Multistage stress classification and cognitive capacity analysis using EEG

Abdul Waleed

A Thesis

in

The Department

of

Concordia Institute for Information Systems Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Information Systems Security) at

Concordia University

Montréal, Québec, Canada

August 2025

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Abdul Waleed**

Entitled: **Multistage stress classification and cognitive capacity analysis using EEG**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Information Systems Security)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Yong Zeng

_____ Examiner
Dr. Ching Suen

_____ Supervisor
Dr. Carol Fung

Approved by _____
Dr. Chun Wang, Chair
Department of Concordia Institute for Information Systems Engineering

_____ 2025

_____
Dr. Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Multistage stress classification and cognitive capacity analysis using EEG

Abdul Waleed

Stress is a physiological and psychological strain caused by mental workload. It is better to detect stress at early stages, which can help in stress management, compared to later stages, which may develop into psychiatric disorders such as anxiety and depression. In this study, we employed Muse S, a four-channel EEG headband, to record participants' EEG data under stressed and control conditions. We utilized mental arithmetic tasks and the Stroop color-word test as stressors to induce stress among our participants. We conducted subject-dependent and subject-independent evaluations by employing 10-fold and LOSO cross-validation strategies, respectively, and analyzed the difference between the two evaluation strategies. We proposed a two-stage deep learning model that comprises a fully connected autoencoder and a bidirectional LSTM model with an attention mechanism to improve the classification metrics for subject-independent evaluation using LOSO cross-validation strategy. We employed our proposed deep learning model to perform both binary and three-stage stress classification, achieving an accuracy of 83% for binary classification, while for three stage stress classification our model reported an accuracy of 66%. We compared the cognitive capacity of our best and worst performers by employing statistical tools such as line graphs and the Mann-Whitney U test. We implemented a regression model using random forest to predict the participants' scores by employing brain waves and response time.

Acknowledgments

I would like to express my sincere gratitude to my supervisor Dr. Carol Fung for her invaluable guidance, insightful feedback, and constant support throughout the course of research. Her knowledge, dedication, and mentorship was indispensable for the research. I am very grateful for all that I have learned under her supervision, and I was very fortunate to have her as my supervisor. I would like to express my heartfelt gratitude to my family for supporting me throughout this journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Stress is the physiological response caused by demanding situations. It can be due to challenging situations such as financial issues, challenges faced in the workplace, difficulties in marital life, and adapting to technological advancements [1]. A mild level of stress elevates focus and leads to higher productivity [2]. A high level of stress leads to psychiatric disorders such as anxiety and depression [3]. People suffering from long-term high stress can become victims of negative emotions that can reduce their ability to focus on their tasks, hence decreasing their output [4] [5].

The fact that stress has an impact on both mental and physical well-being [6] puts emphasis on its detection. People who suffer from stress become victims of negative emotions that degrade their mental well-being. Stress detection in early stages is very important because it can improve mental well-being, prevent burnout, and enhance quality of life [7]. Since stress varies from person to person, before devising personalized measures for stress management, it is very crucial to investigate the reasons behind stress and its impact on mental well-being.

Stress detection is challenging because it varies person to person because of its subjective nature [8]. There are symptoms of stress that can be detected using various methods, such as physiological tests (cortisol levels and blood pressure), questionnaires, and cognitive performance tests [9]. There are some key indicators associated with high stress, for example, high blood pressure, increased heart rate, tense muscles, avoiding eye contact, and increased speech rate [10]. There are

wearable devices such as smart watches that can monitor physiological vital parameters such as blood pressure, heart rate, cortisol levels, and skin temperature [11]. There are smart watches that can even record sleep patterns and utilize physiological parameters to generate reports containing stress scores, enabling them to perform real-time stress detection [12] [13] [14].

EEG is a non-invasive technique that measures electrical potential from the scalp [15]. It is a tool that provides real-time insights about brain activities, making it suitable for stress detection. EEG is widely used in the medical field for detecting various neurological disorders such as schizophrenia [16], epilepsy [17], and sleep disorders [18]. EEG is also used in brain computer interfaces to design devices for physically impaired people that can be controlled with their brainwaves [19]. EEG is also widely employed in literature by studies on emotions, sleep patterns, and stress detection.

EEG has a high temporal resolution that makes it the best choice for stress detection compared to other techniques such as TEP and FMRI. EEG is safe, inexpensive, and more reliable.

Stress detection requires stressors to put participants under controlled stress conditions. Stimuli that cause stress are called stressors. There are various stressors that have been used in the literature to perform stress detection, such as the Stroop color-word test [20], mental arithmetic task [21] [22], cold pressor test [23], and trier social stress test (TSST) [24]. A description of each will be provided in the Background Chapter 3.

There are a lot of studies that have conducted stress detection using EEG. Researchers have utilized models such as Multilayer Perceptron(MLP) [25], Support Vector Machines (SVM) [26], Logistic Regression (LR) [27], and Naive Bayes (NB) [28] to perform stress detection. The mentioned machine learning models reported very high performance evaluation metrics in subject-dependent analysis by employing a K-fold cross-validation strategy.

We recorded EEG data from 50 participants, making sure that each was wearing a Muse S EEG headband while solving stressor questions. We designed our stressors by incorporating the countdown approach by making sure that there would be enough time for the participants to answer each stressor question. After solving each stressor question, the participants could see their current standings, which put them under competition stress. After collecting the EEG data from 50 participants, we performed stress detection by employing subject-dependent and subject-independent evaluation strategies. Subject-dependent evaluation strategy is one in which the same participant data is used

in the training and testing phases, whereas subject-independent evaluation strategy utilizes data from different participants to evaluate the model's performance in the testing phase. For subject-dependent evaluation, we employed 10-fold cross-validation, and for subject-independent evaluation, we used the Leave-One-Subject-Out (LOSO) cross-validation strategy. In the LOSO cross-validation strategy, the model is trained on the remaining subjects after leaving one subject out; the left-out subject is used to test the model's performance in the testing phase. We evaluated five different classifiers, namely Naive Bayes, Random Forest, Support Vector Machine (SVM), Long Short Term Memory (LSTM), and Convolutional Neural Network (CNN), by using both subject-dependent and subject-independent evaluation strategies and investigated the difference in results obtained by the two evaluation strategies. Most of the literature has performed stress detection using EEG by employing subject-dependent evaluation. The biggest drawback that subject-dependent evaluation contains is subject bias, due to which it is not a practical approach. Subject-independent evaluation strategy removes subject bias and is a more practical approach. We proposed a two-stage deep learning model consisting of the fully connected autoencoder and bidirectional LSTM model with an attention-based mechanism to improve the performance evaluation metrics of subject-independent evaluation by employing LOSO cross-validation strategy. We compared the cognitive capacity of the best-performing and worst-performing groups of our study based on the brain waves. We made two groups of 10 participants each, one consisting of best performers, whereas the other group consisted of bad performers. We investigated which brain wave ratio can differentiate between the best and worst performing groups. We employed statistical tools such as line graphs and Mann-Whitney U test to perform this comparison. Kahoot assigns scores to participants based on time taken and is a continuous value. We predicted the scores assigned to participants by Kahoot, and since the assigned scores were continuous, the problem was framed as a regression. We implemented a random forest regressor model to predict the scores of participants assigned by Kahoot and achieved a root mean square value of 1.18.

1.2 Contributions

Our main contributions follow:

- We recorded EEG data from 50 participants to perform EEG-based multistage stress classification. To the best of our knowledge, this is the dataset with the highest number of participants for multistage stress classification.
- The dataset has been shared with the public research community, which is publicly available [29].
- We performed multistage stress classification by employing subject-dependent and subject-independent evaluation strategies and investigated the difference between evaluation strategies.
- We proposed a two-stage deep learning model using a fully connected autoencoder and a bidirectional LSTM model with an attention-based mechanism and achieved a binary classification accuracy of 83% and 66% accuracy for three-stage stress classification.
- We compared the cognitive capacity of the best and the worst-performing groups of our study using their brain waves and established a random forest regressor model to accurately predict the scores of participants.

1.3 Thesis Overview

The rest of the thesis is organized as follows: Chapter 2 covers the literature review that has been conducted on stress detection using EEG. Chapter 3 briefs about the background, throwing light on topics such as brain and brainwaves, stressors, statistical analysis, and machine learning. Chapter 4 illustrates the methodology of our work, explaining topics such as data recording, preprocessing, feature extraction, classifiers, and performance evaluation metrics that have been used. Chapter 5 reports the results obtained, which will be further discussed in Chapter 6. Chapter 7 concludes the thesis while directing towards future work.

Chapter 2

Literature Review

In this chapter, we will cover the literature that has been conducted in the domain of stress detection using EEG. We will also highlight the limitations of the existing works.

2.1 Acute stress detection

This section will cover the literature on short-term stress.

Cambay et al. [30] collected a new dataset from 310 participants who were the victims of an earthquake in Turkey by employing a 14-channel Emotiv Insight EEG headset at a sampling frequency of 128 Hz. The authors proposed a new explainable feature engineering (XFE) model to automatically detect stress. Participants were divided into stressed and control groups, containing 150 and 160 participants, respectively. Participants belonging to the stressed group were shown a 3-minute earthquake video, whereas the participants from the control group were shown a meditation video. This study achieved binary classification accuracies of 92.95% and 73.63%, using 10-fold and LOSO cross-validation strategies, respectively. The authors employed KNN based on the t algorithm (tkNN) as a classifier. Ince et al. [31] proposed a new explainable feature engineering by introducing a novel feature extraction function called Cubic pattern that uses coding channels to form three-dimensional feature vectors. The authors utilized the same dataset and achieved a binary classification accuracy of 96.29% and 76.17% for 10-fold and LOSO cross-validation strategies, respectively.

Ghosh et al. [32] collected a dataset from 40 participants by employing a 32-channel Emotiv Epoch EEG headset at a sampling frequency of 128 Hz. Three stressors, namely the Stroop color-word test, mental arithmetic task, and symmetrical mirror images, were used to induce stress among the participants. Between the stressors, relaxation music was played for the participants to relax before moving to the next stressor. Savitzky-Golay filters were employed by the researchers to remove baseline drifts from the collected EEG data. Artifacts were removed from the data by using wavelet thresholding.

Afify et al. [33] proposed a VGGish-CNN-based model to perform multistage stress classification based on EEG. The authors used the SAM40 [32] dataset, which contains three different stressors and a relaxation state. The study utilized a deep learning VGGish architecture to extract features for the CNN classifier to perform multistage stress classification. The authors achieved an accuracy of 99.25% for four-stage stress classification. This study also investigates the results of five-fold cross-validation to generalize the results.

AlShorman et al. [34] performed a real-time stress detection based on EEG by using SVM as a classifier. EEG data was collected from fourteen participants using 128-channel EGI's Geodesic system. EEG was converted into the frequency domain using Fast Fourier Transform. Participants were divided into stressed and control groups. The participants from the stressed group have to keep their dominant hand in ice water for a duration of 60 seconds. Participants were provided with the depression anxiety stress scale (DASS) both before and after the experiment to self-evaluate their stress. The proposed SVM classifier achieved an accuracy of 98.21% for subject-wise classification.

Phutela et al. [35] collected EEG data from 35 participants using a four-channel Muse EEG headband at a sampling frequency of 256 Hz. Hindi movie clips were used to induce stress among the participants. The authors employed the Fast Fourier Transform signals provided by the Muse device to perform their analysis. A study compared two models, namely LSTM and Multilayer Perceptron (MLP), to perform stress classification. LSTM outperformed MLP and achieved a classification accuracy of 93.17%.

Jun et al. [26] employed an SVM classifier to perform multistage stress classification by using EEG. The data from 10 participants were collected employing a 14-channel Emotiv Epoch EEG headset at a sampling frequency of 128 Hz. The Stroop color-word test and mental arithmetic task

were utilized as stressors. The experiment also had a resting stage where participants could rest by keeping their eyes closed. The authors performed four-fold cross-validation and achieved an accuracy of 75% for three-stage stress classification.

Hou et al. [36] proposed a real-time stress detection system based on EEG by employing SVM as a classifier. The data was collected by using a 14-channel Emotiv Epoch EEG headset from 9 participants at a sampling frequency of 128 Hz. The Stroop color-word was utilized to induce three levels of stress, and a resting state was also part of the experiment. The results were obtained by employing a five-fold cross-validation strategy. SVM achieved a classification accuracy of 67.06% for four levels of stress, for three levels of stress classification, the reported accuracy was 75.22%, and a binary classification accuracy of 85.71% was achieved.

2.2 Stress in response to music track

This section will explore the studies that reported the effects of music on stress.

Asif et al. [27] investigated music as a stimulus to classify stress. EEG data was collected from 27 participants using a four-channel Muse EEG headband at a sampling frequency of 256 Hz. The authors employed five sets of features to train models, namely absolute power values, relative power values, coherence, amplitude asymmetry, and phase lag. Four different classifiers were compared, including logistic regression (LR), multilayer perceptron (MLP), sequential minimal optimization, and stochastic gradient descent. LR outperformed other classifiers by achieving the highest classification accuracies of 98.76% for two-level and 95.06% for three-level stress classification.

Bhatnagar et al. [37] proposed a deep learning model consisting of the EEGnet with CNN and a Relu activation function to perform stress detection. EEG data was collected from 45 participants from frontal, parietal, temporal, and central parts of the brain. The study employed music to detect stress among the participants. The collected EEG data was converted into five bands, such as delta, theta, alpha, beta, and gamma, by utilizing the wavelet transform. EEGnet with CNN achieved a classification accuracy of 99.45% while utilizing the alpha band as a feature.

2.3 Chronic stress detection

This section throws light on papers that performed long-term stress detection.

Saeed et al. [25] collected EEG data from 33 participants in a closed eye position to perform chronic stress classification. EEG data was collected for 3 minutes using a five-channel Emotiv Insight EEG headset at a sampling frequency of 128 Hz. Participants were grouped into stressed and control groups by employing the Perceived Stress Scale questionnaire and expert evaluation. Five different classifiers, namely support vector machine, naive bayes, logistic regression, multilayer perceptron, and k-nearest neighbors, were employed for performing long-term stress classification. SVM and LR achieved the highest classification accuracy of 85.20%.

Arsalan et al. [28] proposed an EEG dataset for perceived stress classification. Data was collected before and after activity from 28 participants by employing a 4-channel Muse EEG headband at a sampling frequency of 256 Hz. In the activity phase, participants have to present a topic in front of the audience. The Perceived Stress Scale (PSS) questionnaire was utilized to label the stress levels of the participants. Three classifiers, namely support vector machine, naive bayes, and multilayer perceptron, were employed by the authors to perform perceived stress classification. MLP outperformed other classifiers by achieving 89.28% for two-level and 60.71% for three-level perceived stress classification using 10-fold cross-validation, and 92.85% for two-level and 64.28% for three-level perceived stress classification using leave-one-out cross-validation.

2.4 Limitations of previous works

Although a lot of studies have performed stress detection using EEG and achieved very high performance evaluation metrics but they employed subject-dependent analysis to evaluate the performance of classifiers. Subject-dependent analysis contains subject-bias, and the results do not generalize to unseen participants' data. Very few papers exist [30], [31] that have performed stress detection employing subject-independent evaluation by using LOSO cross-validation strategy and reported only results for binary classification. We do not only achieved a higher accuracy using subject-independent analysis by employing LOSO cross-validation strategy compared to the literature, but also performed a three-stage stress classification using subject-independent analysis that is

not reported in the literature, to the best of our knowledge.

Chapter 3

Background

This chapter will cover the background by outlining the brain and the brain waves, followed by an explanation of the stressors. We will also explore the basics of statistical analysis and machine learning.

3.1 Brain and Brainwaves

The brain is the central and most important part of the central nervous system. The brain communicates with other parts of the body through neurons. Neurons emit electrical potential when communicating with each other. EEG measures this electrical potential from the scalp.

There are five types of brain waves, namely delta, theta, alpha, beta, and gamma. Delta brain frequency lies between 0.5 Hz to 4 Hz. It is the lowest frequency brain wave that is highly active during deep sleep and low consciousness. Theta is a brain wave whose frequency is higher than the delta wave but lower than the alpha wave. Theta brain wave frequency lies between 4 Hz to 8 Hz. It is highly associated with light sleep, inhibition, and emotional stress in adults. Alpha is a brain wave that is highly associated with relaxation. It has a frequency range between 8 Hz to 12 Hz. It has a lower frequency than a beta wave. Beta wave frequency ranges between 12 Hz to 30 Hz. It is highly active during learning and problem-solving. Gamma is the highest frequency brain wave, having a frequency higher than 30 Hz. It is highly active during complex problem solving and is also associated with intelligence.

3.2 Stressors

The following are the most widely used stressors in the literature.

3.2.1 Stroop color-word test

In the Stroop color-word test [20], the participants are supposed to identify the color of the ink with which the words are printed. It is a very effective psychological test and is widely employed by researchers to put participants under stress. The fact that the participants have to name the color of the ink rather than the words puts them under psychological stress. The following Figures 3.1, 3.2, 3.3, 3.4, and 3.5 depict the examples of Stroop color-word test.

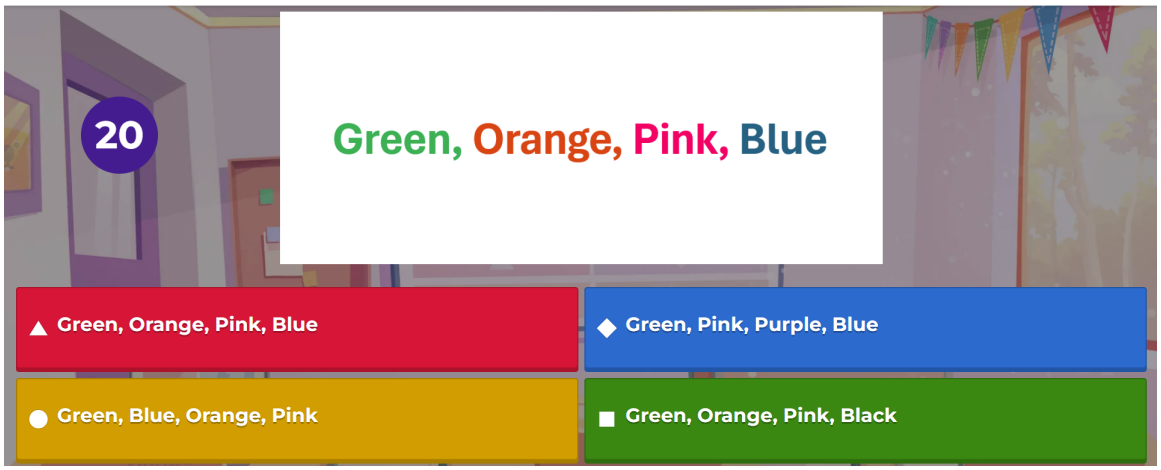


Figure 3.1: Stroop color-word test example 1.

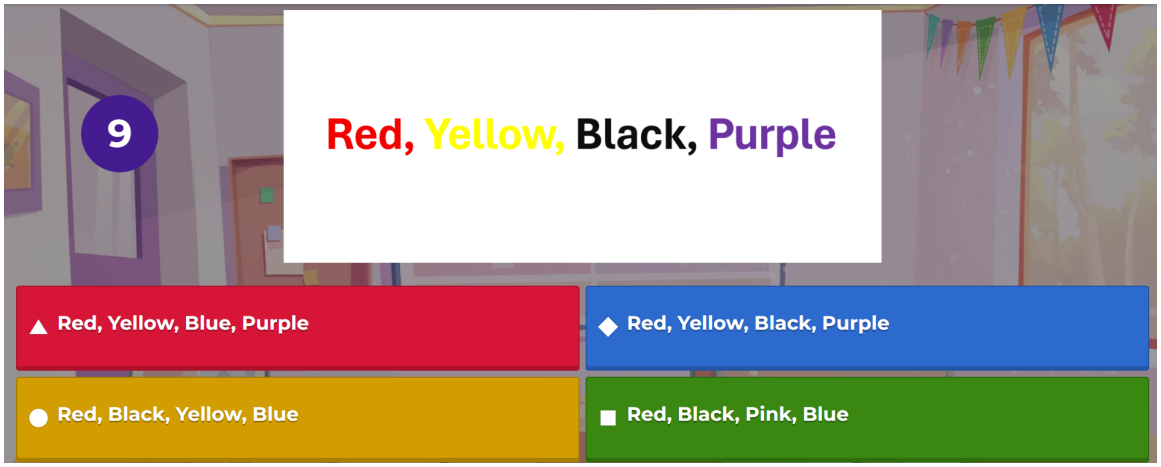


Figure 3.2: Stroop color-word test example 2.

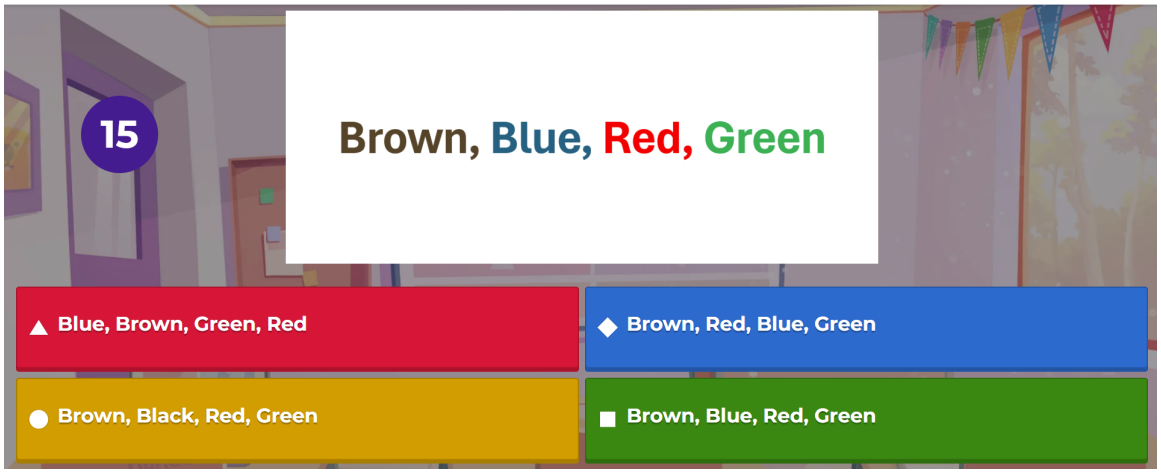


Figure 3.3: Stroop color-word test example 3.

