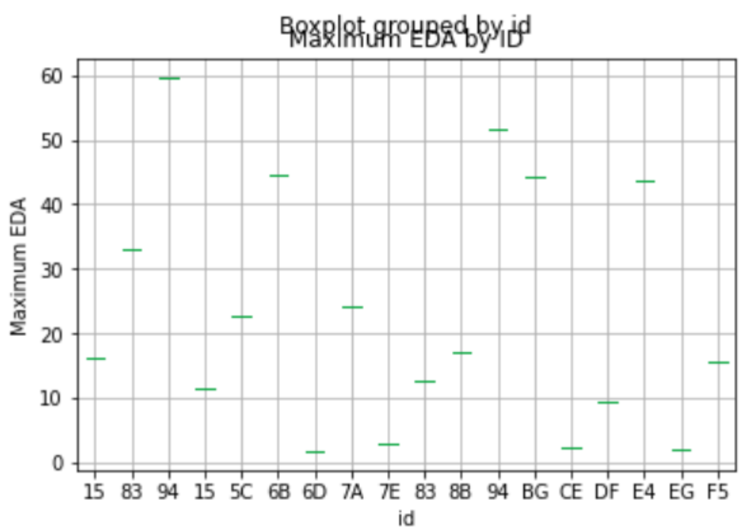
	X	Y	Z	EDA	HR	TEMP	id	datetime	label
0	-13.0	-61.0	5.0	6.769995	99.43	31.17	15	2020-07-08 14:03:00.000000000	2.0
1	-20.0	-69.0	-3.0	6.769995	99.43	31.17	15	2020-07-08 14:03:00.031249920	2.0
2	-31.0	-78.0	-15.0	6.769995	99.43	31.17	15	2020-07-08 14:03:00.062500096	2.0
3	-47.0	-65.0	-38.0	6.769995	99.43	31.17	15	2020-07-08 14:03:00.093750016	2.0
4	-67.0	-57.0	-53.0	6.769995	99.43	31.17	15	2020-07-08 14:03:00.124999936	2.0

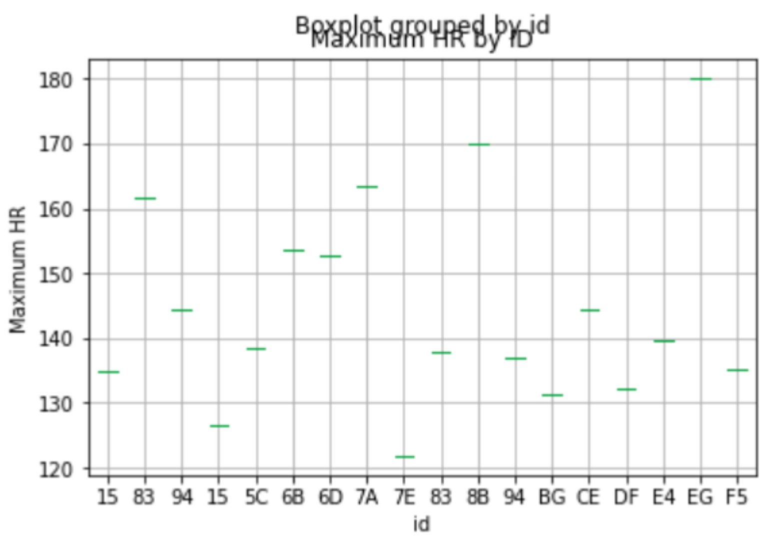
	EDA	HR	TEMP	id	datetime	label
0	6.769995	99.43	31.17	15	2020-07-08 14:03:00.000000000	2.0
1	6.769995	99.43	31.17	15	2020-07-08 14:03:00.031249920	2.0
2	6.769995	99.43	31.17	15	2020-07-08 14:03:00.062500096	2.0
3	6.769995	99.43	31.17	15	2020-07-08 14:03:00.093750016	2.0
4	6.769995	99.43	31.17	15	2020-07-08 14:03:00.124999936	2.0