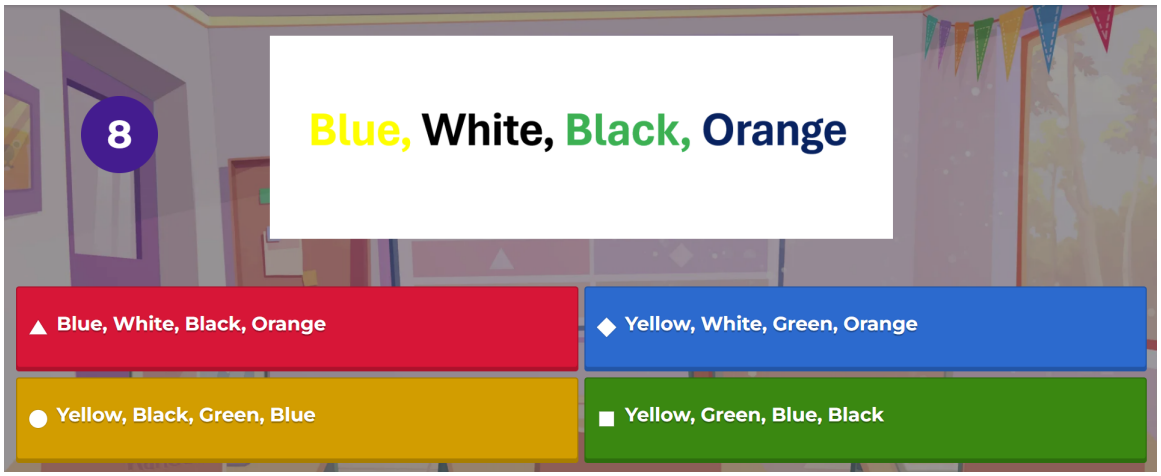


Figure 3.4: Stroop color-word test example 4.

3.2.2 Mental arithmetic test

It is another very widely used stressor in literature [21], [22]. It consists of questions based on mathematical operations such as addition, subtraction, multiplication, and division. It can also contain puzzles based on the same operations. The following Figures 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, and 3.12 depict the examples of mental arithmetic task.

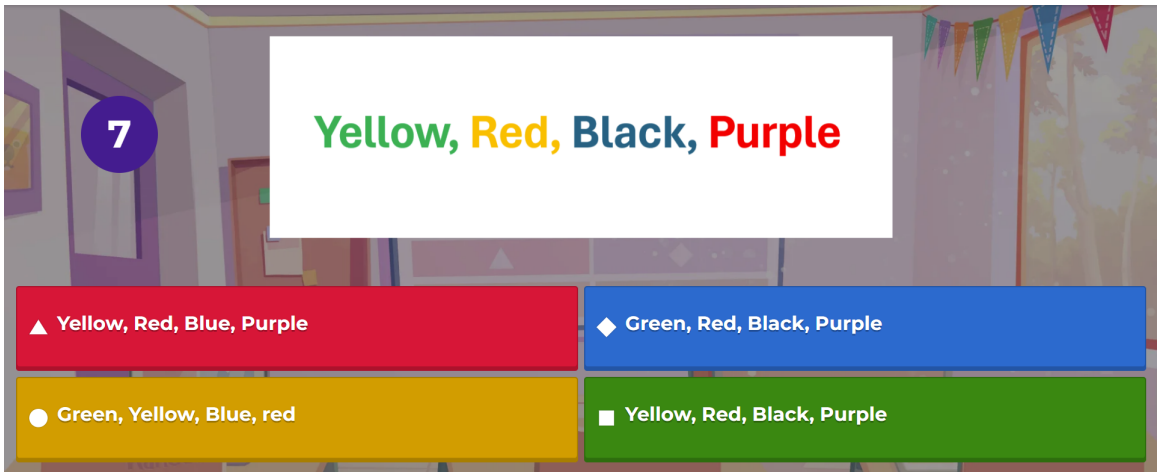


Figure 3.5: Stroop color-word test example 5.

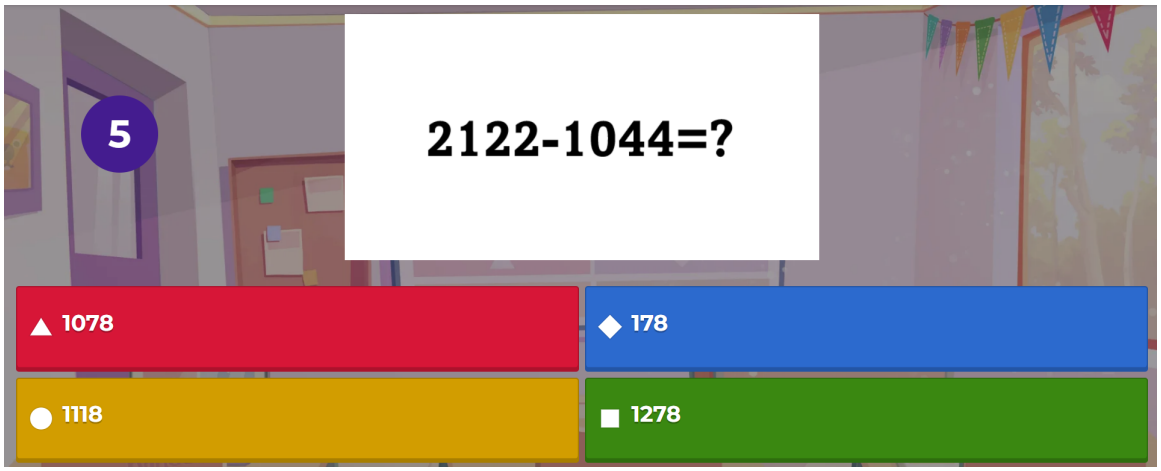


Figure 3.6: Mental arithmetic task containing subtraction.

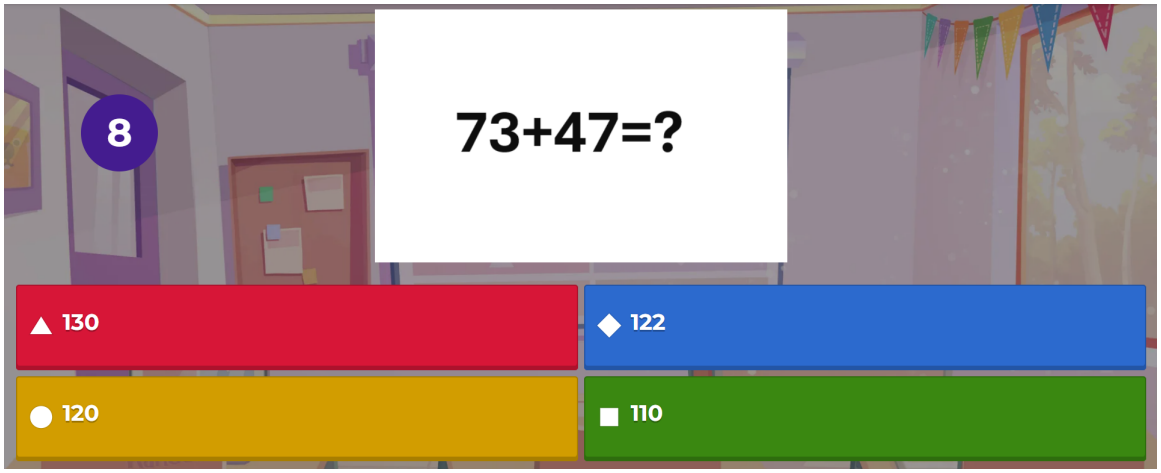


Figure 3.7: Mental arithmetic task containing addition.

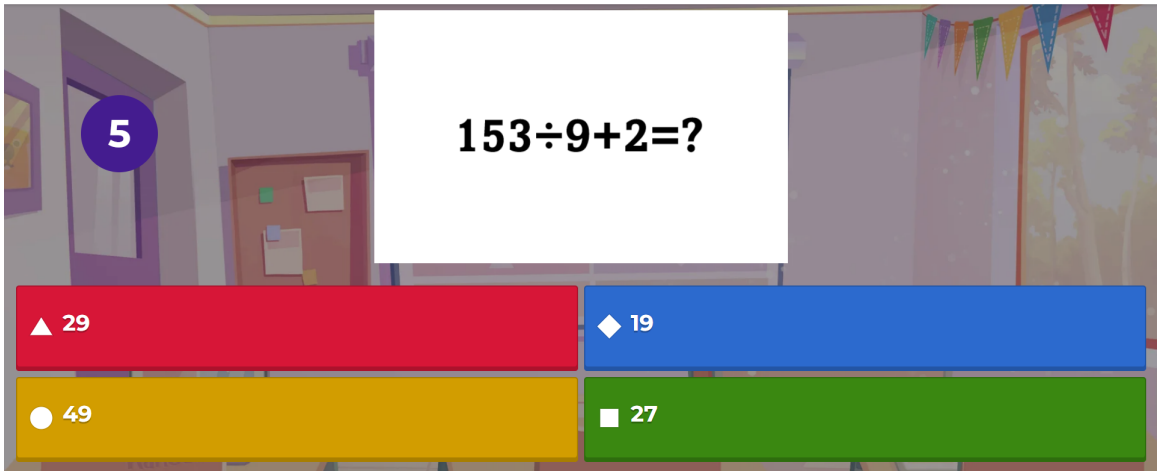


Figure 3.8: Mental arithmetic task containing division and addition.

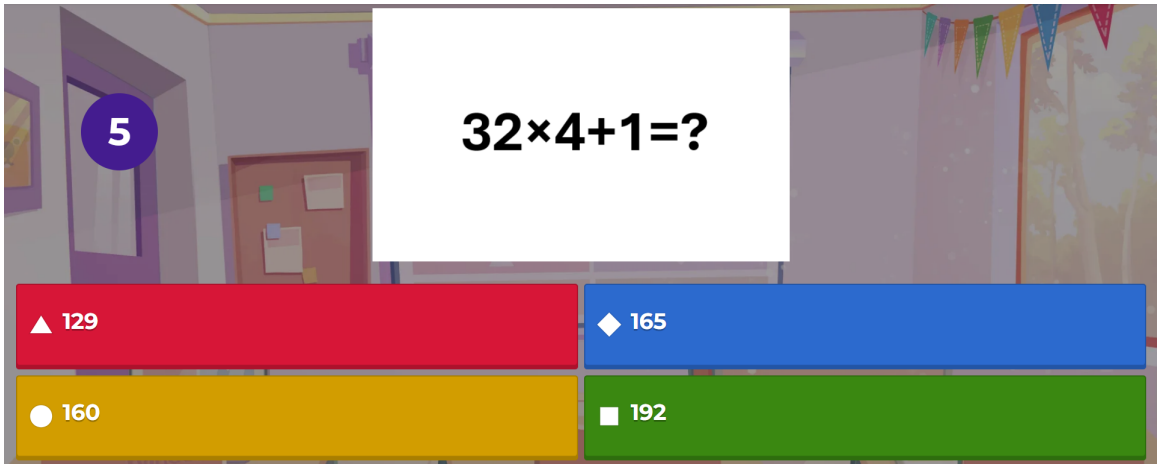


Figure 3.9: Mental arithmetic task containing multiplication and addition.

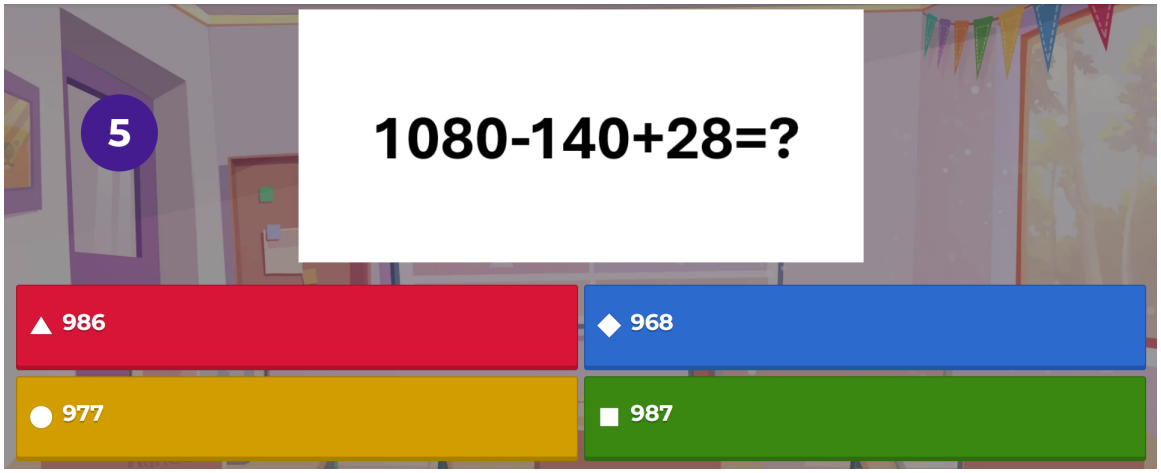


Figure 3.10: Mental arithmetic task containing subtraction and addition.

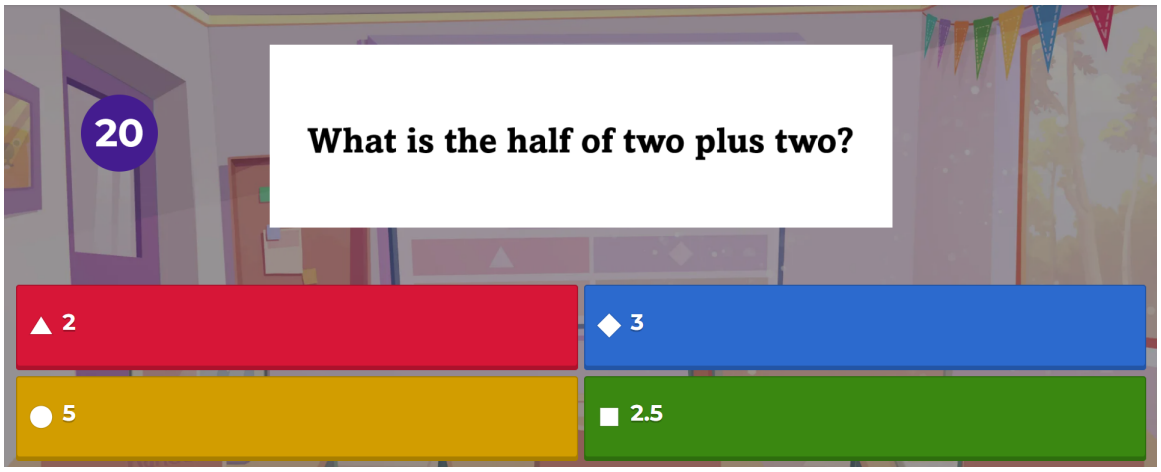


Figure 3.11: Mental arithmetic task containing statement.

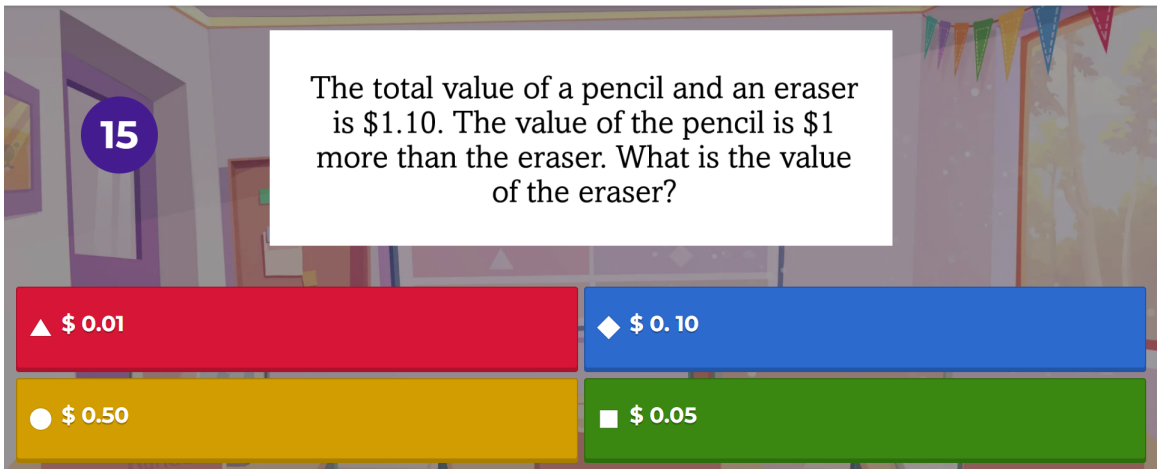


Figure 3.12: Mental arithmetic task containing problem.

3.2.3 Cold pressor test

In the cold pressor test [23], participants keep their hands inside ice-cold water. The duration for which they have to keep their hands depends on the requirements of the experiment. It includes discomfort for the participants, which is why we did not consider it.

3.2.4 Trier social stress test

In the trier social stress test [24], participants prepare a topic and they have to present it in front of the people. The fact that the people will evaluate the presenters puts them under psychological stress.

3.3 Statistical analysis

Statistical tools [38] are used for data analysis, such as mean, mode, and standard deviation. It can also be used to visualize data by employing line graphs. It can be used to verify hypotheses by utilizing tools such as p-values, confidence intervals, and Z-scores.

3.4 Machine learning

Machine learning [39] employs statistical models to learn the patterns present in the data to make decisions automatically. Machine learning can be divided into supervised and unsupervised learning based on the data provided to the model. Supervised machine learning models require the data to be labelled, whereas in the case of unsupervised learning, the data is unlabeled that is provided to the model.

3.5 Muse S EEG headband

We employed Muse S EEG headband to collect the participants' EEG data [40]. Muse S has four EEG sensors located at the positions TP_9, AF_7, AF_8, and TP_10. Each sensor is responsible for collecting EEG data from a specific position of the scalp. Muse S collects EEG data at a sampling

frequency of 256 Hz. Muse headband consists of the headband and sensors illustrated through the Figures 3.13 and 3.14.

Figure 3.13: Muse S EEG Headband.

Figure 3.14: Muse S EEG Headband Assembled.

Chapter 4

Methodology

This Chapter outlines the methodology of our study by illustrating the experimental procedure to collect EEG data and removing noise from the collected data. We will also discuss the feature extraction step, followed by the classifiers employed for the multistage stress classification. At last, we will address the performance evaluation metrics used to evaluate the performance of the employed classifiers.

4.1 Experimental Procedure

We recruited 50 participants for our multistage stress classification study based on EEG. All the participants were part of Concordia University. There were 30 males, 19 females, and 1 unknown. This study was approved by the Human Research Ethics Committee of Concordia University under certificate number 0020206. We employed two different stressors, namely the Stroop color-word test and the Mental arithmetic task. The experiment was divided into two parts: the first part comprising the Congruent Stroop color-word test and an easy mental arithmetic task, whereas the second part contained the Incongruent Stroop color-word test and a difficult mental arithmetic task. Each part consisted of the 15 stressor questions. The stressor questions were hosted in Kahoot [41], which is an online free gaming platform with a countdown approach that we utilized in our experiment to control the difficulty levels of stressor questions. Participants had to answer stressor questions in a limited time, such as 5, 10, or 20 seconds depending on the difficulty level of the stressor question.

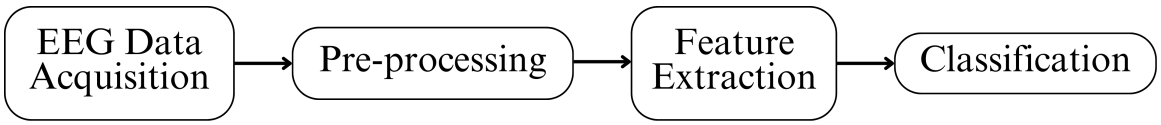


Figure 4.1: Pipeline for stress classification using EEG.

EEG data was collected using Muse S [40] EEG headband at a sampling frequency of 256 Hz. Muse S EEG headband is designed according to the 10-20 international standard for placing the electrodes. Muse S has four channels, namely TP9, AF7, AF8, and TP10. Each of these sensors collects the required EEG data from the specific location of the scalp. All the participants were briefed about the experiment and the Muse EEG headband. Participants were provided with the voluntary informed consent to read and then sign it. After the participants signed the consent forms, the EEG headband was placed on their heads. The good quality of the EEG sensors was confirmed using the Mind Monitor app. The Mind Monitor application connects with the Muse EEG headband with the help of Bluetooth. The Mind Monitor application was also utilized to collect EEG data from the Muse headband and then send it to the drive for further processing. After Mind Monitor confirmed the good quality of all four EEG sensors, we moved towards the next step, which was data collection. Figure 4.1 outlines the pipeline followed by our multistage stress classification based on the EEG signal.

4.2 EEG dataset collection

This section summarizes the EEG dataset collection across different phases, which is publicly available [29].

4.2.1 Voluntary Consent Form

A voluntary consent form was provided to all the participants to read and then sign. The informed consent contained information regarding the purpose of the experiment, procedures involved, risks and benefits of the experiment, and that the participants can withdraw from the experiment at any time.

4.2.2 Warm-up

Our EEG data collection starts with the warm-up phase. It was included in the experiment to familiarize people with the experiment. It consisted of four questions: two questions were based on the Stroop color-word test, and the remaining two stressor questions consisted of a mental arithmetic task. We did not collect EEG data in the warm-up phase. The Figure 4.2 depicts the participant who was ready to start the experiment.

Figure 4.2: Participant Ready to Start the Experiment.

4.2.3 First Relaxation Phase

After the warm-up section, we moved towards the first relaxation phase. Relaxing music [42] was played for 2 minutes, and the participants were asked to listen. The main purpose of this phase was to establish a baseline of relaxation before moving to the first part of the experiment. EEG dataset collection was started from this phase.

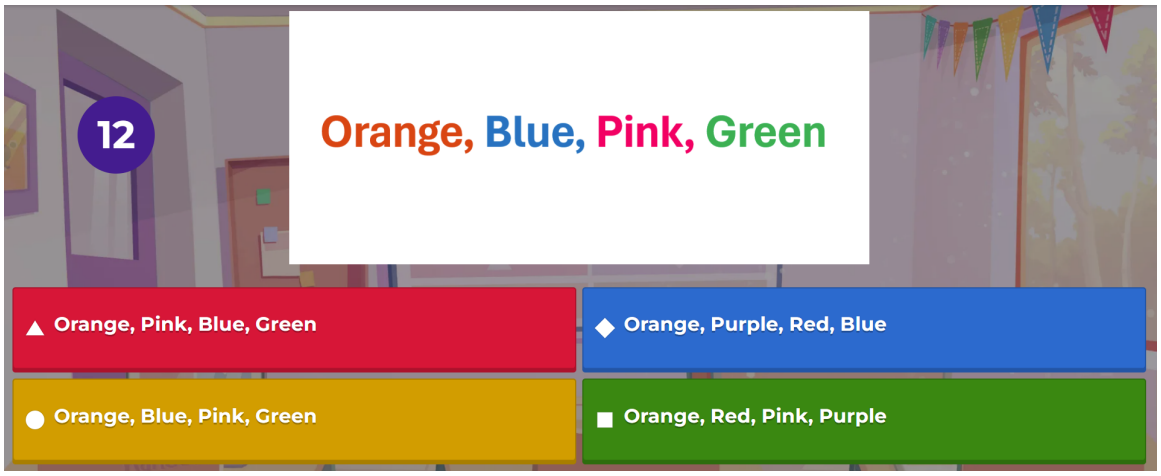


Figure 4.3: Congruent Stroop color-word test.

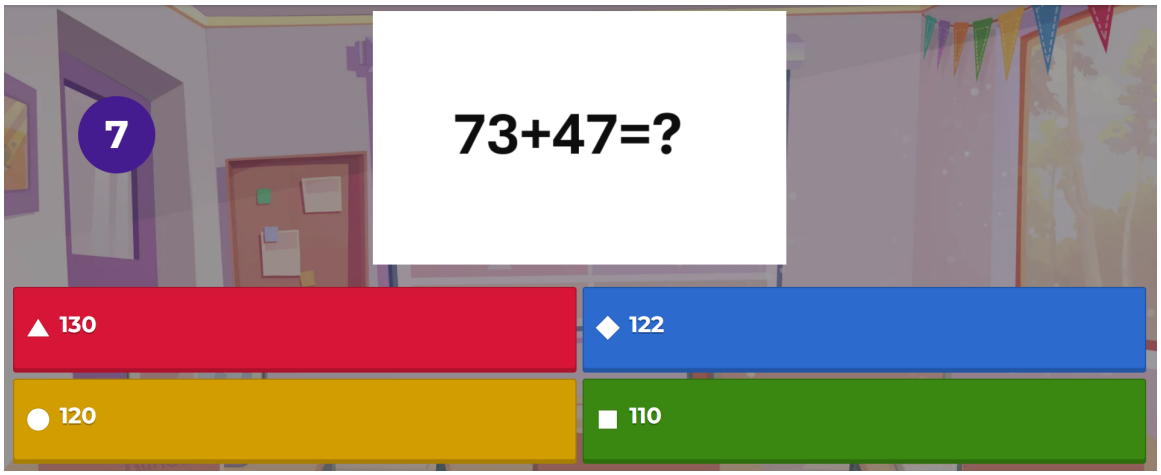


Figure 4.4: Easy mental arithmetic task.

4.2.4 Medium Stress Phase

This was the first part of the experiment. It consists of the congruent Stroop color-word test and an easy mental arithmetic task. The congruent Stroop color-word test is the one in which the color of the ink of the printed word matches the name of the word. It consisted of 15 stressor questions with a countdown approach. The Figure 4.3 presents an example of Congruent Stroop color-word test, whereas Figure 4.4 depicts the easy mental arithmetic task. After this phase, we moved towards the second relaxing phase.

4.2.5 Second Relaxing Phase

After completing the first part of the experiment, we moved towards the second relaxing phase. Relaxing music [43] was played for two minutes, and the participants were asked to listen. Relaxing music was played to establish a baseline of relaxation before moving towards the last part of the experiment.

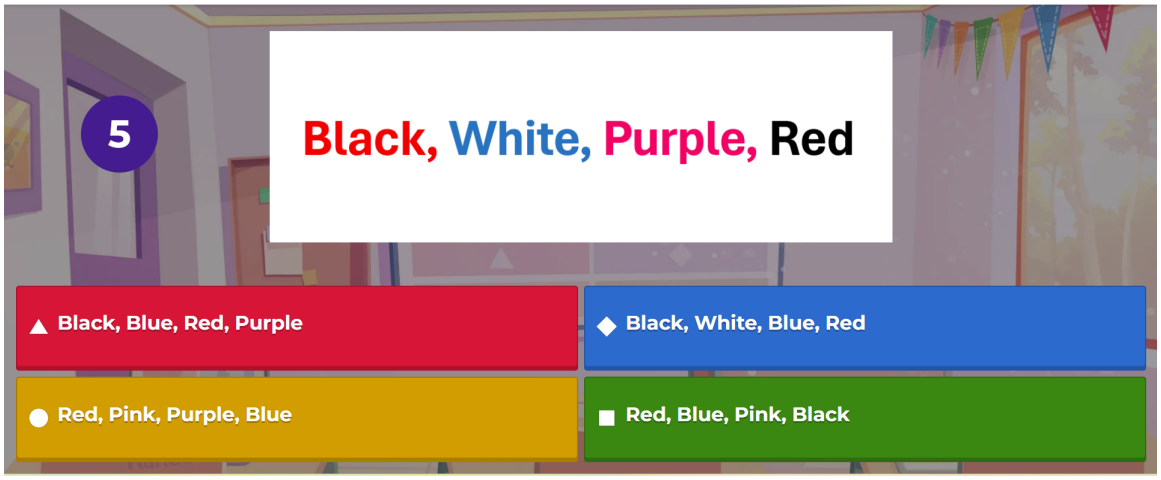


Figure 4.5: Incongruent Stroop color-word test.

4.2.6 High Stress Phase

This was the second part of the experiment. It contained the Incongruent Stroop color-word test and a difficult mental arithmetic task. Incongruent Stroop color-word test is the one in which the color of the ink of the printed word is different from the name of the word. There were a few

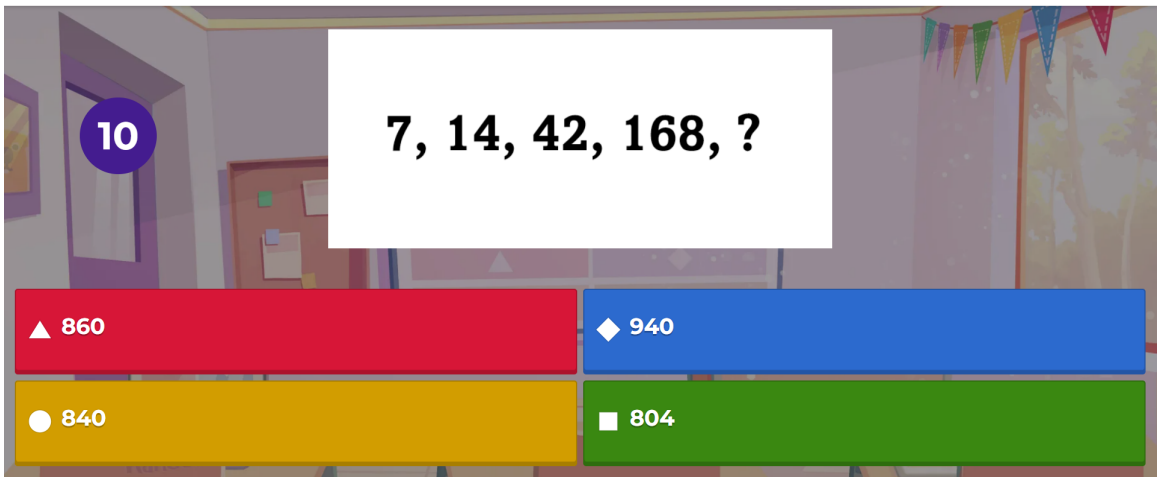


Figure 4.6: Difficult mental arithmetic task.

puzzles in a difficult mental arithmetic task. There were 15 stressor questions with a countdown approach. The countdown time was lower compared to the medium stress phase to induce higher stress. The Figure 4.5 presents an example of the Incongruent Stroop color-word test, whereas Figure 4.6 depicts the difficult mental arithmetic task. After this phase, we moved towards the last stage of the data collection.

4.2.7 Third Relaxing Phase

Again, relaxing music [44] was played for two minutes, and participants were asked to listen. The aim behind this phase was to establish a baseline of relaxation. This phase concludes the data collection. The experiment was performed only once. After the completion of the data collection phase, the EEG headband was removed from the heads of the participants. Participants were asked to fill self-evaluation questionnaire. Self-evaluation questionnaire was designed using Google Forms. Each participant filled out the self-evaluation questionnaire, asking questions regarding the levels of stress they observed during different parts of the experiment. We included filled questionnaires of our five participants in the Appendix A.

We designed our experiment in a way so that participants would perceive no stress during relaxation phases, medium stress during the first part of the experiment, and high stress during the second part of the experiment. In the self-evaluation questionnaire, most of the participants confirmed that they perceived stress in the same order that we designed it. Hence, our EEG dataset contains three

stress states: relaxation, medium stress, and high stress. Relaxation-related data was collected during three relaxation phases of the experiment, medium stress-related data was recorded during the first part of the experiment, and high stress-related data was gathered during the second part of the experiment. After the data collection phase, we moved towards the preprocessing stage.

4.3 Preprocessing

The collected EEG contains artifacts such as jaw clenching, eye blinking, and muscle movements. These artifacts should be removed before data analysis because they decrease the signal-to-noise ratio. Preprocessing makes our data readily available for further analysis. This study utilized the Muse S EEG headband for data collection. Muse S records data at a sampling frequency of 256 Hz. The Mind Monitor application was used to collect EEG data from the Muse headband. It connects with the Muse headband through Bluetooth. The data was sent through the Mind Monitor application for further processing. The Muse headband provides both raw and preprocessed EEG data. Muse has an in-built module for performing preprocessing and removing artifacts from the collected EEG data.

4.3.1 Preprocessing by Muse

Preprocessing removes unwanted noise from the EEG signal and is an important step before performing feature extraction. Muse EEG headband provides both raw and preprocessed EEG data. Muse has an inbuilt notch filter that eliminates electrical power noise from recorded EEG data ranging between 45 Hz to 64 Hz [27]. Muse provides EEG data after removing noise at a sampling frequency of 256 Hz, with a 2uV root mean square value of noise [27]. It has a Driven Right Leg (DRL) circuit between the frontal and Fpz electrode that actively cancels the noise. DRL circuit provides feedback regarding the contact quality between electrodes and the scalp and is responsible for providing clean EEG data. Muse computes three parameters, such as amplitude, variance, and kurtosis, from the EEG data and provides them to the decision tree classifier, which classifies the signals with low values for these parameters as clean. Muse converts raw EEG data from the time domain to the frequency domain by employing Fast Fourier Transform with the help of an onboard

digital signal processing module. It provides five bands in the frequency domain, such as delta, theta, alpha, beta, and gamma.

4.3.2 Preprocessing Phase

This section illustrates the steps followed to preprocess raw EEG data. EEGLab [45] was employed to remove unwanted noise and preprocess raw EEG data. It is an open-source tool that is widely utilized in neuroscience studies to perform preprocessing of the raw EEG data. The EEGLab toolbox is also widely used to visualize EEG data. It can also perform segmentation, plot the frequency spectrum, and be employed for the analysis of EEG data. It contains modules to perform Baseline correction, filtering, interpolation, and Independent Component Analysis (ICA) of the raw EEG data. We performed the following preprocessing steps to remove artifacts and preprocess raw EEG data.

4.3.3 Baseline Correction

It is a very important step in preprocessing and usually the first one that removes unwanted drifts from the raw EEG data. It is a crucial step because it improves the signal-to-noise ratio by eliminating unwanted DC offsets present in the raw EEG data. DC offset can occur due to the poor contact between electrodes and the scalp. It standardizes recorded EEG data by removing the impacts of the impedance of electrodes that can introduce different DC offsets across different trials. EEGLab has a module to perform Baseline correction and remove the unwanted DC offsets present in the raw EEG data.

4.3.4 Filtering

Filtering is another important preprocessing step that removes unwanted frequency components from raw EEG data. Artifacts can be present in the raw EEG data in the form of low-frequency drifts. Electromagnetic interference causes power line noise in the recorded EEG data. Filtering removes both low-frequency drifts and power line noise. It improves the signal-to-noise ratio of the recorded EEG data by removing unwanted artifacts. We designed a bandpass filter to remove the unwanted frequencies and only keep the frequency range that we are interested in. We set the high

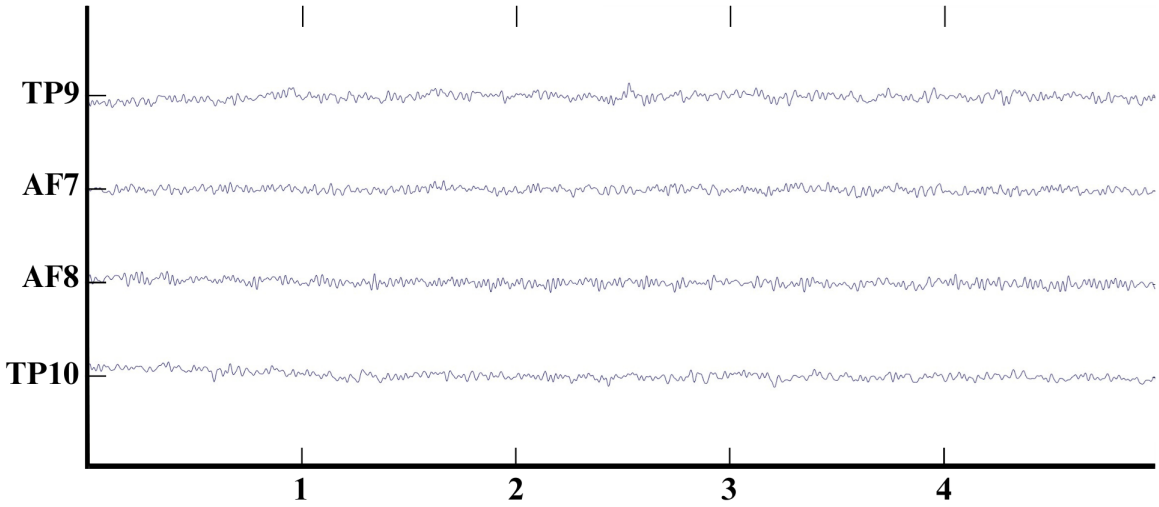


Figure 4.7: Filtered EEG data.

cutoff frequency at 0.5 Hz and the low cutoff frequency at 50 Hz for our bandpass filter. EEGLab has a Finite Impulse Response filter that we utilized to design our bandpass filter. The Figure 4.7 depicts the filtered EEG data.

4.3.5 Interpolation

When recording EEG data, sometimes the contact between the electrodes of the EEG headband and the scalp is lost, which causes the EEG data to be lost. It is indispensable to fill in the lost EEG data before performing analysis on the data. Interpolation is the preprocessing technique that accounts for the problem of lost EEG data during recording. Interpolation can employ various techniques to fill in the empty or lost data present in the recorded EEG data. EEGLab has a module to perform interpolation and allows for two different types, namely linear and spline interpolation. Linear interpolation assumes a straight-line relationship between neighboring electrodes and employs the neighboring sensor values to account for the missing values of the bad sensor. Spline Interpolation employs a smooth curve between neighboring sensors and utilizes the neighboring sensor values to fill in the missing values for the bad sensor. We employed spline interpolation to fill the missing values because it keeps the smoothness of the recorded EEG data intact. Spline interpolation can capture complex spatial relationships between neighboring sensors, which is indispensable for EEG data analysis. Linear Interpolation is only capable of capturing a linear relationship between

sensors.

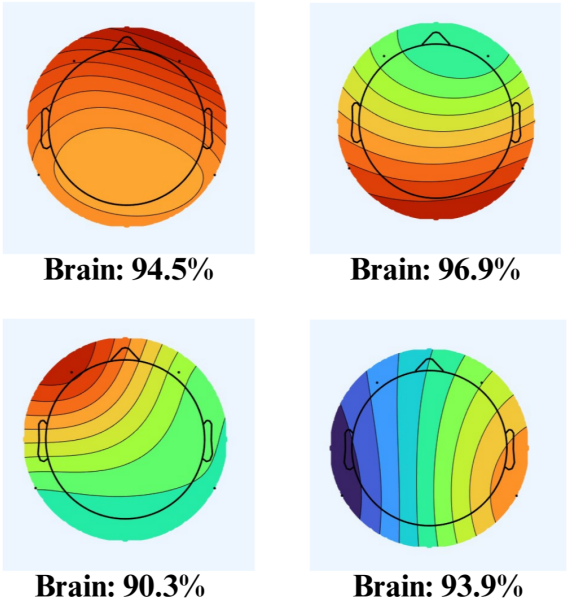


Figure 4.8: Independent Component Analysis (ICLabel).

4.3.6 Independent Component Analysis (ICA)

It is one of the most important steps in the preprocessing because it removes artifacts such as jaw clenching, eye blinking, and muscle movements, etc. from the recorded EEG data. It decomposes a mixed source of signal into its individual components. It improves the signal-to-noise ratio by removing unwanted signals present in the raw EEG data. EEGLab has many functions, such as Jader, Infomax, and SOBI, etc., that allow for the ICA decomposition. We opted for Infomax to perform ICA on the recorded EEG data. EEGLab offers an ICLabel plugin that automatically labels the artifactual components, such as power line noise, eye blinking, and muscle movements, etc., present in the recorded EEG data. It is mandatory to first employ ICA to decompose the EEG signal into its various ICA components; only then ICLabel can be utilized to automatically label those artifactual components. It classifies artifactual components into seven distinct categories, such as Brain, Eye, Heart, Muscle, Channel Noise, Line Noise, and Others. It classifies the components with a probability of more than 90% of being in the muscle or eye artifacts. The artifactual components that were labelled by ICLabel were removed from the recorded EEG data by employing the artifact

rejection module present in EEGLab. The Figure 4.8 depicts the brain components for some of the participants by employing ICLabel. ICA marks the end of the preprocessing phase; next, we moved towards the feature extraction step.

Although we collected data from 50 participants, Muse S could not record data for 4 participants reliably. Muse S malfunctioned while recording data for those 4 participants, and because of that, most of the data was lost, and the recorded data for those 4 participants was not reliable. Because of that, we did not consider those 4 participants while doing the analysis. The analysis of this study is based on the remaining 46 participants' data.

4.4 Feature Extraction

After performing preprocessing, the next step was to extract features required to train classifiers for performing multistage stress classification. The raw EEG data was in the time domain we needed to convert it to the frequency domain. We employed the Discrete Fast Fourier Transform to convert the time domain signal to the frequency domain signal. The Fast Fourier transform provides us with the frequency spectrum of the signal, providing insights about the frequency components. It provides the frequencies and the corresponding amplitudes at those frequencies. We have to compute the power spectral density (PSD) from the provided frequencies and amplitudes. We obtained the PSD using the Pwelch and Pburg [46] methods present in MATLAB. After obtaining the PSD, the power for the specific bands of interest such as delta (0.5Hz – 4Hz), theta (4Hz – 8Hz), alpha (8Hz – 12Hz), beta (12Hz – 30Hz), and gamma (30Hz – 50Hz) can be obtained by integrating the PSD over specified frequency ranges. Muse S EEG headband has four sensors for collecting EEG data located at the positions TP9, AF7, AF8, and TP10. There were five brain wave band power values from each sensor. So, there were 20 power band values in total. After receiving the power band values, we performed a statistical analysis to observe the relationship between different power band values.

4.4.1 Feature extraction for subject-dependent analysis

Statistical features were computed from the power band values by employing statistical measures such as min, max, mean, and standard deviation. Min represents the smallest value present in the dataset, whereas max denotes the largest value present in the dataset. Min and max together provide insight into the range of the dataset, which highlights the spread of data points. Mean provides the average of the data points present in the dataset. It is the measure of the central tendency of data points. Standard deviation illustrates the spread of the data points of the dataset from the mean. Lower standard deviation means most of the data points are closer to the mean, whereas a larger standard deviation means the data points are spread out over a wider range from the mean. The extracted features are represented through the Table 4.1. We computed min, max, mean, and standard deviation over a sliding window of 1 second.

Sr.	All brain waves (Delta, Theta, Alpha, Beta, Gamma)	
1	Minimum	**Features**
2	Maximum	
3	Mean	
4	Standard Deviation	

Table 4.1: Feature Extracted for Subject-Dependent Analysis

4.4.2 Feature extraction for Autoencoder and LSTM model

When we employed statistical features for the subject-independent evaluation, we could not achieve good values for performance evaluation metrics. Hence, we have to compute other features to capture indispensable patterns present in the time series EEG data to train the model. The following were the features computed for the autoencoder and LSTM-based model.

Power band Values: Mean of the power band values of all five brain waves, such as delta, theta, alpha, beta, and gamma, from the four sensors were computed to train the autoencoder and LSTM-based model. **Entropy:** Entropy measures the complexity, unpredictability, and uncertainty present in a signal with respect to time. We employed the Shannon Entropy function to compute Entropy features from time series EEG data. **Phase difference:** It is the difference in phase between signals captured from different sensors. In terms of EEG, it represents the offsets between

brain waves captured from different sensors. We utilized the Hilbert Transform to compute the instantaneous phase of the brain waves. When we subtract the instantaneous phase of one brain wave from the other brain wave, we obtain the phase difference between the two brain waves. **Phase Locking Value:** It implies the consistency of the phase difference between brain waves captured from different sensors over time. We employed the Hilbert transform to calculate phase locking features between brain waves. **Coherence:** It is a correlation between brain signals in the frequency domain. It represents how consistently the phase and amplitude of the brain waves remain at a specific frequency. We employed the SciPy library to compute the coherence features. The extracted features are represented through the Table 4.2. Power band values, entropy, phase difference, phase locking value, and coherence were computed by employing a sliding window of 2 seconds with a step size of 1 second.

Sr.	All brain waves (Delta, Theta, Alpha, Beta, Gamma)	
1	Power band values	**Features**
2	Entropy	
3	Phase difference	
4	Phase locking value	
5	Coherence	

Table 4.2: Feature Extracted for Autoencoder and LSTM model

4.4.3 Feature extraction for score prediction

Statistical features could not perform well when employed for the score prediction. We performed feature engineering to obtain the feature that will generate the least error when employed as a feature to train the random forest regressor model. We started considering different ratios for the power band values. We tried almost all the possible combinations of ratios for brain wave power band values. We performed feature engineering only on a small portion of the EEG dataset to avoid data leakage. When we added alpha and gamma brain waves power band values and divided them by the sum of delta, theta, and beta brain waves, gave us the least error when employed as a feature. There is a reason behind this feature generating the least error. Alpha is highly associated with the focus, and gamma is related to intelligence. Whereas delta, theta are

low-frequency brain waves that are associated with low consciousness and inhibition. The alpha and beta ratio is associated with stress, where a higher beta indicates more stress. Hence, we utilized (Alpha+Gamma)/(Delta+Theta+Beta) as a feature to train the random forest regressor model to accurately predict the scores of the participants.