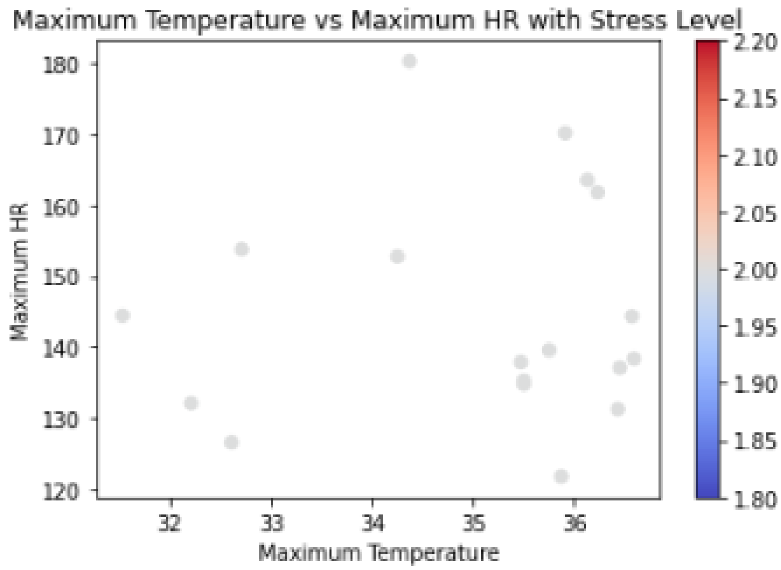


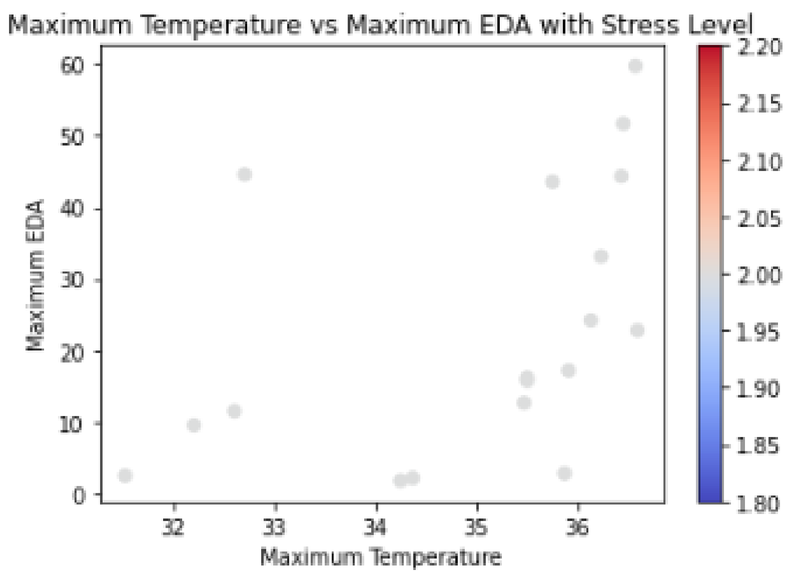

Boxplot grouped by id
Maximum Temperature by ID
Maximum Temperature
36
35
34
33
32
15 83 94 15 5C 6B 6D 7A 7E 83 8B 94 BG CE DF E4 EG F5
id


Boxplot grouped by id
Maximum EDA by ID
Maximum EDA
60
50
40
30
20
10
0
15 83 94 15 5C 6B 6D 7A 7E 83 8B 94 BG CE DF E4 EG F5
id

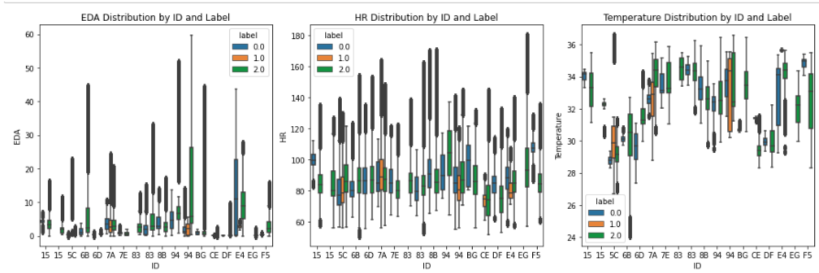


Boxplot grouped by id
Maximum HR by ID
Maximum HR
180
170
160
150
140
130
120
15
83
94
15
5C
6B
6D
7A
7E
83
8B
94
BG
CE
DF
E4
EG
F5
id

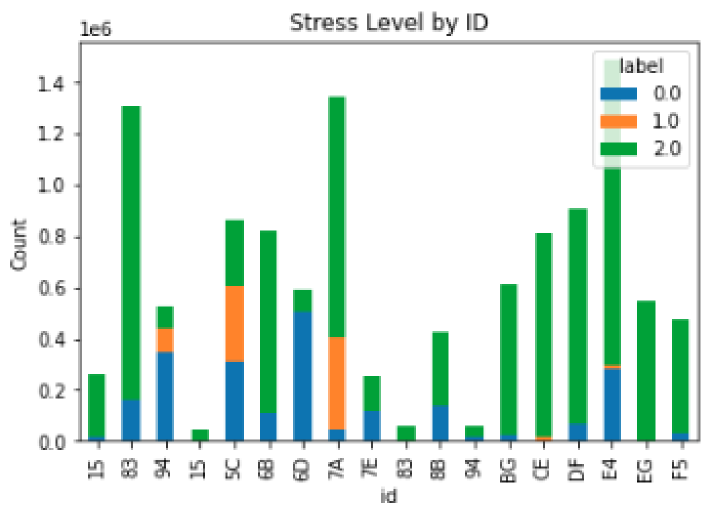

Maximum Temperature vs Maximum HR with Stress Level
Maximum HR
Maximum Temperature
180
170
160
150
140
130
120
32
33
34
35
36
2.20
2.15
2.10
2.05
2.00
1.95
1.90
1.85
1.80