Stressor questions were hosted in Kahoot. Kahoot offers a countdown approach, meaning that the stressor questions should be answered within the specified time. Once the countdown is finished, Kahoot will not take the answer and will assign a zero score for that particular stressor question. Participants will get a score from Kahoot only when they answer correctly within the specified time. Scores also depend on the time taken: If the participants take more time to answer correctly, they will achieve a lower score compared to the participants who take less time to answer correctly. Because of this, the score of participants also depends on the time taken. The fact that Kahoot assigns more score to participants who take less time to answer correctly is related to intelligence. Intelligent participants should take less time to answer correctly compared to less intelligent participants. Intelligence and time have a relation that has been studied in the literature [47], and Kahoot also utilizes the notion. Hence, we considered the time taken as a feature to train the random forest regressor model. We employed a ratio of the power band values of the brain waves and time taken to train the random forest regressor model to predict the scores of the participants. The extracted features are represented through the Table 4.3.

Sr.	Features
1	(Alpha+Gamma)/(Delta+Theta+Beta)
2	Response time

Table 4.3: Features Extracted for Score Prediction

4.5 Machine learning models

This section will discuss the classifiers employed in the study to perform stress detection.

4.5.1 Naive Bayes (NB)

NB [48] belongs to the family of probabilistic models and is a supervised machine learning model. It is employed by researchers to perform classification. It assumes that the features are conditionally independent while performing classification. It accurately classifies when the features have low correlation and are linearly independent. Naive Bayes can be effectively applied to perform stress classification of high-dimensional data, such as EEG, containing data from multiple sensors. We implemented a Naive Bayes classifier using the Scikit-learn library of Python. We employed StandardScaler to normalize the features and standardize the data.

4.5.2 Random Forest (RF)

RF [49] is an ensemble learning model that can be employed for both classification as well as regression problems. It is also a type of supervised machine learning model that works by training multiple decision trees. In classification problems, the target variable is the one with the majority of the votes received from the decision trees, whereas in the regression problem, the output is the average of all the decision trees. It also solves the problem of overfitting associated with decision trees. It employs several decision trees that make it robust to noise. It handles the complex features present in high-dimensional time series EEG data by creating multiple decision trees. We employed Python's Scikit-learn library to implement an RF model. The features were normalized by employing StandardScaler.

4.5.3 Support Vector Machines (SVM)

SVM [50] is also a supervised machine learning algorithm that classifies by finding an optimal line called a hyperplane that maximizes the distance between each class. It is one of the most famous classifiers, widely employed by researchers in classification problems. It performs multistage stress classification by finding a hyperplane that maximizes the distance between each stress class. We implemented an SVM model using Python's Scikit-learn library. Features were standardized by employing the StandardScaler class from the Scikit-learn library. We used a linear kernel while implementing the SVM classifier.

4.5.4 Long Short-Term Memory (LSTM)

LSTM [51] is a type of Recurrent Neural Network (RNN) that accounts for the problem of large data handling in traditional RNNs by only keeping the data that is relevant to the problem at hand. EEG is a high-dimensional time series data containing electrical activities from multiple sensors located at different positions of the scalp. The current electrical activity occurring inside the brain can be dependent on the electrical activities that have already occurred. It makes EEG sequential data, and LSTM is designed to capture temporal patterns present in the sequential data. We implemented a single-layer LSTM model that was responsible for processing the sequential time series EEG data. There were two fully connected layers for performing the classification task. There was a Relu activation function as well to add nonlinearity in the network. We employed PyTorch to implement the LSTM model. StandardScaler was utilized for feature normalization.

4.5.5 Convolutional Neural Network (CNN)

CNN [52] is a deep learning model that is widely employed in image classification problems; now, researchers have successfully applied it to solve classification problems in other domains [53]. It consists of a convolutional layer that applies the convolution operation on the input data. It is very effective in capturing complex spatial relationships between features. EEG contains high-dimensional data from various sensors located at different positions of the scalp. The relationship between these sensors is called a spatial relationship, and CNN can efficiently capture this spatial relationship. It can also capture the temporal patterns present in the time series EEG data. Hence, CNN can accurately classify the high-dimensional EEG data. We implemented a 1D CNN having 2 layers by employing the Pytorch library. Features were normalized by using the StandardScaler class.

4.5.6 Autoencoder and LSTM-based model with Attention Mechanism

When we performed subject-independent evaluation to remove subject bias by employing the LOSO cross-validation strategy, all the classifiers, NB, RF, SVM, LSTM, and CNN, performed

poorly. We implemented a two-stage deep learning model: the first stage consists of the fully connected autoencoder [54], whereas the second stage consists of the bidirectional LSTM model with an attention mechanism. An autoencoder is a deep learning model that can learn the compressed representation of input data. It encodes the input data to a lower dimension and then regenerates the data with minimal loss of information. Input was provided to the autoencoder for feature extraction and dimensionality reduction. Autoencoder while compressing the input data also removes noise and redundant data. It keeps the data that contains patterns that are indispensable for the classification task. It can learn non-linear relationships present in the input data, such as EEG. When we feed the output of the autoencoder, which contains cleaner and reduced-dimensional data, it helps and improves the classification ability of classifiers such as LSTM to capture the temporal patterns present in EEG data. Consequently, it is employed by the researchers to perform stress detection [55].

LSTM is also a deep learning model that we employed as the classifier to perform the stress classification, and we evaluated the model using LOSO cross-validation strategy. We employed a bidirectional LSTM model with two layers to perform the classification task. We also employed an attention mechanism to highlight the indispensable time steps present in the time series EEG data. We implemented a two-stage deep learning architecture with the help of Python's Scikit-learn library. Autoencoder was compressing the input to 64 dimensions, the hidden size for the LSTM model was 128, and the learning rate was 0.001. All these optimal hyperparameters were found using grid search. Relu activation function was used to add nonlinearity, and the Adam optimizer was employed to update the model weights after each batch.

4.6 Classification Metrics

The following are the performance evaluation metrics [56] employed to test the model's performance.

4.6.1 Precision

It tells about how many positive predictions were correct out of all the positive predictions made by the classifier. True Positive (TP) is the ability of the model to accurately predict the positive class, whereas False Positive (FP) depicts that the model inaccurately predict the positive class. False Negative (FN) depicts that the model could not detect the positive class.

$$\text{Precision} = \frac{TP}{TP + FP}$$

4.6.2 Recall

It is the measure of the ability of the model to accurately catch the actual instances.

$$\text{Recall} = \frac{TP}{TP + FN}$$

4.6.3 F1-score

It is the harmonic mean between recall and precision

$$\text{F1-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.6.4 Accuracy

It tells about the model's ability to accurately predict the target variable.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} * 100$$

4.7 Ethics Approval Procedure

To obtain the ethics approval, we completed the TCPS Core Certification. It is mandatory for all researchers who will interact with human participants. The certificates were sent to Concordia's electronic ConRAD system. The Summary Protocol Form was also completed and submitted to Concordia's electronic ConRAD system. The Summary Protocol Form contained questions regarding the usage of data, storage of data, compensation that will be provided to the participants, and risks involved. Flyer of the experiment and voluntary consent form were also submitted to Concordia's electronic ConRAD system. After submitting all these documents, the Human Research Ethics Committee of Concordia University issued us ethics approval under certificate number 0020206.

Chapter 5

Results

This chapter will report the results of the correlation analysis between the extracted features. We will also provide the results of the alpha over beta ratio during relaxation and cognitive task modes, followed by the results of the multistage stress classification using subject-dependent analysis. We will also demonstrate the results of the comparison between SAM40 and our proposed dataset by employing a CNN model. We will then report the classifier's performance using subject-independent analysis by employing the LOSO cross-validation strategy, followed by the results of the two-stage deep learning model. We will also present the results of the comparison between good and bad performers of our study using statistical tools such as line graphs and the Mann-Whitney U test. We will also highlight the results obtained after performing score prediction.

5.1 Correlation between extracted features

We plotted the correlation heatmaps between extracted features. It represents the linear relationship between extracted features. A correlation heatmap was computed using the Pearson correlation coefficient. Correlation heatmap between delta and theta is represented through the Figure 5.1, whereas correlation between beta and gamma is demonstrated through the Figure 5.2.

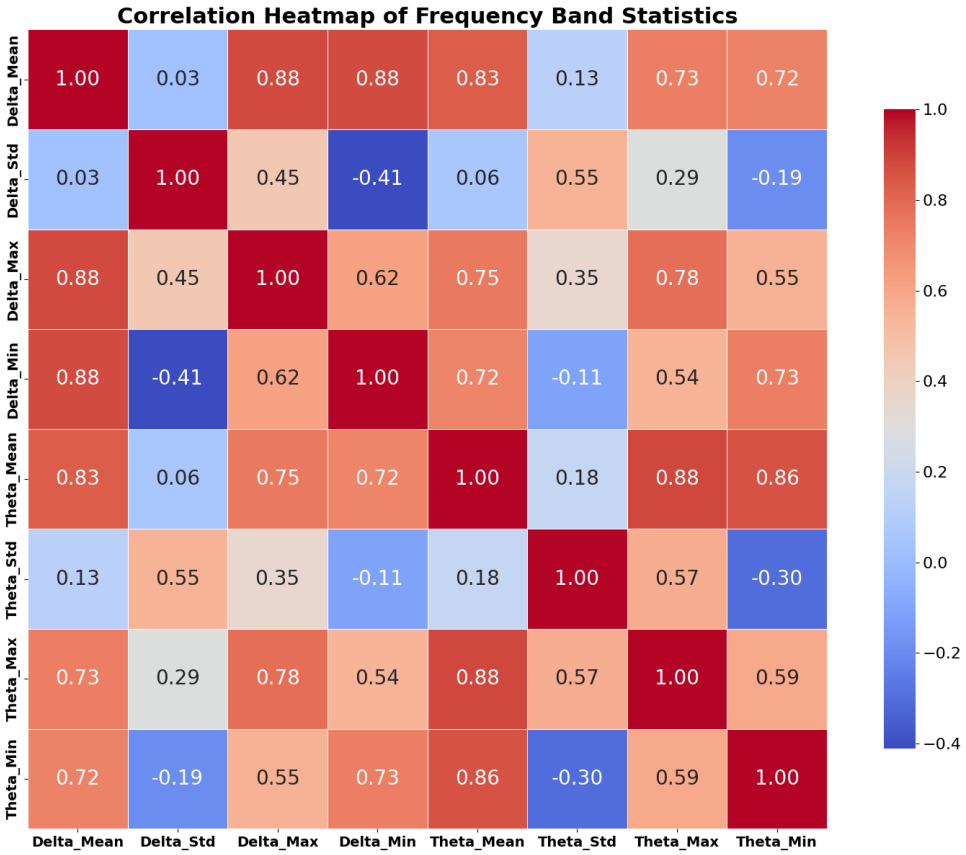


Figure 5.1: Correlation between Delta and Theta brain wave features.

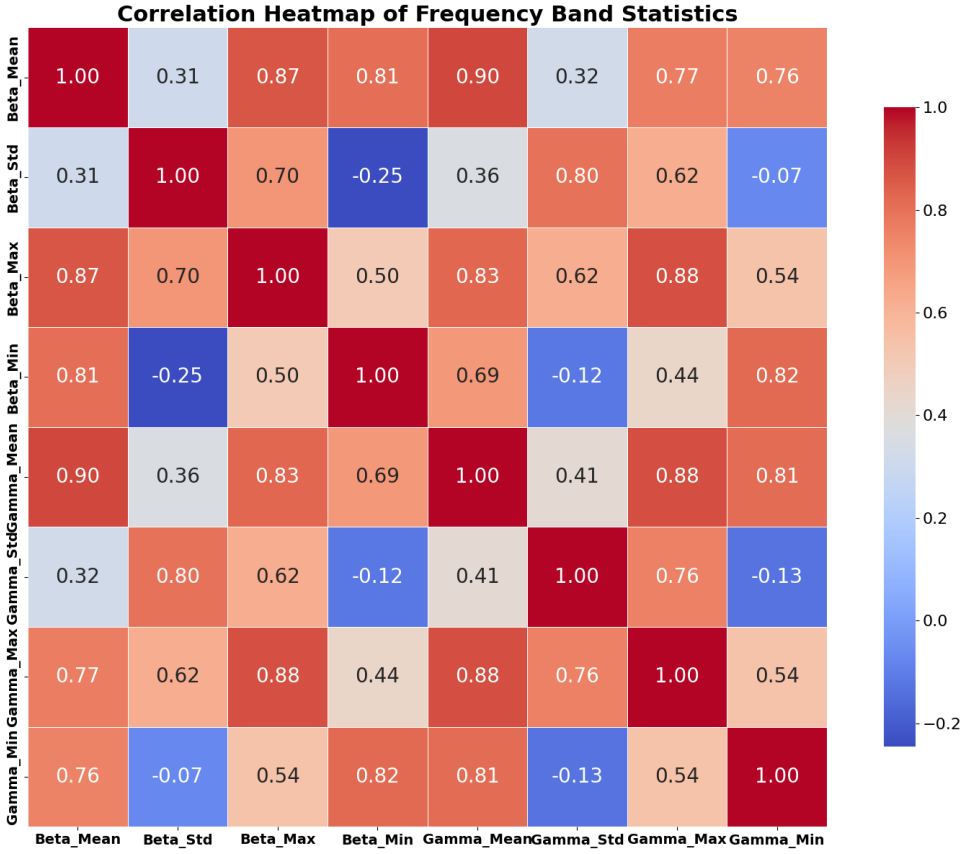


Figure 5.2: Correlation between Beta and Gamma brain wave features.

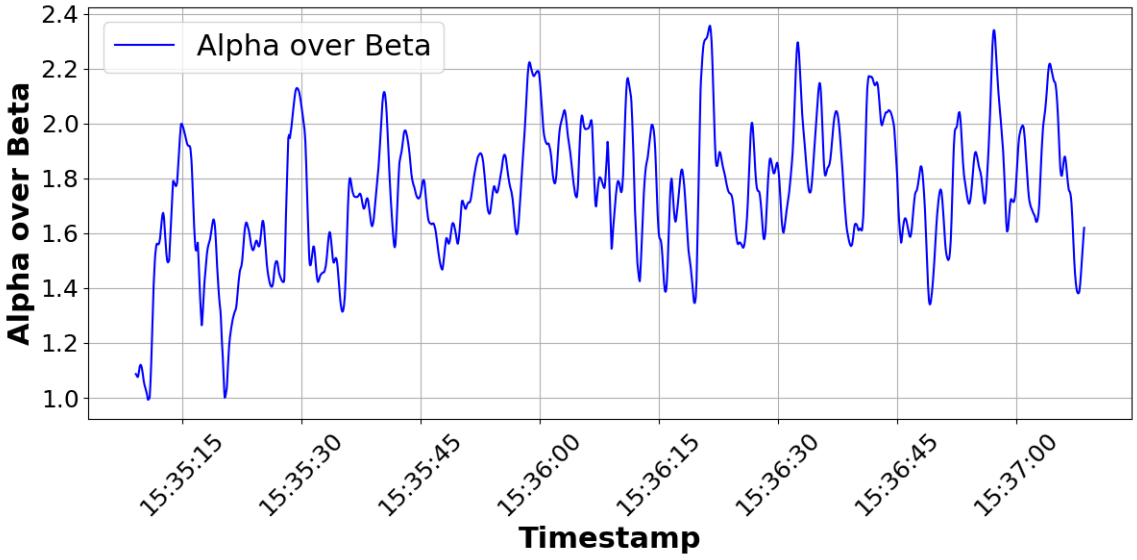


Figure 5.3: Alpha over Beta during relaxation.

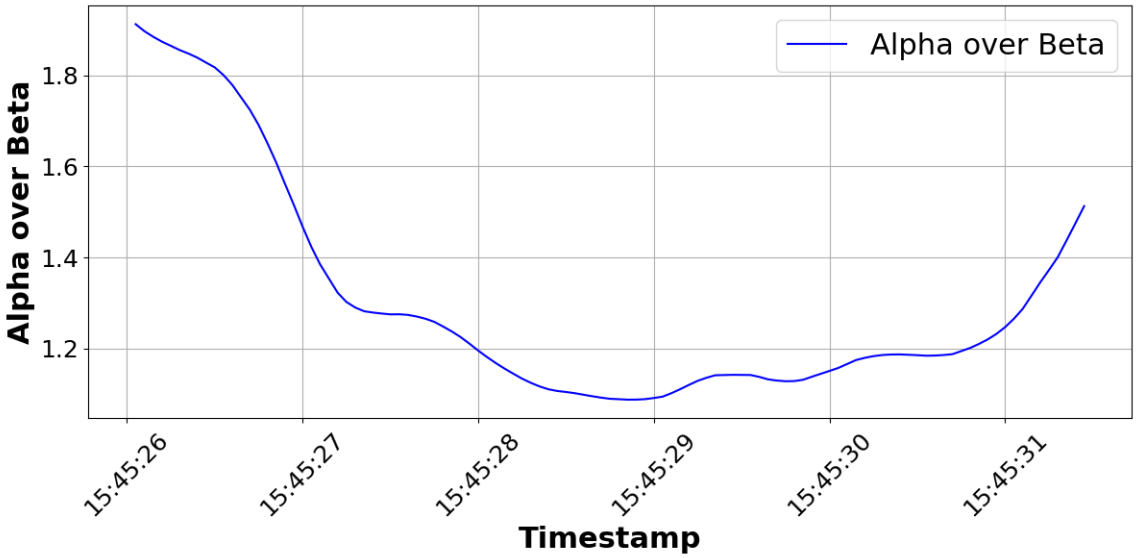


Figure 5.4: Alpha over Beta during a cognitive task.

5.2 Alpha over beta ratio during relaxation and cognitive task mode

The alpha over beta ratio was employed to evaluate the effectiveness of stressors employed in our study. Alpha over beta should be high during relaxation compared to cognitive task mode. The Figure 5.3 demonstrates an alpha over beta ratio during relaxation, whereas the alpha over beta ratio is depicted through Figure 5.4 during cognitive task mode. From the figures 5.3, 5.4, it is obvious that the alpha over beta ratio during relaxation was high compared to the cognitive task mode.

To indicate that there was a statistically significant difference in the alpha over beta ratio during relaxation and cognitive task modes, we employed a Paired t-test with a significance value of 0.05 (alpha = 0.05). We computed the mean values of alpha over beta during relaxation and cognitive tasks for each participant to use in a paired t-test. We computed T-calculated, p-value, effect size, and statistical power as demonstrated in the Table 5.1. Since p is less than 0.05 $p < 0.05$, that confirms that there was a statistically significant difference between the alpha over beta ratio during relaxation and cognitive task modes. The Appendix B contains more examples of line graphs for alpha over beta ratio during relaxation and cognitive task modes.

The paired t-test demonstrates that the participants' alpha over beta ratio was higher during relaxation compared to the cognitive task mode. After proving the effectiveness of stressors in

Significance Level	T-statistic	P-value	Effect Size	Statistical Power
0.05	-2.8871	0.006	-0.4257	0.81

Table 5.1: Paired T-test

inducing stress among participants, we moved to perform multistage stress classification using a 10-fold cross-validation strategy.

5.3 Multistage stress classification using subject-dependent analysis

We performed multistage stress classification using subject-dependent analysis by employing 10-fold cross-validation. We utilized five different classifiers, namely NB, RF, SVM, LSTM, and CNN. We employed performance evaluation metrics such as precision, recall, F1-score, and accuracy to evaluate the performance of the classifiers as depicted in the Table 5.2. From Table 5.2, it is clear that CNN outperformed other classifiers by achieving the highest values for performance evaluation metrics such as precision, recall, F1-score, and accuracy. NB underperformed compared to other classifiers by obtaining the lowest values for classification metrics demonstrated in the Table 5.2.

Model	Accuracy	Precision	Recall	F1-score
NB	73.86	0.68	0.71	0.69
RF	88.46	0.85	0.83	0.84
SVM	95.26	0.94	0.93	0.94
LSTM	99.75	1	1	1
CNN	99.9	1	1	1

Table 5.2: Classifiers Performance Metrics using 10-fold Cross-Validation

To remove the biased results, we shuffled the dataset before performing training. In 10-fold cross-validation, the dataset is divided into ten equal parts; during each iteration, nine folds are employed in the training phase, and the remaining one fold is used to test the model's performance. The process is repeated 10 times until each fold is used to test the model performance once. The evaluation performance metrics are averaged from all 10 iterations of the test set and are provided

as the model's performance. It helps to generalize the model's performance on unseen data because each data point is used once to test the model's performance.

5.4 Comparison between our proposed dataset and the SAM40 dataset

We employed CNN to perform a comparison between our proposed dataset and the SAM40 dataset. The SAM40 dataset contains EEG data from 40 participants and is publicly available. We evaluated the model's performance on both datasets by utilizing classification metrics such as precision, recall, F1-score, and accuracy, as illustrated in the Table 5.3. From the table 5.3, it is evident that CNN obtained higher values for performance evaluation metrics when classifying our proposed dataset compared to the SAM40 dataset.

Classification metrics	Proposed dataset	SAM40
Accuracy	99.9	98.6
Precision	1	0.99
Recall	1	0.97
F1-score	1	0.98

Table 5.3: Comparison between Datasets

5.5 Multistage stress classification employing subject-independent analysis

Although subject-dependent analysis obtained very high values for performance evaluation metrics, the values cannot be generalized to unseen participants' data. We performed a LOSO cross-validation strategy to generalize our results to unseen participants' data. We employed the same performance evaluation metrics, such as precision, recall, F1-score, and accuracy, to report model performance as presented in the Table 5.4. CNN and LSTM achieved the highest joint classification accuracy of 52% as depicted in the Table 5.4. The performance evaluation metrics obtained by all classifiers are very low compared to the classification metrics achieved through 10-fold cross-validation by employing subject-dependent evaluation.

Model	Accuracy	Precision	Recall	F1-score
NB	0.19	0.2	0.18	0.13
RF	0.5	0.48	0.42	0.41
SVM	0.48	0.45	0.43	0.4
LSTM	0.52	0.43	0.43	0.42
CNN	0.52	0.45	0.44	0.43

Table 5.4: Classifiers Performance Metrics using LOSO Cross-Validation

5.6 Autoencoder and LSTM model with Attention Mechanism

We implemented a two-stage deep learning model to improve the performance of subject-independent analysis using the LOSO cross-validation strategy. The first stage consisted of a fully connected autoencoder performing feature extraction and dimensionality reduction. The second stage of the architecture contained a bidirectional LSTM model with an attention mechanism to perform the classification task. We employed an attention mechanism to outline the most valuable timesteps present in the EEG data. The proposed model was utilized to perform both binary and multistage stress classification using subject-independent evaluation. The model obtained an accuracy of 83% for binary classification, as displayed through the Table 5.5, whereas for three-stage stress classification model achieved an accuracy of 66% demonstrated through the Table 5.5. We also reported other performance metrics such as precision, recall, and F1-score, presented in the Table 5.5.

Performance Metrics	Accuracy	Precision	Recall	F1-score
Binary Classification	0.83	0.83	0.82	0.82
Three Level Classification	0.66	0.55	0.55	0.55

Table 5.5: Deep Learning Model Performance Metrics using LOSO Cross-Validation

The confusion matrix and the ROC curve is depicted for the binary stress classification through the Figures 5.5 and 5.6 respectively.

The confusion matrix and the ROC curve is illustrated for the three-stage stress classification through the Figures 5.7 and 5.8 respectively.

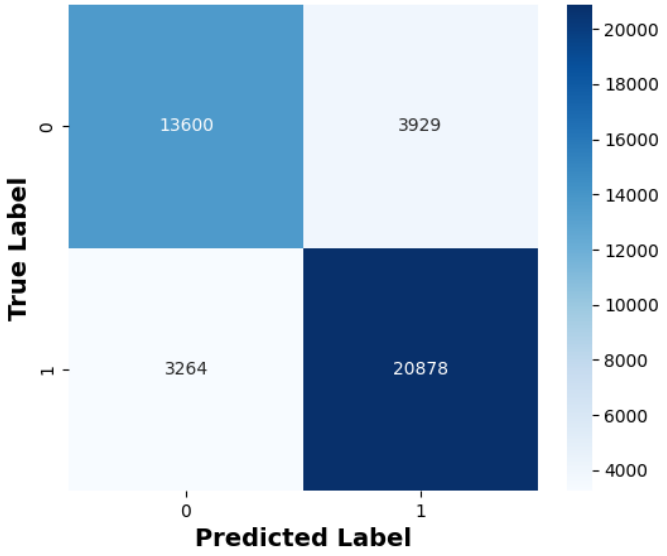


Figure 5.5: Confusion Matrix for Binary Stress Classification.

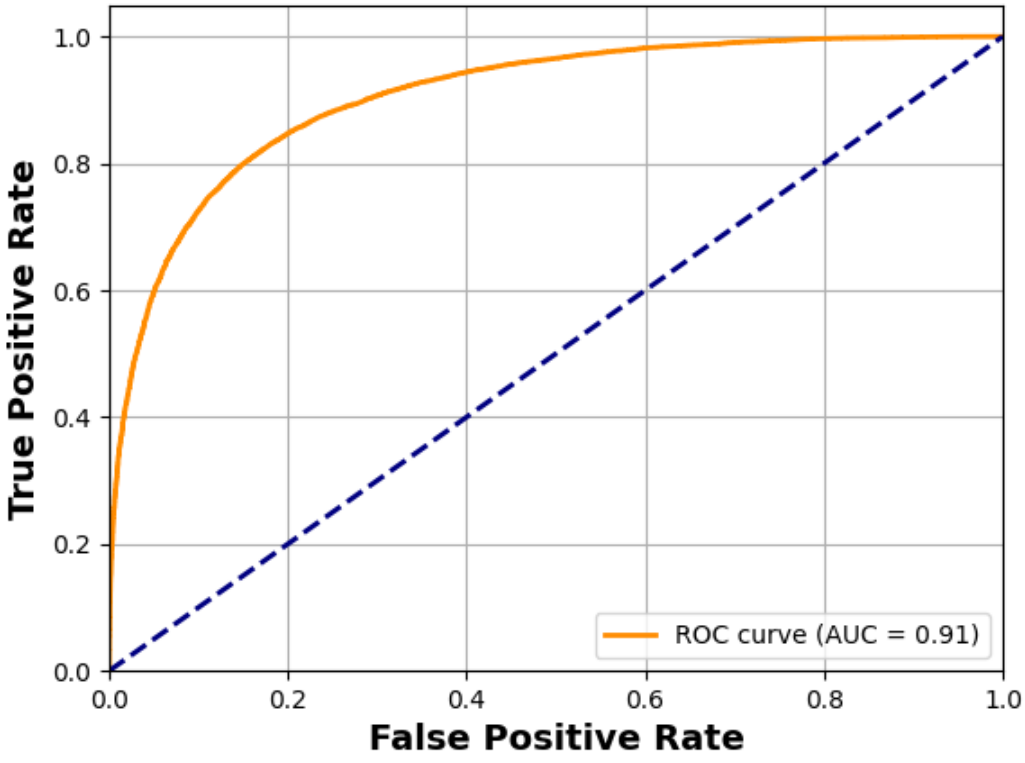


Figure 5.6: ROC Curve for Binary Stress Classification.

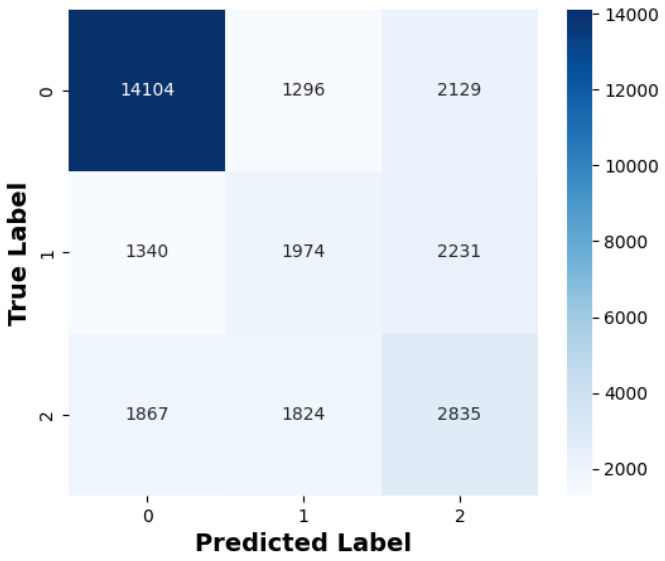


Figure 5.7: Confusion Matrix for Three-Stage Stress Classification.

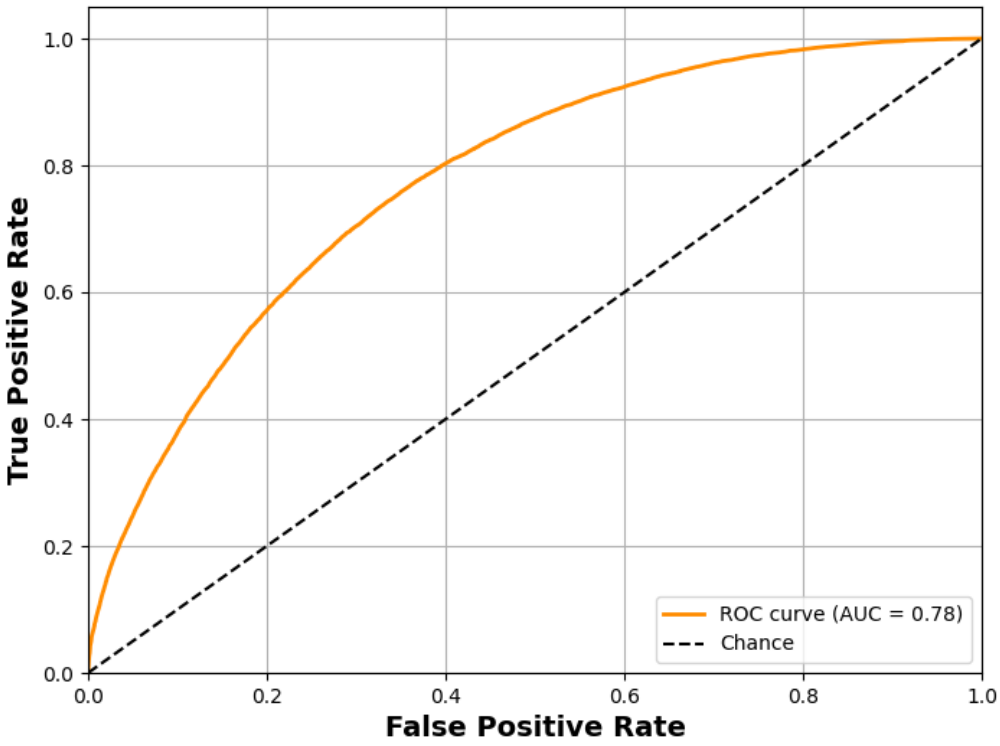
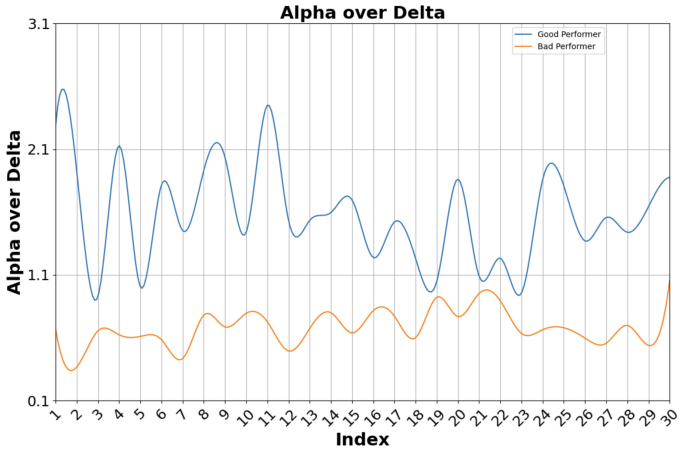
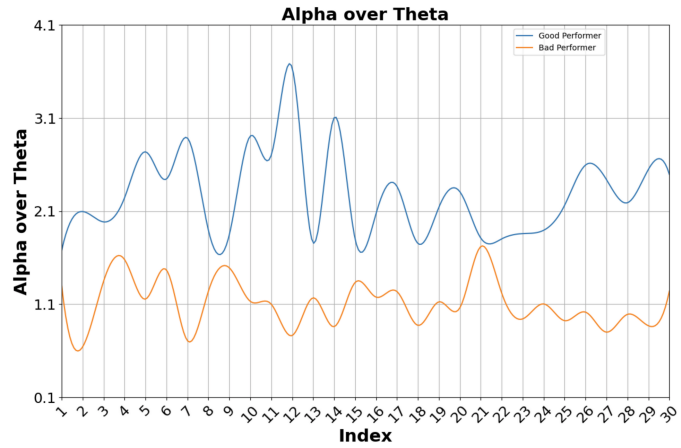


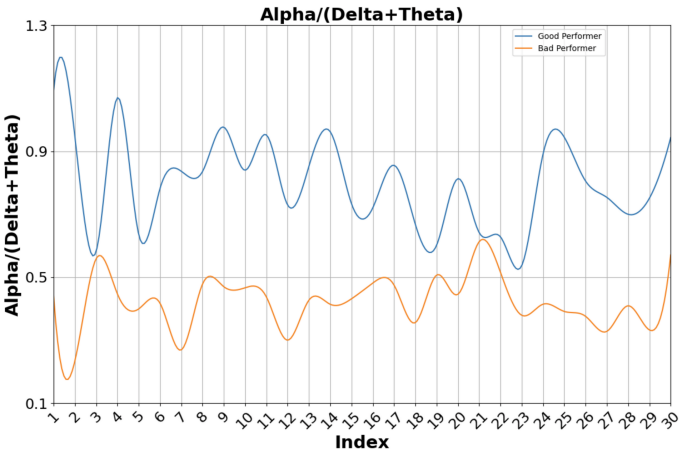
Figure 5.8: ROC Curve for Three-Stage Stress Classification.



(a) Alpha over Delta ratio between good and bad performers.



(b) Alpha over Theta ratio between good and bad performers.



(c) Alpha over Delta and Theta ratio between good and bad performers.

Figure 5.9: Comparison of the brain waves power band values during cognitive task.

5.7 Cognitive capacity analysis

We employed different brain wave ratios to perform a comparison between the brain waves of the best and worst-performing groups of our study. The Figure 5.9a depicts the alpha over delta ratio for one good and one bad performer in our study. The alpha over theta ratio for one good and one bad performer was plotted in Figure 5.9b. We also computed the ratio alpha over delta plus theta for one good and one bad performer of our group, and it is shown in Figure 5.9c. The Figure 5.9 depicts higher values of alpha over delta, alpha over theta, and alpha over delta plus theta for a good performer compared to a bad performer. The Appendix C contains more examples of line graphs for these ratios.

Alpha/Delta		**Alpha/Theta**		**Alpha/(Delta+Theta)**		
U-statistic	P-value	U-statistic	P-value	U-statistic	P-value	Statistical difference
100	0.00017	99	0.00024	100	0.00017	Significant

Table 5.6: Mann-Whitney U test

We made two groups of 10 participants, one group containing the best performers, whereas the other group contained the worst performers of our experiment. We employed the Mann-Whitney U test to prove that there is a statistical difference between the ranks of the brain waves of good and bad performers of our experiment in alpha over delta, alpha over theta, and alpha over delta plus theta. We used a significance level of 0.05 (alpha = 0.05) to perform the Mann-Whitney U test. The Table 5.6 depicts $p < 0.05$, which proves the statistical difference in ranks of the previously mentioned brain wave ratios between good and bad performing groups of our experiment. The effect sizes are 0.89, 0.85, and 0.9 for alpha over delta, alpha over theta, and alpha over delta plus theta, respectively.

5.8 Score prediction

We implemented a random forest regressor model to predict the score of participants based on the brain waves and response time. We evaluated the regression model performance by utilizing

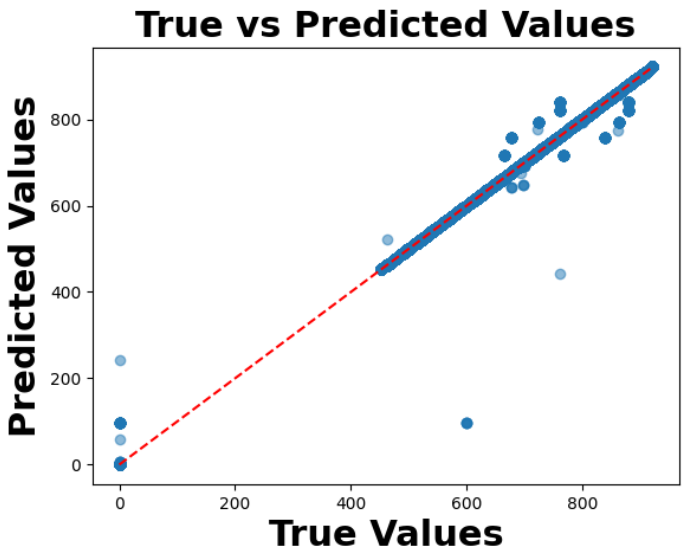


Figure 5.10: Random Forest Regressor Output.

metrics such as mean absolute error, mean square error, root mean square error, and R^2, as demonstrated through Table 5.7. The Figure 5.10 displays the output of the random forest regressor model.

Mean Absolute Error	0.009
Mean Square Error	1.39
Root Mean Square Error	1.18
R^2 score	1

Table 5.7: Random Forest Regression Metrics

5.9 Post-experiment Questionnaire

After the last relaxation music, participants were asked to fill out an online questionnaire. We employed Google Forms to devise our online questionnaire. The questionnaire contained questions regarding the stress levels that the participants perceived during different parts of the experiment. We designed our experiment so that the participants should feel no stress during the relaxation phases, medium stress during the first part of the experiment, and high stress during the second part of the experiment. Most of the participants responded that they were relaxed during relaxation phases, and the stress that they felt during the second part of the experiment was higher compared

to the first part of the experiment. The perceived stress scale questionnaire results for five of the participants are depicted in the Appendix A.1, A.2, A.3, A.4, and A.5.

Chapter 6

Discussion

This chapter will discuss the correlation analysis between extracted features, followed by the explanation of the results of the alpha over beta ratio during relaxation and cognitive task mode. We will elaborate on the results of multistage stress classification using subject-dependent analysis, followed by the illustration of the results obtained after comparison between our proposed dataset and the SAM40 dataset. We will investigate the results obtained after performing subject-independent stress classification. We will analyze the results of the proposed two-stage deep learning architecture using subject-independent analysis by utilizing the LOSO cross-validation strategy. We will explore the results obtained after performing a comparison between the brain waves of the best-performing and the worst-performing groups of our study. We will conclude this section with a discussion of the results of the random forest regression model.

6.1 Correlation between extracted features

Correlation represents the linear relationship between the brain waves. We plotted the correlation heatmaps between the low-frequency brain waves delta and theta, as depicted in the Figure 5.1, and also plotted the correlation heatmap between the high-frequency brain waves beta and gamma, as depicted in the Figure 5.2. Delta is a low-frequency brain wave that is highly associated with deep sleep and low consciousness. Theta is another low-frequency brain wave that is highly active during light sleep, low consciousness, and emotional stress in adults. Since both are low-frequency

brain waves and have similar features, we can see a strong linear relationship between them, as illustrated in the Figure 5.1. Beta is a high-frequency brain wave and is found mainly during learning, thinking, and information processing. Gamma is the highest frequency brain wave that is highly present during complex problem solving, high levels of information processing, and is also related to intelligence. Both brain waves are high frequency and have characteristics in common. Thus, we can see a strong linear relationship between them as depicted through the Figure 5.2.

6.2 Alpha over beta ratio during relaxation and cognitive task mode

Alpha brain wave frequency lies between 8 Hz to 12 Hz and is highly related to relaxation. Beta brain wave frequency ranges between 12 Hz to 30 Hz and is highly active during learning, thinking, and information processing. The fact that alpha is associated with relaxation and beta is present during problem solving, their ratio determines the stress level. The high alpha over beta ratio means that the person is relaxed, and if the alpha over beta ratio is low, that means the person is stressed. We plotted the alpha over beta ratio for one participant during relaxation and cognitive task modes, as demonstrated in the Figures 5.3 5.4 respectively. The alpha over beta ratio is high during relaxation compared to the cognitive task mode, which confirms the effectiveness of stressors in inducing stress during the cognitive task mode. The alpha over beta ratio is employed by physicians in the treatment of anxiety and depression in the neurofeedback training [57].

We employed a Paired t-test [58] to prove that there is a statistically significant difference between the ratio of alpha over beta during relaxation and cognitive task modes, as illustrated through the Table 5.1. The alpha over beta ratio mean values were computed over the interval and used by the paired t-test. Paired t-test compares the mean values from two related groups to see if there is a statistically significant difference between them [58]. The paired t-test compares the data points from related groups that are paired, meaning that it compares the same objects under two different conditions [58]. There are two important parameters in the t-test such as T-statistic and p-value. If the T-statistic is greater than the T-critical, that means there is a statistically significant difference between the mean values of two related groups. The other important parameter, that is, p-value, depicts that the measured difference might happen by chance. If the value of p is less than or equal

to 0.05, that is $p \leq 0.05$, then it means that the observed difference is not by chance and there is a statistically significant difference between the means of two related groups. From the Table 5.1, we can see that p is less than 0.05, which means there is a statistically significant difference in the alpha over beta ratio during relaxation and cognitive task modes. The effect size is -0.4257, which highlights that the alpha over beta ratio decreased as we moved from relaxation to cognitive task mode. As the stress increased during the cognitive task mode, this justifies the effectiveness of stressors employed to induce stress.

6.3 Multistage stress classification using subject-dependent analysis

In this section, we will discuss the results of classifiers after performing multistage stress classification.

6.3.1 Naive Bayes (NB)

Naive Bayes underperformed compared to other classifiers, as depicted in the Table 5.2. NB achieved performance evaluation metrics such as accuracy of 73.86%, precision of 0.68, recall of 0.71, and F1-score of 0.69. NB underperformed because it assumes conditional independence between features while classifying, but there is a strong correlation between extracted features, as depicted through Figures 5.1, 5.2.

6.3.2 Random Forest (RF)

It outperformed NB and achieved performance evaluation metrics such as accuracy of 88.46%, precision of 0.85, recall of 0.83, and F1-score of 0.84 as illustrated through the Table 5.2. RF trains multiple decision trees, and the target class is predicted based on the highest number of votes achieved. The classification metrics depict that RF can be utilized in performing multistage stress classification using 10-fold cross-validation. RF is robust to noise and accounts for the problem of overfitting normally associated with decision trees.

6.3.3 Support Vector Machines (SVM)

It performs classification by finding an optimal line called a hyperplane that maximizes the distance between each class. It outperformed both NB and RF by achieving higher classification metrics, for instance, accuracy of 95.26% and other classification metrics such as precision of 0.94, recall of 0.93, and F1-score of 0.94 as presented in the Table 5.2. The high values for the performance evaluation metrics depict that SVM can be effectively employed to perform multistage stress classification of high-dimensional EEG data.

6.3.4 Long Short Term Memory (LSTM)

It resolves the large memory handling issues associated with the traditional RNNs by keeping only the data that is relevant to the problem at hand. It outperformed NB, RF, and SVM by achieving a higher multistage classification accuracy of 99.75% and higher performance evaluation metrics such as precision, recall, and F1-score of 1, as exhibited through the Table 5.2. It is actively employed by researchers in the EEG classification problems [59] because of its effectiveness in finding the complex temporal relationships present in high-dimensional time series EEG data. The present electrical activity occurring inside the brain can be dependent on the previous electrical activities. LSTM is very effective in identifying such complex temporal patterns in time series EEG data.

6.3.5 Convolutional Neural Network (CNN)

It is widely used in image classification problems, but researchers have successfully applied it to solve classification problems in other fields as well [53]. It consists of a convolutional layer that applies the operation of convolution to the input data. EEG contains data from multiple sensors located at different positions of the scalp. There is a spatial relationship between these sensors, and CNN can effectively find this complex spatial pattern present in the time series EEG data. It can also identify the temporal patterns in time series EEG data. That is the reason it outperformed all the other classifiers, such as NB, RF, SVM, and LSTM, and achieved the highest classification accuracy of 99.9% as displayed in the Table 5.2. The values for other performance evaluation metrics, such as precision, recall, and F1-score, are also one.