Maximum Temperature vs Maximum EDA with Stress Level
Maximum EDA
Maximum Temperature
60
50
40
30
20
10
0
32
33
34
35
36
2.20
2.15
2.10
2.05
2.00
1.95
1.90
1.85
1.80

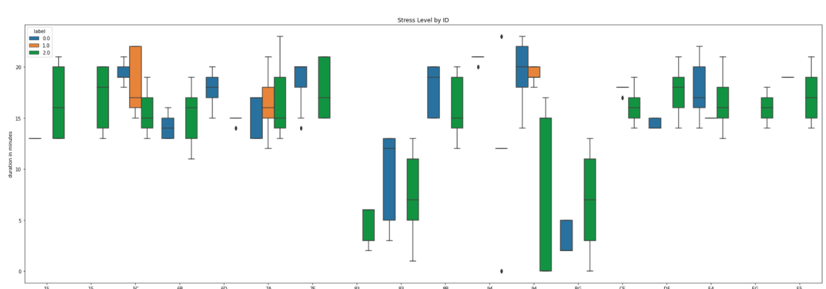

Maximum HR vs Maximum EDA with Stress Level
Maximum EDA
60
50
40
30
20
10
0
120
130
140
150
160
170
180
2.20
2.15
2.10
2.05
2.00
1.95
1.90
1.85
1.80



EDA Distribution by ID and Label
HR Distribution by ID and Label
Temperature Distribution by ID and Label
label
0.0
1.0
2.0
EDA
HR
Temperature
ID

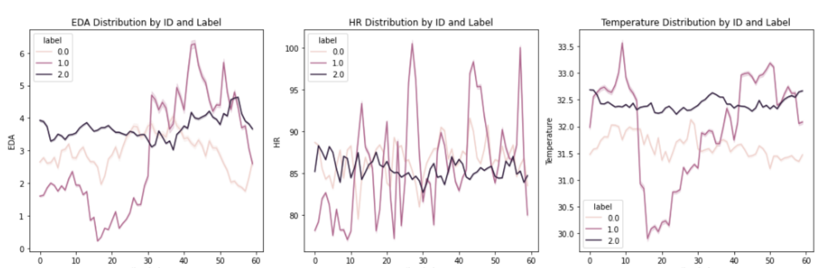


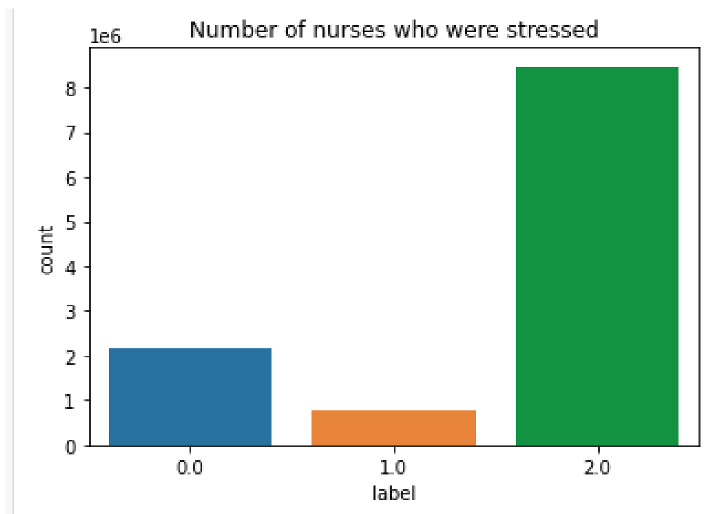
Stress Level by ID
1e6
label
0.0
1.0
2.0
Count
id
15 83 94 15 5C 6B 6D 7A 7E 83 8B 94 BG CE DF E4 EG F5