6.4 Comparison between our dataset and the SAM40 dataset

The SAM40 dataset contains data from 40 participants and is publicly available. SAM40 dataset contains three stressors, namely the Stroop color-word test, the arithmetic task, and symmetrical mirror images. The participants were also provided with relaxation time between the cognitive tasks. The SAM40 dataset was collected by employing a 32-channel Emotiv Epoch EEG headset. When CNN outperformed the rest of the classifiers by achieving the highest multistage stress classification accuracy of 99.9%, we provided the SAM40 dataset as input to the CNN model. We wanted to perform a comparison between our proposed dataset and the SAM40 dataset. The Table 5.3 displays the results of the comparison between our proposed dataset and the SAM40 dataset. It is evident from the Table 5.3 that the CNN model achieved a higher classification accuracy of 99.9% on our dataset compared to the SAM40 dataset, on which the CNN model achieved a three-stage stress classification accuracy of 98.6%. Although the difference is not huge, the fact that our dataset contains data only from four channels, whereas the SAM40 dataset contains data from 32 channels, makes this difference huge. The reason behind better performance on our dataset is due to the better design of our stressors, as we employed a countdown approach. The fact that the participants saw their current standing after answering each stressor question also put them under performance stress.

6.5 Multistage stress classification employing subject-independent analysis

Classifiers achieved high accuracy and other performance evaluation metrics such as precision, recall, and F1-score when performing multistage stress classification using subject-dependent analysis by utilizing a 10-fold cross-validation strategy. The achieved results through subject-dependent analysis do not generalize to new participants' data. The reason behind this is the subject bias that is present in the subject-dependent analysis, and because of this, it is not a practical approach. The subject-independent analysis removes subject bias and is a practical approach. We performed subject-independent analysis by employing the LOSO cross-validation strategy. In the LOSO cross-validation strategy, the model is trained by leaving one subject out; the left-out subject is utilized to

test the model's performance. The procedure is reiterated until each subject is used once to test the model's performance in the testing phase. The classification metrics of all the individual participants are averaged to compute the overall classification metrics. None of the classifiers performed well in the subject-independent analysis, as exhibited in the Table 5.4. The highest multistage classification accuracy of 52% was achieved by LSTM and CNN. We also computed other performance metrics such as precision, recall, and F1-score for all the classifiers, as depicted through the Table 5.4.

6.6 Autoencoder and LSTM model with Attention Mechanism

All the classifiers achieved very low values for the performance evaluation metrics, as displayed through Table 5.4 in subject-independent analysis using the LOSO cross-validation strategy. We designed a two-stage deep learning model to improve the classification metrics in subject-independent analysis. The first stage consisted of a fully connected autoencoder to perform dimensionality reduction of the input features. The output of the autoencoder was passed to the second part of the deep learning model, which consisted of a bidirectional LSTM model with an attention mechanism performing the classification task. An attention mechanism highlights the most important timesteps present in the EEG data.

The deep learning model with only statistical features could not obtain high values for classification metrics. Additional features such as power band values, entropy, phase difference, phase locking value, and coherence were computed to extract more patterns for the model to learn and improve the performance. We came across the problem of class imbalance when performing LOSO cross-validation. We performed oversampling of the minority classes in the training phase only to avoid data leakage. Gaussian Noise was added to oversample minority classes during training with a standard deviation of 0.01. We used the proposed deep learning model for performing the binary as well as three-stage stress classification. The model achieved an accuracy of 83% for binary stress classification, and performance evaluation metrics were precision 0.83, recall 0.82, and F1-score 0.82 as presented through the Table 5.5. When performing the three-stage stress classification model achieved an accuracy of 66% and other classification metrics were precision 0.55, recall 0.55, and F1-score 0.55, as demonstrated through Table 5.5.

Macro average was used to compute classification metrics to depict the model's performance equally across all the classes. To improve the performance evaluation metrics for three-stage stress classification, discrete wavelet transform features were computed. The discrete wavelet transform features were only computed for multistage stress classification to extract additional patterns to improve performance. The random chance or guessing provides us 33% for three-stage stress classification. Our deep learning model achieved twice that accuracy, depicting that our model can capture temporal patterns present in time series EEG data. The accuracy of the multistage stress classification can further be increased by increasing the number of participants. To prove that we plotted a line graph between accuracy and the number of participants, as depicted in Figure 6.1. It is clear from Figure 6.1, that the accuracy increases with the increase in the number of participants.

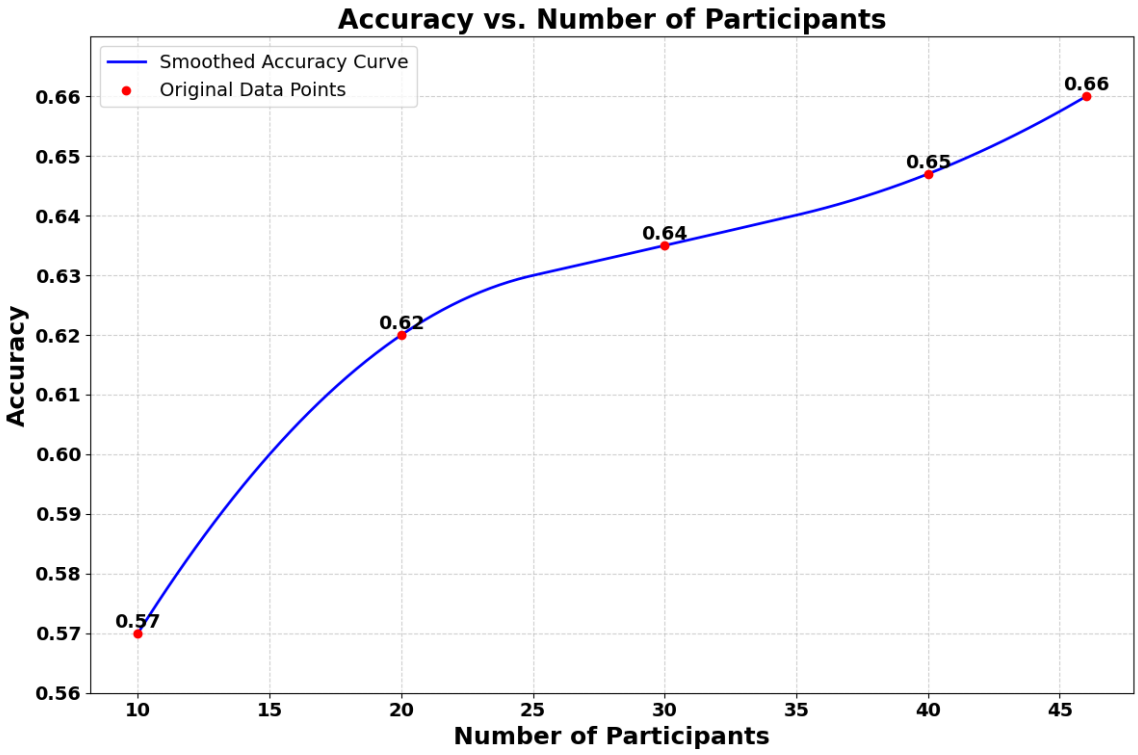


Figure 6.1: Relation between accuracy and number of participants

6.7 Cognitive capacity analysis

We made two groups of ten participants each, one group containing the best performers of our experiment, whereas the other group contained the worst performers of our experiment. We employed the Mann-Whitney U test [60] to prove that there is a statistical difference between the ranks of two different groups. We compared the alpha over delta, alpha over theta, and alpha over delta plus theta brain wave ratios of the best and worst-performing groups. We also plotted the mentioned brain wave ratios through line graphs as depicted through the Figures 5.9a, 5.9b, 5.9c respectively. From Table 5.6, since p is less than 0.05, that means there is a statistically significant difference between the ranks of the mentioned brain wave ratios between the best and the worst groups of our experiment. The alpha brain wave ratio is highly present during relaxation and a state of focus. Delta is a low-frequency brain wave that is associated with deep sleep, low consciousness, and inhibition. Theta is another low-frequency brain wave that is associated with light sleep, low consciousness, and emotional stress in adults. The presence of theta brain waves is also found during disappointment and frustration. We already saw a strong linear relationship between delta and theta brain waves through the correlation heatmap displayed in the Figure 5.1. In the best performers, alpha brain wave was found to be high as expected because alpha brain wave is associated with relaxation and a state of focus, depicting that the best performers were relaxed and focused during the cognitive task, which is the reason behind their good performance. The high delta and theta brain waves found among the worst performers of the experiment indicate that they were also focused during the early stages of the experiment. The presence of high delta waves avoided external distractions by repressing nonessential brain activities. The low-performing participants became frustrated and disappointed by low scores, which further increased the delta and theta brain waves and decreased the alpha brain wave and relaxation. Thus higher the alpha over delta, alpha over theta, and alpha over delta plus theta, means more focus and relaxed, and hence higher performance.

6.8 Score prediction

We implemented a random forest-based regression model to predict the scores obtained by participants for each stressor question based on the brain waves and the time taken to answer the stressor

question. Random forest trains multiple decision trees, and the target variable is predicted based on the average of all the decision trees. It also resolves the problem of overfitting normally associated with the decision tree. The Table 5.7 depicts the results of the random forest regression model after predicting the scores assigned to participants from Kahoot. The proposed model obtained a root mean square error of 1.18, illustrated in the Table 5.7, indicating that, on average, the predicted values can deviate 1.18 units from true values.

6.9 Data storage and source code

The participants' EEG data are saved in the CSV files separately for each participant. In the CSV file, there is a label column that contains labels 0, 1, and 2. Label 0 represents the instances related to relaxed state. Label 1 represents the instances of medium stress and Label 2 represents the instances of high stress.

The source code for all the classifiers is written in Python by employing libraries such as Scikit-learn for classical and PyTorch for deep learning models. All the codes and the dataset are publicly available [29].

Chapter 7

Conclusion and Future Work

This research investigated multistage stress classification by employing both subject-dependent and subject-independent analysis. To perform subject-dependent analysis, we employed 10-fold cross-validation, whereas to perform subject-independent analysis, we used the LOSO cross-validation strategy. We also analyzed the difference between subject-dependent and subject-independent analysis. We designed a two-stage deep learning model to perform stress classification and evaluated its performance by employing the LOSO cross-validation strategy. The proposed model achieved a binary classification accuracy of 83% and an accuracy of 66% for three-stage stress classification. We compared the cognitive capacity of the best and the worst performers of our experiment and observed that for better performance, alpha brain wave should be high, whereas delta and theta brain waves should be low. We also predicted the scores achieved by participants with an RMSE of 1.18.

The future work may focus on including more puzzles, problem-solving, and reasoning-type questions to broadly investigate the cognitive ability and intelligence of participants based on brain waves. We will recruit a higher number of participants in the experiment to generalize our findings.