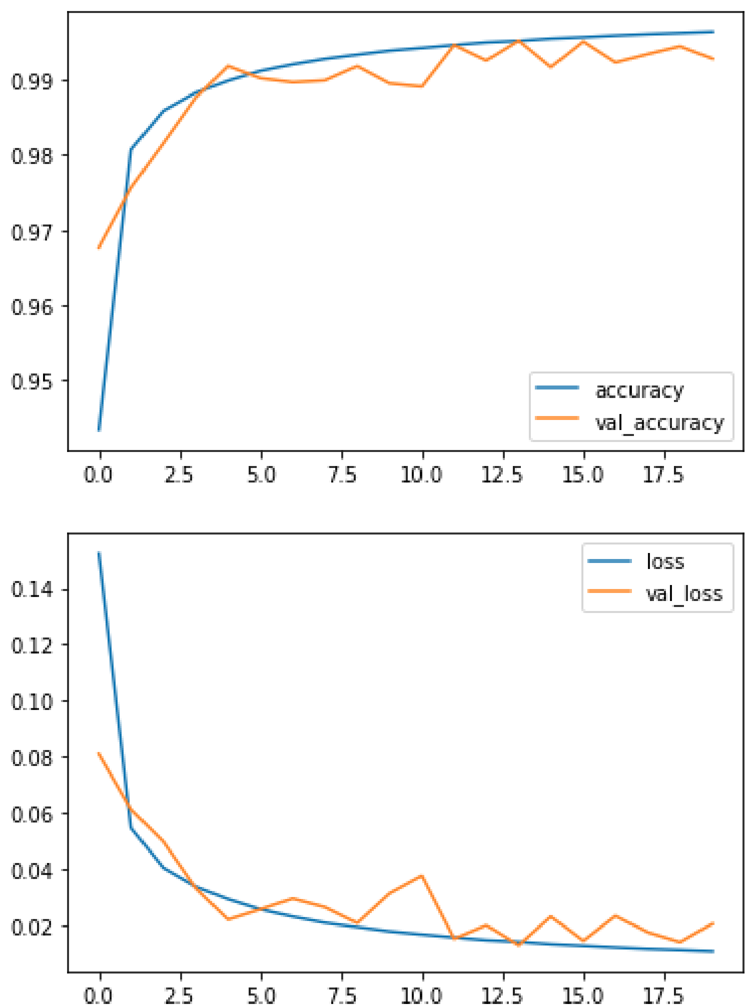


Stress Level by ID

label	0.0	1.0	2.0
id			
15	13441.0	NaN	248703.0
83	163204.0	NaN	1147516.0
94	348873.0	87458.0	87956.0
15	NaN	NaN	46987.0
5C	309121.0	295685.0	261124.0
6B	107522.0	NaN	710346.0
6D	503041.0	NaN	88322.0
7A	46082.0	355876.0	942743.0
7E	117124.0	NaN	136323.0
83	NaN	NaN	62099.0
8B	138243.0	NaN	285520.0
94	17279.0	NaN	44529.0
BG	21122.0	NaN	587533.0
CE	NaN	19201.0	794886.0
DF	65282.0	NaN	842886.0
E4	281190.0	15361.0	1189399.0
EG	NaN	NaN	549124.0
F5	30721.0	NaN	441623.0

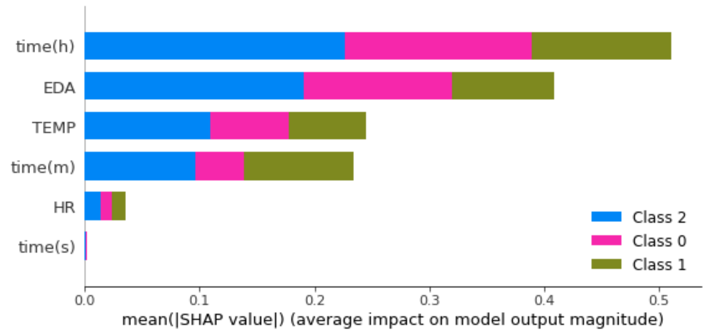


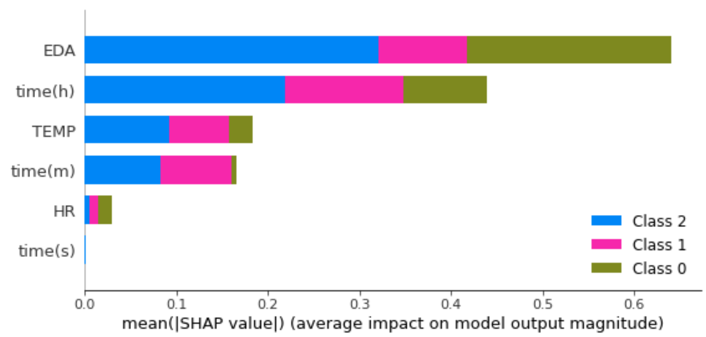

Number of nurses who were stressed
1e6
8
7
6
5
4
3
2
1
0
count
0.0
1.0
2.0
label



0.99
0.98
0.97
0.96
0.95
accuracy
val_accuracy
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
0.14
0.12
0.10
0.08
0.06
0.04
0.02
loss
val_loss
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

```
shap.summary_plot(shap_values, X_train, feature_names=features, plot_type="bar")
shap.summary_plot(shap_values_test, X_test, feature_names=features, plot_type="bar")
```