Appendix A

Questionnaire

Perceived stress scale Questionnaire follows

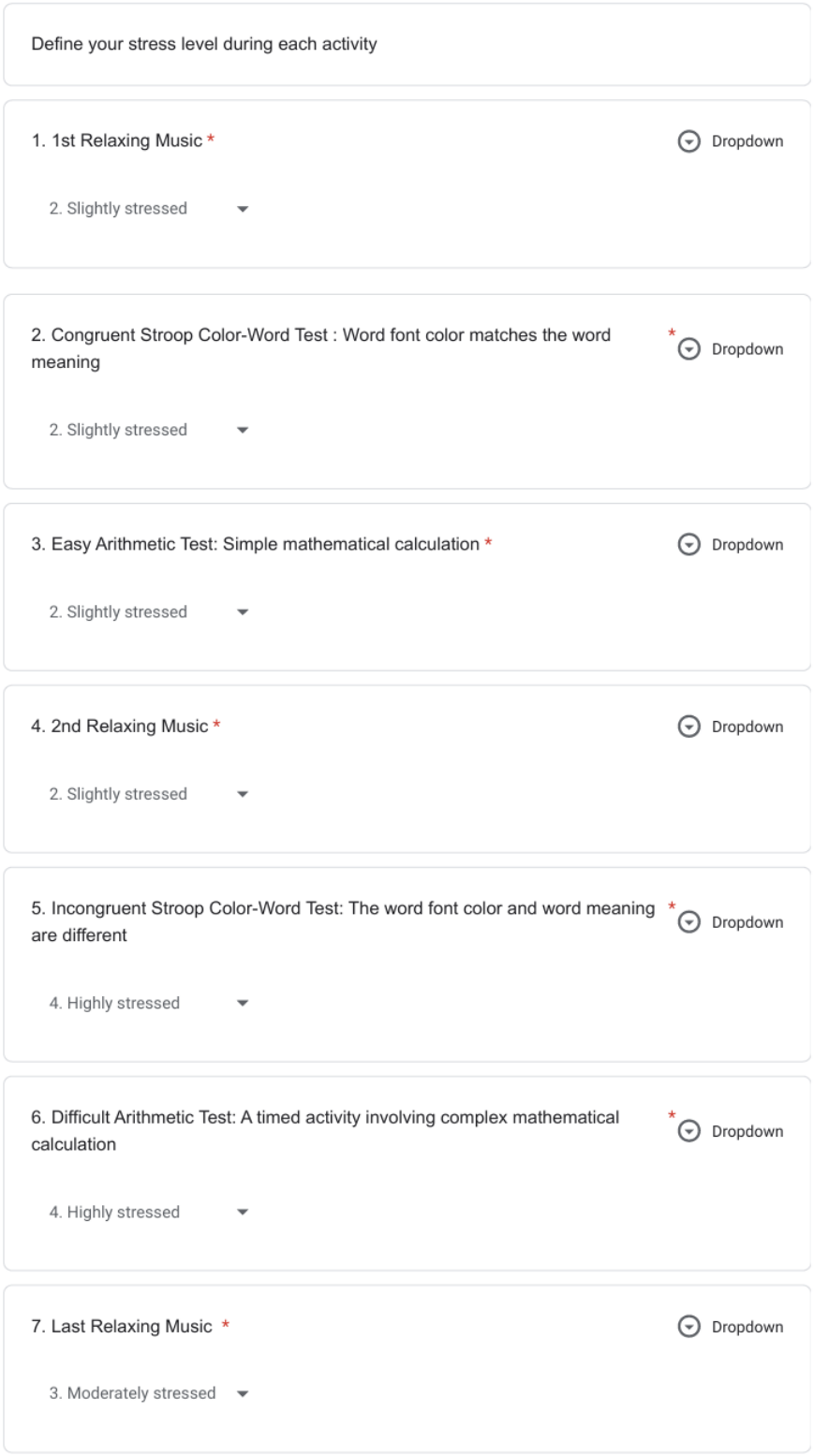


Figure A.1: Questionnaire 1

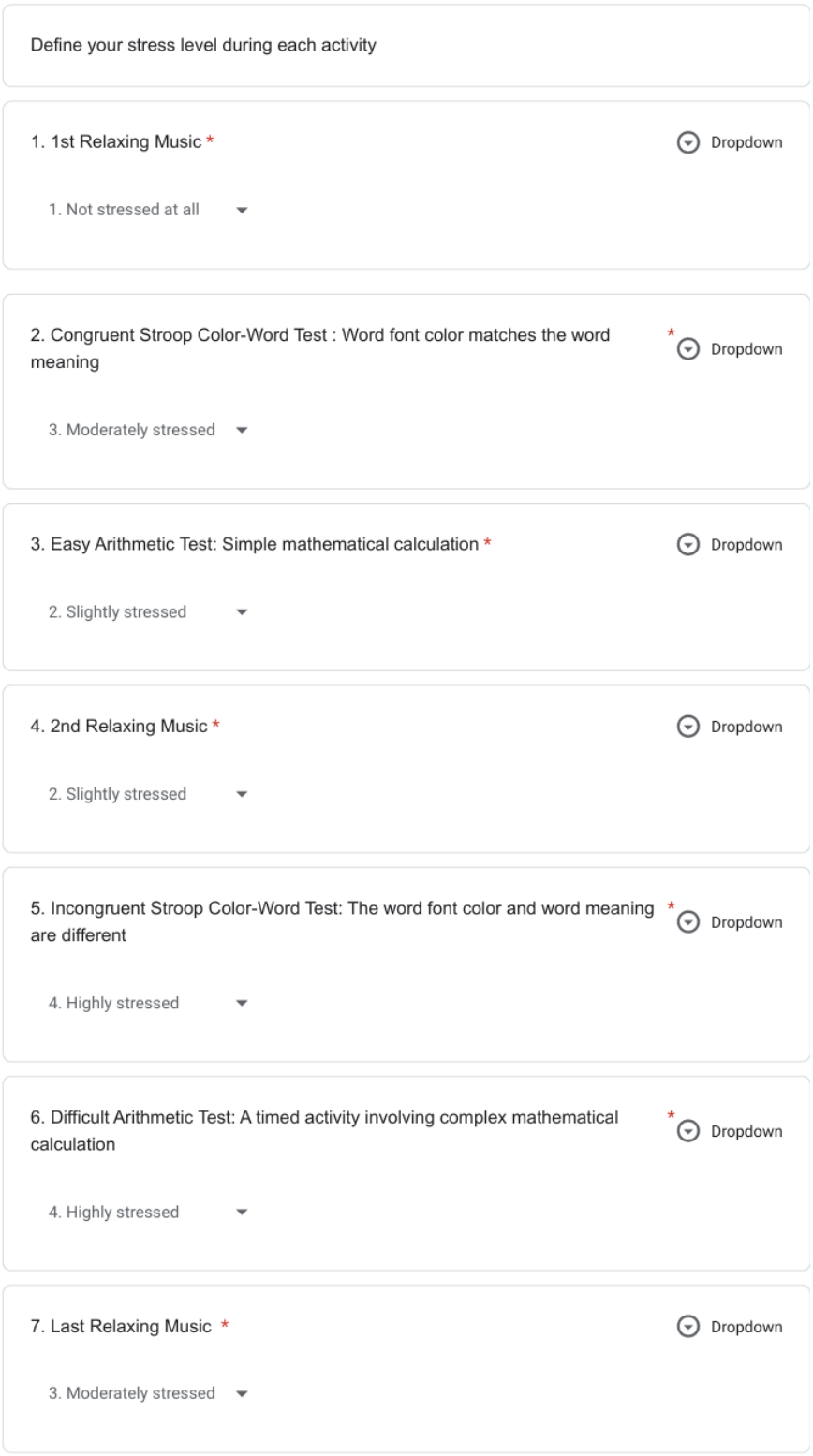


Figure A.2: Questionnaire 2

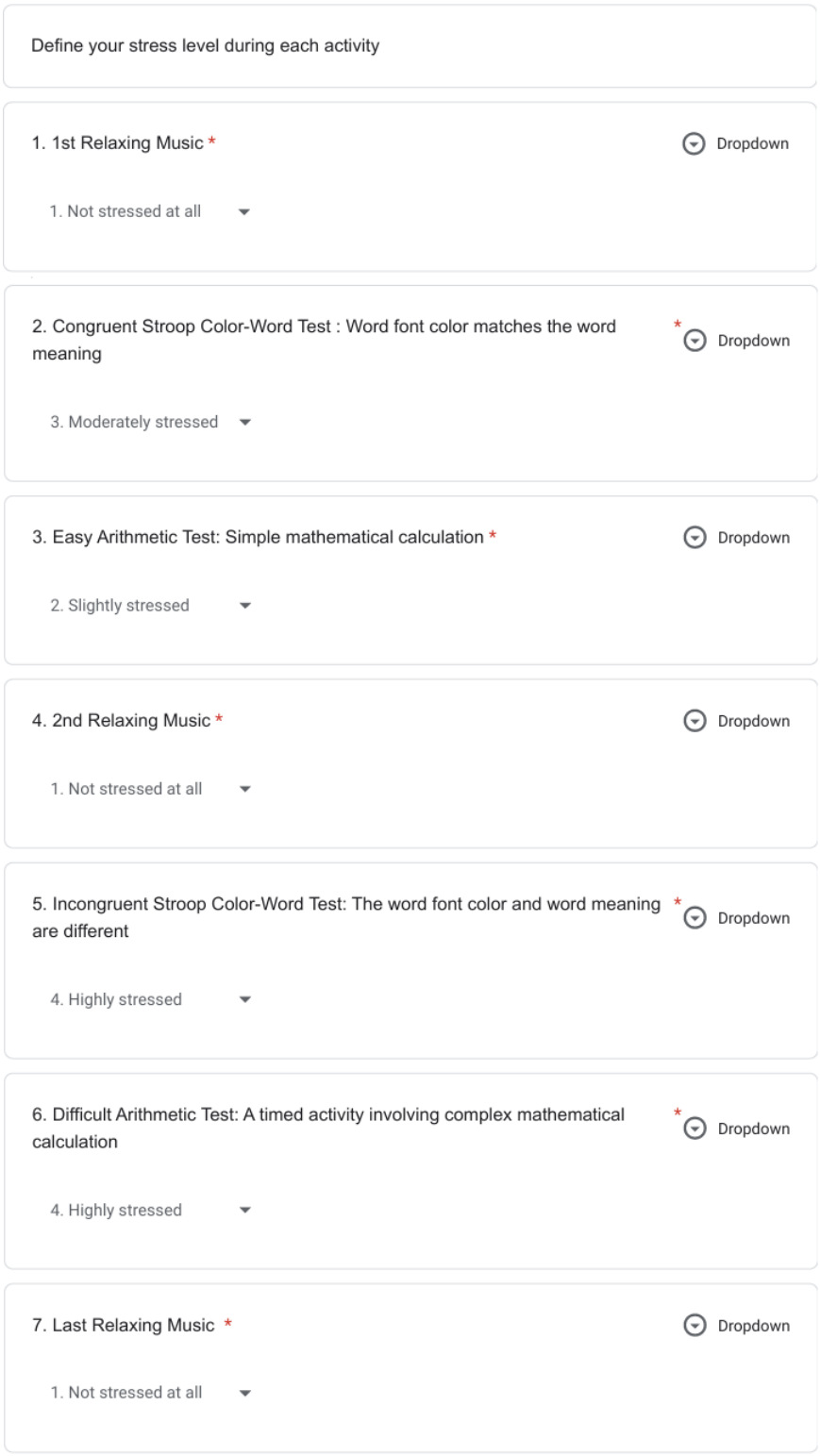


Figure A.3: Questionnaire 3

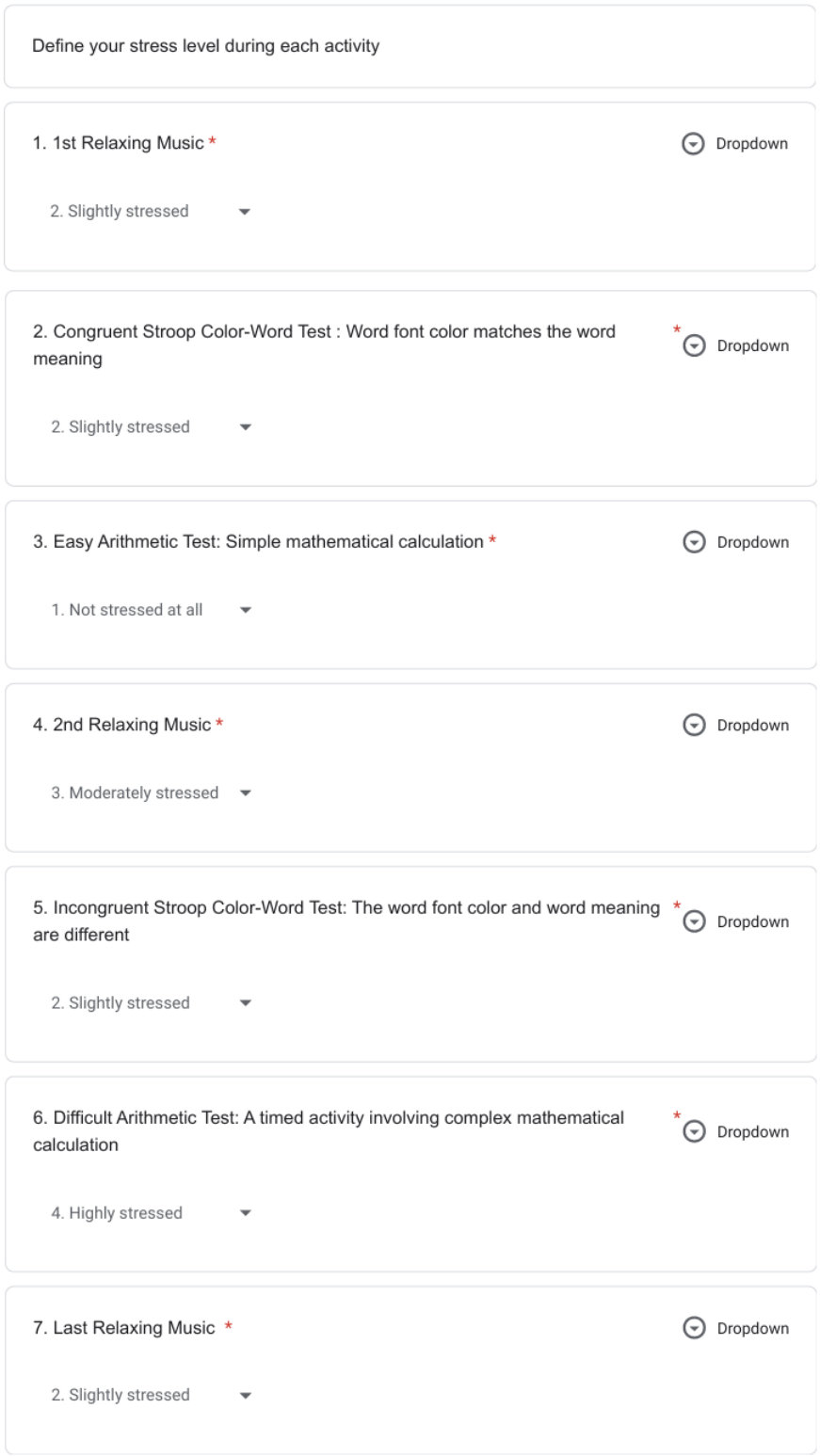


Figure A.4: Questionnaire 4

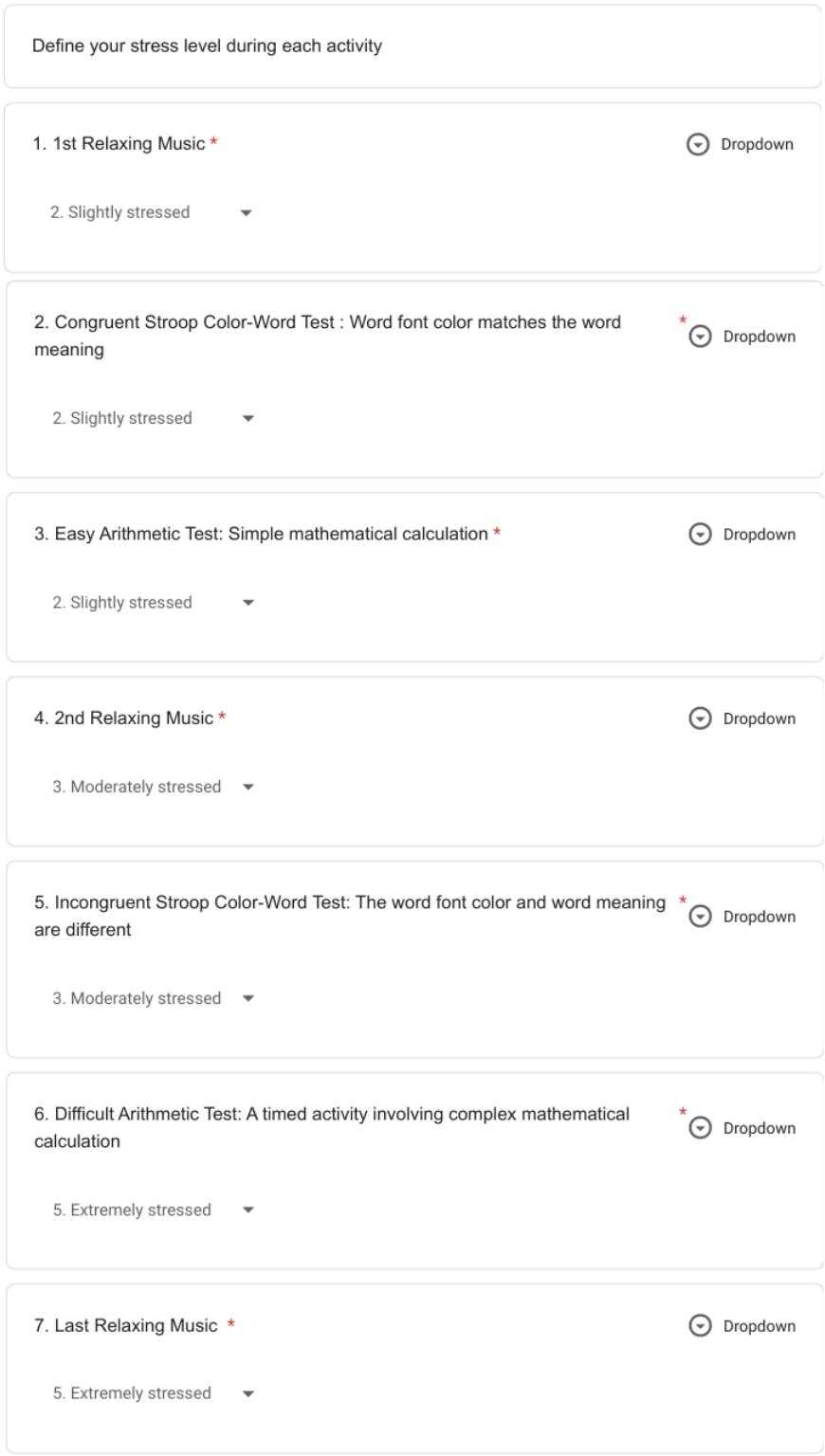


Figure A.5: Questionnaire 5

Appendix B

Alpha over Beta line graphs

Following are the line graphs for alpha over beta ratios during relaxation and cognitive task modes.

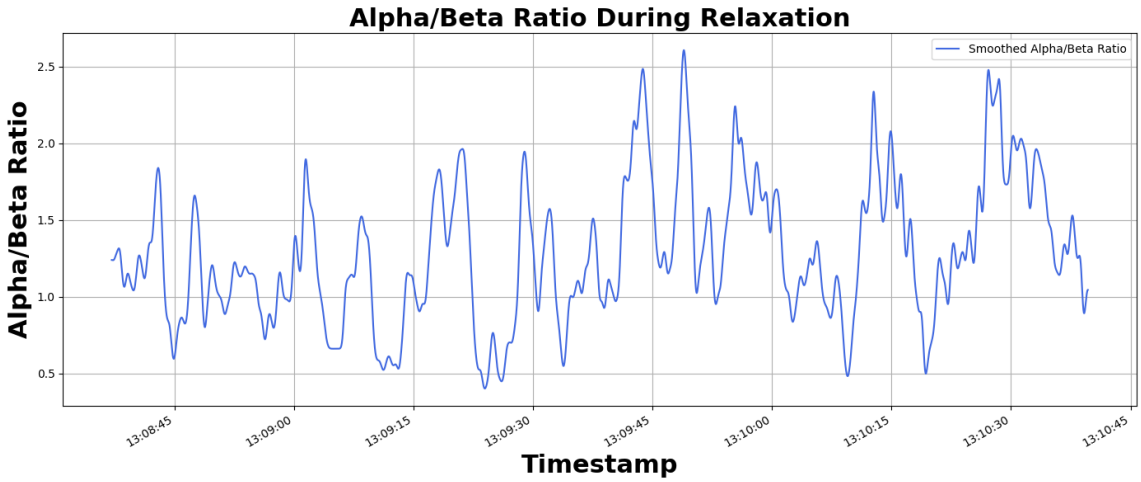


Figure B.1: Alpha over Beta ratio during relaxation first participant.

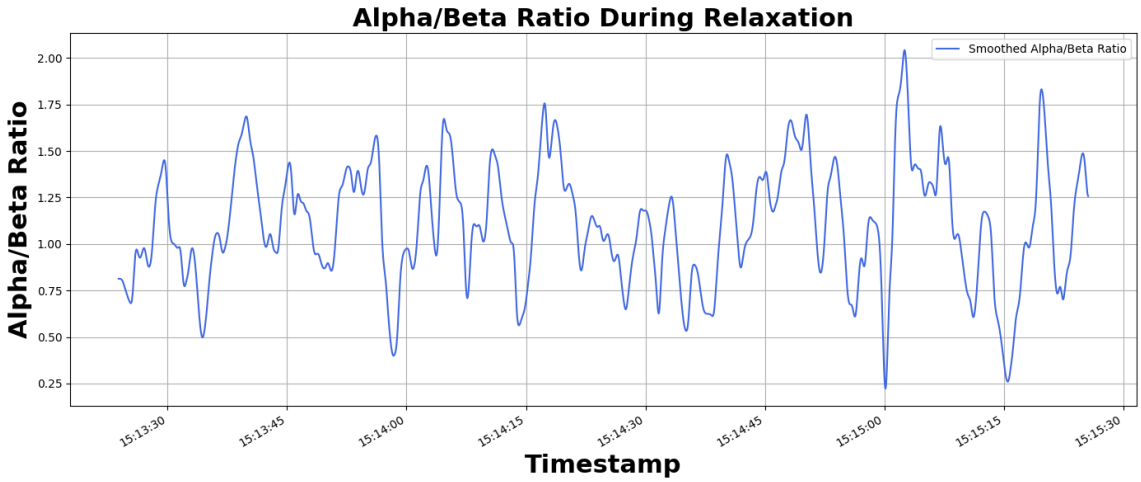


Figure B.2: Alpha over Beta ratio during relaxation second participant.

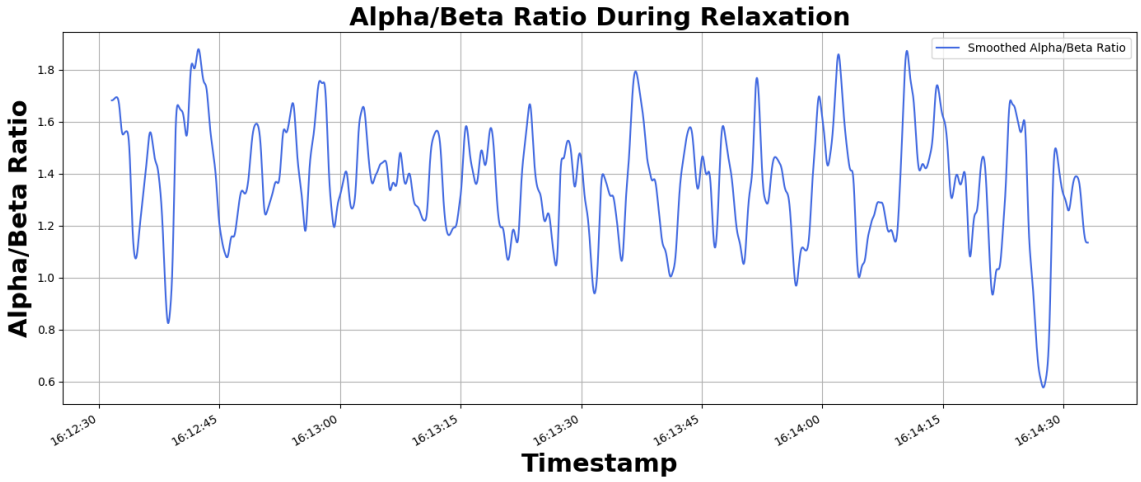


Figure B.3: Alpha over Beta ratio during relaxation third participant.

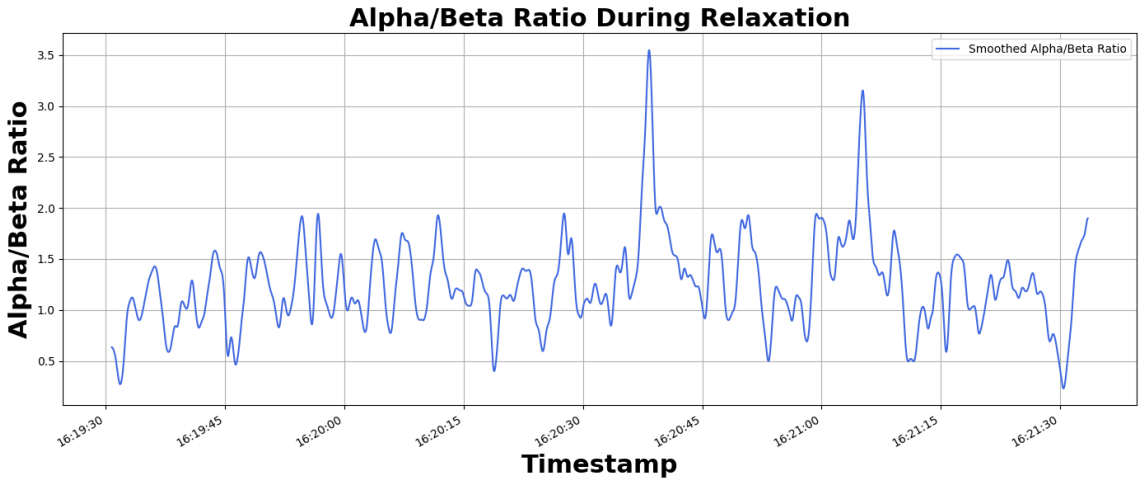


Figure B.4: Alpha over Beta ratio during relaxation fourth participant.

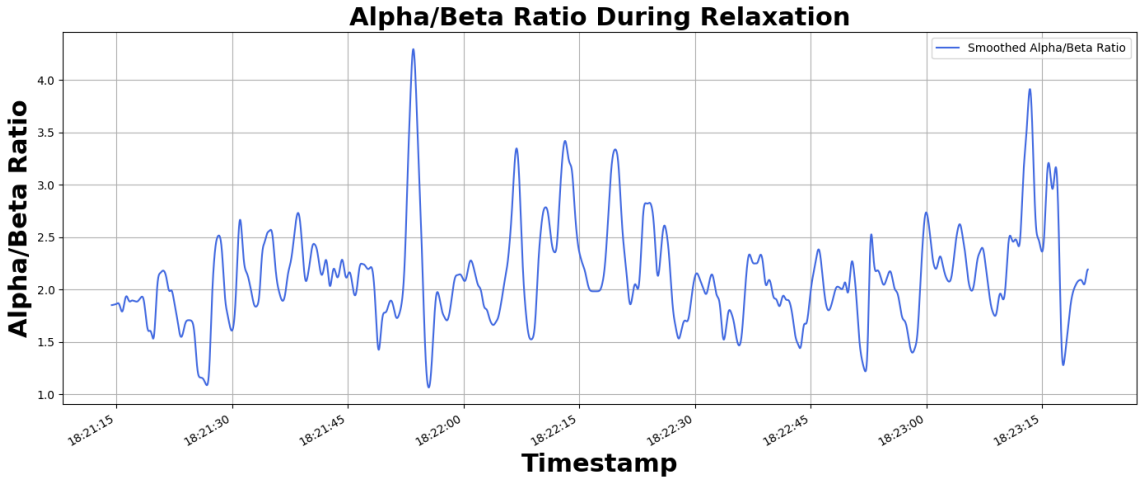
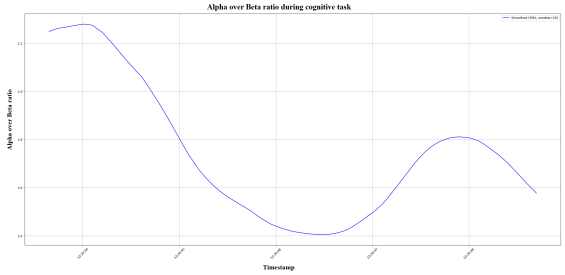
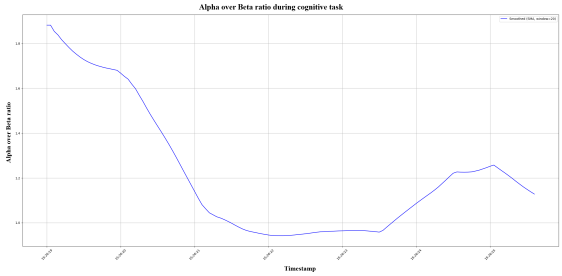


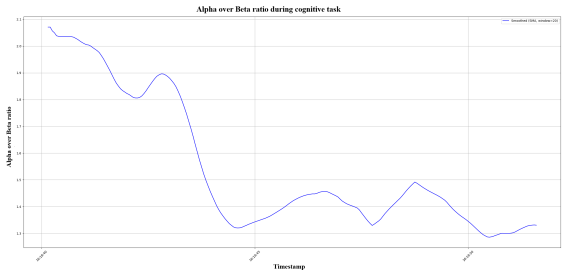
Figure B.5: Alpha over Beta ratio during relaxation fifth participant.



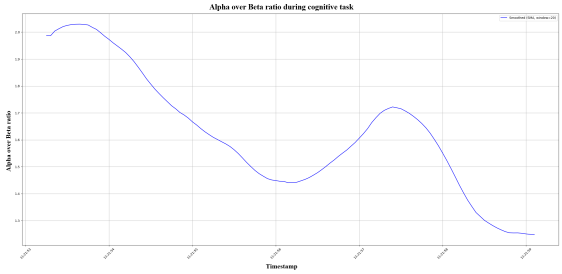
(a) Participant 1



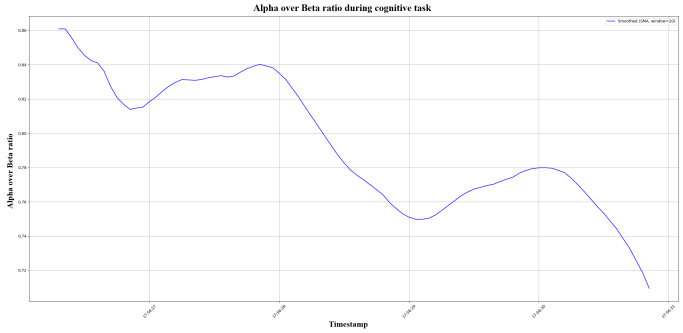
(b) Participant 2



(c) Participant 3



(d) Participant 4



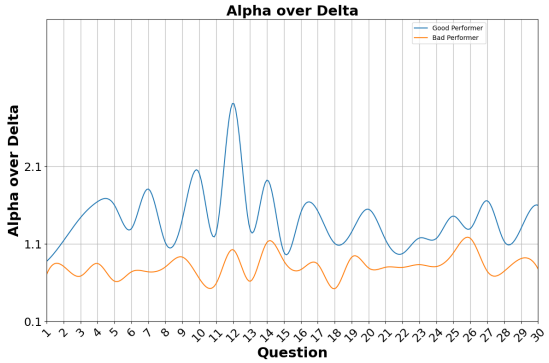
(e) Participant 5

Figure B.6: Alpha over Beta ratio during a cognitive task for five participants.

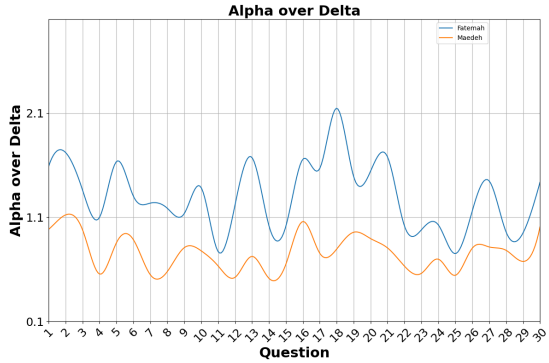
Appendix C

Alpha over Delta, Alpha over Theta, and Alpha over Delta plus Theta line graphs

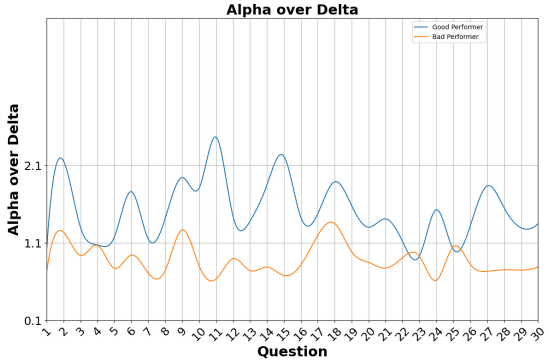
Following are the line graphs for alpha over delta, alpha over theta, and alpha over delta plus theta ratios.