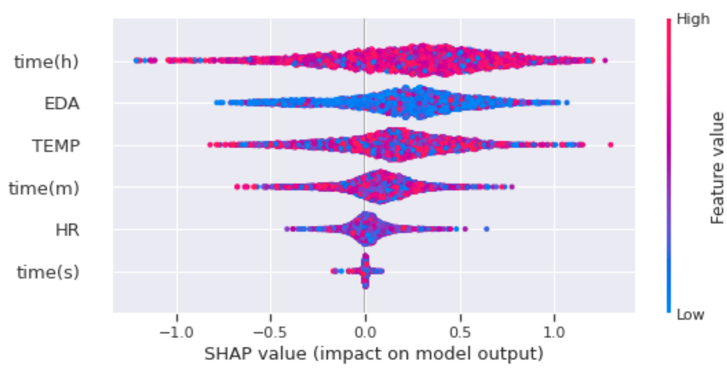





```
y_train_xgb_pred = fitted_xgb_model.predict(X_train)
y_test_xgb_pred = fitted_xgb_model.predict(X_test)
RMSE_xgb_train = math.sqrt(\
                    metrics.mean_squared_error(y_train,\
                                               y_train_xgb_pred))
RMSE_xgb_test = math.sqrt(\
                    metrics.mean_squared_error(y_test,
                                               y_test_xgb_pred))
R2_xgb_test = metrics.r2_score(y_test, y_test_xgb_pred)
print('RMSE_train: %.4f\tRMSE_test: %.4f\tr2: %.4f' %\
                    (RMSE_xgb_train, RMSE_xgb_test, R2_xgb_test))
```

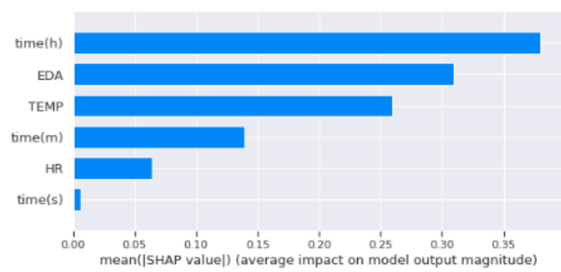
```
RMSE_train: 0.1894      RMSE_test: 0.2348       r2: 0.9121
```

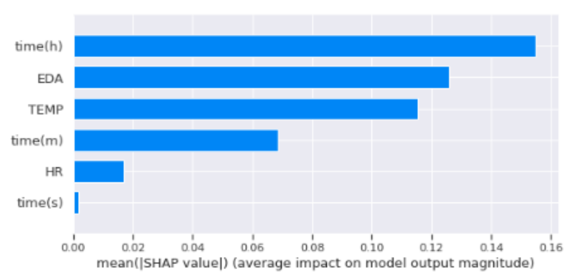

time(h)
EDA
TEMP
time(m)
HR
time(s)
−1.0
−0.5
0.0
0.5
1.0
SHAP value (impact on model output)
High
Feature value
Low


time(h)
EDA
TEMP
time(m)
HR
time(s)
−1.0
−0.5
0.0
0.5
1.0
SHAP value (impact on model output)
High
Feature value
Low

```
shap.summary_plot(shap_nn_values_test[0], X_test[random_indices2],feature_names=features, plot_type="dot")
```


```
shap.summary_plot(shap_xgb_values_test, X_test[random_indices2],feature_names=features, plot_type="bar")
shap.summary_plot(shap_nn_values_test[0], X_test[random_indices2],feature_names=features, plot_type="bar")
```





	X	Y	Z	ERD	HR	TEMP	Time(s)	Time(m)	Time(h)	id	Exam_type	Score
0	-3.0	-62.0	12.0	0.002563	84.00	22.51	0.241551	0.004026	0.000067	s1	midterm	78
1	-3.0	-61.0	12.0	0.019221	85.00	22.51	0.362327	0.006039	0.000101	s1	midterm	78
2	-2.0	-62.0	12.0	0.021784	86.00	22.51	0.483103	0.008052	0.000134	s1	midterm	78
3	-3.0	-63.0	12.0	0.023065	86.75	22.51	0.603878	0.010065	0.000168	s1	midterm	78
4	-2.0	-61.0	12.0	0.024347	87.40	22.51	0.724654	0.012078	0.000201	s1	midterm	78

	EDA	HR	TEMP	Time(s)	Time(m)	Time(h)	Score	Performance
0	0.002563	84.000000	22.51	0.241551	0.004026	0.000067	78.0	2
1	0.019221	85.000000	22.51	0.362327	0.006039	0.000101	78.0	2
2	0.021784	86.000000	22.51	0.483103	0.008052	0.000134	78.0	2
3	0.023065	86.750000	22.51	0.603878	0.010065	0.000168	78.0	2
4	0.024347	87.400000	22.51	0.724654	0.012078	0.000201	78.0	2
...	...	...	...	...	...	...	...	...
244562	0.021781	104.040514	22.51	10799.063841	179.984397	2.999740	58.0	1
244563	0.020500	104.040514	22.53	10799.180861	179.986348	2.999772	58.0	1
244564	0.021781	104.040514	22.53	10799.297881	179.988298	2.999805	58.0	1
244565	0.021781	104.040514	22.53	10799.414901	179.990248	2.999837	58.0	1
244566	0.020500	104.040514	22.53	10799.531920	179.992199	2.999870	58.0	1

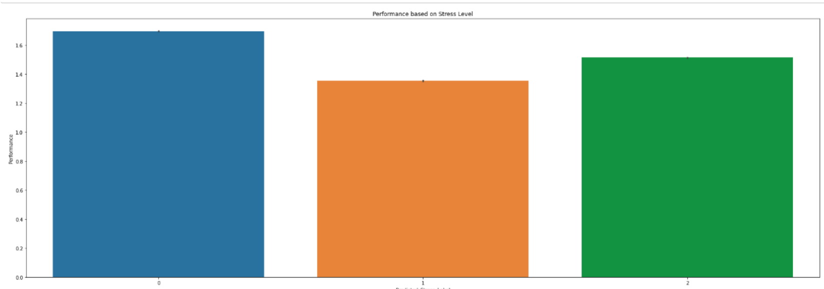
1623561 rows × 8 columns

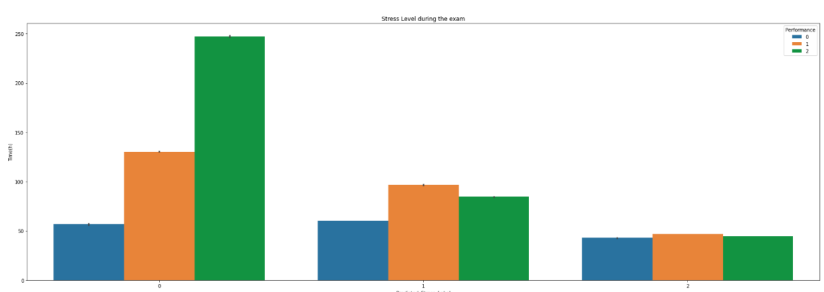
```
df4['Performance'].value_counts()
```

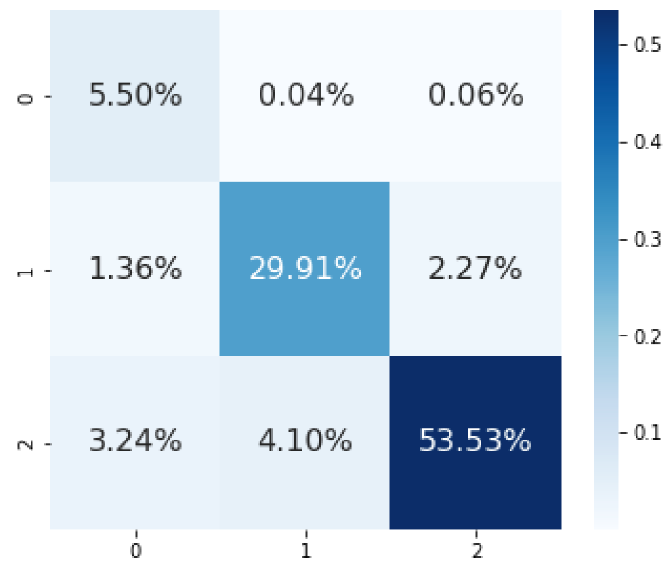
```
2    988389
1    542198
0     92974
Name: Performance, dtype: int64
```

		Predicted_Stress_Label
Score	Performance	
33.0	0	1.721112
39.0	0	1.786766
55.0	1	1.183123
58.0	1	0.969400
63.0	1	1.238244
64.0	1	1.817483
67.0	1	1.390833
71.0	1	1.435079
74.5	1	1.157452
75.0	2	1.786750
77.0	2	1.710758
78.0	2	1.445975
78.5	2	1.379248
80.0	2	1.198147
82.0	2	1.245239
85.0	2	1.714600
87.5	2	0.973603
88.0	2	1.492671
89.0	2	1.366828
90.0	2	1.719909
91.0	2	0.936194
92.0	2	1.205093
94.0	2	0.023634