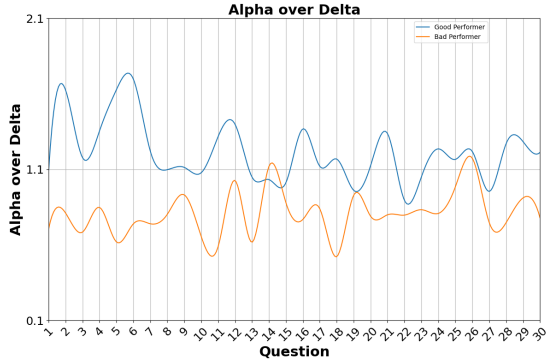
(a) Example 1



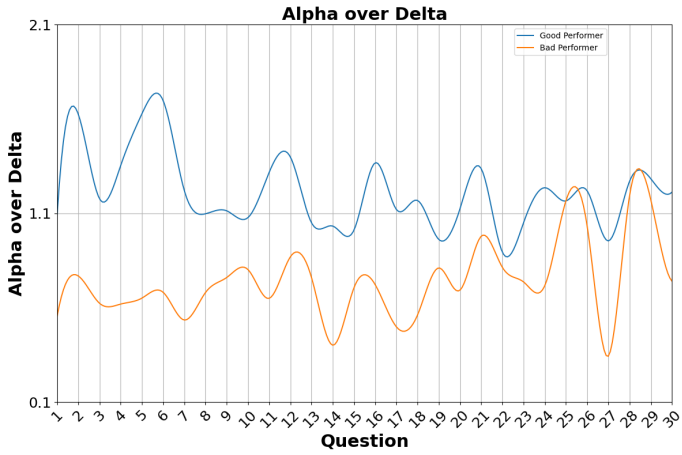
(b) Example 2



(c) Example 3

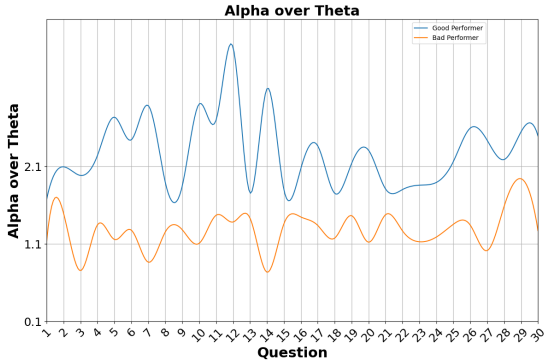


(d) Example 4

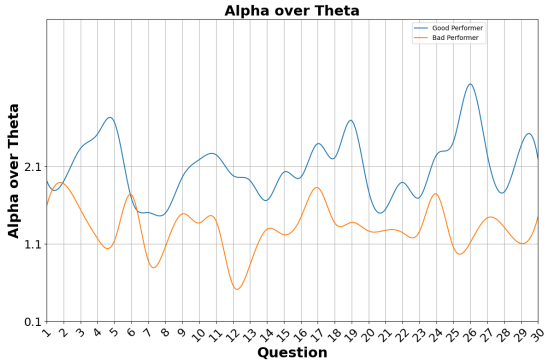


(e) Example 5

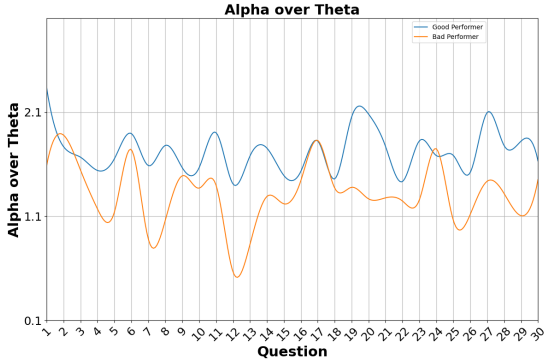
Figure C.1: Alpha over Delta ratio for good and bad performing participants across five examples.



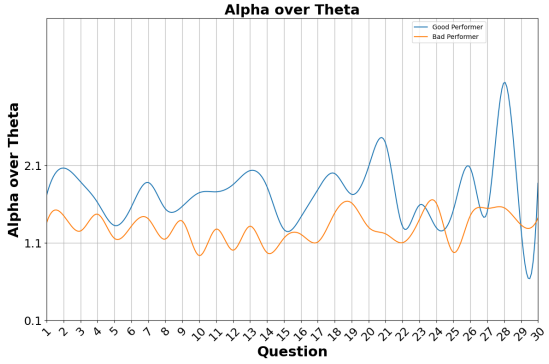
(a) Example 1



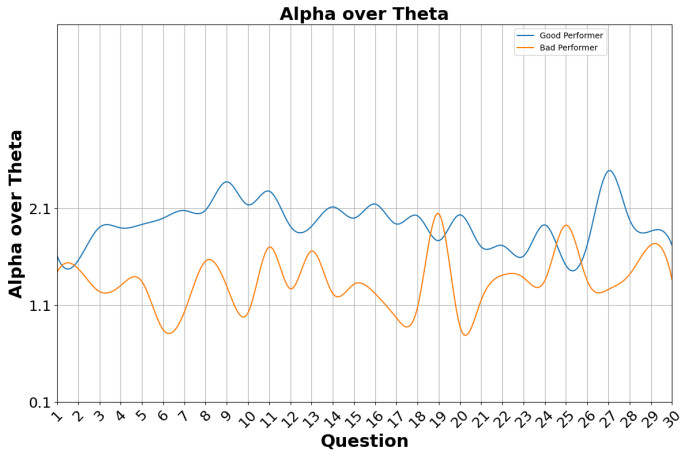
(b) Example 2



(c) Example 3

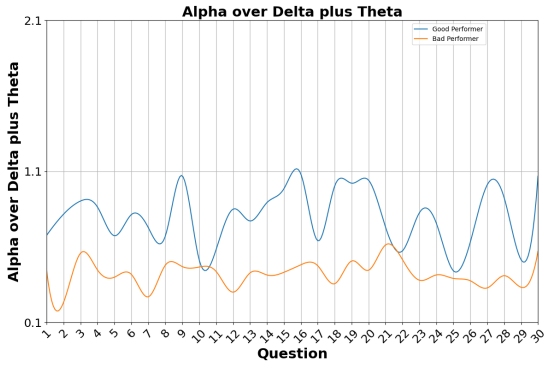


(d) Example 4

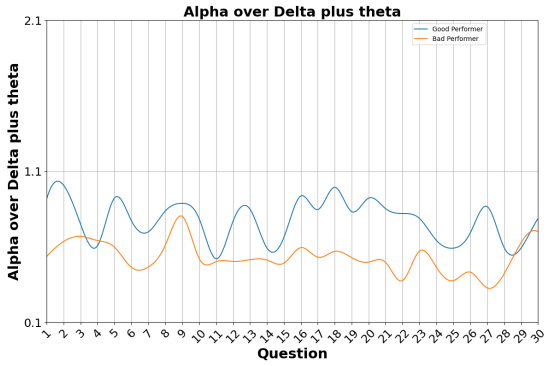


(e) Example 5

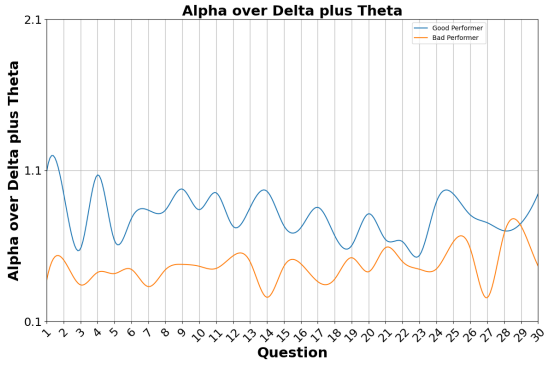
Figure C.2: Alpha over Theta ratio for good and bad performing participants across five examples.



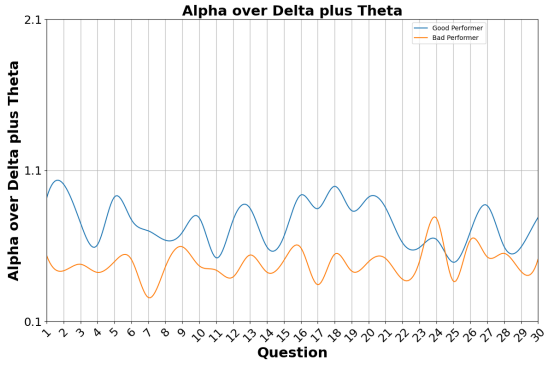
(a) Example 1



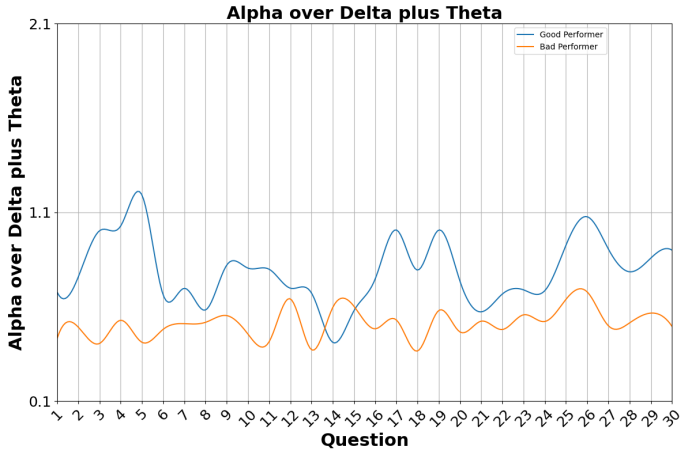
(b) Example 2



(c) Example 3



(d) Example 4



(e) Example 5

Figure C.3: Alpha over Delta plus Theta ratio for good and bad performing participants across five examples.

Appendix D

Publication

- Abdul Waleed, Nusrat J. Nitu, and Carol Fung. ”Multistage Stress Classification and Cognitive Capacity Analysis Using EEG.” 2025 The 21st International Conference on Data Science (ICDATA’2025).

Bibliography

[1] D. Bhargava and H. Trivedi, "A study of causes of stress and stress management among youth," *IRA-International Journal of Management & Social Sciences*, vol. 11, no. 03, pp. 108–117, 2018.

[2] F. Ali, A. Farooqui, F. Amin, K. Yahya, N. Idrees, M. Amjad, M. Ikhlaq, S. Noreen, and A. Irfan, "Effects of stress on job performance," *International Journal of Business and Management Tomorrow*, vol. 1, no. 2, pp. 1–7, 2011.

[3] H. M. Van Praag, "Can stress cause depression?," *Progress in neuro-psychopharmacology and Biological Psychiatry*, vol. 28, no. 5, pp. 891–907, 2004.

[4] N. Talib and M. Zia-ur Rehman, "Academic performance and perceived stress among university students," *Educational Research and Reviews*, vol. 7, no. 5, pp. 127–132, 2012.

[5] M. J. Khan, S. Altaf, and H. Kausar, "Effect of perceived academic stress on students' performance.," *FWU Journal of Social Sciences*, vol. 7, no. 2, 2013.

[6] G. P. Chrousos, "Stress, chronic inflammation, and emotional and physical well-being: concurrent effects and chronic sequelae," *Journal of Allergy and Clinical Immunology*, vol. 106, no. 5, pp. S275–S291, 2000.

[7] C. Norra, E. C. Skobel, M. Arndt, and P. Schauerte, "High impact of depression in heart failure: early diagnosis and treatment options," *International journal of cardiology*, vol. 125, no. 2, pp. 220–231, 2008.

[8] K. A. Leger, S. T. Charles, N. A. Turiano, and D. M. Almeida, "Personality and stressor-related affect.," *Journal of personality and social psychology*, vol. 111, no. 6, p. 917, 2016.

[9] M. Dol, E. McDonald, and M. A. Ferro, "Psychometric properties of the cesd, stai-t, and pss among parents of children with mental illness," *Journal of Family Studies*, vol. 28, no. 4, pp. 1416–1432, 2022.

[10] 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.

[11] B. A. Hickey, T. Chalmers, P. Newton, C.-T. Lin, D. Sibbritt, C. S. McLachlan, R. Clifton-Bligh, J. Morley, and S. Lal, "Smart devices and wearable technologies to detect and monitor mental health conditions and stress: A systematic review," *Sensors*, vol. 21, no. 10, p. 3461, 2021.

[12] M. Adil, I. Atiq, and S. Younus, "Effectiveness of the apple watch as a mental health tracker," *Journal of Global Health*, vol. 14, p. 03010, 2024.

[13] K. Nielsen, K. Mobley, R. Culbreth, J. Palmier, A. Nabulya, and M. H. Swahn, "Feasibility and acceptability of wearable devices and daily diaries to assess sleep and other health indicators among young women in the slums of kampala, uganda," *Digital Health*, vol. 10, p. 20552076241288754, 2024.

[14] E. N. Smith, E. Santoro, N. Moraveji, M. Susi, and A. J. Crum, "Integrating wearables in stress management interventions: Promising evidence from a randomized trial.," *International Journal of Stress Management*, vol. 27, no. 2, p. 172, 2020.

[15] M. Teplan *et al.*, "Fundamentals of eeg measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.

[16] R. Boostani, K. Sadatnezhad, and M. Sabeti, "An efficient classifier to diagnose of schizophrenia based on the eeg signals," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6492–6499, 2009.

[17] F. Rosenow, K. M. Klein, and H. M. Hamer, "Non-invasive eeg evaluation in epilepsy diagnosis," *Expert review of neurotherapeutics*, vol. 15, no. 4, pp. 425–444, 2015.

[18] M. M. Siddiqui, S. Rahman, S. H. Saeed, and A. Banodia, "Eeg signals play major role to diagnose sleep disorder," *International Journal of Electronics and Computer Science Engineering (IJECSE)*, vol. 2, no. 2, pp. 503–505, 2013.

[19] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of eeg-based brain–computer interface paradigms," *Journal of neural engineering*, vol. 16, no. 1, p. 011001, 2019.

[20] J. Tulen, P. Moleman, H. Van Steenis, and F. Boomsma, "Characterization of stress reactions to the stroop color word test," *Pharmacology Biochemistry and Behavior*, vol. 32, no. 1, pp. 9–15, 1989.

[21] N. Sulaiman, M. N. Taib, S. Lias, Z. H. Murat, S. A. M. Aris, M. Mustafa, and N. A. Rashid, "Development of eeg-based stress index," in *2012 International Conference on Biomedical Engineering (ICoBE)*, pp. 461–466, IEEE, 2012.

[22] C. Poh, T. Hershcovici, A. Gasiorowska, T. Navarro-Rodriguez, M. Willis, J. Powers, N. Ashpole, C. Wendel, N. Noelck, and R. Fass, "The effect of antireflux treatment on patients with gastroesophageal reflux disease undergoing a mental arithmetic stressor," *Neurogastroenterology & Motility*, vol. 23, no. 11, pp. e489–e496, 2011.

[23] L. Schwabe, L. Haddad, and H. Schachinger, "Hpa axis activation by a socially evaluated cold-pressor test," *Psychoneuroendocrinology*, vol. 33, no. 6, pp. 890–895, 2008.

[24] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1-2, pp. 76–81, 1993.

[25] S. M. U. Saeed, S. M. Anwar, H. Khalid, M. Majid, and U. Bagci, "Eeg based classification of long-term stress using psychological labeling," *Sensors*, vol. 20, no. 7, p. 1886, 2020.

[26] G. Jun and K. G. Smitha, "Eeg based stress level identification," in *2016 IEEE international conference on systems, man, and cybernetics (SMC)*, pp. 003270–003274, IEEE, 2016.

[27] A. Asif, M. Majid, and S. M. Anwar, "Human stress classification using eeg signals in response to music tracks," *Computers in biology and medicine*, vol. 107, pp. 182–196, 2019.

[28] A. Arsalan, M. Majid, A. R. Butt, and S. M. Anwar, "Classification of perceived mental stress using a commercially available eeg headband," *IEEE journal of biomedical and health informatics*, vol. 23, no. 6, pp. 2257–2264, 2019.

[29] A. Waleed, "Waleed-eeg/eeg-dataset." https://github.com/Waleed-EEG/EEG-Dataset, Jul 2025.

[30] V. Y. Cambay, I. Tasci, G. Tasci, R. Hajiyeva, S. Dogan, and T. Tuncer, "Quadtpat: Quadruple transition pattern-based explainable feature engineering model for stress detection using eeg signals," *Scientific Reports*, vol. 14, no. 1, p. 27320, 2024.

[31] U. Ince, Y. Talu, A. Duz, S. Tas, D. Tanko, I. Tasci, S. Dogan, A. Hafeez Baig, E. Aydemir, and T. Tuncer, "Cubicpat: Investigations on the mental performance and stress detection using eeg signals," *Diagnostics*, vol. 15, no. 3, p. 363, 2025.

[32] R. Ghosh, N. Deb, K. Sengupta, A. Phukan, N. Choudhury, S. Kashyap, S. Phadikar, R. Saha, P. Das, N. Sinha, *et al.*, "Sam 40: Dataset of 40 subject eeg recordings to monitor the induced-stress while performing stroop color-word test, arithmetic task, and mirror image recognition task," *Data in Brief*, vol. 40, p. 107772, 2022.

[33] H. M. Afify, K. K. Mohammed, and A. E. Hassanien, "Stress detection based eeg under varying cognitive tasks using convolution neural network," *Neural Computing and Applications*, pp. 1–15, 2025.

[34] O. AlShorman, M. Masadeh, M. B. B. Heyat, F. Akhtar, H. Almahasneh, G. M. Ashraf, and A. Alexiou, "Frontal lobe real-time eeg analysis using machine learning techniques for mental stress detection," *Journal of integrative neuroscience*, vol. 21, no. 1, p. 20, 2022.

[35] N. Phutela, D. Relan, G. Gabrani, P. Kumaraguru, and M. Samuel, "Stress classification using brain signals based on lstm network," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, p. 7607592, 2022.

[36] X. Hou, Y. Liu, O. Sourina, Y. R. E. Tan, L. Wang, and W. Mueller-Wittig, "Eeg based stress monitoring," in *2015 IEEE international conference on systems, man, and cybernetics*, pp. 3110–3115, IEEE, 2015.

[37] S. Bhatnagar, S. Khandelwal, S. Jain, and H. Vyawahare, "A deep learning approach for assessing stress levels in patients using electroencephalogram signals," *Decision Analytics Journal*, vol. 7, p. 100211, 2023.

[38] Y. Skaik, "The bread and butter of statistical analysis "t-test": Uses and misuses," *Pakistan journal of medical sciences*, vol. 31, no. 6, p. 1558, 2015.

[39] 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.

[40] F. M. Garcia-Moreno, M. Badenes-Sastre, F. Expósito, M. J. Rodriguez-Fortiz, and M. Bermudez-Edo, "Eeg headbands vs caps: How many electrodes do i need to detect emotions? the case of the muse headband," *Computers in Biology and Medicine*, vol. 184, p. 109463, 2025.

[41] Q. Zhang and Z. Yu, "A literature review on the influence of kahoot! on learning outcomes, interaction, and collaboration," *Education and Information Technologies*, vol. 26, no. 4, pp. 4507–4535, 2021.

[42] L. Music, "Dreamy horizon." https://www.youtube.com/watch?v=KkbJA5HwBKM, Oct 2024.

[43] G. Luke, "Summer memories." https://www.youtube.com/watch?v=tyWMCwIRLVs.

[44] K. MacLeod, "Clear waters." https://www.youtube.com/watch?v=kBtvkDAzG70.

[45] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[46] H. Petkar, S. Dande, R. Yadav, Y. Zeng, and T. A. Nguyen, "A pilot study to assess designer's mental stress using eye gaze system and electroencephalogram," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 48999, pp. 899–909, 2009.

[47] L. D. Sheppard and P. A. Vernon, "Intelligence and speed of information-processing: A review of 50 years of research," *Personality and individual differences*, vol. 44, no. 3, pp. 535–551, 2008.

[48] T. Bayes, "Naive bayes classifier," *Article Sources and Contributors*, pp. 1–9, 1968.

[49] D. R. Edla, K. Mangalorekar, G. Dhavalikar, and S. Dodia, "Classification of eeg data for human mental state analysis using random forest classifier," *Procedia computer science*, vol. 132, pp. 1523–1532, 2018.

[50] M. Sha'Abani, N. Fuad, N. Jamal, and M. Ismail, "knn and svm classification for eeg: a review," in *InECCE2019: Proceedings of the 5th International Conference on Electrical, Control & Computer Engineering, Kuantan, Pahang, Malaysia, 29th July 2019*, pp. 555–565, Springer, 2020.

[51] G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang, and A. Etemad, "Classification of hand movements from eeg using a deep attention-based lstm network," *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3113–3122, 2019.

[52] W. Mao, H. Fathurrahman, Y. Lee, and T. Chang, "Eeg dataset classification using cnn method," in *Journal of physics: conference series*, vol. 1456, p. 012017, IOP Publishing, 2020.

[53] B. TaghiBeyglou, A. Shahbazi, F. Bagheri, S. Akbarian, and M. Jahed, "Detection of adhd cases using cnn and classical classifiers of raw eeg," *Computer Methods and Programs in Biomedicine Update*, vol. 2, p. 100080, 2022.

[54] P. Li, Y. Pei, and J. Li, "A comprehensive survey on design and application of autoencoder in deep learning," *Applied Soft Computing*, vol. 138, p. 110176, 2023.

[55] A. Oskooei, S. M. Chau, J. Weiss, A. Sridhar, M. R. Martínez, and B. Michel, "Destress: deep learning for unsupervised identification of mental stress in firefighters from heart-rate variability (hrv) data," in *Explainable AI in healthcare and medicine: building a culture of transparency and accountability*, pp. 93–105, Springer, 2020.

[56] G. Naidu, T. Zuva, and E. M. Sibanda, "A review of evaluation metrics in machine learning algorithms," in *Computer science on-line conference*, pp. 15–25, Springer, 2023.

[57] N. C. Moore, "A review of eeg biofeedback treatment of anxiety disorders," *Clinical electroencephalography*, vol. 31, no. 1, pp. 1–6, 2000.

[58] X. Manfei, D. Fralick, J. Z. Zheng, B. Wang, M. T. Xin, and F. Changyong, "The differences and similarities between two-sample t-test and paired t-test," *Shanghai archives of psychiatry*, vol. 29, no. 3, p. 184, 2017.

[59] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "Lstm-based eeg classification in motor imagery tasks," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 26, no. 11, pp. 2086–2095, 2018.

[60] P. E. McKnight and J. Najab, "Mann-whitney u test," *The Corsini encyclopedia of psychology*, pp. 1–1, 2010.