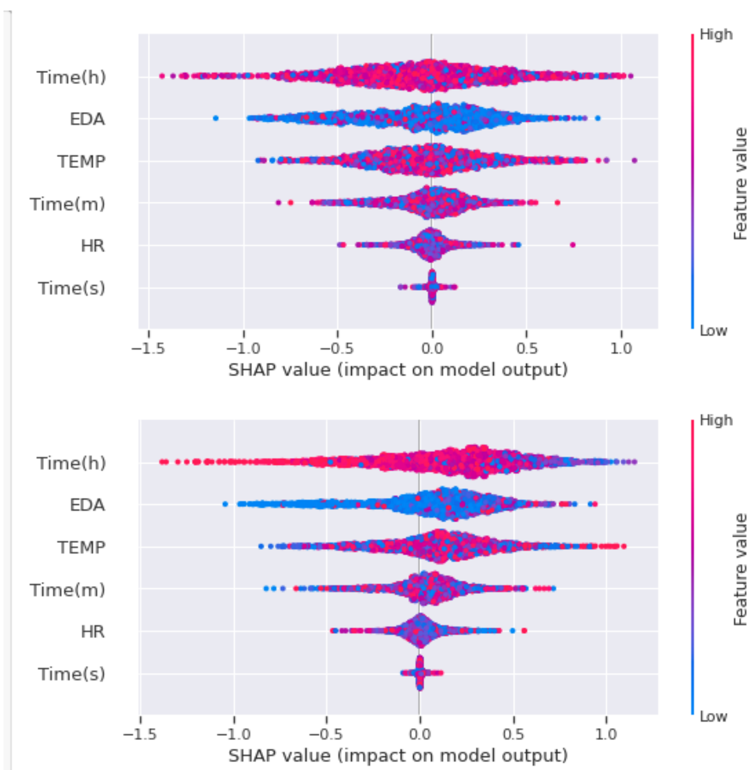
Performance	Score	Predicted_Stress_Label
0	36.208639	1.756222
1	63.643141	1.373780
2	85.523173	1.207945



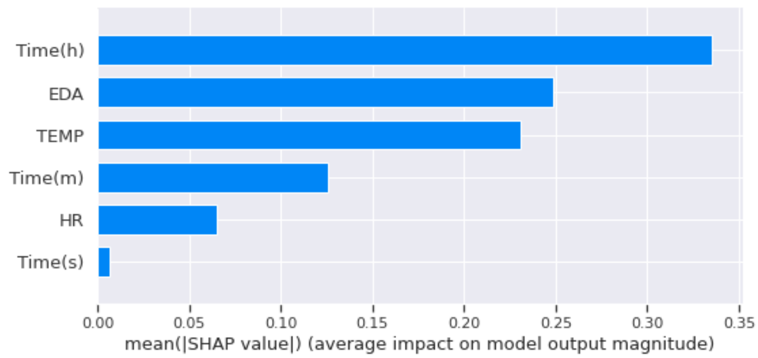


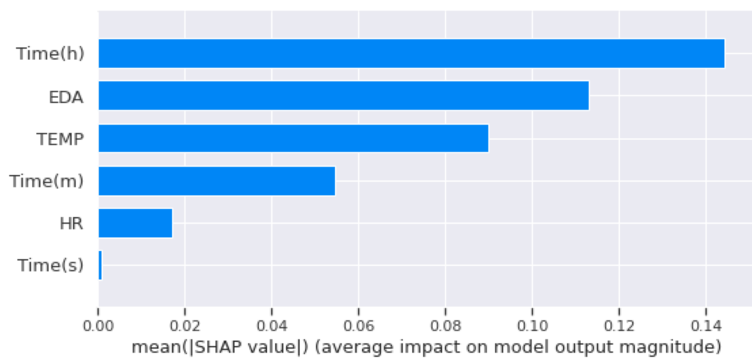


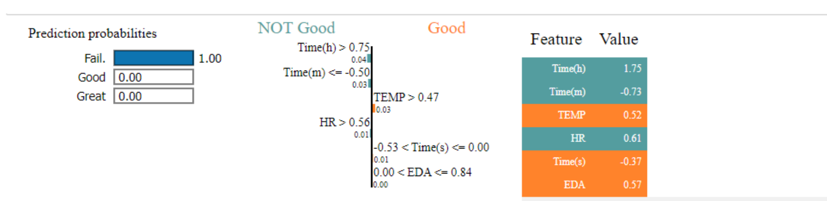
5.50%
0.04%
0.06%
1.36%
29.91%
2.27%
3.24%
4.10%
53.53%
0
1
2
0.5
0.4
0.3
0.2
0.1

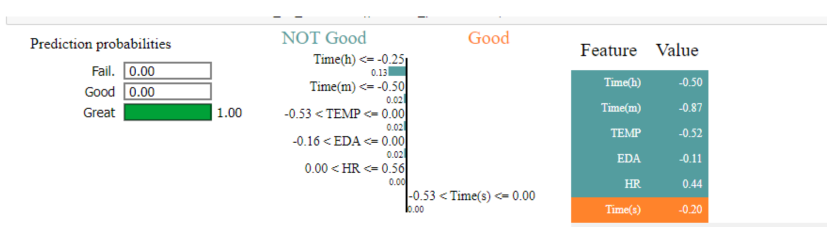


Time(h)
EDA
TEMP
Time(m)
HR
Time(s)
−1.5
−1.0
−0.5
0.0
0.5
1.0
SHAP value (impact on model output)
High
Feature value
Low


Time(h)
EDA
TEMP
Time(m)
HR
Time(s)
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35
mean(|SHAP value|) (average impact on model output magnitude)

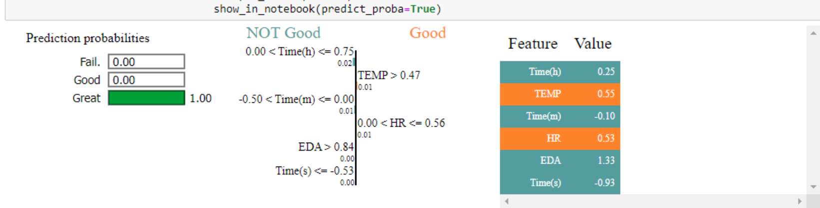

Time(h)
EDA
TEMP
Time(m)
HR
Time(s)
0.00 0.02 0.04 0.06 0.08 0.10 0.12 0.14
mean(|SHAP value|) (average impact on model output magnitude)


Prediction probabilities
Fail. 1.00
Good 0.00
Great 0.00
NOT Good
Good
Time(h) > 0.75
Time(m) <= -0.50
TEMP > 0.47
HR > 0.56
-0.53 < Time(s) <= 0.00
0.00 < EDA <= 0.84
Feature Value
Time(h) 1.75
Time(m) -0.73
TEMP 0.52
HR 0.61
Time(s) -0.37
EDA 0.57

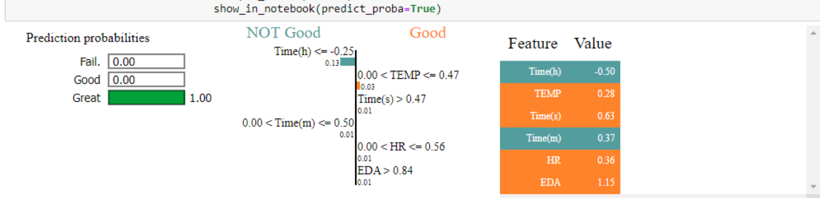

Prediction probabilities
Fail. 0.00
Good 0.00
Great 1.00
NOT Good
Good
Time(h) <= -0.25
Time(m) <= -0.50
-0.53 < TEMP <= 0.00
-0.16 < EDA <= 0.00
0.00 < HR <= 0.56
-0.53 < Time(s) <= 0.00
Feature Value
Time(h) -0.50
Time(m) -0.87
TEMP -0.52
EDA -0.11
HR 0.44
Time(s) -0.20


Prediction probabilities
Fail. 1.00
Good 0.00
Great 0.00
NOT Good
Good
0.00 < Time(h) <= 0.75
-0.53 < TEMP <= 0.00
-0.16 < EDA <= 0.00
HR <= -0.44
-0.53 < Time(s) <= 0.00
-0.50 < Time(m) <= 0.00
Feature Value
Time(h) 0.75
TEMP -0.38
EDA -0.05
HR -0.81
Time(s) -0.13
Time(m) -0.37


```
lime_explainer.explain_instance(X_test[0],\
                                 pre_model.predict).\
                    show_in_notebook(predict_proba=True)
```




```
lime_explainer.explain_instance(X_test[1],\
                                 pre_model.predict).\
                    show_in_notebook(predict_proba=True)
```



		Predicted_Performance
id	label	
15	0.0	1.288595
	2.0	1.198807
83	0.0	1.654151
	2.0	1.660499
94	0.0	1.042654
	1.0	1.139210
	2.0	1.282619
15	2.0	0.647413
5C	0.0	1.231395
	1.0	1.064822
	2.0	1.039514
6B	0.0	1.029148
	2.0	1.015847
6D	0.0	1.137545
	2.0	1.148106
7A	0.0	1.240289
	1.0	1.165001
	2.0	1.258998
7E	0.0	0.874654

id	label	Predicted_Performance
15	1.897453	1.203411
83	1.750970	1.659708
94	0.502339	1.099018
15	2.000000	0.647413
5C	0.944572	1.116654
6B	1.737068	1.017596
6D	0.298707	1.139123
7A	1.666811	1.233481
7E	1.075752	0.966068
83	2.000000	1.894105
8B	1.347546	1.025967
94	1.440881	1.089972
BG	1.930595	1.699058
CE	1.976414	0.928320
DF	1.856234	1.112073
E4	1.611198	1.294014
EG	2.000000	1.045276
F5	1.869921	1.230876

	label	Predicted_Performance
month		
4	1.877031	1.312775
5	1.215060	1.100981
6	1.316576	1.112017
7	1.465043	1.124384
8	2.000000	1.100060
10	1.558812	1.554684
11	1.659822	1.353381
12	1.965187	1.621680

Appendix B

APPENDIX B: Data

Data sets, experiment and other findings provided: `https://drive.google.com/drive/folders/1gvwG-7GkMsc3GuGjK8DSBytAEGGGs921?usp=sharing`